Dynamics studies of luciferase using elastic network model: how the sequence distribution of luciferase determines its color

Yi Mao

Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA

To whom correspondence should be addressed.
E-mail: ymao@nimbios.org

Received June 9, 2010; revised October 6, 2010; accepted November 3, 2010

Edited by Bruce Tidor

Luciferase exhibits a broad range of emitting frequencies. Light emission from the bioluminescence of luciferase makes it an excellent tool for monitoring gene expression, thus the control of its bioluminescence color has great bio-analytical applications. Here I use an elastic network model to examine how the sequence distribution of luciferase is related to bioluminescence multicolor emission. Based on the open and closed forms of crystal structures for luciferase, several computational analysis tools are applied to characterize the functionally relevant dynamical features within luciferase, and probe the dynamical mechanisms underlying the interactions between luciferin (a light-emitting substrate) and luciferase. Perturbation-based correlation analysis is used to identify hot-spot residues that are dynamically coupled to the active site of luciferase, and the results show that the sequence region of subdomain B of luciferase is largely responsible for determining the emitting color of bioluminescence. Moreover, the mode decomposition analysis reveals that the lowest frequency mode is the major contributor to the dynamical couplings between the hot-spot residues and the binding site in luciferase.

Keywords: bioluminescence/correlation/emitting color/luciferase/network model

Introduction

Bioluminescence refers to the light emitted from a living organism due to a chemical reaction that converts chemical energy into optical energy. Firefly luciferase is one of the best characterized bioluminescence systems (Deluca, 1976), and has been extensively studied for more than 50 years (Fraga, 2008). Luciferase is the enzyme that catalyzes the light-emitting reactions of luciferin in bioluminescent organisms. Typically these reactions involve two-step oxidations of luciferin in the presence of ATP, Mg2+ and oxygen molecule to produce light, oxyluciferin, CO2 and AMP (Deluca and McElroy, 1974).

Bioluminescence has received enormous research attention in recent years, largely due to its application as a molecular imaging tool for studying gene expression and gene regulation in real time (Gould and Subramani, 1988; Gloskrezutz and Schenborn, 1997; Greer and Szalay, 2002). Imaging techniques that rely on bioluminescence (bioluminescence imaging) are based on the sensitive detection of visible light produced during bioluminescence when the enzyme is expressed in vivo as a molecular reporter (Contag and Bacmann, 2002). These powerful techniques enable the study of ongoing biological processes in living organisms without high cost and invasion (Naylor, 1999; Rettig et al., 2006). They have been applied successfully to mouse models of lung inflammation/injury, bacterial pneumonia, and tumor growth and metastasis (Squirrel et al., 2002; Wang et al., 2003).

Despite the great success of bioluminescence imaging in a variety of applications, there is still room for improvement in advancing the technology. One important issue is the range of the light frequency emitted by luciferase. More specifically, luciferase-catalyzed oxidation of luciferin results in bioluminescence over a range of wavelengths from 530 to 640 nm with a peak at 562 nm (Wang et al., 2003). In contrast, tissue attenuation of optical wavelength photons is minimized in red to near-infrared wavelengths (600–900 nm; Contag and Bacmann, 2002). As a result, luciferase could be more easily used as a gene reporter in animal tissues if the luciferase enzyme could be tuned such that it caused emission of a larger percentage of bioluminescence photos at slightly longer wavelengths. Thus, the challenge posed for protein engineering is how to design a luciferase with the desired emitting frequency (i.e. red-color luciferase).

