How to Turn the Reaction Coordinate into Time

Anthony Auerbach

Department of Physiology and Biophysics, SUNY at Buffalo, Buffalo, NY 14214

What are the internal motions of an ion channel protein when it “gates” between closed and open-channel conformations? In 2005, Zhou, Pearson, and Auerbach (ZPA) (Biophys. J. 89:3680–3685) showed that \( \Phi \), a parameter derived from the forward and backward rate constants of a chemical reaction, may in some cases provide temporal information about the moving parts of a protein. This “primer” is intended for ion channel biologists and others who seek an intuitive understanding of how rate constants might reflect the relative timing of the internal motions of proteins. Here, I will focus on the method of \( \Phi \)-value analysis and, specifically, the basis for the temporal interpretation of \( \Phi \).

The Transition Region

The gating isomerization of an ion channel is a chemical reaction in which the reactant (C for closed) is separated from the product (O for open) by a potential energy barrier. Ion channels are typically large membrane proteins and many bonds are likely to change position in the C\( \leftrightarrow \)O reaction. Undoubtedly, the protein adopts numerous short-lived, intermediate structures during its passage across the energy barrier that separates C from O. In patch clamp experiments there is little direct sign of these intermediate states because their lifetimes are too brief to be detected (regardless of whether or not they can conduct ions).

When we write down a reaction scheme like C\( \leftrightarrow \)O, the letters symbolize the stable end state ensembles and the arrows symbolize passage through the ensemble of the separating barrier. The highest point along the lowest energy trajectory across this barrier has been called the “transition state.” However, for a reaction as big and complex as channel gating there may be no obvious such single state. Rather, the barrier may be populated by many short-lived microstates. I will therefore refer to the separating barrier as the “transition region” (TR) of the reaction.

The C\( \leftrightarrow \)O reaction can be viewed as an energy diagram, where the y axis is energy and the x axis is the “reaction coordinate,” which can be thought of as the extent to which the reaction has occurred, on a scale from 0 to 1. There is no a priori reason to assume any particular overall shape of the TR energy barrier, but the simplest picture is a parabolic peak at the intersection of the parabolas that define the end-state energy wells. However, other barrier shapes are also consistent with an apparent two-state reaction, as long as events that occur during crossing are too brief to be measured.

Conformational Dynamics

If we want to understand the “mechanism” of gating then a good place to start is to know which amino acids change position between C and O. If we had high resolution structures of the end states, we could see which atoms moved. Unfortunately, for most channels we do not yet have such complete pictures, but we do have a number of tools that can be used to infer residue motion.

In the cysteine accessibility method, a change in the rate at which a sulfhydryl reagent reacts with a target cysteine assesses if the residue has a different accessibility in C vs. O. Fluorescence measurements can show if a residue experiences a change in its dielectric environment between C and O, and electron spin resonance can measure a change in side chain mobility or accessibility. Similarly, a mutation-induced change in the equilibrium constant between open and closed (\( K_{eq} \)) indicates that the amino acid experiences a change in energy in the reaction, because \( K_{eq} \) is related (exponentially) to the energy difference between C and O. For instance, consider a mutation that destabilizes C more than O (or stabilizes O more than C), and which therefore increases \( K_{eq} \). Energy and structure are related, so this change in \( K_{eq} \) implies that there is a difference in the structure at that residue between C and O, and, hence, a relative gating “motion” in the vicinity of the mutation. An absence of signal (in the equilibrium constant, accessibility or fluorescence) does not necessarily imply the absence of motion, but simply that the environment of the probe did not change much.

For example, if an entire helix moves as a unit, the relative positions of the residues do not change but all the residues do move.

