FPMOD: A Modeling Tool for Sampling the Conformational Space of Fusion Proteins

Jason Chiang, **Isaac Li**, **Elizabeth Pham**, **Kevin Truong***, Member, IEEE

Abstract— **Fusion proteins are an important class of proteins with diverse applications in biotechnology. They consist of 2 or more rigid domains joined by a flexible linker. Understanding the conformational space of fusion proteins conferred by the flexible linkers is important to predicting its behavior. In this paper, we introduce a modeling tool called FPMOD (Fusion Protein MODeller) which samples the conformational space of fusion proteins by treating all domains as rigid bodies and rotating each of them around their flexible linkers. As a demonstration, FPMOD was used to predict the fluorescence resonance energy transfer (FRET) efficiency of three different fusion protein biosensors. The simulation results of the FRET efficiency prediction were consistent with the *in vitro* experimental data, which verified that FPMOD is a valid tool to predicting the behavior of fusion proteins.**

I. Introduction

FUSION proteins consist of a tandem fusion of multiple domains. They are used in various applications such as drug discovery [1], fluorescence markers [2] and biosensors [3]. For example, by designing different fusion protein biosensors, multiple signaling pathways can be monitored simultaneously [4]. The conformational space of fusion proteins in presence of water is important to the understanding its behavior. Thus, we developed a modeling tool called FPMOD to sample the possible conformations of fusion proteins by treating each domain as a rigid body and rotating the domain with respect to their flexible linkers. Furthermore, FPMOD is capable of creating fusion protein models given solved atomic structures of individual protein domains. Several protein modeling tools are available for molecular mechanics, molecular dynamics, homology modeling, and protein structure visualization such as Biochemical Algorithms Library[5], Visual Molecular Dynamics [6], MODELLER [7], and DeepView Swiss PDB viewer [8], respectively. However, none easily samples the conformational space of fusion proteins.

The ability to sample fusion protein conformational space helps us understand the behavior of fluorescence resonance energy transfer (FRET) Ca^{2+} biosensors. FRET is a natural phenomenon which occurs via resonance between two fluorophores (one is called the donor and the other, the acceptor) with a spectral overlap between the donor emission and the acceptor excitation. The strength of FRET depends on the relative distance and orientation between the donor and acceptor. FRET Ca^{2+} biosensors are typically created by sandwiching a Ca^{2+}-sensing target protein with a cyan fluorescent protein (CFP) donor [9] and yellow fluorescent protein (YFP) acceptor [10]. Upon binding Ca^{2+}, the Ca^{2+} sensing component undergoes a conformational change that causes a change of FRET, which is correlated to a change in $[Ca^{2+}]$. In this paper, we first discuss the algorithms employed in FPMOD and then demonstrate its use in predicting the behaviors of 3 different FRET Ca^{2+} biosensors. Lastly, we verify the predicted behavior against *in vitro* experimental data.

II. Simulation Process

To sample the conformational space of fusion proteins given the atomic structure for each domain, FPMOD requires several features such as fusion protein construction and rigid-body domain rotation. Both of these features are discussed as well as a special case where the conformational space of a domain was greatly reduced by its neighboring domains. Lastly, we discuss the prediction of FRET efficiency of fusion protein biosensors.

Fusion Protein Construction and Sampling the Conformational Space

Given the solved atomic structure of each domain in the standard Protein Data Bank (PDB) format, fusion protein models are created by joining each domain with flexible linkers using three sub-routines: residue insertion, residue deletion, and domain fusion. The residue insertion sub-routine is mainly used for adding linker residues, whereas the residue deletion is necessary for protein truncation. Note that it is necessary to add linker residues because they are introduced during subcloning process or, optionally, they are added for increasing the conformational space. Note also that the residue deletion function is called when only a portion of the domain in a PDB file is used for protein fusion. After the linker residues are added and the deletion function is called, if necessary, the fusion proteins are then created by recursively joining two domains by the fusion sub-routine.

This work was supported by the Banting Foundation and the Natural Sciences and Engineering Research Council of Canada (NSERC) under Research Grant RG-PIN 276250.

JC, IL and EP are with the department of Electrical and Computer Engineering (ECE) and Institute of Biomaterials and Biomedical Engineering (IBBME), University of Toronto, Toronto, ON, Canada.

