B13. Analysis of UV-vis spectra as linear combinations of pure spectra.

Although x-ray diffraction experiments give us an amazingly detailed view of an enzyme, they provide only static snapshots of the protein and leave a lot of questions about it's mechanism. In order to follow what is happening in the active site of the enzyme in real time we would like to use more dynamic techniques.

One of the obvious features of many redox enzymes is that they are colored. The same delocalized orbital systems that permit addition or loss of electrons at moderate energy levels, allowing a protein to cycle through two or more redox states during catalysis, often lead to visible absorption spectra that change in a characteristic way when the redox state changes. In addition to this, the chromophores and the UV-absorbing amino acids in the protein itself often undergo spectral changes as their environment changes due to conformational changes or ligand binding. Finally, the reaction catalyzed, or side reactions involving substrate, product, or inhibitors, may involve spectral change due to introduction or rearrangement of double bonds. This makes Vis/UV spectroscopy a powerful approach for studying, for example, cytochrome systems. The application of electronic technology to spectroscopic instrumentation after World War II by such pioneers as Britton Chance led to a revolution in our understanding of these systems. However that revolution seemed to reach a plateau in the 70's, seemingly having reached its potential, and little advance has been made since.

The difficulty with using Vis/UV spectroscopy with complex systems is that the spectral features are broad (with respect to the wavelength range over which they are distributed), hence spectral overlap is the norm. Techniques such as Mass spectroscopy, NMR, EPR, or infrared spectroscopy are characterized by sharp lines or bands, so that quantitation of a feature is often as measuring the peak height at that position. (Vibrational spectroscopy of complex systems such as proteins does involve a lot of overlap, and might benefit from the approach described here.) Dealing quantitatively with such spectra requires (a) a procedure for decomposing experimental spectra into a linear combination of "basis" spectra of pure compound (or pure spectral changes), and (b) a way to obtain those basis spectra applicable to the problem at hand.

Problem a is not at all difficult, however it requires computation and hence for all practical purposes requires digital data. In the 60's and 70's spectrophotometers were almost universally analog, so digital data aquisition or digitization required a major effort. Although this must have been a serious dampening effect, it has to be said that most of the pioneering work was done in the days of analog spectrophotometers, when only a select few talented researchers had the wherewithal to bring digital computation to their spectroscopy problems. In the 1980's the microprocessor revolution brought digital spectroscopy to the masses, but had surprisingly little effect on the way spectroscopists use their data for biological applications. For the most part spectra are recorded digitally. printed out on a plotter or ink-jet, and pasted into the notebook in the same way a plot recorded on an analog XY recorder would have been in the 60's.

A robust procedure for decomposing spectra into their basis components was described in 1960 by Sternberg et al. [1]. This procedure can be described either as least-squares fitting of the observed spectra to the basis, or as generalized matrix inversion: it is a least-squares problem that can be easily solved using the generalized inverse of the nonsquare matrix whose columns are the basis spectra. The justification for least-squares is statistical: if the standard deviation for absorption measurement is the same at each wavelength, that solution (i.e. set of concentrations) which minimizes the sum of squared residuals is the solution which would have the greatest probability of resulting in the observed spectrum. The method can readily be extended to cases where the standard deviation is not constant, and the standard deviation at each point in a spectrum can be estimated from the spectrum itself if finely sampled. However in

practice this is almost never worth doing. The spectra either fit; in which case they fit everywhere and the relative weighting at different wavelengths makes no difference, or they don't fit; in which case the basis spectra are not appropriate and no kind of weighting will give the right answer. (There are exceptions like for instance where the experimental spectrum goes offscale and gets clipped within a wavelength range, and one assigns weight zero to that range).