Experimentally most efforts aimed towards finding red-emitting luciferase variants were directed at examining the optical properties of the various mutants of the luciferase created through random mutagenesis (Eames et al., 1999; Branchini et al., 2004b). Mutant luciferase proteins were used to locate several ‘hot’ residues—amino acids whose mutation led to an observable shift in the bioluminescence spectrum of the luciferase-catalyzed reaction (Kajiyma and Nakano, 1991). Most notably, the mutation S286N of Japanese firefly (Luciola cruciata) produced a bioluminescence maximum at 605 nm, which is a red shift of 45 nm from the wild-type (Nakatsu et al., 2006). In addition, there were also extensive chimeric studies on the bioluminescence properties of luciferase systems. Through these studies it was found that chimeric luciferases derived from pairs of different color-emitting homologous luciferases exhibited either the bioluminescence activity found in one or the other parent, or lost bioluminescent activity altogether (Ohmiya et al., 1996; Viviani et al., 2004). Based on these outcomes it was hypothesized that one or two sequence regions, corresponding to distinct subdomains in luciferase, solely or jointly determined the emitting color of the luciferase-catalyzed reaction. Despite the wealth of information on the spectral properties of mutant and chimeric luciferases, the
relationship between the wide range of bioluminescent colors and the sequence of luciferase remains an open question. Obviously, the answer to this question would be of immense value to bioluminescence probe design.

Molecular mechanism of color variation in firefly bioluminescence has been intensively studied both experimentally and theoretically (McCapra et al., 1994; Orlova et al., 2003; Branchini et al., 2004a; Nakatani et al., 2007; Liu et al., 2008; Navizet et al., 2010). Earlier studies were focused on how the structure of the light emitter (oxyluciferin) affects the emitting frequency. The color shift in the bioluminescence was attributed either to the C2–C2' angle rotation of oxyluciferin (McCapra et al., 1994), or to the resonance structure of the excited oxyluciferin in keto and enolate forms (Branchini et al., 2004a). Recently high-resolution crystal structures of wild-type and mutant S286N in complex with an intermediate analog became available (Nakatsu et al., 2006). The crystallographic data showed that the substrates in both complexes have similar structures (the root mean-square deviations between the two structures are 0.32 Å for Cα atoms and 0.04 Å for ligand molecules, respectively), and the protein conformational change induced by the mutation was localized on the side chain of Ile288. This experimental evidence inspired a 'microenvironment mechanism' for the color variation in firefly bioluminescence, which stated that the key to color variation in luciferase is the polar environment of oxyluciferin (Orlova et al., 2003; Nakatani et al., 2007; Liu et al., 2008; Navizet et al., 2010). However, those mechanistic investigations on multicolor bioluminescence were focused on the static structure and local interaction analysis (structure-based approach), and did not take into consideration luciferase dynamics. It has been argued that large conformational changes of luciferase play a role in the reaction catalysis (McElroy et al., 1967; Gulick, 2009). Luciferase belongs to the superfamily of acyl-adenylate forming and thioester-forming enzymes, which adopt two different conformations to catalyze two half reactions; and it uses a large-scale C-domain rotation for the different partial reactions (Gulick et al., 2004; Branchini et al., 2005). The role of conformational dynamics of luciferase such as the large-scale C-domain rearrangement in bioluminescence color determination is yet to be elucidated.

It has been established in recent years that the study of protein enzymatic activities needs going beyond the view of static protein structures, and dynamical properties of the enzymes play an important role in the catalytic processes (Miller and Agard, 1999; Eisenmesser et al., 2005; Henzler-Wildman et al., 2007, Kong and Karplus, 2009). Dynamic-based computational approaches have been used to characterize the flow of energy and information in protein function such as allostery and signal transduction. Using molecular dynamic simulations, a general framework was formulated for computing frequency-dependent communication maps between protein residues (Leitner, 2009), and two continuous interaction pathways of PDZ domains were discovered by a hierarchical clustering analysis of the residue correlation matrix (Kong and Karplus, 2009). Moreover, because global motions of proteins, one of the major functionally important dynamical features, are insensitive to structural details, a variety of coarse-grained models have been developed to explore the relation between protein dynamics and function more efficiently. Particularly elastic network models in conjunction with normal modes and perturbation analyses examine protein motions at residue-level detail, provide insight into enzyme specificity and regulation, signal transduction mechanism of large membrane proteins and identify residues with key roles in protein functions (Attilgan et al., 2001; Delarue and Sanjeouand, 2002; Zheng and Doniach, 2003; Ma et al., 2005). The findings from these studies also suggest that the sequence distribution of the functionally important residues is closely tied to the dynamical motions of the enzyme.