*KT is also with ECE and IBBME, University of Toronto, Toronto, ON, Canada (kevin.truong@utoronto.ca)

Once the fusion protein models are created, FPMOD then samples their conformational space. This is done by treating all domains as rigid bodies and rotating them around their flexible linkers. The flexible linkers are located between domains and do not have any secondary structure such as α-helices or β-strands. Torsion angles (also known as dihedral angles) are defined as relative angles between the two planes A-B-C and B-C-D where A, B, C, and D were adjacent atoms. There are three torsion angles in each residue (named Φ, ψ, and ω) (Fig. 1). During the residue rotation, for each flexible residue, we rotate all the atoms in the fusion protein preceding the N atom along the N-C_α bond by the torsion angle Φ and then rotate all the atoms after the C atom along the C_α –C bond by the angle ψ. Note that the angles Φ and ψ do not have any restriction (ranging from -180 to 180 degrees), while the angle ω was fixed at 180 degrees as this trans-configuration is most common in nature. The equations associated with this rotation operation are:

$$P_{updated} = [M] \otimes P_{current} + P_{ref} \quad (1)$$

$$[M] = \begin{bmatrix} A^2(1-\cos(\theta))+\cos(\theta) & AB(1-\cos(\theta))-C\sin(\theta) & AC(1-\cos(\theta))+B\sin(\theta) \\ AB(1-\cos(\theta))+C\sin(\theta) & B^2(1-\cos(\theta))+\cos(\theta) & BC(1-\cos(\theta))-A\sin(\theta) \\ AC(1-\cos(\theta))-B\sin(\theta) & BC(1-\cos(\theta))+A\sin(\theta) & C^2(1-\cos(\theta))+\cos(\theta) \end{bmatrix} \quad (2)$$

$P_{updated}$ and $P_{current}$ are the locations of atoms after and before the rotation; P_{ref} is the pivot point (N and C atoms for the N- C_α and C_α -C bond rotations, respectively); (A, B, C) are the (x, y, z) coordinates of the direction cosine of N- C_α and C_α -C bonds; θ is the torsion angle. All computations are calculated in the Cartesian coordinate system. After rotating around all the flexible residues, one possible conformation of the fusion protein model is created. The model is then saved in the PDB format and then the whole process is repeated until the number of models is representative of the conformational space of the fusion protein of interest (Fig. 2).

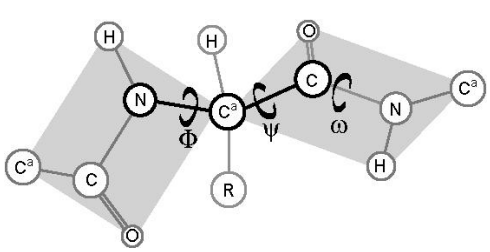

Figure 1 Cartoon diagram for the torsion angles Φ, ψ, and ω. The gray parallelograms are the rotation planes and the thick black bonds are the rotation axes.

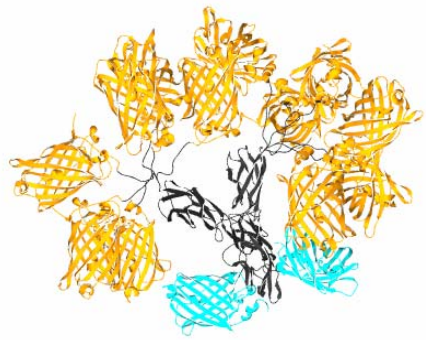

Figure 2 The possible conformations of YFP (in yellow) in the CEcad12Y biosensor with Ca^{2+}

B. Sandwiched Domain Rotation Algorithm

Here, we discuss a special case when the conformational space of a domain is greatly reduced because both of its terminals were locked by the neighboring domains. Frequently, fusion proteins are designed such that all the domains are in series and therefore the possible conformations can be sampled by the algorithm discussed in section 2-A. However, when two domains bind each other and form a complex, the possible conformations of the third domain will be greatly restricted if it is sandwiched in between the complex (Fig. 3a). To address this case, FPMOD follows the below algorithm:

1) Isolate the sandwiched domain along with all flexible residues and two residues from each of the two neighboring domains (called overlapping residues) (Fig. 3b). Those two overlapping residues are used for matching purpose.
2) Rotate the flexible residues and the overlapping residues as described in section 2-A.
3) Calculate the relative distance of the overlapping residues from the isolated domain and compare the distance with that from the overlapping residues in the domain complex, if a match is present, then a new model is created successfully.
4) Repeat steps 1 to 3.