What is the timing of all of the atomic movements through the TR? Remarkably, it may be possible to probe the landscape of this invisible world simply by examining the kinetics of the C\( \leftrightarrow \)O reaction. Just as a change in the equilibrium constant following a perturbation informs us of a change in the energy (structure) of the protein, it might also inform us of the timing of the motions. How can we use the rate constants derived from the reaction scheme C\( \leftrightarrow \)O to understand this timing? One method is to transform the reaction coordinate into a temporal parameter called \( \Phi \) following a method described by ZPA. Although there are many different \( \Phi \) parameters one can derive from the reaction scheme for C\( \leftrightarrow \)O, \( \Phi_{eq} \) seems to be the most intuitive overall indicator of gating motion because it is the parameter derived from the rate constants. For a two-state reaction, \( \Phi_{eq} \) is simply the equilibrium constant for the reaction, which is the ratio of forward and backward rate constants: \( \Phi_{eq} = K_{eq} = \frac{K_{f}}{K_{b}} \).

Abbreviations used in this paper: TR, transition region.
near the residue in going from C to O, changes in the opening and closing rate constants inform us of the movement of residues in the TR.

REFER Analysis
We start with an empirical observation. For many different kinds of chemical reactions, including ion channel gating, changes in the opening (k_o) and closing (k_c) rate constants are often correlated. For example, if a mutation increases K_eq (=k_o/k_c), then commonly k_o will increase and k_c will decrease. There is no fundamental thermodynamic reason why this must be so, it just often happens. The analysis of the correlation between k_o (or k_c) and K_eq goes by a number of names: Φ-value analysis, Brønsted analysis, linear free energy relationship (LFER) analysis, or rate-equilibrium free energy relationship (REFER) analysis. In all cases, log k_o vs. log K_eq are plotted for a family of mutations of a single residue. The slope of a straight line fit of the data points is called Φ, which typically varies from 0 to 1. If Φ = 0, then the change in K_eq was not caused by a change in the closing rate constant k_c, but instead entirely by a change in the opening rate constant k_o. If Φ = 1, then the change in K_eq was caused entirely by a change in k_o, and a fractional Φ means that both k_o and k_c contributed to the net change in K_eq.

Not only are the changes in the rate and equilibrium constants often correlated for mutation of a single residue, but frequently the relative fold-changes in k_o vs. k_c are approximately the same for all mutants in the family. This pattern gives rise to an approximately linear REFER. What does a linear REFER imply? Log(k_o), the y axis of the REFER, is a function of the energy difference between the closed state and the top of the barrier. Log(K_eq), the x axis, is a function of the energy difference between the bottom of the C and O wells. Therefore, a line of constant slope implies that all the members of the mutation family saw similar changes in energy relative to those of the ground states when the channel reached the TR. Φ measures the extent to which mutations of a residue change the opening and closing rates and provides a glimpse of the TR energy landscape and the structure of the protein when it is between C and O.

Physical Interpretations of Φ
Three different physical interpretations of fractional Φ values have been proposed. (1) Φ reflects the fractional structure of the perturbed residue at the TR (see www.pitb.de/biophys/bp23/). The perturbed side chain is somewhere between its stable positions in C and O, to an extent that is quantified by Φ. Φ is a snapshot of the residue’s energy and, hence, structure at the TR. In protein folding reactions, Φ is often thought to report the fraction of “native” contacts of the perturbed residue at the TR. (2) Φ reflects structural heterogeneity arising from the existence of multiple reaction pathways across multiple TRs. The experimental value of Φ will depend on the probability of taking each pathway. (3) Φ reflects the fractional time during the process of channel opening when the perturbed residue changes its local structure, completely and in an all-or-none fashion, from C to O. At this moment, some parts of the protein are fully C-like and others, fully O-like. Φ is a snapshot of the C vs. O structure of the entire protein at the TR. The molecule wanders through a series of intermediate states in the TR and finally arrives in the open state, and, as we shall see, the reaction coordinate has become time.