Here I study the sequence-induced color variation of luciferase by examining its dynamics. For the sake of computational efficiency, an elastic network model with a coarse-grained representation of protein and a harmonic potential describing the interactions of its components is used (Tirion, 1996). Despite its simplicity, this type of model has, in the past, been highly successful in capturing the collective dynamics of proteins (Zheng et al., 2006; Yang et al., 2007; Eyal and Bahar, 2008). More specifically with respect to luciferase, network model-based dynamical analyses have generated new insights into the mechanism of anesthetic inhibition of luciferase enzymatic activities (Szarecka et al., 2007). For the study of inhibition of luciferase, the protein backbone network was perturbed by the introduction of additional nodes representing anesthetics. The study showed that the molecular mechanism of the anesthetic inhibition of the firefly luciferase can be related to the effects of the anesthetics on luciferase's global motion, and the major action of anesthetics is to disrupt the modes of motion essential to luciferase functions. Inspired by its success, here I apply the elastic network modeling in conjunction with various analysis tools to investigate how the sequence distribution of luciferase is related to bioluminescence multicolor emission. Again the link between luciferase dynamics and function is used as a platform to examine the protein's intrinsic susceptibility to mutation and consequent functional effects, and more importantly it allows probing the sequence basis for bioluminescence color variation without a full mechanistic understanding of color variation. The paper is organized as follows: first, the propensity of individual residues to influence the bioluminescence spectrum is evaluated in terms of the residues dynamical coupling with the active site of luciferase; secondly, a mode decomposition analysis is performed to find the dominant modes that account for the dynamical correlations between the binding site and the residues identified as functionally important. The results support the hypothesis that a limited number of global motions are sufficient to capture the functional dynamics of the protein, and they are found to be in good agreement with previous experimental findings. Finally suggestions for future experiments on the mechanistic issue of bioluminescence color determination are proposed.

Materials and Methods

X-ray structures of luciferase

The conformations of an isolated luciferase from Photinus pyralis (Conti et al., 1996) and the ligand-bound (AMP and oxyluciferin) from L. cruciata (Nakatsu et al., 2006) were chosen to represent luciferase in open and closed forms (PDB codes 1LCI and 2D1R, respectively).

Y. Mao
Elastic network model

The elastic network model simplifies the native structure of a protein by only retaining the coordinates of the Cα atoms. In this way each residue is represented by its Cα atom and becomes a node in the network model. The model is homogeneous (all the nodes are identical), under the assumption that the sequence information is already encoded in the native structure. Any pair of nodes within a given distance $R_c$ is assigned a harmonic potential. The total energy in the elastic network representation of a protein is

$$E_{\text{network}} = \frac{1}{2} \sum_{d_{ij} < R_c} \gamma (d_{ij} - d_{ij}^0)^2,$$

where $\gamma$ is the uniform force constant, $d_{ij}$ is the distance between the $i$th and $j$th Cα atoms and $d_{ij}^0$ is the corresponding distance in the native structure. The summation runs over all the pairs within the cutoff distance $R_c$. In this study $R_c$ is set to be 10 Å, and the results were not sensitive to the choice of $R_c$. The equilibrium correlations between the fluctuations $\Delta R_i$ and $\Delta R_j$ are given as:

$$\langle \Delta R_i \Delta R_j \rangle = \frac{k_BT}{\gamma} \left[ H^{-1}_{aa} + H^{-1}_{bb} + H^{-1}_{cc} \right],$$

where $H^{-1}$ is the pseudoinverse of the Hessian matrix of the elastic model with six degrees of a rigid body motion removed. $a$, $b$, $c$ and $a'$, $b'$, $c'$ are indices for $x$, $y$ and $z$ components of the Cartesian coordinates of the Cα atoms of residues $i$ and $j$, respectively. The correlation is further normalized as:

$$C_{ij} = \frac{\langle \Delta R_i \Delta R_j \rangle}{\left( \langle \Delta R_i^2 \rangle \langle \Delta R_j^2 \rangle \right)^{1/2}}.$$

Despite the drastic simplification in this model, it has been shown to successfully reproduce the B-factor in X-ray diffraction experiments and NMR experiments, locate the crucial dynamical correlations attributable to enzyme specificity, and discover the important pathways in signaling proteins (Zheng and Doniach, 2003; Zheng et al., 2006; Yang et al., 2007; Eyal and Bahar, 2008). More detailed information on this model can be found elsewhere (Tirion, 1996; Hinsen, 1998; Attilgan et al., 2001). The online servers [http://iggmltest.ccb.pitt.edu/cgi-bin/anm/anm1.cgi (Eyal et al., 2006) and http://enm.lobos.nih.gov/ (Zheng and Doniach, 2003)] were used for part of the calculations. For the ligand-bound protein, only the Cα atoms of the protein are represented by the elastic network model and the ligand is not incorporated into the model.

