



Article

Red-Edge Excitation Shift Spectroscopy (REES): Application to Hidden Bound States of Ligands in Protein–Ligand Complexes

Md Lutful Kabir ¹, Feng Wang ² and Andrew H. A. Clayton ^{1,*}

¹ Optical Sciences Center, Department of Physics and Astronomy, School of Science, Swinburne University of Technology, Melbourne, VIC 3122, Australia; mdlutfulkabir@swin.edu.au

² Center for Translational Atomaterials, Department of Chemistry and Biotechnology, School of Science, Swinburne University of Technology, Melbourne, VIC 3122, Australia; fwang@swin.edu.au

* Correspondence: aclayton@swin.edu.au

Abstract: Ligand-protein binding is responsible for the vast majority of bio-molecular functions. Most experimental techniques examine the most populated ligand-bound state. The determination of less populated, intermediate, and transient bound states is experimentally challenging. However, hidden bound states are also important because these can strongly influence ligand binding and unbinding processes. Here, we explored the use of a classical optical spectroscopic technique, red-edge excitation shift spectroscopy (REES) to determine the number, population, and energetics associated with ligand-bound states in protein–ligand complexes. We describe a statistical mechanical model of a two-level fluorescent ligand located amongst a finite number of discrete protein microstates. We relate the progressive emission red shift with red-edge excitation to thermodynamic parameters underlying the protein–ligand free energy landscape and to photo-physical parameters relating to the fluorescent ligand. We applied the theoretical model to published red-edge excitation shift data from small molecule inhibitor–kinase complexes. The derived thermodynamic parameters allowed dissection of the energetic contribution of intermediate bound states to inhibitor–kinase interactions.

Keywords: fluorescence spectroscopy; red-edge excitation shift; protein–ligand interactions; free energy landscape; kinase inhibitors

Citation: Kabir, M.F.; Wang, F.; Clayton, A.H.A. Red-Edge Excitation Shift Spectroscopy (REES): Application to Hidden Bound States of Ligands in Protein–Ligand Complexes. *Int. J. Mol. Sci.* **2021**, *22*, 2582. <https://doi.org/10.3390/ijms22052582>

Academic Editors: Lukasz Kurgan and Mahmoud Ghomi

Received: 15 January 2021

Accepted: 26 February 2021

Published: 4 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Ligand–protein binding is a key regulatory process that controls the vast majority of bio-molecular functions [1]. Evolution has shaped the constituents of living matter so that these macromolecular recognition events are largely robust and specific. On the other hand, disease states can result from improper protein–ligand interactions, which drive malfunctions in cellular behavior. Ligand–protein binding is also of interest to drug companies, where often the goal is to develop drugs that bind with increasing affinity and specificity to a particular drug target. These companies also use structural techniques such as x-ray crystallography or Nuclear magnetic resonance spectroscopy (NMR) to characterize the structure of the ligand–protein complex and to define the environment surrounding the ligand in the protein-binding pocket. Structure-based design aims to increase the affinity of the drug for the drug target based on this structural knowledge [2].

Drug efficacy, on the other hand, depends on the dynamics of protein–ligand interactions [3]. For example, the population of transient or “hidden”, ligand-bound states during ligand release from the protein–ligand complex can increase ligand residency times at or near the ligand-binding pocket. Ligand residency time is thought to be an important parameter in drug efficacy because a long ligand residency time provides a temporal discrimination against shorter-lived off-target interactions [3]. The importance of intermediate states can be understood in the context of the induced fit model of protein–ligand

interaction [3]. In this scenario, the ligand first associates with the protein surface (as an encounter complex) to form an (or more intermediate) initial bound state(s) from where the final protein–ligand complex geometry is reached. Ligand release occurs in an opposite way to ligand binding, but requires the population of one or more intermediate bound states prior to ligand dissociation from the ligand–protein complex. At each stage of intermediate state population, the ligand can either repopulate the bound state or move toward a surface state. The determination of these intermediate states is experimentally challenging [4] because they often constitute a minor population (at steady-state equilibrium) and require specialized techniques to track them [5,6].

Red-edge excitation spectroscopy (REES) is undergoing a renaissance [7–11] as one method that could potentially determine the presence of minor (sub) populations in an ensemble of protein conformational ensembles. First described by Weber [12], Galley [13], and Rubinov [14], REES relies on a distribution of fluorophore–solvent interactions, referred to as in-homogeneous broadening, which arises from a distribution of microenvironments around a population of fluorophores. The peak of the $S_0 \rightarrow S_1$ transition in the absorption spectrum represents the most populated fluorophore–solvent microenvironment, whereas excitation at the red-edge of the absorption spectrum selectively excites less populated microenvironments where the ground state is destabilized (in energy) and the excited state is stabilized relative to the ground state. Provided the solvent environment relaxes slowly compared to the excited-state emission, then the presence of the red-edge excited species is unveiled from a characteristic red shift in the emission from the fluorophore population as the excitation wavelength is moved from the peak of the absorption spectrum to the red-edge of the absorption spectrum. If the ligand is an environment-sensitive fluorophore and the protein represents the solvent (and is static on the fluorescent timescale), then red-edge excitation spectroscopy should, in principle, be able to detect intermediate ligand bound states of ligand–protein complexes.

In the present work, our motivation was to link the red-edge excitation shift phenomenon to the properties of intermediate ligand-bound states (number, population, and energetics). To do this, we built a thermodynamic model of a fluorescent ligand interacting with a number of microstates in a protein. We used Marcus electron transfer theory [15,16] to describe the absorption and emission processes of the fluorescent ligand in each microstate. The population distribution of microstates along the protein–ligand coordinate is derived from a protein–ligand free energy landscape (Figure 1).

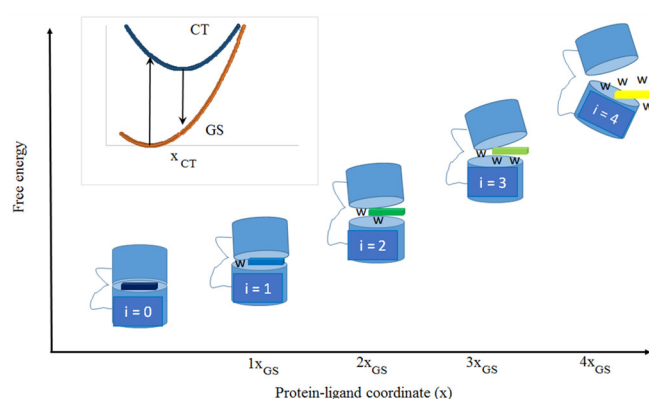


Figure 1. Idealized thermodynamic model for ligand–protein interactions. Free energy of protein–ligand microstates ($i = 0$ to 4) as a function of generalized protein–ligand coordinate (x). In this model, the protein–ligand microstates are equally spaced (x_{GS}). The blue bi-cylindrical shapes depict the protein, ligand is represented by the colored rectangles, while bound water molecules are represented by the letter w . The change in color of the ligand along coordinate x represents the change in spectroscopic parameters of the ligand (i.e., red-shift in absorption and emission). Inset: Marcus free energy surfaces for the ligand ground state (GS) and ligand excited-state (CT) with x_{CT} (the change in geometry of the ligand in the CT state) indicated.

The rest of this paper is organized as follows. In Section 2, we describe a theory of the red-edge excitation shift for a fluorescent ligand bound to a protein. In Section 3.1, we use the model to simulate the influence of the number, population, and energetics of the ligand-bound states on the shape of the red-edge excitation shift curve. In Section 3.2, we used the model to fit the experimental red-edge excitation shift data [17] from a solvatochromic kinase inhibitor [18], under conditions of restricted protein motion [19], bound to two different enzymes [20,21]. The thermodynamic parameters derived from the model allow energetic dissection of the contribution of intermediate bound states to the ligand binding process.

2. Theory

2.1. Requirements of the Red-Edge Excitation Spectroscopy (REES) Effect

For REES to occur, there are a general set of requirements [7,11,13], which we will outline. First, the fluorophore requires a ground state and an electronic excited state, which both differ in the interaction with the environment. For this to occur, there needs to be a dipole moment difference (magnitude and/or direction) between ground and excited-states. Second, the fluorophore needs to be immersed in a range of environments, such that there is a distribution of fluorophore–environment interactions. Third, the dynamics of the environment needs to be comparable or slower than the time-scale of the fluorescence emission. We will make an additional stipulation that the dynamics of the environment is slower than the rate of fluorescence emission. This stipulation means that we only need to specify the distribution of environments in our model. We will next place these requirements in the context of a quantitative model for the REES effect.

2.2. Absorption and Emission from a Ligand in a Single Microstate

Consider a two-level ligand, L , with an electronic ground state, GS, and a single electronic excited-state, termed a charge-transfer or CT state. The term CT implies that the excited-state has a charge-transfer character (i.e., a dipole moment, which is different in the direction and/or magnitude to the electronic ground state), which is a requirement for the REES effect. We will treat the absorption and emission processes using classical Marcus theory such as CT absorption and CT emission, respectively.

We denote the free energy of the GS as a function of a collective coordinate x as $G_{GS}(x)$ and the corresponding free energy of the CT state, as $G_{CT}(x)$, Equation (1).

From Marcus electron transfer theory [15,16], we can write the local free energies as,

$$G_{GS}(x) = \frac{kx^2}{2} \quad (1a)$$

$$G_{CT}(x) = \frac{k(x-x_{CT})^2}{2} + \Delta G_{CTGS} \quad (1b)$$

where k is a constant and x_{CT} is the displacement of the excited-state free energy surface along the coordinate x (with respect to the ground state free energy surface), and ΔG_{CTGS} is the vertical displacement of the minima of the ground-state and excited-state free energy surfaces. For simplicity, we have taken k to be equal in the GS and CT states (ignore the different impact on k in GS and CT states).

Assuming thermal equilibrium in the ground-state, the probability of the system being at coordinate x , $P(x)$, is given by the Boltzmann distribution (Equation (2)),

$$P(x) = \sqrt{2\pi T k_b} e^{-\left(\frac{G_{GS}(x)}{k_b T}\right)} \quad (2a)$$

$$P(x) = \sqrt{2\pi T k_b} e^{-\left(\frac{0.5kx^2}{k_b T}\right)} \quad (2b)$$

In Equation (2b), $P(x)$ is a Gaussian function centered at $x = 0$; k_b is the Boltzmann constant; and T is the temperature. Note that $G_{GS}(x)$ is given in Equation (1a).