Employing this algorithm, the conformational space of the sandwiched domain can be sampled.

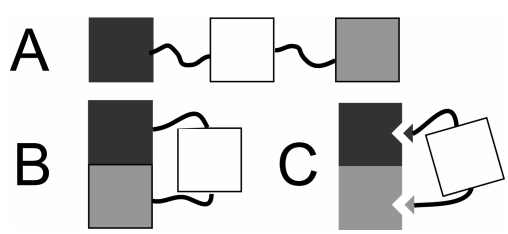

Figure 3 Cartoon diagram for sandwiched domain rotation algorithm. A) a regular fusion protein with all domains (boxes) in series; B) protein complex between two domains (dark and light gray boxes), sandwiched domain (white box) along with flexible residues (thick black lines) and overlapping residues (arrowheads); C) one possible conformation of the fusion protein

C. Fluorescence Resonance Energy Transfer (FRET) efficiency Prediction

FPMOD was used to predict the FRET efficiency (E%) changes of fluorescent protein Ca^{2+} biosensors. FRET efficiency is defined as the percentage of energy transfer between the donor-acceptor fluorophore pair and is a function of fluorophore pair distance (R), Förster distance factor (R_0), and the orientation factor (κ^2):

$$E\% = \frac{R_0^{\ 6}}{R_0^6 + R^6} \tag{3}$$

$$R_0 = 9.78\times10^3 \times \left(Q_d \kappa^2 n^{-4} J\right)^{\frac{1}{6}} \text{ Å} \tag{4}$$

$$\kappa^2 = \left[\sin(\theta_D)\sin(\theta_A)\cos(\alpha) - 2\cos(\theta_D)\cos(\theta_A)\right]^2 \tag{5}$$

Based on donor-acceptor fluorophore pair of the biosensor (in this case, CFP donor and YFP acceptor), several parameters had constant values: quantum yield (Q_d) = 0.42, refractive index (n) = 1.4, and overlap integral (J) = 1.4618e-13. Other parameters included the following: the angles between the acceptor/donor fluorophore dipoles and the joining vector (θ_A and θ_D respectively), and the angle between the fluorophore pair planes (α) (Fig. 4). Note that the common assumption of constant value κ^2 = ⅔ did not apply because the linkers within biosensors were not in isotropic motion after Ca^{2+} binding [11].

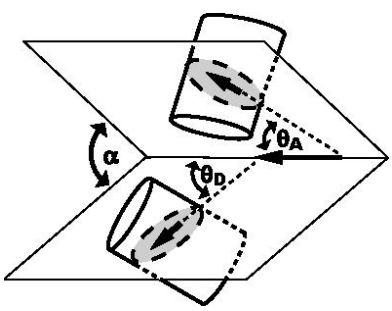

Figure 4 Illustration of θ_D, θ_A, and α that are required for the κ^2 calculation. The upper barrel indicates the acceptor and the lower barrel indicates the donor. θ_D and θ_A are the angles between the donor / acceptor dipoles (the arrows within the donor and the acceptor) and the joining vector (the arrow along the intersection line between two fluorophore dipole planes), respectively. α is the angle between the two fluorophore dipole planes.

FPMOD was used to sample the conformational space of the biosensor. The E% can, then, be calculated for each conformation using the equations (3)-(5). From the conformational space, a distribution of E% can be created to estimate the expected E% of the biosensor. Therefore, by sampling the possible conformations of a FRET Ca^{2+} protein biosensor before and after the Ca^{2+} induction, the change of E% can be predicted.