We can apply each of these interpretations to a hypothetical REFER. Suppose we measure k_o and k_c for 10 different substitutions of a residue and measure the slope of the REFER, Φ = 0.3. That is, a mutation that increased K_eq by 10-fold did so by increasing k_o by twofold and decreasing k_c by fivefold (log 2/log 10 = 0.3). In this instance, the changes in K_eq were driven by unequal contributions from the rate constants. Using the first interpretation of Φ, we imagine that in the TR the residue has moved ~30% of the way between its stable structure in C vs. O. Using the second interpretation of Φ we imagine that the residue is either fully O-like or fully C-like in structure in either of two TRs, with probabilities of 0.3 and 0.7, respectively. Using the third interpretation of Φ we imagine that the mutated residue switches instantaneously from a C-like structure to an O-like structure ~70% of the way in time between C and O.

REFERs are not always linear. They may have points that have more scatter than can be accounted for by measurement errors. For example, if mutations change k_o and k_c equally there is no change in K_eq and the REFER plot becomes a vertical line (Φ = ∞). Also, in some cases the slope of the REFER is not constant but may decrease as K_eq increases. These complexities whet the appetite for more rigorous analyses and may someday allow us to distinguish between the physical interpretations of Φ, but in this primer we will follow the interpretation of linear REFERs with a slope between 0 and 1.

Temporal Interpretation of Φ
In this section I expand on only one of these interpretations, that Φ may reflect the relative timing of the change in structure of an amino acid. In this picture, the protein changes its shape in steps, making incremental changes in its conformation. In terms of an energy diagram, this is a random walk along a lumpy TR. In the simplest case, it is a one dimensional walk through sequential states. The molecule jumps thermally from C to the edge of the TR and then diffuses through the microstates of the TR until it either falls back to C or falls forward into O. While the net rate of crossing the TR is faster than the patch clamp can reveal, the channel may fall back to C several times before it reaches O. Sometimes a “walk” in the TR will even cross the transition state divide, only to return to C. The fraction of all
sojourns in the TR that reach O is called the transmission coefficient \( k_{C \to O} \) and is a number between 0 and 1. The inverse of the transmission coefficient is the average number of times the TR is visited before stable O is achieved \( n_{C \to O} \).

We can understand, qualitatively, how the number of TR visits and \( \Phi \) are related. Considering only the forward reaction, perturbations that influence the number of visits will influence \( k_{o} \), and, hence, the y axis of the REFER. However, the x axis, log \( K_{eq} \), depends only on the energy difference between C and O and is independent of the number of visits. Assume that a perturbation changes the energy of just one of the microstates structures of the TR. The position of the perturbation influences the magnitude of \( n_{C \to O} \) and, hence, \( k_{o} \). Perturbations near the C side of the barrier will influence nearly all of the random walks and will therefore have a significant effect on the average number of visits (“crossings”). By the same logic, perturbations near the O side of the barrier won’t have much effect on the number of crossings because most of the “unproductive” random walks (which end up back at C) rarely sample this part of the TR. However, an energy change will affect \( K_{eq} \) to the same extent, regardless of its position in the TR. Therefore, because of the random walk, perturbations near the O side of the barrier won’t have much effect on the number of crossings because most of the “unproductive” random walks (which end up back at C) rarely sample this part of the TR. However, an energy change will affect \( K_{eq} \) to the same extent, regardless of its position in the TR.

For quantitative analysis we will consider the number of times the channel must visit the TR before reaching O. A one-dimensional, sequential pathway between C and O is equivalent to the Markov chain: \( C \leftrightarrow X_{1} \leftrightarrow X_{2} \ldots \leftrightarrow X_{n} \leftrightarrow O \). The X states are the brief intermediate states of the TR. If we assume that the total lifetime in the TR is small compared with the lifetimes in C and O, there is an approximate solution for the average number of TR crossings required for a full C-to-O transit:

\[
 n_{C \to O} \approx (1 + r_{1} + r_{1}r_{2} + r_{1}r_{2}r_{3} + \ldots + [r_{1}r_{2} \ldots r_{n}]). \tag{1}
\]