The absorption energy ($h\nu_{abs}$) of a photon required to make L transition from the ground state GS to the excited state CT is given by the expression in Equation (3),

$$h\nu_{abs}(x) = G_{CT}(x) - G_{GS}(x) = \Delta G_{CTGS} + \frac{k(x_{CT})^2}{2} - kx(x_{CT}) \quad (3)$$

Equation (3) shows that the absorption energy is a linear function of x . By inference, the absorption spectrum will also be a Gaussian function of the absorption energy with a single peak (mode) at $x = 0$ and a width related to temperature and the constant k .

The absorption maximum is located at $x = 0$, i.e., and from Equation (4), we can see that

$$h\nu_{abs}(x = 0) = G_{CT}(0) - G_{GS}(0) = \Delta G_{CTGS} + \frac{k(x_{CT})^2}{2} \quad (4)$$

The value $(1/2) k(x_{CT})^2$ is called the reorganization energy in Marcus theory.

To describe the emission process from the excited CT state to the GS ground state on the ligand L , we will assume that thermal (vibrational) relaxation occurs in the local CT state (i.e., a single microstate) prior to emission. By analogous reasoning, the emission spectrum is given as a Gaussian function with the mean (and mode) position of the excited CT state at $x = x_{CT}$. The peak (and mean value) of emission spectrum is given by Equation (5),

$$h\nu_{em}(x = x_{CT}) = G_{CT}(x_{CT}) - G_{GS}(x_{CT}) = \Delta G_{CTGS} - \frac{k(x_{CT})^2}{2} \quad (5)$$

Combining Equations (4) and (5), we arrive at the expected results from Marcus theory, that the reorganization energy for the CT to GS transition is half the Stokes shift and that the free energy for the GS to CT transition is given as the average of the energies of the absorption and emission peaks, Equation (6a,b).

$$\frac{h\nu_{abs}(x = 0) - h\nu_{em}(x = x_{CT})}{2} = \frac{k(x_{CT})^2}{2} \quad (6a)$$

$$\frac{h\nu_{abs}(x = 0) + h\nu_{em}(x = x_{CT})}{2} = \Delta G_{CTGS} \quad (6b)$$

The reorganization energy is normally composed of two parts, an inner-sphere contribution, which relates to changes in the geometry of the ligand L (and closely associated solvent molecules) (i.e., changes in bond lengths) and an outer-sphere contribution, which relates to changes in the solvent dipoles to accommodate the charge transfer state. As stipulated above, we will assume that the dominant contribution to relaxation in the excited CT state is due to vibrational relaxation in the ligand (i.e., inner sphere reorganization) with no or negligible reorganization of the protein or water dipoles during the excited-state lifetime.

Note that because we have assumed vibrational/inner-sphere relaxation in the local CT state, the emission will be independent of the excitation energy within the single microstate. For the single GS , single CT state considered here, the red-edge excitation shift will be zero.

2.3. Absorption and Emission from a Ligand in a Collection of Discrete Protein Microstates

We now consider the ligand L distributed among a collection of N discrete microstates. The motivation for considering discrete states, as opposed to a continuum, is from experimental and sophisticated computational studies, revealing the existence of discrete (yet transient) bound states of ligands in proteins [4,5,22–24]. For simplicity, we will assume that the microstates are equally-spaced along coordinate x by a displacement of x_{GS} . We also define a protein–ligand free energy landscape FEL (i), which is a parametric function of the microstates. This definition means that for N states, we only need to define one energy parameter, as opposed to N individual parameters for the N states. We assume functions of the form Ai^b where A and b are constants ($b = 1$ linear, $b = 2$ quadratic, etc.) and i is the i^{th} microstate (starting from $i = 0$ to $i = N$). For example, $b = 1$ describes a free energy ramp, $b = 2$ is a quadratic free energy function.

With these prescriptions, the ground state free energy surfaces of the 0th, 1st, ..., i^{th} , ..., and Nth microstates are given by the expressions in Equation (7a–d).

$$\begin{aligned} G_{GS0}(x) &= \frac{kx^2}{2} \\ G_{GS1}(x) &= \frac{k(x-x_{GS})^2}{2} + A \\ G_{GSi}(x) &= \frac{k(x-ix_{GS})^2}{2} + Ai^b \\ G_{GSN}(x) &= \frac{k(x-Nx_{GS})^2}{2} + AN^b \end{aligned} \quad (7a-d)$$

We now have to consider the CT states (i.e., excited-states) of the ligand in the different microstates. We denote the free energy difference between ground-state and CT state in the 0th microstate as ΔG_{CTGS0} , and the corresponding quantity for the Nth microstate as ΔG_{CTGSN} . The free energy surfaces of the 0th, 1st, ..., i^{th} , ..., and Nth, CT states are then given by the expressions in Equation (8a–d),

$$\begin{aligned} G_{CT0}(x) &= \frac{k(x-x_{CT})^2}{2} + \Delta G_{CTGS0} \\ G_{CT1}(x) &= \frac{k(x-x_{GS}-x_{CT})^2}{2} + A + \Delta G_{CTGS0} + \left(\frac{1}{N}\right)(\Delta G_{CTGSN} - \Delta G_{CTGS0}) \\ G_{CTi}(x) &= \frac{k(x-ix_{GS}-x_{CT})^2}{2} + Ai^b + \Delta G_{CTGS0} + \left(\frac{i}{N}\right)(\Delta G_{CTGSN} - \Delta G_{CTGS0}) \\ G_{CTN}(x) &= \frac{k(x-Nx_{GS}-x_{CT})^2}{2} + AN^b + \Delta G_{CTGSN} \end{aligned} \quad (8a-d)$$

In Equations (9a–d), we have implicitly assumed that the reorganization energy for the CT to GS transition is independent of the microstate. This assumption is in accordance with the stipulation of limited outer-sphere reorganization of the protein environment during the excited-state lifetime of the ligand (i.e., only vibrational relaxation of the ligand contributes to the Stokes shift). Note that in the construction of the CT surfaces of the microstates, three additional parameters are needed: x_{CT} , ΔG_{CTGS0} , and ΔG_{CTGSN} . We will see later that these quantities can be determined from experiment.

We can immediately distinguish different behaviors of the microstates depending on the different parameter values characterizing the FEL.

(i) When $\Delta G_{CTGSN} < \Delta G_{CTGS0}$ and $A > 0$, $b > 1$, as we proceed along the FEL(i) in the 0 to N direction, the ground state increases in energy, while the CT states become stabilized in energy with respect to the GS. This is the qualitative behavior of the red-edge excitation effect (i.e., red-edge excitation leading to a red-shift in the emission).

(ii) When $\Delta G_{CTGSN} > \Delta G_{CTGS0}$ and $A > 0$, $b > 1$, as we proceed along the FEL(i) in the 0 to N direction, the ground state increases in energy, while the CT states become destabilized in energy with respect to the GS. This is the qualitative behavior of the red-edge excitation, leading to increasing blue shifted emission (i.e., an unorthodox red-edge effect).

(iii) Combinations of (i) and (ii) can lead to essentially either a red-edge effect or blue emission upon increasing red-excitation.

Having defined the FEL, we now consider the behavior of the ligand L within the ensemble of the N microstates (i.e., at thermal equilibrium). The population distribution of ligand L, among the microstates, depends on the FEL and temperature via the Boltzmann equation. The populations of state 0, 1, i, and N are given by the expressions in Equation (9a–d),

$$\begin{aligned} P_{GS0}(x) &= e^{-\left(\frac{G_{GS0}(x)}{k_b T}\right)} \\ P_{GS1}(x) &= e^{-\left(\frac{G_{GS1}(x)}{k_b T}\right)} \\ P_{GSi}(x) &= e^{-\left(\frac{G_{GSi}(x)}{k_b T}\right)} \\ P_{GSN}(x) &= e^{-\left(\frac{G_{GSN}(x)}{k_b T}\right)} \end{aligned} \quad (9a-d)$$

The relative weights of microstates 0, 1, ..., i, ... and N are defined in Equations (10a–e),

$$\begin{aligned} W_{GS0}(x) &= P_{GS0}(x)/Z(x) \\ W_{GS1}(x) &= P_{GS1}(x)/Z(x) \\ W_{GSi}(x) &= P_{GSi}(x)/Z(x) \\ W_{GSN}(x) &= P_{GSN}(x)/Z(x) \end{aligned} \quad (10a-e)$$

$$Z(x) = P_{GS0}(x) + P_{GS1}(x) + P_{GSi}(x) + \cdots + P_{GSN}(x)$$

It is useful to define the absorbance energy for the GS to CT transition of each microstate. The absorbance energy for each microstate 0, 1, ..., i, ..., and N are denoted by Equation (11a–d),

$$\begin{aligned}
h\nu_{abs0}(x) &= \Delta G_{CTGS0} + \frac{k(x_{CT})^2}{2} - kx(x_{CT}) \\
h\nu_{abs1}(x) &= \left(\frac{1}{N}\right)(\Delta G_{CTGSN} - \Delta G_{CTGS0}) + \Delta G_{CTGS0} + \frac{k(x_{CT})^2}{2} - kx(x_{CT}) \\
h\nu_{absi}(x) &= \left(\frac{i}{N}\right)(\Delta G_{CTGSN} - \Delta G_{CTGS0}) + \Delta G_{CTGS0} + \frac{k(x_{CT})^2}{2} - kx(x_{CT}) \\
h\nu_{absN}(x) &= \Delta G_{CTGSN} + \frac{k(x_{CT})^2}{2} - kx(x_{CT})
\end{aligned} \tag{11a-d}$$

The emission from each microstate is considered to occur from the equilibrated energy minimum of the CT state within each microstate, but not between microstates (no transfer between microstates during the excited-state lifetime). The position of the emission peak (and average emission energy) of each microstate is given by Equation (12a–d),

$$\begin{aligned}
h\nu_{em0}(x) &= \Delta G_{CTGS0} - \frac{k(x_{CT})^2}{2} - kx(x_{CT}) \\
h\nu_{em1}(x) &= \left(\frac{1}{N}\right)(\Delta G_{CTGSN} - \Delta G_{CTGS0}) + \Delta G_{CTGS0} - \frac{k(x_{CT})^2}{2} - kx(x_{CT}) \\
h\nu_{emi}(x) &= \left(\frac{i}{N}\right)(\Delta G_{CTGSN} - \Delta G_{CTGS0}) + \Delta G_{CTGS0} - \frac{k(x_{CT})^2}{2} - kx(x_{CT}) \\
h\nu_{emN}(x) &= \Delta G_{CTGSN} - \frac{k(x_{CT})^2}{2} - kx(x_{CT})
\end{aligned} \tag{12a-d}$$

We now turn to the determination of how the ensemble behaves in photon absorption and emission.