Perturbation method

Recently, a structural perturbation method was introduced in the context of elastic network modeling to decipher the network of residues that are dynamically coupled with a functional site in biological nanomachines (Zheng Brooks, 2005; Zheng et al., 2007, 2005).

In this method, the energetic perturbation $\delta E_t$ is introduced at residue $i$ in the form of a change to the force constants of those springs connecting residue $i$ to its neighbors (residue $k$) within $R_c$.

$$\delta E_t = \frac{\delta \gamma}{2} \sum_{d_{ik} < R_c} (d_{ik} - d_{ik}^0)^2.$$

The response at a subset of residues (denoted as residue $j$) to the perturbation then is measured in terms of the changes in the mean-square fluctuations $\delta \langle \Delta R_j \Delta R_j \rangle$. The more details of the method can be found elsewhere (Zheng et al., 2007).

In practice, the residues of the active site in a protein are residues $j$, and all the other residues in the protein are residues $i$. The top 10% among residues $i$ with the largest percentage change in the fluctuations of residues $j$ are selected as dynamically important residues. These residues are the so-called hot-spot residues, whose mutations may have the potential to impact the protein’s function (Zheng et al., 2007). In the literature a very different definition of hot residues (kinetically hot residues) in proteins can be found (Demirel et al., 1998; Rader et al., 2004). Kinetically hot residues are identified by their crucial role played in protein folding and stability. They are generally located in the tightly packed core of protein (Rader et al., 2004). Dynamically they are involved in the fast motions of proteins, and often found to be highly conserved. The main distinction between kinetically hot residues in the literature and dynamically hot-spot residues in this study lies in the different results of mutagenesis experiments at these residues. Mutations of these kinetically hot residues lead to protein misfolding, and mutations of dynamically hot-spot residues result in the modulation of protein functions while keeping the protein structure intact. In addition, dynamically hot-spot residues are not necessarily to be highly conserved, as shown below. Their locations are not restrained to the protein core and can be found on the protein surface and loops.

Results

Normal mode analysis: the capture of global conformational change

The catalytic role of firefly luciferase in the bioluminescence oxidation of luciferin is accomplished through conformational changes from the open to closed forms. Here we choose the high-resolution X-ray structure of the product-bound (AMP and oxyluciferase) luciferase from the Japanese L. cruciata as the closed form (Nakatsu et al., 2006), and the substrate-free luciferase from the American firefly P. pyralis as the open form (Conti et al., 1996) (Fig. 1). These two luciferase from two different species share 67% sequence identity and >80% sequence similarity. They are considered to be highly homologous, and thus compose a good open/closed pair for structural study of enzymatic activities. The conformational change from the open to closed form of luciferase involves a large downward movement of the C-terminal towards the N-terminal, with the N-terminal remaining largely fixed. It is accomplished primarily through a hinged, rigid-body motion of the C-terminal. It has been reported elsewhere that the tendency of the small C-terminal to move towards the large N-terminal is triggered by ATP binding, which facilitates the interactions between the substrate and the protein (Gullick et al., 2003).
We first examine the conformational change that occurs in the crystal structures of luciferase on going from the open to closed form, then compare this change with the elastic network model derived normal modes of the open form luciferase. The results are computed as the overlaps between the two (Zheng and Doniach, 2003), illustrated in Fig. 2. Apart from the rest of the eigenmodes, the two lowest frequency modes of the open form luciferase correlate strongly with the direction of the luciferase conformational change. This is consistent with the findings from previous studies that the functionally important conformational changes are well described by a limited number of low-frequency normal modes (Zheng et al., 2006). The eigenvectors of these two lowest frequency modes are plotted in Fig. 3, where the displacements of individual C\(_\alpha\) atoms that occur as a result of the conformational change are shown to be very close to these eigenvectors. Both modes involve large range motions of the C-terminal: the first and slowest mode of the isolated luciferase is dominated by a hinge bending and colliding type of motion from the C-terminal, while the second slowest mode of the isolated luciferase is the rotation of the C-terminal with respect to the N-terminal (Fig. 4). The validation of applying the elastic modeling method to the mechanistic study of luciferase is confirmed by the adequate capture of the open to closed conformational change by these two lowest frequency modes. Similar results and conclusions can be obtained by projecting the conformational changes onto the normal modes of the closed form of luciferase.