D. FRET Ca^{2+} Biosensor Demonstration and Result Comparison

To demonstrate the effectiveness of FPMOD in predicting E% changes, we designed three FRET Ca^{2+} biosensor models, simulated their E% changes due to Ca^{2+} induction, and compared the simulation results to the *in vitro* experimental data. We chose Epithelial Cadherin repeats 1-2 (Ecad12) [12] and Calmodulin (CaM) as the Ca^{2+} sensing components because their Ca^{2+} binding mechanisms are well studied. For Ecad12, the linker between the repeats is floppy when no Ca^{2+} is present. After Ca^{2+} induction, each linker binds 3 Ca^{2+} ions and adopts extended and rigid conformation. Subsequently, two Ecad12 molecules bind each other to form a stable assembly [13]. We designed an Ecad12 FRET Ca^{2+} biosensor (named CEcad12Y biosensor) by labeling CFP and YFP at the *N*- and *C*- terminals of the Ecad12 molecule, respectively. In contrast, upon binding to Ca^{2+}, CaM encompasses its target peptide CaM-dependent Kinase Kinase (CKKp) and forces CKKp to undergoing a conformational change. As a result, the CaM-CKKp complex was formed [14]. Based on the CaM-CKKp binding mechanism, we designed two biosensors by labeling the fluorophore pair at different locations. For one biosensor CFP and YFP were labeled at the *N*- and *C*- terminal of the CaM-CKKp fusion protein (named CCaMCKKpY), respectively. For the other biosensor, we changed the order of the domains to YFP-CKKp-CFP-CaM to simulate the conformational space of the sandwiched CFP (named YCKKpC-CaM) (Fig. 5).

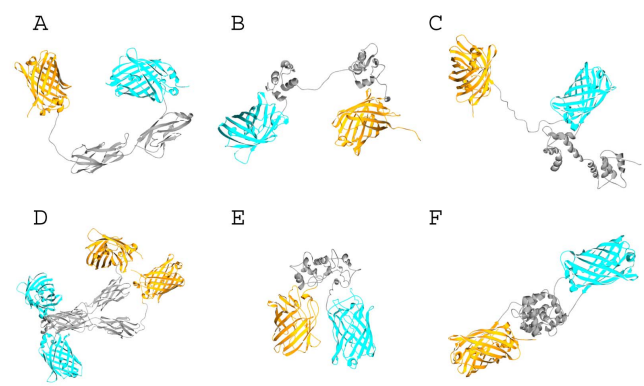

Figure 5 Cartoon diagram for possible conformations of the CEcad12Y, CCaMCKKpY, and YCKKpC-CaM biosensors under different conditions (in order). **A, B, C,** with no Ca^{2+}. **D, E, F**, with Ca^{2+}. For all figures, YFP is in yellow, CFP is in cyan, and Ca^{2+} sensing components are in gray.

From the simulation results, CEcad12Y and CCaMCKKpY biosensors showed increases in E% whereas YCKKpC-CaM biosensor showed a decrease upon binding Ca^{2+}. The simulation results were consistent with the *in vitro* experimental data. Using FPMOD, the E% for the (CEcad12Y, CCaMCKKpY, YCKKpC-CaM) biosensors before and after Ca^{2+} induction were (1.7%, 5.8%, 17.0%) and (3.1%, 16.5%, 12.3%), respectively (Table I). Therefore, the changes in E% during Ca^{2+} induction were (1.4%, 10.7%, -4.7%). Since from the *in vitro* experiment only the FRET emission ratio (the division of acceptor emission by donor emission after donor excitation) can be calculated, it was necessary to find the relationship between the E% and emission ratio. An increase of X% in FRET

efficiency means an X% energy loss from the donor and an X% energy gain to the acceptor. The change in FRET emission ratio, R, was estimated by the following equation:

$$R = \frac{(1+X\%)}{(1-X\%)} - 1 \qquad (6)$$

Substituting variable X with the changes in E% from the each biosensors, the corresponding changes in FRET emission ratios were (2.82%, 24.0%, -9.7%). To verify with the simulation results, we then created the physical FRET Ca^{2+} biosensors using the cassette technology introduced by Truong *et al* [15], and from the *in vitro* experimental data the (CEcad12Y, CCaMCKKpY, YCKKpC-CaM) biosensors had FRET emission ratio changes of (14.1%, 95.2%, -16.8%) during Ca^{2+} induction, respectively. Thus, the emission ratio changes were qualitatively consistent with the simulation results, which verified that the rigid body sampling of possible protein conformations using FPMOD was a valid approach to predict the direction of FRET changes of FRET Ca^{2+} biosensors.

	CEcad12Y	CCaMCKKpY	YCKKpC-CaM
no Ca^{2+}	1.7%	5.8%	17.0%
With Ca^{2+}	3.1%	16.5%	12.3%
E% change	1.4%	10.7%	-4.7%
R	2.82%	24.0%	-9.7%
in vitro	14.1%	95.2%	-16.8%

Table I. The E% with and without Ca^{2+}, E% changes, corresponding emission ratios (R), and in vitro experimental data for CEcad12Y, CCaMCKKpY, and YCKKpC-CaM biosensors.

