TIME CORRELATION FUNCTIONS.

We think of variables as correlated if when one variable has a large value, the other also has a large value, and if this parallel behavior persists for some period of time. Intuitively, we understand that after some longer period of time the correlation will vanish. A formal expression of the correlation between two variables for a system at equilibrium is the time correlation function,

$$C_{\beta\alpha}(t) = <\beta(t)\alpha(0)> \qquad \text{(A.1)}$$

where $\alpha(t)$ and $\beta(t)$ are time-dependent values for the two variables given as deviations from their means, and the angle brackets indicate ensemble average. The extremum of $C_{\beta\alpha}(t)$ is the limit value as $t \rightarrow 0$, $C_{\beta\alpha}(0)$. Over a sufficiently long time $C_{\beta\alpha}(t)$ decays to zero.

Correlation functions are sometimes normalized as

$$C^N_{\beta\alpha}(t) = \frac{<\beta(t)\alpha(0)>}{<\alpha(0)\alpha(0)>^{1/2}<\beta(0)\beta(0)>^{1/2}} \qquad \text{(A.2)}$$

Correlations within sets of variables are described by a matrix of correlation functions, $\mathbf{C}(t) = <\alpha(t)\alpha^T(0)>$, where α is a column vector of variables and α^T its transpose.

The time correlation function describes averaged properties of a system at equilibrium.

(a) $C(0)$ *describes fluctuations at equilibrium.* For a single variable, the zero-time autocorrelation function $C_{\alpha\alpha}(0)$ is equal to the mean squared deviation of the variable from its average value, equal to the variance for a Gaussian distribution of values of the variable. For a set of two variables, the off-diagonal elements of the zero-time correlation matrix, $C_{\beta\alpha}(0)$, are equal to the covariance for the variable pair. If $C_{\beta\alpha}(0) = 0$, there is no correlation between the variables (they are statistically independent). The

joint probabilities for values of correlated variables at equilibrium can be calculated from the zero-time correlation matrix.

(b) $C(t)$ *describes relaxation of the fluctuations.* For a single variable, $C_{\alpha\alpha}(t)$ defines a characteristic time (τ_α), under the often good assumption that relaxation (decay) of a fluctuation back to the equilibrium (time-average) value is a first order process. For two variables, $C_{\beta\alpha}(t)$ defines a characteristic time, $\tau_{\beta\alpha}$, for the loss of correlation.

The time correlation function $C(t)$, although it is determined by the behavior of a system at equilibrium, contains information on the dynamics and properties of similarly-constituted nonequilibrium systems. The connection between equilibrium and nonequilibrium systems is given by the fluctuation-dissipation theorem: variables of a nonequilibrium system not far from equilibrium relax and respond on average in the same way as fluctuations in the same variables relax on average in the equilibrium system.

(a) $C(0)$ *describes static response:* the response of one variable after long time to a constant perturbation of the same or another variable; the response observed after the system perturbed from its original equilibrium has reached its new equilibrium. For one variable, $C_{\alpha\alpha}(0) = k_B T/K$, where K is the force constant describing a region of the potential surface, generally a well bottom, within which the correlation function was measured. Thus $C(0)$ relates the response of a variable to a static external force that was applied to the same or another variable and that perturbed the system from equilibrium.

(b) $C(t)$ *describes dynamical properties of the nonequilibrium system:* the relaxation of a nonequilibrium system to equilibrium. $C_{\alpha\alpha}(t)$ defines τ_α for relaxation of a nonequilibrium state just as it does for relaxation of equilibrium fluctuations. $C_{\alpha\alpha}(t)$ also describes the opposite of relaxation to equilibrium, the time response to an external force that displaces the variable α and drives the system away from equilibrium. Interpretation of $C_{\beta\alpha}(t)$ is similar: it describes the time response of the variable β with external force applied to displace the variable α.

(c) *The response may lag the force.* The lag, due to friction and inertial effects, is greater for high rate of change of a variable and is absent for the zero-frequency response, $C(0)$, which reflects only the potential function. Energy is dissipated when work is done by an external force. Dissipation is related to the lag in response and to the contributions of friction and inertial effects.