**Perturbation-based correlation analysis: identification of hot-spot residues**

The relationship between protein sequence and function is normally probed by identifying functionally important residues, which is accomplished experimentally by point mutation experiments. In this study, the sequence basis for bioluminescence color determination is examined by locating those residues, which may change the emission frequency by mutations. A computational analog to point mutation experiments in the context of protein elastic network modeling is a perturbation method (Zheng et al., 2007). The idea is to introduce a change in the force constant of a residue, and then measure the subsequent changes in the fluctuations of the residues selected to represent a functional site. Those
residues generating the largest fluctuation changes of the selected residues are considered to be the hot-spot residues. Because the hot-spot residues are dynamically strongly coupled with the functional site, their mutations have functional consequences. Thus, they are functionally important. Here this method is used to identify these hot-spot residues of luciferase whose mutations have the potential of modulating its function and influencing the frequency of emitting light.

Luciferase’s binding pocket consists of $\alpha 8$ (residues 248–260), $\beta 12$ (286–289), $\beta 13$ (313–316), $\beta 14$ (339–342), $\beta 15$ (351–353) and a loop (343–350) (Nakatsu et al., 2006). In this hydrophobic binding pocket, the closest contact between the enzyme luciferase and the ligand oxyluciferin is more than 3 Å away. In this study, five residues in the binding pocket with the close distances to the substrate (Phe249, Ile288, Arg339, Gly341 and Ala350) are chosen to represent the binding site, and the resulting hot-spot residues based on the crystal structure of the closed form of luciferase are listed in Table I (the crystal structure of luciferase in the open form generated very similar results). The hot-spot residues cover a large span of spatial distances from the binding site: some residues are located as far as 16 Å away from oxyluciferin (Fig. 5). These residues constitute an extensive network that operates cooperatively to ensure the functionality of luciferase. Overall the distribution of hot-spot residues is concentrated in the sequence region of residues 228–356. There are a few clusters of residues formed around the five residues at the binding site: residues 252–255 around residue 249; residues 286–294 around residue 288; residues 338–342 around residue 339 and residues 350–356 around residue 350. This suggests that local interactions play a significant role in deciding the coupling between the binding site and the residue with the potential to impact on the spectral properties of the luciferase. On the other hand, the isolated hot-spot residues tend to locate rather far away from the binding site. Their interactions with the binding site must therefore be achieved through the chain of interacting residues. These indirect couplings are further analyzed by mode decomposition of the correlations as described below.

Extensive mutagenesis experiments have been carried out on firefly luciferase (Kajiyama and Nakano, 1991; Branchini et al., 2003; Tafreshi et al., 2008). With respect to $L. cruci- ate$, five single mutants were discovered to possess a broad range of spectral shifts (Kajiyama and Nakano, 1991). The mutation sites in these five mutants are all located far from the binding site. Three of them are discovered by the perturbation analysis performed here. While the following detailed atomistic level analysis has revealed the long-range

<table>
<thead>
<tr>
<th>Hot-spot residues</th>
</tr>
</thead>
</table>

Fig. 4. Two lowest frequency modes of the open form of luciferase. (a) The corresponding secondary structure for the elastic network model. (b) The lowest frequency mode as represented by the elastic network model. (c) The second lowest frequency mode as represented by the elastic network model.