III. Conclusion

None of the existing modeling tool easily samples the conformational space of fusion proteins. Thus, we developed a modeling tool called FPMOD which samples the possible conformations of fusion proteins using rigid-body domain rotation algorithms. FPMOD was then used to predict the E% changes of FRET Ca^{2+} biosensors. We discovered the simulation results were indeed consistent with *in vitro* experimental data. In the future, each fusion protein conformation will be evaluated based on free energy, where lower energy is favored. Therefore, the sampled conformational space would account for protein forces such as van der Waals and electrostatic forces.

ACKNOWLEDGEMENT

This work was supported by grants from the Canadian Foundation of Innovation (CFI) and the National Science and Engineering Research Council (NSERC).

REFERENCES

[1] G. Milligan, G. J. Feng, R. J. Ward, N. Sartania, D. Ramsay, A. J. McLean, and J. J. Carrillo, "G protein-coupled receptor fusion proteins in drug discovery," *Curr Pharm Des*, vol. 10, pp. 1989-2001, 2004.

[2] J. Wiedenmann, S. Ivanchenko, F. Oswald, F. Schmitt, C. Rocker, A. Salih, K. D. Spindler, and G. U. Nienhaus, "EosFP, a fluorescent marker protein with UV-inducible green-to-red fluorescence conversion," *Proc Natl Acad Sci U S A*, vol. 101, pp. 15905-10, 2004.

[3] J. J. Chiang and K. Truong, "Computational modeling of a new fluorescent biosensor for caspase proteolytic activity improves dynamic range," *Nanobioscience, IEEE Transactions on*, vol. 5, pp. 41-45, 2006.

[4] J. J. Chiang and K. Truong, "Using co-cultures expressing fluorescence resonance energy transfer based protein biosensors to simultaneously image caspase-3 and Ca2+ signaling," *Biotechnol Lett*, vol. 27, pp. 1219-27, 2005.

[5] O. Kohlbacher and H. P. Lenhof, "BALL--rapid software prototyping in computational molecular biology. Biochemicals Algorithms Library," *Bioinformatics*, vol. 16, pp. 815-24, 2000.

[6] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," *J Mol Graph*, vol. 14, pp. 33-8, 27-8, 1996.

[7] A. Fiser and A. Sali, "Modeller: generation and refinement of homology-based protein structure models," *Methods Enzymol*, vol. 374, pp. 461-91, 2003.

[8] W. Kaplan and T. G. Littlejohn, "Swiss-PDB Viewer (Deep View)," *Brief Bioinform*, vol. 2, pp. 195-7, 2001.

[9] J. Hyun Bae, M. Rubini, G. Jung, G. Wiegand, M. H. Seifert, M. K. Azim, J. S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber, and N. Budisa, "Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins," *J Mol Biol*, vol. 328, pp. 1071-81, 2003.

[10] A. Rekas, J. R. Alattia, T. Nagai, A. Miyawaki, and M. Ikura, "Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity," *J Biol Chem*, vol. 277, pp. 50573-8, 2002.

[11] R. M. Clegg, "Fluorescence resonance energy transfer," *Curr Opin Biotechnol*, vol. 6, pp. 103-10, 1995.

[12] B. Nagar, M. Overduin, M. Ikura, and J. M. Rini, "Structural basis of calcium-induced E-cadherin rigidification and dimerization," *Nature*, vol. 380, pp. 360-4, 1996.

[13] J. R. Alattia, H. Kurokawa, and M. Ikura, "Structural view of cadherin-mediated cell-cell adhesion," *Cell Mol Life Sci*, vol. 55, pp. 359-67, 1999.

[14] M. Osawa, H. Tokumitsu, M. B. Swindells, H. Kurihara, M. Orita, T. Shibanuma, T. Furuya, and M. Ikura, "A novel target recognition revealed by calmodulin in complex with Ca2+-calmodulin-dependent kinase kinase," *Nat Struct Biol*, vol. 6, pp. 819-24, 1999.

[15] K. Truong, A. Khorchid, and M. Ikura, "A fluorescent cassette-based strategy for engineering multiple domain fusion proteins," *BMC Biotechnol*, vol. 3, pp. 8, 2003.