The ensemble absorption energy along coordinate x is given as the weighted sum of the absorbances of the individual microstates (Equation (13)),

$$h\nu_{abstot}(x) = h\nu_{abs0}(x)W_{GS0}(x) + h\nu_{abs1}(x)W_{GS1}(x) + \cdots + h\nu_{absN}(x)W_{GSN}(x) \tag{13}$$

The average emission energy from the ensemble along coordinate x is given as the weighted sum of the average emissions from the individual microstates (Equation (14)),

$$h\nu_{emtot}(x) = h\nu_{em0}W_{GS0}(x) + h\nu_{em1}W_{GS1}(x) + \cdots + h\nu_{emN}W_{GSN}(x) \tag{14}$$

A plot of the LHS of Equation (14) as a function of LHS of Equation (13) is the average emission energy as a function of the absorbance energy (i.e., the red-edge excitation shift in energy space). The REES curve depends on the free energy landscape parameters (A and b), the microstate spacing (x_{GS}), the number of microstates N , and L parameters (reorganization energy $((1/2))kx_{CT}^2$), and min and max free energy gaps $(\Delta G_{CTGSN}, \Delta G_{CTGS0})$.

The photo-physical parameters pertaining to the ligand L dictate the separation in energy between the highest energy blue emission and the lowest energy red emission as well as the peak (blue) absorption and red absorption. The parameters thus dictate the positions and ranges of observables, but do control the shape of the REES curve. The reorganization energy and free energy gaps can be obtained from spectral data provided the peak absorbance and blue/red emissions can be isolated.

From the absorbance and emission energies of the highest energy or lowest energy states, the reorganization energy can be obtained from half the Stokes shift and the free energy gaps obtained from the average of the absorbance and emission energies (Equations (15a–d)),

$$\begin{aligned}
\frac{k(x_{CT})^2}{2} &= \frac{h\nu_{absN}(max) - h\nu_{emN}(max)}{2} = \frac{h\nu_{abs0}(max) - h\nu_{em0}(max)}{2} \\
\Delta G_{CTGS0} &= \frac{h\nu_{abs0}(max) + h\nu_{em0}(max)}{2} \\
\Delta G_{CTGSN} &= \frac{h\nu_{absN}(max) + h\nu_{emN}(max)}{2} \\
(\Delta G_{CTGS0} - \Delta G_{CTGSN}) &= h\nu_{em0}(max) - h\nu_{emN}(max)
\end{aligned}
\tag{15a-d}$$

In practice, one would normally assign $h\nu_{abs0}(max)$ from the $S_0 \rightarrow S_1$ absorbance peak of the ligand in the ligand–protein complex (or possibly the equivalent quantity from the fluorescence excitation spectrum) and $h\nu_{em0}$ from the corresponding emission peak. $h\nu_{emN}$ could be inferred from the emission peak upon red-excitation or alternatively from data fitting. We will discuss the modeling in the next section.

Naturally, the parameters that control the energies and positions of the microstates do influence the shape of the REES curve. To show this, we will simulate different conditions that may occur in actual ligand–protein-complexes. These simulations are presented in Section 3.1.

2.4. Modeling Approach

Equations (1)–(15) present the mathematical description of the model at hand and at first sight contain many parameters. In this sub-section, we explain our approach to fitting a dataset, how we obtained parameters from experiment, and how we evaluated and justified different models. As described previously [17], we found that a sigmoidal fit was sufficient to describe REES data from kinase inhibitor–kinase complexes. This model contains four parameters, two which control the end-points of the curve and two parameters that control the mid-point and the slope of the mid-point in the transition region. This sigmoidal curve provides a mathematical representation of the REES data, which is used in subsequent model fitting and interpretation. The parameters in the REES model for the ligand can be obtained from the end-points of the sigmoidal fit to the REES data and either the absorption maximum (or the excitation spectrum maximum). Specifically, from Equation (15b), ΔG_{CTS0} can be obtained from absorbance maximum and the blue-emission end-point in the REES curve, while ΔG_{CTSN} can be obtained from the blue and red endpoints and ΔG_{CTS0} . The reorganization energy (and thus x_{CT}) can be obtained from the absorbance maximum and the blue end-point from the sigmoidal fit. Thus, while there are three parameters associated with the ligand, all of them can be obtained from the experiment. To model the REES data using the FEL model, we first modeled the REES data with one intermediate (i.e., $N = 1$). This model has four free energy surfaces, one pair (ground and excited state) for the bound state and one pair (ground and excited state) for the intermediate state. This model has been published previously to describe red-edge effects in other biological macromolecules [25]. This model contains the least number of parameters (three parameters for the ligand and two parameters for the intermediate state). Note that because the ligand parameters were fixed from the experiment, the number of parameters needed to fit the REES transition region for this model was two. We performed this fit minimizing the sum least squares from the sigmoid fit and the theoretical model.

The next level of complexity involves models with two or more intermediate states. We created models with $N = 2$ up to 20 intermediate states. Each model was a separate model (in the same way that a fit of fluorescence decay curve to one exponential decay is different from a model that is a fit to two exponential decays). These models contained two extra parameters that define the FEL of the microstates (compared to the single intermediate model), these parameters are N (the number of intermediate states) and b . We found that $b = 2$ worked best and this agreed well with the free energy functional from a

recent study of a kinase inhibitor–kinase complex [4]. Quadratic free energy functionals abound in biology, physics, and chemistry and for the present model, imply that the probability of the N^{th} state has a Gaussian probability function with N . Nevertheless, b and N were both varied to obtain an optimum fit.

The increase in the number of parameters with the $N > 1$ models needs to be justified on the basis of some criterion. We used the Akaike information criterion (AIC), which allows a comparison of models with different numbers of parameters and balances the need for good data fitting with over-parameterization. The formula for the AIC value is given in Equation (16),

$$AIC = 2k + n \ln LSQ \quad (16)$$

where k is the number of parameters; n is the number of datapoints; LSQ is the sum of squared residuals; and \ln is the natural logarithm. For example, for the p38a-inhibitor complex, with $N = 1$, $LSQ = 10$, and with $N = 10$, $LSQ = 4$. With 10 datapoints, $n \ln(LSQ) = 23$ ($N=1$ model) and $n \ln(LSQ) = 13.8$ ($n = 10$), which was a larger difference than $2k = 4$ (penalty for the greater number of parameters) with the $N = 10$ model.

3. Results

3.1. Simulations of REES Curves for Different Protein–Ligand Free Energy Landscapes

To understand how the theoretical REES curve depends on parameters of the protein–ligand free energy landscape, we have carried out simulations for a defined system consisting of 10 ligand–protein microstates. The spectroscopic parameters of the ligand L were set to ($x_{CT} = 0.89$ (eV)^{0.5}, $\kappa = 1$, $\Delta G_{CTGS0} = 3.346$ eV, $\Delta G_{CTGSN} = 3.223$ eV), which corresponds to an emission peak in the range of 420nm to 438nm (note that these values are somewhat arbitrary).

Figure 2 contains a series of REES curves as a function of the energy parameters of the protein–ligand free energy landscape. It is notable that these curves are sigmoidal functions with emission wavelength limits matching those input into the model, as expected. In Figure 2a, x_{GS} was fixed and b was set to 1 ($b = 1$ (linear ramp) and A was variable. This defines the FEL as a linear ramp function. As can be seen in Figure 2a, increasing the value of the energy parameter A (i.e., the energy gradient), causes a significant change in the position of the mid-point of the sigmoidal REES function along the excitation wavelength axis (i.e., to longer wavelengths). In Figure 2b, A and b were fixed, and the influence on the spacing of the free energy surfaces of the microstates along coordinate x (i.e., the parameter x_{GS}) was determined. Increasing the value of x_{GS} changed the shape of the REES curve, causing the gradient near the mid-point to increase significantly.