Fig. 5. The hot-spot residues identified by the perturbation-based correlation analysis (shown in the ball-and-stick representation) are plotted on the conformation of the substrate-bound luciferase (PDB code 2D1R).
interaction chain that links these residues to the binding site, it is important to note that such detailed information is absent in the coarse-grained network model used for the current analysis. Nevertheless, this simple network model is still able to capture the long-range correlations sufficiently to accurately predict these hot-spot residues, and there is a consistency between the interaction analysis at the atomic level and the correlation analysis performed at the coarse-gained level.

(i) Ser286: The mutant S286N luciferase emits red light at 607 nm. Ser286 is located about $\sim 9 \, \text{Å}$ away from oxyluciferin. As part of $\beta 12$, it mainly interacts with the binding site through Ile 288. The side chain of Ile288 is in close contact with the benzothiazole ring of oxyluciferin. The crystal structure of luciferase mutant (S286N) intermediate has been resolved at a high resolution (Nakatsu et al., 2006). It shows the replacement of Ser286 by Asn only induced local structural reorganization, mostly affecting the interactions between the side chain of Ile288 and the benzothiazole ring of oxyluciferin. In particular the close van der Waals contacts between three carbon atoms of the side chain of Ile288 and the benzothiazole ring no longer exist in the mutant structure, due to the hydrogen-bonding network around Asn286. It has been hypothesized that these changes in the interactions between the side chain of Ile288 and the benzothiazole ring are responsible for the observed red shift in the spectrum of this mutant protein (Nakatsu et al., 2006).

(ii) Gly326: The mutant G326S produces red light with a peak at 609 nm. The location of Gly326 is at least $\sim 15 \, \text{Å}$ from oxyluciferin. This long-distance signal a chain of interactions that allows the propagation of the mutation effect to the binding site. Further correlation analysis reveals that the coupling between Gly326 and the binding site is mainly attributed to the correlation between Gly326 and Arg339. The distance between Gly326 and Arg339 ($\sim 9 \, \text{Å}$) does not suggest a close contact between them. Instead their coupling is mediated through Val338. There are close van der Waals contacts between the carbonyl group of Gly326 and one side-chain carbon atom of Val338. Therefore, the link between Gly326 and the binding site is through a network of residues consisting of Arg339 and Val338.

(iii) Val239: The mutant V239I has a moderate shift in the spectrum with an emission frequency at 565 nm. The location of Val239 is found to correlate well with Ile288, which is located at a distance of at least 6 Å from Val239. In the meantime, Val239 interacts with another predominant mutation site Ser286, a neighbor of Ile288, by forming van der Waals contacts at the backbone between the oxygen atom of Val239 and the carbon atom of Ser286. It is thus likely that the impact of Val239 on oxyluciferin goes through Ile288 and Ser286.

The hot-spot residues in Table 1 consist of a small fraction of all residues ($\sim 10\%$). It is estimated that on average at least 10% of the total number of residues play a significant role in protein function. So far less than 10 residues of luciferase whose mutations could lead to bioluminescence spectral shift have been discovered by random mutagenesis experiments, and no complete mutagenesis experiments have been performed on luciferase yet. The list of hot-spot residues discovered by the elastic network analysis provides a useful starting point for experimentalists to explore the sequence space of luciferase in search for other mutants, which produce different color of light.