The term "least squares fitting" has a number of negative connotations associated, however these involve nonlinear least-squares fitting. The Sternberg procedure uses linear least-squares. Non-linear least squares involves iterative approach to a solution, sometimes slow convergence; sometimes to a false local minimum. Linear least squares have a single minimum (at the bottom of a multi-dimensional parabola) and are solved by a single matrix multiplication. People criticizing least-square fits often refer to the ability to "fit an elephant" given sufficient parameters. In fitting spectra the number or parameters is the number of absorbing species (basis spectra) plus 1 - 3 baseline parameters. The number of data is the number of points, typically 10 per nm for several hundred nm. For fitting 10-20 components, the system is ridiculously over-determined. (This is a bit of an exaggeration, because data points taken at 0.1 nm intervals are not really independent, due to the smoothness of UV-vis spectra. Still the problem has been well-determined in my experience, except where two of the basis spectra are nearly identical). Theoretically the number of wavelength points needs to be greater than the number of components being analyzed. Sternberg et al. used a rather modest number of data points, presumably due to the difficulty in digitizing the data. In this case the choice of wavelengths becomes critical. Today when spectra are recorded digitally, the most straightforward approach is to use the entire spectrum, and it can be shown that this is as good or better than using the optimized wavelengths. The additional wavelengths that contribute little do not "dilute out" the information from the most discriminating wavelengths.

The second problem, obtaining the basis spectra, can be as simple as buying the pure compounds and taking their spectra (as in Sternberg's case), or may require a great deal of careful work. In cases of cytochrome complexes where the individual chromophores cannot be isolated without irreversibly altering their spectral properties, it involves taking spectra while manipulating the system so that as many as possible permutations of the possible spectral states are obtained. Spectra can be seen as vectors in n-dimensional space, where **n** is the number of wavelengths. However **m** different components can only span an m-dimensional subspace. The first step of problem 2 is thus to obtain a basis spanning the subspace mapped out by the experimental spectra. Unfortunately this basis is not unique. One solution is the correct biological basis, i.e. the spectra (or difference spectra) of all the contributing species (or changes). However multiplying this by any non-singular mxm matrix gives an equally appropriate basis, from a mathematical viewpoint. The Graham-Schmidt procedure gives an orthogonal basis for the spectra. The Singular Value Decomposition procedure gives another orthogonal basis which has a nice property: the recurring spectral changes segregate into the first, largest vectors, whereas unique changes like noise segregate into the later, smaller vectors. Thus a set of 100 spectra of a system with 5 spectral changes will, because of noise, occupy a space of dimension 100. The SVD procedure will give 100 basis vectors which in linear combination will exactly reproduce any of the spectra. However the first 5 of these will contain virtually all the significant information and will reproduce every spectrum to within the noise level. Their span can then be taken as the basis of the 5-dimensional space spanned by the spectra.

If the spectral change (at any one wavelength) of the spectra taken to define the basis space can be fit to an analytic equation of some parameter such as time (during a

reaction course) or pH or Eh (during a titration), then by fitting the equation "globally", i.e. at all wavelengths at the same time, using the same parameters (rate constants, pK_a's, E_m's), then the fitted absorbance change at every wavelength gives the spectrum of each transition. In the case of pH or E_h titrations, this is just the difference spectrum of the titrating component. For kinetic fits to systems of multiple first-order reactions, an additional step is required to go from the absorption change associated with each time constant to the changes for the individual reactions, however this is straightforward if the kinetic model is known. Making the fit global greatly increases the robustness of the fit, because the same parameters have to fit at different wavelengths where the relative absorbance of the different components are different. The global fit can be combined with SVD to remove noise and reduce the dimension. Linear filters such as subtracting the average spectrum from all, subtracting the best-fit straight line from each spectrum, and taking the component orthogonal to some interferants, can be applied before SVD and/or curve-fitting.