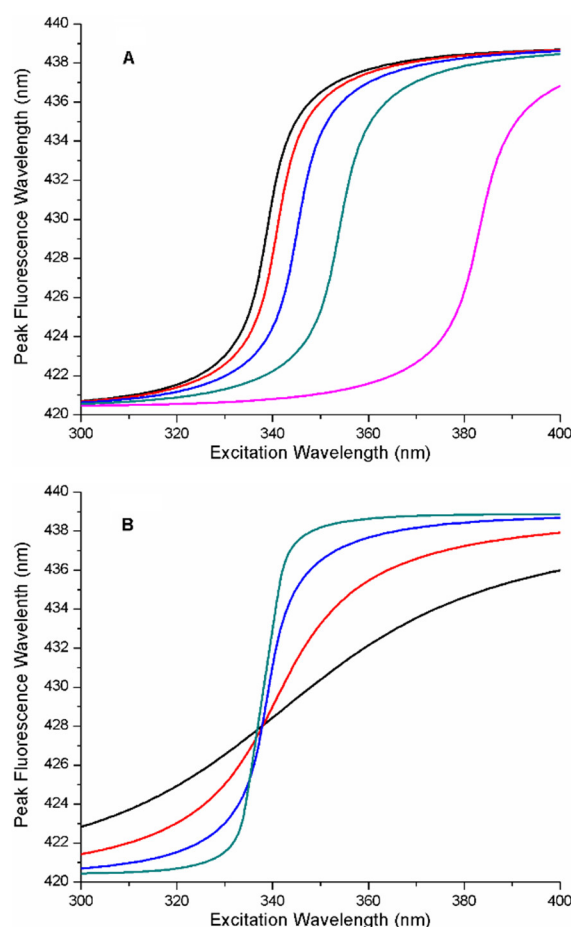


Figure 2. Influence of protein–ligand free energy landscape on red-edge excitation spectroscopy (REES) plots. **(A)** REES curves for a linear ramp free energy landscape ($b = 1$). Plots from left to right represent the effect on REES of increasing free energy gradient ($A = 0.0025$, $A = 0.005$, $A = 0.01$, $A = 0.02$, $A = 0.05$ eV). All other parameters were fixed ($x_{GS} = 0.1$ eV^{0.5}, $b = 1$). The photo-physical parameters for the ligand, L were fixed ($x_{CT} = 0.89$ eV^{0.5}, $\kappa = 1$, $\Delta G_{CTGS0} = 3.346$ eV, $\Delta G_{CTGSN} = 3.223$ eV). Plots were generated with the specified parameters using Equations (13) and (14) from the theory section (the range of x was from -1 to 2). **(B)** REES curves for a linear ramp free energy landscape ($b = 1$). Plots represent the effect, on REES, of increasing spacing between microstate free energy surfaces along the reaction coordinate from shallowest to steepest near the transition mid-point ($x_{GS} = 0.025$, $x_{GS} = 0.05$, $x_{GS} = 0.1$, $x_{GS} = 0.2$ eV^{0.5}). All other parameters were fixed ($b = 1$, $A = 0.0025$ eV). The photo-physical parameters for the ligand, L were fixed ($x_{CT} = 0.89$ eV^{0.5}, $\kappa = 1$, $\Delta G_{CTGS0} = 3.346$ eV, $\Delta G_{CTGSN} = 3.223$ eV). Plots were generated with the specified parameters using Equations (13) and (14) from the theory section (the range of x was from -1 to 2).

In the linear ramp model for the free energy landscape ($b = 1$), the free energy difference between the successive minima of the microstate free energy surfaces is given by Equation (17),

$$G_{GS1}(\min) - G_{GS0}(\min) = \Delta G_{GSi+1,i} = A \quad (17)$$

And the ground-state reorganization energy, RE (Equation (18)) is

$$RE = \frac{k(x_{GS})^2}{2} \quad (18)$$

From Marcus theory, the activation energy between microstates is (in the linear ramp model) is given by Equation (19),

$$\Delta G_{GSi+1,i}^{act} = \frac{(\Delta G_{GSi+1,i} + RE)^2}{4RE} \quad (19)$$

In the context of a linear protein–ligand FEL, shifts in the mid-point of the REES curve indicate changes in the relative thermodynamic stability of the most populated state relative to the least populated state. Changes in the steepness of the REES transition (all other things being equal) would indicate a change in the roughness of the FEL (i.e., increase or decrease in activation energy between successive microstates). A rougher FEL produces a steeper REES transition.

We also considered more complex protein–ligand energy landscapes such as quadratic (or harmonic) free energy as a function of the microstate. The results of a simulation on the influence of changing the A parameter on the REES curves are shown in Figure 3 for the quadratic FEL (i.e., the i^{th} microstate free energy proportional to i^2). The simulations reveal that increasing A (which is analogous to a Hook’s law spring constant) causes both a shift in the mid-point of the REES curve (along the excitation wavelength axis to longer wavelengths) and a consequential change in the steepness of the REES transition (more shallow appearing transition). In Figure 3b, A and b were fixed, and the influence on the spacing of the free energy surfaces of the microstates along coordinate x (i.e., the parameter x_{GS}) was determined. Increasing the value of x_{GS} , changes the shape of the REES curve, causing the gradient near the mid-point to increase significantly.

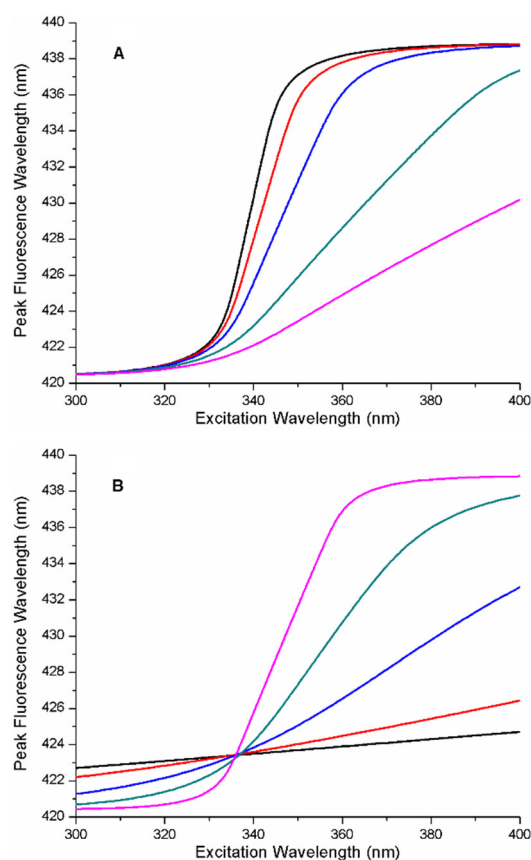


Figure 3. REES curves for a quadratic free energy landscape ($b = 2$). **(A)** Plots from left to right represent the influence of increasing the free energy parameter A, on REES curves ($A = 0.001$ eV, $A = 0.002$ eV, $A = 0.005$ eV, $A = 0.01$ eV). All other parameters were fixed ($x_{GS} = 0.15$ eV $^{0.5}$, $b = 2$). The photo-physical parameters for the ligand, L were fixed ($x_{CT} = 0.89$ eV $^{0.5}$, $\kappa = 1$, $\Delta G_{CTGS0} = 3.346$ eV, $\Delta G_{CTGSN} = 3.223$ eV). Plots were generated with the specified parameters using Equations (13) and (14) from the theory section (the range of x was from -1 to 2). **(B)** Plots represent the effect on REES of increasing spacing between microstate free energy surfaces along the reaction coordinate from shallowest to steepest near the transition mid-point ($x_{GS} = 0.01$, $x_{GS} = 0.02$, $x_{GS} = 0.05$, $x_{GS} = 0.1$, $x_{GS} = 0.2$ eV $^{0.5}$). All other parameters were fixed ($b = 2$, $A = 0.0025$ eV). The photo-physical parameters for the ligand, L were fixed ($x_{CT} = 0.89$ eV $^{0.5}$, $\kappa = 1$, $\Delta G_{CTGS0} = 3.346$ eV, $\Delta G_{CTGSN} = 3.223$ eV). Plots were generated with the specified parameters using Equations (13) and (14) from the theory section (the range of x was from -1 to 2).

If we examine the free energy landscape for the quadratic model ($b = 2$), we note that the free energy difference between the successive minima of the microstate free energy surfaces is (Equation (20)),

$$G_{GSi}(min) - G_{GSi+1}(min) = \Delta G_{GSi+1,i} = A(2i + 1) \quad (20)$$

From Marcus theory, the activation energy between successive microstates is (in the quadratic FEL model) given by Equation (21),

$$\Delta G_{GSi+1,i}^{act} = \frac{(A(2i + 1) + RE)^2}{4RE} \quad (21)$$

By inspection of Equation (20), we see that in the quadratic FEL model, the successive free energy differences between the microstate minimum increase with the i^{th} microstate, and consequently the Marcus activation energy between microstates is also no longer independent of the microstate (Equation (21)). According to the quadratic FEL model (Equation 21), the activation barrier between successive microstates increases with increasing i , while the activation barriers decrease in the opposite direction.

To explore the effect of different numbers of microstates on the REES curves, we fixed all the parameters in the protein–ligand FEL and ligand L (parameters were $b = 2$, $A = 0.0025$, $x_{GS} = 0.15$) and varied N from 5, 10, 15, and 20 microstates. The influence of the number of microstates, N , on the REES curves is depicted in Figure 4. An increase in the number of microstates, N , is accompanied by a substantial change in shape of the REES curves. In general, both the position of the mid-point and the slope of the curve at the mid-point changed with N . Increasing N decreased the slope near the transition mid-point and increased the excitation wavelength at which the mid-point transition occurred.

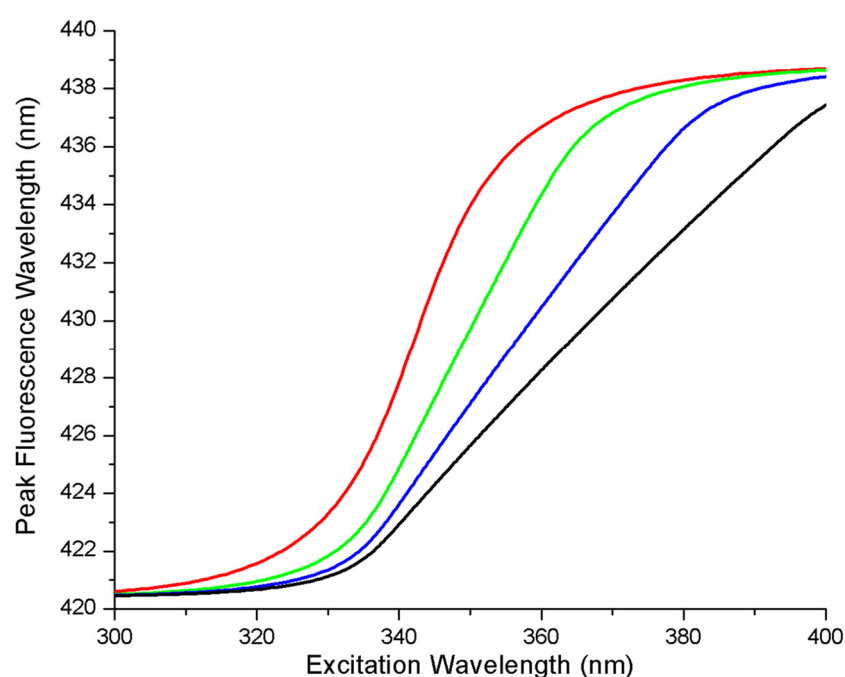


Figure 4. REES curves for a quadratic energy landscape ($b = 2$) as a function of the number of microstates, N . Plots from left to right represent the influence of increasing the number of microstates, N , on the REES curves ($N = 5$, $N = 10$, $N = 15$, $N = 20$). All other parameters were fixed ($b = 2$, $A = 0.0025$ eV, $x_{GS} = 0.15$). The photo-physical parameters for the ligand, L, were fixed ($x_{CT} = 0.89$, $\kappa = 1$, $\Delta G_{CTGS0} = 3.346$ eV, $\Delta G_{CTGSN} = 3.223$ eV). Plots were generated with the specified parameters using Equations (13) and (14) from the theory section (the range of x was from -1 to 2).