In summary, the perturbation method in elastic network modeling is able to identify some amino acid residues in luciferase whose substitutions produce the change in its bioluminescence color. This method based on the correlation analysis provides a good tool for detecting the long-range networks that link the remote mutation sites to the functional site. A mode decomposition technique is applied to further analyze the nature of such long-range correlations in luciferase. It examines the contribution of each eigenmode to the total correlation between two residues. The contribution from the $k$th mode to the correlation between the $i$th and $j$th residues are represented as $v_{ik}v_{jk}/\lambda_k$, where $\lambda_k$ is the eigenvalue of mode $k$, and $v_{ik}$ and $v_{jk}$ are the $i$th and $j$th components of the eigenvector of mode $k$, respectively. These contributions are normalized by dividing them by a constant $\sqrt{(\sum_{m < M} (\sum_{jm} (v_{jm}/\lambda_m)^2))}$, where $M$ is the cutoff mode (Ma et al., 2005). Figure 6 shows the normalized contributions from eigenmodes up to the mode 100, to the pairwise correlations between the selected hot-spot residues (Ser286, Gly326 and Val239) and the binding site residues. It is clear that the low-frequency modes make dominant contributions to the couplings between those dynamically important residues and the binding site. Moreover, higher modes are involved in the couplings, but their contributions vary with the hot-spot residues, and modes with an index greater than 40 have little impact. The notion that low-frequency motions correlate with protein function has been discussed and supported by many studies (Ma et al., 2005; Zheng et al., 2006; Eyal and Bahar, 2008), and the results presented here provide further evidence in this regard. Furthermore, the mode contributions to the correlations of these residue pairs in Fig. 6 are all dominated by the lowest frequency mode, suggesting that the long-range correlations between the hot-spot residues and the binding site are mostly through indirect couplings in the collective mode of the lowest frequency. This finding is in line with the previous discovery that the slowest motions of luciferase make the predominant contributions to the protein’s enzymatic function (Szarecka et al., 2007). My study also demonstrates that the modes of motion accomplishing the large C-domain rearrangement and essential for luciferase function (Gulick, 2009) are involved in the color determination of luciferase as well.

**Conservation analysis**

It has been speculated that the residues, which are dynamically important and whose mutations lead to changes in enzyme function, are evolutionarily strongly conserved (Zheng et al., 2005). To test this idea, the conservation of the residues in luciferase is examined. A conservation score is assigned to every residue based on the multiple sequence alignments of homologous protein sequences found in the ConSurf server [http://consurf.tau.ac.il/ (Landau et al., 2005)]. The lowest (highest) conservation score represents the most (least) conserved residue position. The conservation score of the residues of luciferase range from $-1.39$ to $2.28$. The average conservation scores for the protein and
dynamically important residues found by the elastic network model are 0.055 and 0.033, respectively. The six experimentally verified dynamically important residues have a wide range of conservation scores: Ser286(0.91), Ile288(0.79), Gly326(0.52), Val239(0.24), His433(−0.24) and Pro452(−0.63). Since these six residues were all found to have a noticeable influence on the emitting frequency of luciferase (Kajiyama and Nakano, 1991), it is fair to say that so far there is no evidence to link residues influencing the emitting frequency of luciferase with high conservation. In luciferase, the most conserved residues are located at the joints between different structural domains, such as Ile444(−1.39) and His247(−1.39) at the helix/coil transition, and Arg439(−1.37) and Asp438(−1.31) at the hinge regions. In previous studies, large variations in the conservations of functionally important residues have been discovered and discussed as well (Zheng and Tekpinar, 2009). One possible explanation is that the residues that compose the networks or pathways in allosteric enzymes tend to have strong coevolving relationships, despite their weaker conservation. Supporting evidence for this hypothesis comes from the great success of statistical coupling analysis methods based on the correlated evolution pattern of pairwise amino acids in locating the long-range communication pathways of various proteins (Lockless and Ranganathan, 1999; Lee et al., 2008; Halabi et al., 2009).

Discussions

The identification of protein domains important for protein function has long been an active research area. Here the relationship between luciferase’s sequence and function is examined in terms of its dynamics. In a previous similar study (Szarecka et al., 2007), the lowest frequency modes of luciferase, dominated by the motions of the C-domain, were found to be the dynamical features with direct consequences on the enzymatic activities; and in particular, it was shown that disruption of these modes by means of ligand docking at the interface of N- and C-domains can lead to the loss of function of the protein. However, in this study the focus is not on the inhibition of luciferase function but on its modification. Instead questions involved include what aspects of luciferase’s dynamics are responsible for the variation in the emitting color of luciferase? What is the network of residues or protein domains that play a dominant role in determining the emitting color? Based on the correlation-based perturbation analysis of a coarse-grained elastic model, an overall picture of how certain residues and domains of luciferase influence the outcome of luciferase bioluminescence emerges.

Luciferase possesses a large N-domain (residues 1–436) and a small C-domain (residues 440–539), connected by a flexible linker (residues 437–439). The N-domain is further divided into subdomains A, B and C (Conti et al., 1996). The luciferin-binding site is located at a depression formed by the packing of subdomain B against the subdomains A and C. The functional implications of these structural domains are examined by the results from dynamical studies of luciferase.