As an example of the correspondence between equilibrium and nonequilibrium systems, consider a chemical reaction. Let $\alpha(t)$ be the concentration of one of the two species in an isomerization equilibrium. Then $C_{\alpha\alpha}(t)$ is the time correlation function for concentration fluctuations at equilibrium. The characteristic time for the relaxation of fluctuations described by $C_{\alpha\alpha}(t)$ in the equilibrium system is equal, by the fluctuation-dissipation theorem, to the relaxation time for a system prepared to have a nonequilibrium concentration of the α species. This relaxation time for the nonequilibrium system is a simple function of the forward and reverse rate constants for the isomerization reaction. Thus the rate constant for a nonequilibrium process is directly related to the concentration fluctuations at equilibrium and to $C(t)$. Why is this important? The central point is the correspondence between reaction rate and $C(t)$. This is easy to make plausible as done above. It can also be shown in other ways, such as by use of the Mori-Zwanzig formalism. Once one knows that $C(t)$ defines the rate constant, effective molecular dynamics algorithms for equilibrium systems can be used to compute $C(t)$ and thus the reaction rate. The modeling of the molecular dynamics gives a deep understanding of the microscopic events that determine the rate constant and underlay the macroscopic process described by the rate constant: friction; barrier and well shapes; collective motions, especially of the environment (protein and solvent for an enzyme active site); time scales and coupling to the environment; cross correlations and dependence of the reaction rate on other variables; etc.

CORRELATION -- TIME SCALE SEPARATION, RELAXATION, CROSS CORRELATION, DYNAMIC AND STATIC RESPONSE.

We want to make general statements about the time dependence of $C(t)$, the mathematical form of the cross-correlation $C_{\beta\alpha}(t)$, and the relation of the cross-correlation characteristic time $\tau_{\beta\alpha}$ to the characteristic times of the separate motions of the α and β variables. The intent is to understand coupling between protein motions of the same or different time scale as it applies to protein processes, such as enzyme rate processes.

A system at equilibrium but transiently in a fluctuant state relatively far from equilibrium will evolve by relaxing back toward equilibrium (decay of the fluctuation). For example, a fluctuant value for a coordinate of a system in a potential well will evolve toward the value at the well bottom. Considered within the framework of the Generalized Langevin equation, the characteristic time for relaxation of some particular system variable is determined by interactions (i) with other variables of the same set **A** and (ii) with variables that are sensed only as part of the environment; the effects of these interactions are described, respectively, by the factors $i\hat{\omega}$ and $K(t)$ of the Generalized Langevin equation.

An external force, acting upon some variable to displace it from its equilibrium value, elicits a response in the variable displaced and also in the other variables of the set **A**. By the fluctuation-dissipation theorem, the time course of this response is on average exactly that for relaxation of the corresponding fluctuant quantity of the equilibrium system.

For some dynamical variable that relaxes and responds on a particular time scale, we may expect there to be certain other variables that relax and respond on approximately the same time scale, and there to be yet others with dynamics on much faster or slower time scales. Considering, for example,s a particular torsional libration of a protein that relaxes on the 0.1-1 ps time scale, we would expect there to be other torsions relaxing and

responding on the same time scale, with the hard degrees of freedom of bond vibration, bond bending, etc., being on faster time scales, and on slower time scales, flips between torsional minima, charge redistribution, cooperative structural events, etc. Some of the motions, such as torsional libration and bond vibration, are motions within wells of a high-dimensional potential-energy surface; others, like flips between torsional minima, involve crossing a barrier between potential-energy wells. The event of barrier crossing is typically on the 0.1-1 ps time scale, with the slowness of a process such as torsional flip, charge movement, etc., being due to the low probability of a crossing event (a high barrier).

We can group dynamical variables as fast or slow according to time scale of relaxation and response. If we pick a particular time-scale group, such as torsional librations for a protein, and we observe the system on the time scale of this group (0.1-1 ps), then we can call "fast" any variables relaxing on a much shorter time scale (<0.1 ps), and "slow" those variables relaxing on the observation time scale. It is proper as well as convenient to ignore, for the present, variables with much longer characteristic times, because these are invariant on the observation time scale. The fast variables, insofar as they influence the slow, comprise the random force driving fluctuations of slow variables about their average trajectories, the term $f(t)$ of the Generalized Langevin equation, and cause damping of the motion, the factor $K(t)$ or $\Gamma(0)$ of the Generalized Langevin equation.

For a particular process, such as catalysis of a reaction by some enzyme, one can identify, at least in principle, a group of slow variables, the set **A(t)**, that fully describe the process. Different processes are described, of course, by different sets **A** of slow variables.