The parameters chosen in these simulations were in the range of what we might expect for intermediate states in protein–ligand complexes. The free energy difference between the most stable state in the protein–ligand complex and the least stable bound state

in the protein ligand complex was 10A (for $N = 10$ and for the linear model, $A = 0.0025$ – 0.05 eV). This corresponds to a free energy value range from 0.025 to 0.5 eV or 2.5–50 KJ/mol or about 0.6–12 kcal/mol ($2 k_B T$ to $20 k_B T$ (k_B is Boltzmann constant)). By way of comparison, the total free energy between the lowest energy bound state and a totally-free ligand state between $6.9 k_B T$ and $20.7 k_B T$ corresponds to an equilibrium dissociation constant of between 1 mM and 1 nM (i.e., the range of equilibrium constants for the majority of protein–ligand interactions).

3.2. Application to REES Data from Kinase Inhibitor–Kinase Complexes

To provide a concrete example of the application of the theoretical models to real data, we refer to recent published data from the anilino-quinazoline small molecule kinase inhibitor (AG1478) in complex with a kinase (MAPK14 or p38 α kinase) [17]. AG1478 is an intrinsically-fluorescent kinase inhibitor and exhibits environmentally-sensitive fluorescence with a Stokes shift as large as 100 nm in highly polar solvents [18]. Therefore, AG1478 is a suitable ligand for REES studies. Spectroscopic studies on the AG1478-p38 α kinase complex revealed a progressive red-shift in the emission spectrum with increasing excitation wavelength, consistent with a REES-type effect. The magnitude of the total red-edge shift was also largely independent of temperature (in the range 283–313 K) [17], which is an indication [19] that the motions of the environment are restricted on the nanosecond time-scale. In this case, we can apply the theory outlined here to analyze the REES data. Figure 5 depicts the REES curve as a solid line represented by the Boltzmann sigmoid fitting function used in the publication [17]. Individual datapoints corresponding to the REES data are indicated by the open circles in the inset to Figure 5. To fit this data to our thermodynamic model, we first optimized the photo-physical parameters of the AG1478 ligand in our model to provide a good fit to the extremes of the REES curve (i.e., to match the initial and final emission values in the REES curve). In our analysis, we then fixed the values for the ligand photo-physics (see Table 1 for parameter values). We then chose the form of the FEL function (linear or quadratic or cubic) and then performed a least-squares fit to the experimental curve by adjusting two parameters, the energy parameter A and the microstate spacing parameter x_{GS} . The number of microstates was initially fixed to $N = 10$ per simulations in the previous section. Table 2 outlines the results of each analysis, together with the sum of squares for the fit. Plots of the simulated REES curves are shown in the inset of Figure 5 to afford visual comparison. The linear model provided a reasonable fit with a SLS value of 53.5. The cubic model also provided a reasonable fit to the data with a SLS value of 204. However, the quadratic model outperformed both models with a SLS value of 3.5 (i.e., by a factor of 10 improvement). All models had the same number of adjustable parameters; therefore, we rejected the linear and cubic models, in favor of the quadratic model.

Table 1. Fit photophysical parameters for ligand from red-edge excitation spectroscopy (REES) curve. (Reorg denoted reorganization energy, other abbreviations/symbols defined in main text)

Ligand	Protein	Type FEL	ΔG_{CTGS0} (eV)	ΔG_{CTGSN} (eV)	x_{CT} (eV) ^{0.5}	Reorg (eV)
AG1478	P38alpha	Quadratic	3.347	3.223	0.891	0.397

Table 2. Fit parameters for REES curve according to different thermodynamic models. (SLSQ refers to the sum of least-squares, a measure of the quality of the model fit to the experiment).

Ligand	Protein	Type FEL	A (eV)	x_{GS} (eV) ^{0.5}	ΔG_{GS90} (eV)	SLSQ
AG1478	P38alpha	Linear	0.0083	0.078	0.074	53
AG1478	P38alpha	Quadratic	0.00178	0.153	0.144	4
AG1478	P38alpha	Cubic	0.00049	0.305	0.357	205

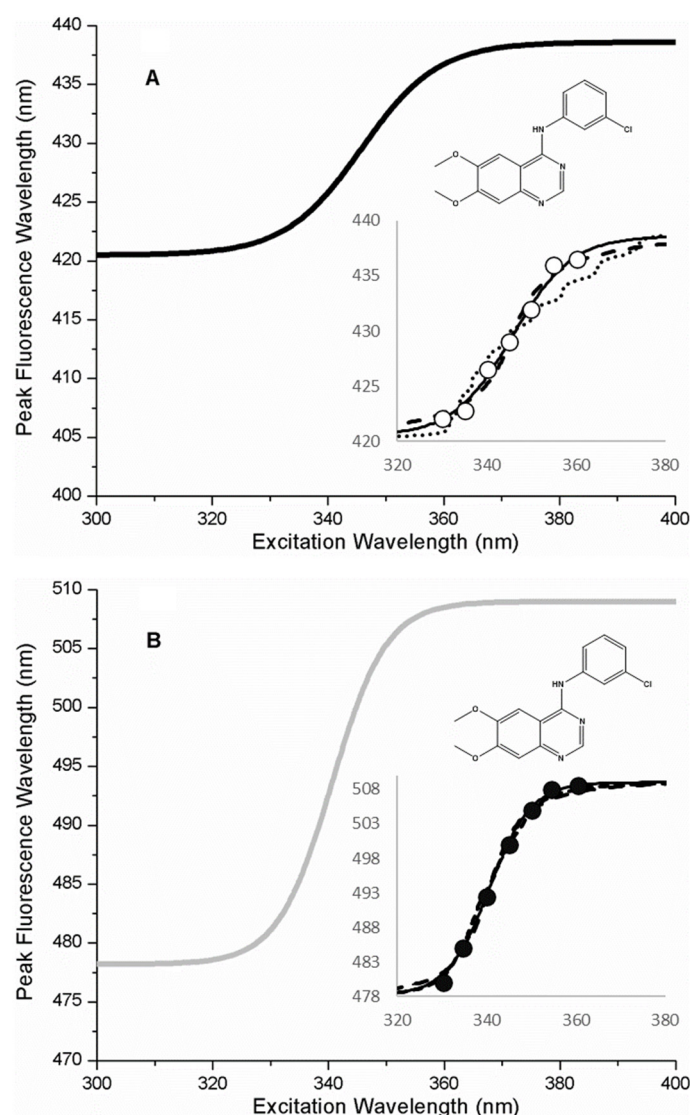


Figure 5. REES curves for kinase inhibitor–kinase complexes. **(A)** p38 α kinase-AG1478 complex (sigmoidal fit, black symbols). Inset: REES data (open circles), fit to quadratic FEL model (solid line), fit to linear FEL model (dashed line), fit to cubic FEL model (dotted line). **(B)** APH2 aminoglycosidase-AG1478 complex (sigmoidal fit, grey line). Inset: REES data (filled circles), fit to quadratic model (solid line), fit to linear model (dashed line), fit to cubic model (dotted line). REES data (peak emission wavelength as a function of excitation wavelength) were from [17]. Structure of the ligand, AG1478, is indicated in the figure.

In fitting the REES curve for the AG1478-p38 α kinase complex, we initially assumed that the number of microstates was $N = 10$. We repeated the analysis of the REES curve for different values of N for the quadratic FEL model as described above, (range of $N = 3$ to 20). The goodness-of-fit parameter as a function of the number of microstates is depicted in Figure 6a. As can be seen in Figure 6a, the goodness-of-fit is dependent on the number of microstates in the model, with $N = 10$ providing the best fit to the data.

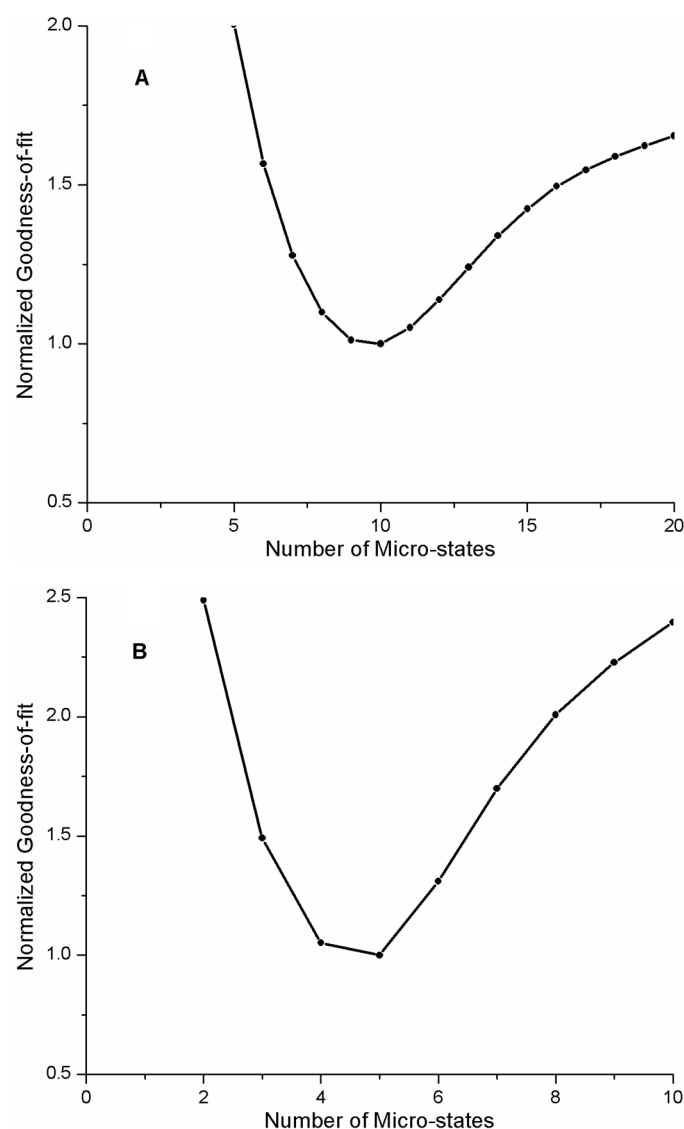


Figure 6. Goodness-of-fit as a function of number of microstates, N . (A) p38 α kinase-AG1478 complex. (B) APH2 aminoglycosidase-AG1478 complex.