Subdomain B: the major player in bioluminescence color determination

The perturbation method identifies the residues that have the closest ties to the binding site. Subdomain B (residues 228–356) turns out to be the sequence region with a high concentration of these hot-spot residues, and the structural explanation is that the binding site of luciferin is located in subdomain B, and subdomain B contains many important residues and motifs directly interacting with excited oxyluciferin. Presumably the interactions of luciferase and luciferin dictate the properties of bioluminescence, and a large part of those interactions occurs at or near the binding site.

The importance of this sequence region has been discovered through the extensive chimerical studies on several luciferases by different groups. These chimerical studies were carried out on the Hotaria parvula and Pyrococcus furiosus fireflies (yellow- and green-light emitting luciferases, respectively; Ohmiya et al., 1996). The emission spectra of the chimerical luciferases revealed that “the active sites for the color determination between green and yellow must be located in the fragment (208–318) of P. furiosus and H. parvula luciferase”. A similar chimerical study on Phrixotrix railroad worm luciferase concluded that residues 220–344 of the luciferase had a major effect on color determination (Viviani et al., 2004). It is likely that similarly significant sequence region exists for all proteins in the luciferase family, and that this region is typically located in the middle of the N-domain, which embraces the active site.

The lowest frequency mode: the major contributor to long-range correlations

Some hot-spot residues in subdomain B are not located in the direct vicinity of the active site of luciferase. The dynamical correlations between these sites and the active site of the protein are long range and occur through indirect couplings. Mode decomposition analysis of these long-range couplings (Fig. 6) reveals the dominance of a single mode (the lowest frequency mode). The simplicity revealed by this analysis offers a simple mechanical picture of the complicated long-range couplings, and confirms the critical role of the lowest frequency mode in luciferase enzymatic activities. This finding also adds further support to the hypothesis that large-scale motions discovered by normal mode analysis in elastic network modeling have functional implications. In addition, the network of residues involved in the inhibition of luciferase function (Szarecka et al., 2007) and the network of residues participating in the color-modification of luciferase are at different residue locations, but both networks communicate through the same low-frequency mode. This suggests that the functional dynamics of proteins captured by a simple mode can be manifested in different ways.

The hinge region

A small hinge region (residues 437–439) connects a large N-domain and a small C-domain in the luciferase protein, and affects the relative motion of these two domains. The binding of the light-emitting molecule (luciferin) to luciferase leads to the bending of the C-domain towards the N-domain along the hinge. Therefore, it is not surprising that perturbations at the hinge region can affect the binding of luciferin, and thereby disrupt the function of luciferase. This is exactly the finding from the anesthetic docking study of luciferase using network models (Szarecka et al., 2007). While the hinge region and the C-domain may be crucial for the binding of the light-emitting molecule (luciferin) to luciferase, my correlation analysis above suggests that the hinge
The role of luciferase dynamics in color determination

The success of applying dynamics-based approach to locating the hot-spot residues of luciferase validates the important role of luciferase dynamics played in the firefly bioluminescence reaction, consistent with the findings based on detailed kinetic and structural investigations of luciferase and other adenylating enzymes (Gulick, 2009). More specifically, large-scale C-domain movements are identified as the important dynamical features in luciferase color determination. So far, the conformational dynamics of luciferase has not received a lot of research attention in the mechanistic study of multicolor emission of bioluminescence. Hopefully my work will encourage the future investigations to study the issue from the perspective of dynamics. The connection between the mutational effects on the frequency of the emitting light and the global dynamical properties of luciferase is particularly worth pursuing. The differences in dynamical properties of the wild and mutant types of luciferase can be investigated by direct solution nuclear magnetic resonance measurements experimentally (Wolf-Watz et al., 2004) and atomic molecular dynamics simulations computationally, and the results will shed light on how residue-specific dynamical properties of luciferase are correlated with spectral shifts.

Acknowledgements

The author thanks Dr Sharon Bewick for proof reading the paper.

Funding

This work is partially supported by National Center for Supercomputing Applications (NCSA) under MCB100001.

References

Dynamics studies of luciferase using elastic network model