Pitfalls- Some problems arise but are easily dealt with. A small baseline drift, and especially appearance of a slanted baseline, will completely destroy what would otherwise be a good fit This can be easily dealt with by fitting a 1^{st} - 3^{d} order polynomial together with the unknowns. This amounts to adding 1-3 new components: a constant, a slanted line, and a parabola; and introduces 1-3 parameters. Wavelength error is a problem when using standards obtained with one spectrophotometer to fit experimental spectra obtained on another. The spectra need to be calibrated within about 0.1 nm for best results. Although wavelength shift is a nonlinear affect on the spectra, it can be approximated by a Taylor's series- the spectrum of a shifted sample. is equal to the correct spectrum, plus the derivative multiplied by the shift, plus the second derivative multiplied by ½ the shift squared, third derivative by 1/6 the shift cubed, etc. For shifts of ~0.1 nm the first derivative is sufficient. So if a derivative term is included in the basis spectra, the spectrum auto-corrects to the basis wavelength. Furthermore the amount of derivative term used is a very accurate measure of the amount of shift. One nice thing about digital spectra is that the wavelength calibration can be changed easily after the fact, at least in increments of the sampling interval. So the spectra can be back-shifted to the basis calibration to within the sampling interval (typically 0.1 nm), and any remaining error taken care of by a derivative term.

So why hasn't this procedure been widely applied? Why doesn't every spectrophotometer come with software for running the Sternberg procedure? One particularly unsuccessful but well-published[2-4] attempt can be looked to for "lessons learned". This study set out to ascertain difference spectra and midpoint potentials for all the titrating spectral components of mitochondria. The spectra they obtained did not look like the cytochromes expected, the number of components seemed wrong, and the work has been largely ignored since. Unfortunately they used a spectrophotometer with the monochromator after the sample, which means light from the sample had to fall on the entrance slit in order to be recorded. Such an arrangement is exquisitely sensitive to turbidity. Mitochondrial suspensions are turbid, and vary their turbidity with swelling and shrinking or changes in the suspension medium. Looking at the raw data or principal component spectra, it is clear that they are dominated by baseline changes due to turbidity. A great deal of this interference could have been removed by a linear filter, subtracting the best quadratic curve from each spectrum before SVD, but nothing like this was attempted. Finally the fit was not done in a global way, i.e. fitting all the eigenspectra simultaneously to the same set of Em's, rather each was fit separately resulting in a different set of Em's for each Eigenspectrum titration.

An example of the way we think this should be done was published with little comment as part of a purification paper in 1991 [5]. Admittedly the data was much less

challenging, since we were titrating an optically clear solution of purified potato bc_1 complex, using a spectrophotometer with the PM tube end-on to the cuvet to catch scattered light and minimize the effect of turbidity. Eigenspectra were obtained from SVD on spectra obtained in a non-potentiometric titration (eliminating the need for mediators), adjusting the redox level with dithionite and ferricyanide. These eigenspectra, uncontaminated by mediator spectra, together with difference spectra for pure mediators, were used to fit the potentiometric titration to determine the linear combinations of basis spectra making up each real spectral transition. Unfortunately a referee didn't let us publish the resolved spectra, favoring the raw spectra from the titrations and refusing to allow both in an already overly-long paper. Only a table of difference extinction coefficients at selected wavelength pairs was provided. But in fact a printed spectrum would not do much good- the digital data would be needed to put them to use. There is a real need for a public database of spectra.

1. Sternberg, J., H. Stillo, and R. Schwendeman, *Spectrophotometric Analysis of Multicomponent Systems Using the Least Squares Method in Matrix Form.* Analytical Chemistry, 1960. **32**(1): p. 84-90.
2. Reddy, K.V. and R.W. Hendler, *Complete analysis of the cytochrome components of beef heart mitochondria in terms of spectra and redox properties. The b-type cytochromes.* J Biol Chem, 1983. **258**(14): p. 8568-81.
3. Reddy, K.V. and R.W. Hendler, *Complete analysis of the cytochrome components of beef heart mitochondria in terms of spectra and redox properties. The c1-cytochromes.* Biophys J, 1986. **49**(3): p. 693-703.
4. Reddy, K.V., R.W. Hendler, and B. Bunow, *Complete analysis of the cytochrome components of beef heart mitochondria in terms of spectra and redox properties. Cytochromes aa3.* Biophys J, 1986. **49**(3): p. 705-15.
5. Berry, E.A., L.S. Huang, and V.J. DeRose, *Ubiquinol-cytochrome c oxidoreductase of higher plants. Isolation and characterization of the bc1 complex from potato tuber mitochondria.* J Biol Chem, 1991. **266**(14): p. 9064-77.