An important concern is whether by pre-selecting a quadratic FEL, we are biasing the fitting in some way. Therefore, we repeated the analysis of the p38 α -kinase-AG1478 REES data, allowing all the parameters of the FEL to vary. The goodness-of-fit parameter versus number of microstates revealed a clear minimum (near $N = 9$ – 10) with the FEL exponent close to 2 ($b = 1.94$). To examine the robustness of our fitting procedures and model in more detail, we simulated REES data, fit the simulated REES curve to a sigmoidal function (as we do for experimental data), then fit the resulting sigmoid to a FEL model. Table 3 compares the input parameters with output parameters for FELs over an energy range of 0.04–0.24 eV. It is noted that the output parameters were within 15% of the input parameters.

Table 3. Input versus output for selected REES and free energy landscape (FEL) models.

INPUT			OUTPUT		
ΔG_{GSN0} (eV)	N	b	ΔG_{GSN0} (eV)	N	b
0.041	10	2	0.047	8,9	2.1
0.081	10	2	0.080	9,10	1.9
0.162	10	2	0.150	9,10	1.9
0.240	10	2	0.210	9,10	2.3

A plot of the free energy landscape of the protein–ligand complex AG1478-p38 α kinase is shown in Figure 7a, depicting the relative free energies of the 10 protein–ligand microstates as a function of the generalized protein–ligand coordinate. According to the quadratic FEL model and the fit parameters, the free energy difference between the most stable bound state and the least stable bound state in the AG1478-p38 α kinase complex was -0.144 eV (or -5.6 k $_B$ T or -14.4 kJ/mol or -3.5 kcal/mol). The reported inhibition constant for the AG1478-p38 α kinase complex was 0.560 μ M (20). A 0.6 μ M equilibrium dissociation constant corresponded to a free energy difference between bound and free AG1478 of -14.4 k $_B$ T. According to our analysis, the intermediate states of AG1478-p38 α kinase could contribute as much as $\sim 40\%$ ($= 5.6 / 14.4$) to the total free energy of binding.

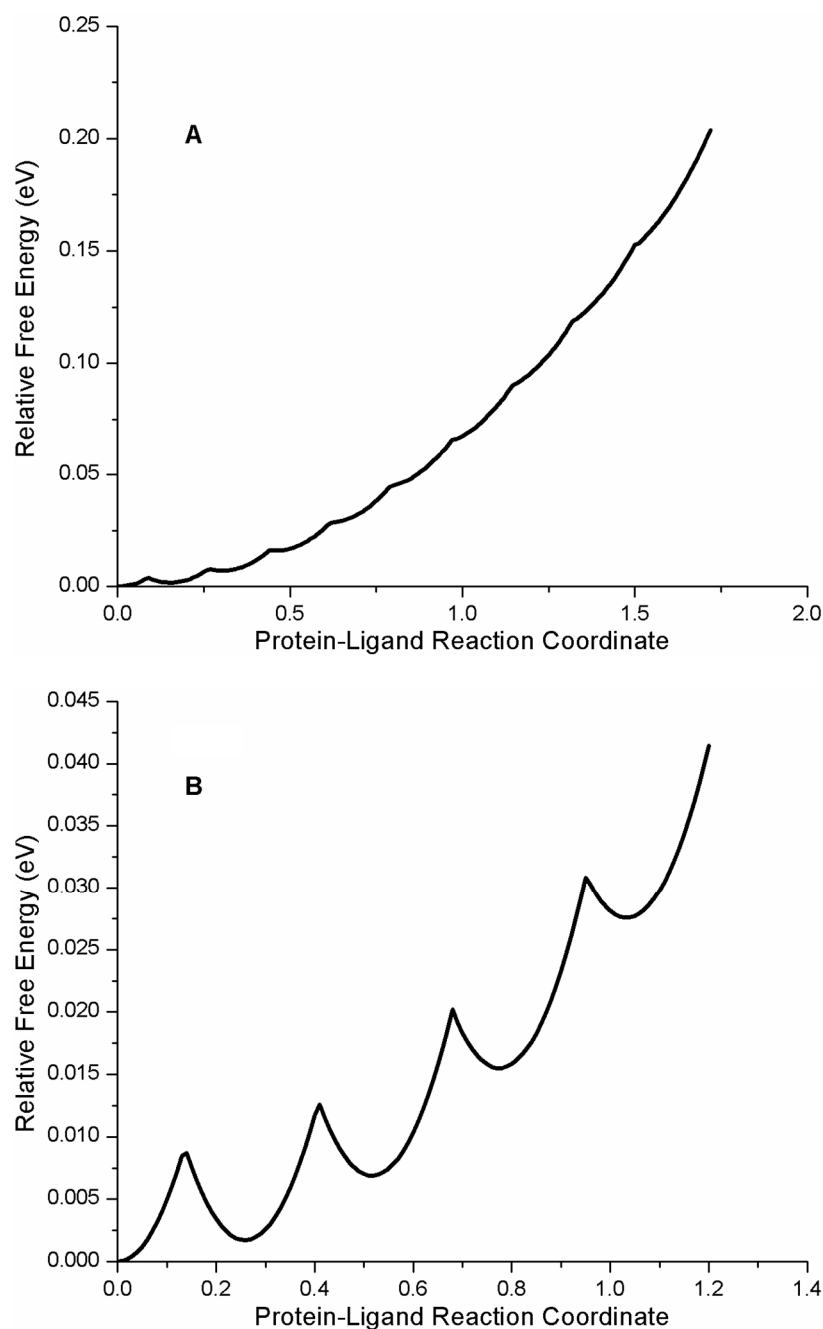


Figure 7. Free energy landscapes for AG1478-p38 α kinase and AG1478-APH2 derived from REES data. Vertical axes represents the free energy difference (relative to zero for the most stable state) in units of eV and horizontal axis is the protein–ligand reaction coordinate, x . (A) FEL for AG1478-p38 α kinase. (B) FEL for AG1478-APH2.

To test this idea further, we also analyzed the published REES data from AG1478 bound to another kinase, aminoglycophosphokinase (APH) [17] which is a threonine-serine kinase derived from bacteria, but has a similar fold to eukaryotic protein kinases. A plot of the REES curve for the AG1478-APH complex is depicted in Figure 5b, REES data in the Figure 5b inset, photophysical parameters in Table 4, and analysis of different models collected in Table 5. We found the best fit was to a quadratic model (Table 4) for the FEL surface with five microstates (i.e., $N = 5$, Figure 6b). From the model fit, the free energy difference between the most stable state and the least stable state in the AG1478-APH kinase complex was -0.03 eV or approximately -1 k_BT.

Table 4. Fit photophysical parameters for ligand from the REES curve. (Reorg denoted reorganization energy, other parameters/symbols are defined in the main text)

Ligand	Protein	Type FEL	ΔG_{CTGS0} (eV)	ΔG_{CTGSN} (eV)	χ_{CT} (eV) ^{0.5}	Reorg (eV)
AG1478	APH	Quadratic	3.172	3.015	1.076	0.579

Table 5. Fit parameters for REES curve according to different thermodynamic models. (SLSQ refers to the sum of least-squares, a measure of the quality of the model fit to the experiment).

Ligand	Protein	Type FEL	A (eV)	χ_{GS} (eV) ^{0.5}	ΔG_{GS40} (eV)	SLSQ
AG1478	APH	Linear	0.00558	0.207	0.022	58
AG1478	APH	Quadratic	0.00172	0.258	0.028	6.01
AG1478	APH	Cubic	0.000403	0.268	0.026	55

A comparison of the free energy landscapes for the two kinase-AG1478 complexes is shown in Figure 7 (generated using Equation (7)). It is notable that as a function of the protein–ligand reaction coordinate (x in Equations (1)–(14)), the p38 α -AG1478 free energy profile was steeper while the APH-AG1478 the free energy profile was shallower. From a physics analogy of a Harmonic spring, one would deduce that it is harder (i.e., costs more energy) to displace AG1478 from its most stable bound state in the p38 α kinase complex along the reaction coordinate (to its least stable bound state) than to displace the most-stable bound state in the corresponding APH-AG1478 complex because the AG1478-p38 α kinase complex interaction has a larger effective spring constant. This qualitatively agrees with the relative inhibition constants of the two proteins. The inhibition dissociation constant for p38 α kinase-AG1478 was $K_i(\text{p38}\alpha \text{ kinase}) = 0.560$ μM and represents “tighter” binding than for APH-AG1478 interaction, with an inhibition dissociation constant of $K_i(\text{APH}) = 15.6$ μM (21).