(This is Eq. 7 in ZPA.) Each \( r \) is the ratio of the backward/forward rates out of the corresponding X state \( (r_{1} = k_{X_{1} \to C}/k_{X_{1} \to X_{2}}, \text{etc.)} \). This ratio depends only on the relative heights of the microbarriers of the TR and is independent of the depths of the microstate wells. Remember that the time spent in the TR is negligible compared with the time spent in C or O, so the depths of the TR wells are irrelevant, as long as they are small compared with the stable end states. The depth of these wells determines the effective “friction” across the TR but not the number of crossings.

Notice that \( r_{1} \) appears in almost every term in Eq. 1 whereas \( r_{n} \) appears only in one. This means that the backward/forward ratio out of the first microstate in the TR will have a greater influence on the number of crossings than the backward/forward ratio out of the last microstate, since the channel-opening process starts in microstate 1. The weight that each \( r \)-value contributes to \( n_{C \to O} \) decreases as the system progresses across the TR. The correlation of \( \Phi \) with time is a natural consequence of walking through the states of TR. Perturbations of “early” structural transitions will have a greater influence on \( k_{o} \) and a higher \( \Phi \) than do “later” ones. The reaction coordinate may not be linear in time, but it is monotonic.

Examples

A surprising and useful aspect of this model is that it is invertible. That is, one can use the experimentally determined \( \Phi \) values to calculate \( r \) values and, hence, the relative heights of the energy barriers along the potential energy surface in which the random walk occurs.

Suppose we measure \( k_{o} \) and \( k_{c} \) for a family of mutations of one residue and fit the resultant REFER to obtain \( \Phi \). Next, imagine that we repeat this process for many different residues throughout the protein and obtain many such \( \Phi \) values. Finally, we bin these \( \Phi \) values and count the number of populations. If there are \( N \) \( \Phi \) populations, we would conclude that there is a microstate in the TR corresponding to each value and that there are \( N \) intermediate states. If we include both C and O in our accounting, this makes \( N + 2 \) total states in the Markov scheme for the reaction. In the following examples we will ignore the real experimental issues of errors (in the estimates of the rate constants, \( \Phi \) values, and \( \Phi \) population means) and REFER curvature.

The quantitative relationship between \( \Phi \) and number of crossings is

\[
 \Phi_j = 1 - \frac{n_{C \to X_j}}{n_{C \to O}}. \tag{2}
\]

(This is Eq. 16 in ZPA). \( \Phi_j \) is the experimentally observed value of \( \Phi \) following a perturbation of the jth intermediate step (\( X_{j-1} \leftrightarrow X_{j} \)), and \( n_{C \to X_j} \) is the average number of crossings required to reach \( X_{j} \) starting from C (=1 + \( r_{1} + r_{1}r_{2} + \ldots + [r_{1}r_{2} \ldots r_{j}] \)). The observed value of \( \Phi \) following a perturbation is, simply, the fractional position of the perturbed step in the TR (i.e., the reaction coordinate) expressed as the fractional number of crossings.

Now we can relate the experimentally observed \( \Phi \) values to the shape of the TR energy landscape. If we expand Eq. 2 by expressing the number of crossings in terms of \( r \) values and then rearrange terms, we arrive at a simple quantitative relationship between \( \Phi \) and \( r \):

\[
 r_j = \frac{\Phi_j - \Phi_{j+1}}{\Phi_{j-1} - \Phi_j}, \tag{3}
\]

which is \( r_{1} = (\Phi_{1} - \Phi_{2})/(\Phi_{0} - \Phi_{1}), r_{2} = (\Phi_{2} - \Phi_{3})/(\Phi_{1} - \Phi_{2}), r_{3} = (\Phi_{3} - \Phi_{4})/(\Phi_{2} - \Phi_{3}) \ldots \) We now define \( \Phi_{0} = 1 \) and \( \Phi_{n+1} = 0 \). Given experimentally determined
Φ values, it is easy to calculate r values and (by Eq. 1) the average number of TR crossings required for a full C-to-O transit. Further, we can calculate the exit rate constant from C (O) into the TR just by multiplying the apparent opening (closing) rate constant by \( n_{C \rightarrow O} \) (\( n_{O \rightarrow C} \)).