If we specify some one slow dynamical variable β of the set **A**, say a particular torsional libration of a protein, a time correlation function $C(t)$ describes its relaxation. The mathematical form of $C(t)$ is a sum of exponentials in the time:

$$C(t) = \sum_{i=1}^{n} R_i \, e^{\gamma_i t} \qquad \text{(D.1)}$$

where the γ_i are the poles (singularities) of the Laplace transform of $C(t)$ and the R_i are the residues at the n poles. The γ_i can be complex valued. The real part of γ_i if negative defines a relaxation on the time scale $\tau_i = 1/Re(\gamma_i)$. There must be least one pole, corresponding to the slow variable β, for which $\tau >> \tau_c$, the correlation time of the random force (the fast variables). If there are other poles corresponding to a value of $\tau >> \tau_c$, these represent other slow variables potentially coupled with the slow variable β. Cross correlations $C_{\beta\alpha}(t)$ as well as single-variable autocorrelatons $C_{\alpha\alpha}(t)$ have the mathematical form of Eq. (D.1). For a particular set of slow variables **A**, every correlation function is described by the same set of exponentials in the time, one exponential term for each variable, with the summations differing in the coefficients R_i of Eq. (D.1). We want to understand the rules for selecting which exponential term dominates for a particular correlation function. The following section is a summary of some relevant points.

COMMENTS -- TIME SCALE SEPARATION, RELAXATION, CROSS CORRELATION, DYNAMIC AND STATIC RESPONSE.

(1) An analysis based on Eq. (D.1) can be used as a guide to understanding the structure of time correlation functions, particularly two-variable cross-correlation functions, $C_{\beta\alpha}(t)$.

(a) The correlation function is a sum of exponentials in the time.

(b) Each exponential term contributes to the time dependence of the correlation either as a relaxation (a decay toward zero), or as a divergence if the exponent is positive.

(c) If the exponent has an imaginary part, the term has a periodic factor (an oscillation superimposed on the relaxation).

(d) The number of terms is less than or equal to the number of slow variables that are coupled. Two slow variables, such as a coordinate and a velocity, may be associated with even an apparently simple, single process.

(e) The relative weight of each term and its relaxation time τ are functions of parameters of all the coupled slow variables. However, one variable or a small number of variables can dominate, and the sum of exponentials can reduce to one or two terms, or equivalently, the relaxation can be determined by one or a few variables of similar τ.

(2) Protein motions cover a range of time scales. It is important to consider coupling between processes of widely different τ. For the correlation function $C_{\beta\alpha}(t)$, let β be the slower process and α the faster.

(a) For wide time-scale separation, the relaxation of $C_{\beta\alpha}$ will be controlled by τ_β, the slower characteristic time. If observation is on the time scale of the faster α process, then no relaxation of the β variable or of

$C_{\beta\alpha}$ will be seen.

(b) This can be stated also in terms of response of the β variable to displacement of the α variable by an external force. If displacement is carried out on the α process time scale, there can be no response of the β variable. If displacement of the α variable is carried out slowly and on the β process time scale, the β variable will respond.

(c) If the external force is applied over a very long time, the β variable will respond fully. This long-time response is the static, or equilibrium, response. The size of the static response is proportional to $C_{\beta\alpha}(0)$, the covariance of the equilibrium fluctuations.

(d) The symmetries of the generalized Langevin equation require that $C_{\beta\alpha}(t) = \pm C_{\alpha\beta}(t)$, with the sign negative if the α and β processes differ in time reversal symmetry.

(e) Consequently, $\tau_{\beta\alpha} = \tau_{\alpha\beta}$. The α variable when driven by displacement of the β variable responds on the time scale of the β variable. This can come as no surprise, since the β variable cannot be displaced in times shorter than τ_β. Because $\tau_\alpha << \tau_\beta$, the faster α variable shows instantaneous response (static or equilibrium response) to change in the β variable.

(3) The statements of paragraph (2) have several consequences with regard to analysis of an enzyme rate process.

(a) A slow enzyme process cannot be coupled to a fast process on the time scale of the fast process. For example, a pK shift cannot be coupled to barrier crossing, the time scale of the latter being 0.1-1 ps and of the former, μs or longer.

(b) For a long-lifetime ES complex, there is the possibility for static coupling (correlation) among all processes with characteristic times shorter than the lifetime of the ES complex.

(c) During barrier crossing, those variables with characteristic times greater than that of barrier crossing maintain the quasi-equilibrium values of the ES complex from which barrier crossing was initiated. The distribution of these values is described by the static correlation function $C_{\beta\alpha}(0)$ for those variables with characteristic time less than the lifetime of the ES complex.

(d) The nature of the distribution of quasi-equilibrium correlations carried into the transition state is a factor in determining height of the barrier. This contribution should be viewed as an equilibrium solvent effect. Such correlations are almost certainly important for enzyme catalysis: consider the distribution of protons among charged groups in the transition state for an enzyme RDS to which acid-base catalysis contributes.