At equilibrium, the total binding free energy for the protein–ligand interaction ($\Delta G_{\text{protein-ligand}}$) is the sum of the free energy for ligand association with protein (ΔG_{assoc}) to a surface state (or encounter complex) and a free energy for ligand intrusion ($\Delta G_{\text{intrusion}}$) from a surface-bound state to the interior-bound state. If we equate ($\Delta G_{\text{intrusion}}$) with the free energy difference between the most-stable bound state and the least-stable bound state, then this quantity may be extracted from the protein–ligand FELs derived from REES data. From the FELs for the AG1478 complexes with APH and p38 α kinase, we estimated ($\Delta G_{\text{intrusion}}$)_{APH-AG1478} = -0.7 kcal/mol and ($\Delta G_{\text{intrusion}}$)_{p38 α -AG1478} = -3.3 kcal/mol. The total free energies for binding (from the inhibition constants) were ($\Delta G_{\text{protein-ligand}}$)_{APH-AG1478} = -6.5 kcal/mol for the APH-AG1478 interaction and ($\Delta G_{\text{protein-ligand}}$)_{p38 α -AG1478} = -8.3 kcal/mol for the p38 α -AG1478 interaction. From $\Delta G_{\text{intrusion}}$ and $\Delta G_{\text{protein-ligand}}$, we estimated (ΔG_{assoc})_{APH-AG1478} = -5.8 kcal/mol and (ΔG_{assoc})_{p38 α -AG1478} = -5.0 kcal/mol. Comparing the differences between the two proteins, we can see that the difference in total free energies for binding (($\delta \Delta G_{\text{protein-ligand}}$)_{p38 α /APH} = 1.84 kcal/mol) was more than accounted for by the difference in the free energies for intrusion (($\delta \Delta G_{\text{intrusion}}$)_{p38 α /APH} = 2.6 kcal/mol). In contrast, the difference

in free energies for surface association ($(\delta\Delta G_{\text{assoc}})_{\text{p38}\alpha/\text{APH}} = -0.8$ kcal/mol) makes a comparatively smaller contribution. Thus, the energetics of the ligand-bound states of the protein–ligand complexes is a differentiating factor in explaining the different inhibition constants of AG1478 for two different kinases.

The protein–ligand free energy landscapes, derived from the REES data, also provide estimates of activation barriers for transitions between the different ligand–protein microstates (along the reaction coordinate, x). As eluded to above, for the quadratic FEL, the activation barrier is related to the free energy difference and the reorganization energy between microstates (see Equation 20). For the APH-AG1478 complex, the largest activation barrier between the 4th and 5th microstate was 0.016 eV. For the p38 α kinase-AG1478 complex, the largest activation barrier between successive microstates was between the 9th and 10th microstates and was 0.031 eV. Thus for both complexes, the successive barrier heights were not significantly greater than the thermal energy (i.e., $k_B T = 0.025$ eV at 298K). From the model, one would also predict that the total magnitude of the REES effect is not very dependent on temperature, which agrees qualitatively with published observations over the temperature range 283–308 K [17].

Our focus thus far has been on the properties of the FEL (i.e., in the electronic ground states of the ligand in the microstates defined by the protein–ligand interaction). The extrema of the REES curves and associated analysis provide useful information on the CT states, and by inference, on the types of environments encountered by the ligand in the protein microstates (i.e., polarity of the amino acids and/or relative hydration). The relative displacement of the CT state relative to the GS free energy surfaces, x_{CT} , indicates the change in structure of the ligand (and possibly environment) between the GS and CT states. From this displacement x_{CT} , we can determine the Marcus reorganization energy for the CT to GS transition, equivalent to half the Stokes shift. Comparing the x_{CT} values for the two protein–ligand complexes (c/f Table 1 and Table 3), the displacement associated with the CT–GS transition in AG1478-APH ($x_{\text{CT}} = 1.1$) was slightly larger than the corresponding quantity in the AG1478-p38 α complex ($x_{\text{CT}} = 0.9$). Consequentially, the reorganization energy for the CT–GS transition in AG1478-APH was 0.6 eV and for AG1478-p38 α , it was 0.4 eV. The reorganization energy consists of two parts, the inner-sphere contribution (from changes in the geometry of the AG1478 and tightly-associated molecules), and an outer-sphere part, due to changes in dipolar reorganization of the environment around the AG1478. The inner-sphere term can be determined from the fluorescence of AG1478 in a non-polar solvent (where the outer-sphere contribution due to solvent dipolar reorganization is zero). From the published emission of AG1478 in non-polar solvents (toluene and dioxane), we estimated the inner-sphere reorganization energy in AG1478 to be 0.4 eV [18]. Thus, in the AG1478-p38 α kinase complex and the AG1478-APH complex, the inner-sphere reorganization of the ligand is the dominant contribution to the reorganization energy. This suggests that the dipolar reorganization of the protein environment during the excited state lifetime is restricted in the AG1478-APH complex and absent (or undetected) in the AG1478-p38 α kinase complexes. The free energy gap between CT and GS provide valuable information on the stabilization of the CT state relative to the GS by the protein environment and water. In the AG1478-APH complex, the CT states were more stabilized with respect to the GS ($\Delta G_{\text{CTGS0}} = 3.18$ eV, $\Delta G_{\text{CTGSN}} = 3.02$ eV) than the corresponding AG1478-p38 α kinase complex ($\Delta G_{\text{CTGS0}} = 3.35$ eV, $\Delta G_{\text{CTGSN}} = 3.22$ eV). These observations suggest that AG1478 experiences a more polar environment along its microstate trajectory in the APH protein compared to the p38 α kinase protein. It is tempting to speculate that the higher affinity of AG1478 for p38 α than APH is due to the more hydrophobic binding environments in p38 α , perhaps contributed in part by the nature and number of water molecules and the amino acids comprising the ligand binding pocket. However, the polarity gradients along the microstate trajectories ($\Delta G_{\text{CTGSN}} - \Delta G_{\text{CTGS0}} = -0.16$ eV (APH); $\Delta G_{\text{CTGSN}} - \Delta G_{\text{CTGS0}} = -0.13$ eV (p38 α kinase)) were similar in direction and magnitude in both proteins. This may imply that least populated ligand states are more

hydrated than the most populated ligand state, a view that is congruent with the idea that ligands become more hydrated as they leave the protein.

4. Discussion

The goal of this work was to link the properties of the ligand–protein free energy landscape to spectroscopic properties of a fluorescent ligand, namely the red-edge excitation shift. Our model enabled us to determine the number and energetics (i.e., populations) of intermediate bound states as well as the associated environments.

There are several assumptions underlying this work. First, we assumed that a thermodynamic model would provide an appropriate framework to link spectroscopy with populations and energetics. The assumption holds, provided that protein–ligand intermediates (microstates) are at thermal equilibrium in the ground electronic state. Spectroscopic measurements of proteins and ligands in closed cuvettes and under constant temperature regulation ensure that any changes in temperature or concentration of reagents are too small during the total measurement time to invalidate this assumption.

The second assumption is that each microstate retains its identity during the process of photon absorption and fluorescence. In other words, exchange between microstates is slower than the rate of fluorescence emission. Dynamic averaging between microstates would lead to a reduction in the REES effect, a result that would underestimate the number of distinct states and invalidate the derived thermodynamic parameters. Dynamic averaging can be evaluated by changing the lifetime of the probe or changing the dynamics of the protein. For example, a decrease in the red-edge effect upon heating indicates that environmental motions are on a time scale comparable to fluorescence. A temperature-independent REES (provided the protein is still in its native state) is one hallmark of restricted environment motion on the fluorescence timescale [19]. For the ligand–protein complexes studied in this paper, REES was found to be temperature independent and so this assumption likely holds in the present case. For cases where dynamic averaging is suspected, a different approach to the one posited here is required.

Our model for the free energy landscape emphasizes a small (i.e., 5–20 microstates) number of discrete protein–ligand microstates. This model is in accordance with recent molecular dynamics simulations showing hidden poses of ligands bound at various locations to a protein, along a protein–ligand binding trajectory. For the interaction of the PP1 drug with c-Src kinase, the Sugita lab identified one distinct fully-bound ligand pose as well as four semi-bound, four intermediate, and four encounter complexes [4]. While for the release of Dasatinib from c-Src, the Berne laboratory identified six states corresponding to five ligand-bound poses as well as an unbound state [21]. The number of intermediate bound states for AG1478 with APH was found to be five in the current work, all very close in energy (i.e., within thermal energy) to the most stable bound state. This analysis compared remarkably well with the number of ligand poses described in the x-ray derived structure of the APH–AG1478 complex. Between 9–10 bound states were found for AG1478 with p38 α kinase from our model. Docking studies with a hydrophobic ligand found 19 potential binding poses in p38 α kinase [23]. We wish to stress that the number of intermediates obtained is model-dependent (not absolute) and in our model, it is limited by the number of spectroscopically-distinguishable classes of states that we can determine. Thus the number of intermediate states extracted from the analysis here may be different to those reported by other means.

One possible interpretation of our results is that as the ligand proceeds along the protein–ligand trajectory away from the most stable binding configuration toward the protein surface, it encounters more polar or hydrophilic environments that stabilize the CT state with respect to the GS state. Our reaction coordinate then might correspond to the (relative) displacement in the center of masses of the protein and the ligand. This coincides with the view that ligands become more hydrated as they gradually leave the protein. However, our approach did not allow us to determine the kinetic path of ligand binding (or release), only the relative energetics and environments of these different states.

The free energy values obtained in the current study can be compared with values obtained by other labs for kinase–drug interactions. For the interaction of a PP1 drug with c-Src kinase ($IC_{50} = 0.17\mu M$), the Sugita lab [4] used molecular dynamics simulations and arrived at a value for the free energy for intrusion of $\Delta G_{intrusion} = -4.1$ kcal/mol. This value is similar to the experimental value obtained in the present study of $\Delta G_{intrusion} = -3.5$ kcal/mol for the p38 α –AG1478 complex, which also has a sub-micromolar inhibition constant. Sugita estimated the free energy change associated with the encounter complex step to be $\Delta G_{association} = -5$ kcal/mol for PP1–Src kinase, which is also close to our estimated value of $(\Delta G_{assoc})_{p38\alpha-AG1478} = -5.0$ kcal/mol for the p38 α –AG1478 association. The Kern lab [24] has revealed that the physical association step of the drug Glevec with Abl is associated with a free energy change of about -6.5 kcal/mol, while there was an additional free energy of intrusion (associated with a conformational change) step of -4.5 kcal/mol. Collectively, these results suggest that the free energy for encounter complex formation accounts for about -5 – -6.5 kcal/mol of the total binding free energy. Thus, high affinity binding, in these cases at least, is largely dictated by additional steps after encounter-complex formation.