In the first example, suppose we measure only one population with \( \Phi = 0.5 \). According to our simple model of the TR, one \( \Phi \) population implies that the TR has one microstate and that the appropriate scheme has three states: \( C \leftrightarrow X \leftrightarrow O \) (each arrow is a TR barrier). We now solve Eq. 3, \( r_1 = (1 - 0.5) / (0.5 - 0) \), and find that \( r_1 = 1 \).

We also learn from Eq. 1 that the average total number of crossings for a full C-to-O transit is \( n_{C \rightarrow O} = 1 + 1 = 2 \). This is easy to understand because the backward and forward transitions from the single intermediate X state are made with equal probabilities. As a consequence, the escape rate constant from C is twice the apparent opening rate constant. We do not have any information about the lifetime or conductance of the intermediate state because it is too brief to be observed directly by the patch clamp.

In the second example we assume that we have measured two \( \Phi \) populations, 0.8 and 0.4. The model now has two intermediate states (\( C \leftrightarrow X_1 \leftrightarrow X_2 \leftrightarrow O \)). From Eq. 3 and Eq.1 we compute \( r_1 = 2.0, r_2 = 1.0, \) and \( n_{C \rightarrow O} = 5 \). Here, the C side of the TR energy landscape is tilted upwards, leading to a significant number of recrossings. The exit rate constant from C is now five times faster than the apparent opening rate constant.

Solving the equations becomes more difficult with a larger number of \( \Phi \) populations but it is still easily accomplished using a symbolic math program. In the next two cases we will assume that four \( \Phi \) populations have been detected. Suppose that these are 0.8, 0.6, 0.4, and 0.2. We calculate (Eq. 3) that all of the r values are 1, which means that the TR barriers are perfectly flat with the backward/forward exit rate constants from all of the intermediates being equal. We also calculate that C (or O) must be revisited an average of five times before a full, observable “gating” transition takes place. In general, with a flat (but corrugated) TR the number of crossings is proportional to the number of microstates.

In the last example, we will use 0.98, 0.75, 0.25, and 0.02 for the measured values of \( \Phi \). We first compute the four r values: 11.5, 2.17, 0.46, and 0.09. That is, the first half of the TR landscape is uphill (the backward rate constants are larger then the forward ones, which makes \( r > 1 \)) and second half is downhill (\( r < 1 \)). This pattern for \( \Phi \) traces out a parabolic TR that requires, on average, \(~49\) visits for each full transit.

**Summary**

A simple Markov model of the TR passage, and a consideration that multiple crossings of the TR can contribute to the magnitude of an apparent rate constant, leads to a temporal interpretation of \( \Phi \). One useful aspect of this approach is that it allows us to consider TRs that are not parabolic and, importantly, to use experiments (the rate constants) to probe the overall shape of the TR barrier. Certainly, reaction mechanisms for protein conformational change are more complex than is encoded in the simple model. It is likely that the microscopic structural transitions of the TR are not obligatorily sequential. There may be multiple pathways over the separating barrier. Perturbations of the TR may change more than a single microscopic equilibrium constant. It still uncertain whether the sequential, Markov scheme for the TR is a sufficient approximation for some reactions, and, if not, whether this approach can be extended to incorporate more complex mechanisms of protein conformational change.

Yu Zhou derived Eq. 3. I thank Fred Sachs, Chris Lingle, and Joe Henry Steinbach for comments and suggestions.

Olaf S. Andersen served as editor.

Submitted: 4 October 2007
Accepted: 6 November 2007