The quadratic model was the preferred FEL model (compared with linear or cubic) for the two protein–ligand complexes tested (free floating the FEL exponent also resulted in values close to two). The quadratic model also agreed quite well with the published free energy landscape derived from the molecular dynamics simulations for PP1 and c-Src kinase [4] near the minimum of the free energy landscape and in the region of the bound and semi-bound ligand poses. Free energy functions derived from molecular dynamics studies revealed that the free energy was approximately quadratic as a function of the radial distance of the drug from the binding site [4]. Since we expect microstates experiencing more polar environments to be further from the binding site toward the protein surface, a quadratic dependence of the free energy on microstate number N seems a reasonable first order approximation. Note that we have attempted more complex free energy functionals such as the Morse potential function, but these have more parameters with no improvement in the fit quality. Our mathematical model for the free energy landscape function—linear, quadratic-, or cubic- is of course somewhat simplistic. Real free energy landscapes are complex and can be very bumpy, containing many hills and valleys. As we learn about real protein–ligand FELs, more complex functions may be used to describe the energetics of the protein–ligand microstates and can be easily incorporated into the current model.

One of the main results of this paper is to provide a theoretical shape for the REES effect curve under different idealized conditions. Our simulations suggest that the full REES curve is a sigmoidal-like function. This is expected as binding models derived from thermodynamic considerations for many states are also often sigmoidal or Hill-type functions [26]. In the REES experiment, changing the excitation wavelength to longer values is similar to adding a ligand (or denaturant) to a protein (i.e., shifting the (observed) population distribution). An important difference with REES is that the population of states is not perturbed during the experiment, only the observation window. However, there are conditions where a full sigmoidal curve may not be obtained from a REES experiment and only some part of a sigmoidal curve may be detected. Thus, an apparent linear, exponential, or Gaussian REES curve may be measured under certain circumstances. The model presented here allows simulation of under what conditions these different types of REES curves may be observed. However, we caution that extracting parameters from REES curves that are not strictly sigmoidal may be problematic (i.e., solving the inverse problem in this case). This issue is similar to trying to find binding constants or denaturation free energies from incomplete binding or denaturation data.

Perhaps the main virtue of the approach presented here is to allow for a comparison of REES curves from different protein–ligand complexes or under different conditions to be placed under the same conceptual framework. For example, applying the model to real REES data allowed us to compare the properties and stabilities of the intermediate ligand-

bound states in two different proteins bound to an intrinsically fluorescent small molecule inhibitor. Encouragingly, the relative energetics deduced from our model correlated well with previously published relative inhibition constants as well as related thermodynamic data from other kinase–inhibitor complexes.

The REES approach may not necessarily be restricted to proteins with well-defined tertiary folds and secondary structural elements. REES is an attractive approach for examining dynamics and structure of disordered states, molten globule states, and intrinsically-disordered proteins, which may exhibit multiple conformational states along a rich conformational protein landscape [7,8,10].

5. Conclusions

The present statistical mechanical model allows for the interpretation of red-edge excitation shift spectroscopy (REES) curves from protein–ligand complexes under different conditions. The approach leads to a novel way to study ligand–protein binding from the determination of less populated, intermediate, and transient bound states of protein–ligand complexes, which has been experimentally challenging. The model revealed the relationship between the progressive emission red shift with red-edge excitation and the protein–ligand free energy landscape, and relates thermodynamic and photo-physical parameters to the fluorescent ligand. We applied the theoretical model to available red-edge excitation shift data from the small molecule inhibitor–kinase complexes AG1478–APH and AG1478–p38 α . It was discovered that a quadratic protein–ligand free energy landscape model provided a good description of the data from both inhibitor–kinase complexes. The derived thermodynamic parameters allowed dissection of the energetic contribution of intermediate bound states to inhibitor–kinase interactions. A further study using quantum mechanical calculations in this direction is under investigation.

Abbreviations

GS	Ground-state
CT	Charge-transfer state
FEL	Free energy landscape

Author Contributions: A.H.A.C. and F.W. conceived and supervised the research. M.L.K. carried out the research. A.H.A.C., F.W. and M.L.K. edited and contributed to the preparation of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: M.L.K. acknowledges a scholarship in the form of the Swinburne University Postgraduate Research Award (SUPRA).

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: M.L.K. acknowledges the Swinburne University Postgraduate Research Award (SUPRA).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Janin, J. Protein–protein recognition. *Prog. Biophys. Mol. Biol.* **1995**, *64*, 145–166, doi:10.1016/S0079-6107(96)00001-6.
2. Blundell, T.L. Protein crystallography and drug discovery: Recollections of knowledge exchange between academia and industry. *IUCrJ* **2017**, *4*(Pt 4), 308–321.
3. Copeland, R.A. Conformational adaption in drug–target interactions. *Future Med.Chem.* **2011**, *3*, 1491–1501.
4. Re, S.; Oshima, H.; Kasahara, K.; Kamiya, M.; Sugita, Y. Encounter complexes and hidden poses of kinase–inhibitor binding on the free-energy landscape. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 18404–18409.
5. Iwahara, J.; Clore, G.M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **2006**, *440*, 1227–1230.

6. Kinz-Thompson, C.D.; Gonzalez, R.L., Jr. smFRET studies of the 'encounter' complexes and subsequent intermediate states that regulate the selectivity of ligand binding. *FEBS Lett.* **2014**, *588*, 3526–3538.
7. Chattopadhyay, A.; Halder, S. Dynamic Insight into Protein Structure Utilizing Red Edge Excitation Shift. *Acc. Chem. Res.* **2014**, *47*, 12–19.
8. Catci, D.A.; Amos, H.E.; Yang, Y.; van den Elsen, J.M.; Pudney, C.R. The red edge excitation shift phenomenon can be used to unmask protein structural ensembles: Implications for NEMO-ubiquitin interactions. *FEBS J.* **2016**, *283*, 2272–2284.
9. Jones, H.B.; Wells, S.A.; Prentice, E.J.; Kwok, A.; Liang, L.L.; Arcus, V.L.; Pudney, C.R. A complete thermodynamic analysis of enzyme turnover links the free energy landscape to enzyme catalysis. *FEBS J.* **2017**, *284*, 2829–2841.
10. Guha, S.; Rawat, S.S.; Chattopadhyay, A.; Bhattacharyya, B. Tubulin conformation and dynamics: A red edge excitation shift study. *Biochemistry* **1996**, *35*, 13426–13433.
11. Demchenko, A.P. The red-edge effects: 30 years of exploration. *Luminescence* **2002**, *17*, 19–42.
12. Weber, G.; Shinitzky, M. Failure of Energy Transfer between Identical Aromatic Molecules on Excitation at Long Wave Edge of Absorption Spectrum. *Proc. Natl. Acad. Sci. USA* **1970**, *65*, 823.
13. Galley, W.C.; Purkey, R.M. Role of Heterogeneity of Solvation Site in Electronic Spectra in Solution. *Proc. Natl. Acad. Sci. USA* **1970**, *67*, 1116.
14. Rubinov, A.; Tomin, V. Bathochromic luminescence in solutions of organic dyes at low temperatures. *Opt. Spektrosk.* **1970**, *29*, 578.
15. Marcus, R.A. Relation between charge transfer absorption and fluorescence spectra and the inverted region. *J. Phys. Chem.* **1989**, *93*, 3078–3086.
16. Marcus, R.A. Exchange reactions and electron transfer reactions including isotopic exchange. Theory of oxidation-reduction reactions involving electron transfer. Part 4. — A statistical-mechanical basis for treating contributions from solvent, ligands, and inert salt. *Discuss. Faraday Soc.* **1960**, *29*, 21–31.
17. Khattab, M.; Wang, F.; Clayton, A.H.A. Conformational Plasticity in Tyrosine Kinase Inhibitor–Kinase Interactions Revealed with Fluorescence Spectroscopy and Theoretical Calculations. *J. Phys. Chem. B* **2018**, *122*, 4667–4679.
18. Khattab, M.; Wang, F.; Clayton, A.H.A. UV–Vis spectroscopy and solvatochromism of the tyrosine kinase inhibitor AG-1478. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2016**, *164*, 128–132.
19. Demchenko, A.P. Fluorescence Molecular Relaxation Studies of Protein Dynamics—The Probe Binding-Site of Melittin Is Rigid on the Nanosecond Time Scale. *FEBS Lett.* **1985**, *182*, 99–102.
20. Cumming, J.G.; McKenzie, C.L.; Bowden, S.G.; Campbell, D.; Masters, D.J.; Breed, J.; Jewsbury, P.J. Novel, potent and selective anilinoquinazoline and anilinopyrimidine inhibitors of p38 MAP kinase. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5389–5394, doi:10.1016/j.bmcl.2004.08.007.
21. Shakya, T.; Stogios, P.J.; Waglechner, N.; Evdokimova, E.; Ejim, L.; Blanchard, J.E.; McArthur, A.G.; Savchenko, A.; Wright, G.D. A small molecule discrimination map of the antibiotic resistance kinome. *Chem. Biol.* **2011**, *18*, 1591–1601, doi:10.1016/j.chem-biol.2011.10.018.
22. Tiwary, P.; Mondal, J.; Berne, B.J. How and when does an anticancer drug leave its binding site? *Sci. Adv.* **2017**, *3*, e1700014.
23. Yang, L.; Sun, X.; Ye, Y.; Lu, Y.; Zuo, J.; Liu, W.; Elcock, A.; Zhu, S. p38 α Mitogen-Activated Protein Kinase Is a Druggable Target in Pancreatic Adenocarcinoma. *Front. Oncol.* **2019**, *9*, 1294, doi:10.3389/fonc.2019.01294. eCollection 2019.
24. Agafonov, R.V.; Wilson, C.; Otten, R.; Buosi, V.; Kern, D. Energetic dissection of Gleevec's selectivity towards human tyrosine kinases. *Nat. Struct. Mol. Biol.* **2014**, *21*, 848–853, doi:10.1038/nsmb.2891.
25. Johnson, N.P.; Ji, H.; Steinberg, T.H.; von Hippel, P.H.; Marcus, A.H. Sequence-Dependent Conformational Heterogeneity and Proton-Transfer Reactivity of the Fluorescent Guanine Analogue 6-Methyl Isoxanthopterin (6-MI) in DNA. *J. Phys. Chem. B.* **2015**, *119*, 12798–12807, doi:10.1021/acs.jpbc.5b06361.
26. Garcia, H.G.; Kondev, J.; Orme, N.; Theriot, J.A.; Phillips, R. Thermodynamics of biological processes. *Methods Enzymol.* **2011**, *492*, 27–59.