COMPUTATIONAL DESIGN OF A GENETICALLY ENCODED IRON BIOSENSOR

Melike Berksöz, Canan Atılgan

Faculty of Natural Sciences and Engineering, Sabancı University
Tuzla 34956 Istanbul, Turkey
phone: + (90) 05315759001 email: mberksoz@sabanciuniv.edu, canan@sabanciuniv.edu

ABSTRACT

Genetically encoded fluorescent biosensors (GEFB) proved to be reliable tracers for many metabolites and cellular processes. In the simplest case, a fluorescent protein (FP) is genetically fused to a sensing protein which undergoes a conformational change upon ligand binding. This drives a rearrangement in the chromophore environment and changes the spectral properties of the FP. Structural determinants of successful biosensors are determined only in hindsight when the crystal structures of both ligand-bound and ligand-free forms are available. This makes the development of new biosensors for desired analytes a long trial-and-error process. In this work, we propose a novel design strategy that combines Alphafold2 (AF2) and all-atom molecular dynamics (MD) simulations as predictors of GEFB structure and dynamics. *H. influenza* ferric binding protein (FBP) is selected as a model sensor protein. Hydrogen bond occupancies around the chromophore are analyzed in apo and holo states of fused FBP. Hydrogen bond profile of intact GFP alone with neutral and anionic chromophore was used as a reference for dark and bright states respectively.

1. Introduction

GEFBs are protein-based indicators that can be used for dynamic imaging of practically any molecule or biochemical process in living cells. These proteins can be easily expressed in the cytoplasm and targeted to desired subcellular locales by use of peptide tags. Broadly, GEFBs have a modular structure whereby a fluorescent reporter is genetically fused to a sensing domain. They fall under two main categories; Förster resonance energy transfer (FRET)-based biosensors which consist of two fluorophores (donor and acceptor) linked by a sensing domain and single FP based biosensors where fluorescence of a single FP is allosterically modulated. FRET based biosensors usually require a sensing domain that undergoes a significant conformational change upon ligand binding that effectively changes the distance between the two fluorescent proteins. This requirement limits the number of sensing proteins that can be used in FRET sensors. In experiments with FRET sensors, two emission wavelengths are observed. This constrains the time frame of observable events to the time required for image acquisition [1]. In that respect, single FP biosensors are more user friendly. Additionally, their small size provides an opportunity for rational design. Allosteric communication between the two modules is most commonly achieved by inserting a circularly permuted FP into the sensing domain. In most cases, mechanism of allosteric modulation of fluorescence is speculated based on crystal structures. We know that the local protein environment is important for the chromophore since the denatured FPs or the isolated chromophore are not fluorescent [2]. More specifically, hydrogen bond network around the chromophore is decisive on fluorescence efficiency. Structure prediction algorithms and MD simulations can potentially increase the success of biosensor design by providing an insight into the hydrogen bonding dynamics, stability and efficiency of the chimeric protein.

2. Background

In a seminal paper, Baird. *et al.* showed that yellow FP preserves its fluorescence when Y145 is replaced with Calmodulin or Zinc finger domain [3]. In this architecture, the chromophore is exposed to nearby residues of the inserted protein. The region between residues 145-148 is where FP β barrel slightly bulges towards the solvent to accommodate the chromophore (Fig 1.a). Crystal structure of GFP (PDB: 1EMA) shows that the side chains of two residues in the bulge flank towards the exterior, disrupting the ordered hydrogen bonding network within the 11 stranded β barrel structure. Therefore this region is permissive to creation of new termini or insertion of a new protein without compromising the fluorescence. Circular permutation or insertion of a sensing domain exposes the tyrosine derived phenol moiety, rendering its pKa sensitive to local changes in the hydrogen bond network Uncharged phenol form of both GFP and red FP absorbs shorter wavelength light (395 nm) and is less fluorescent whereas anionic phenolate form absorbs longer wavelength light (475 nm) is much more fluorescent [4]. Therefore, shifting the phenol-phenolate equilibrium as a result of a conformational change in the sensing domain is the working principle of most single FP based biosensors.

Crystal structures of biosensors reveal that the bulge region is exclusively preferred for sensing domain insertion. The two residues flanking the bulge region are termed as ‘gate post residues’, which are Y145 and H148 in GFP [5]. Gate post residues as well as the linkers connecting the sensing domain and FPs are diverse among reported biosensors.

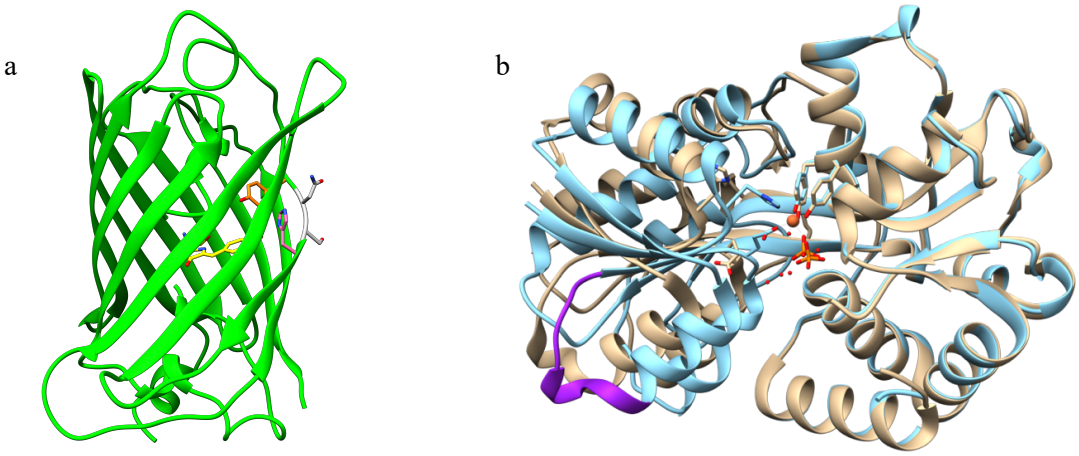

Figure 1. a. GFP from *Aequorea Victoria.* Bulge region is colored grey. Y145 (purple) and H148 (orange) gate post residues are adjacent to the bulge. **b. *H.Influenza* FbP.** Apo (tan) and holo (blue) states are superimposed on the fixed domain (residues 83-87, 102-225, 277-307). Fe^{3+} in the binding site is shown as an orange sphere. Allosteric 44-52 loop is colored purple.

During the development, a pool of sensors with randomized gate posts and linkers are screened for highest analyte-dependent fluorescence change. Linkers are kept as short as possible to allow for efficient coupling between the chromophore and the sensing domain but not too short that folding of either partner is compromised.

A prominent example of single FP based biosensors is GCaMP, a Ca^{2+} indicator where cpGFP is inserted between Calmodulin (CaM) and the M13 peptide [6]. Crystal structures of apo and holo forms of this sensor revealed that the solvent exposure of the chromophore decreases dramatically when Ca^{2+} bound-CaM wraps around the M13 peptide. Ca^{2+} bound bright state of this sensor is attributed to the deprotonated phenolate moiety [7].

A number of single FP biosensors were derived from bacterial periplasmic binding proteins (PBPs) [1], [8]–[10].These proteins are primarily involved in transport of various critical solutes such as metals, ions, sugars and vitamins from extracellular space into the cytoplasm. The typical structure of PBPs consists of two domains linked by a two or three stranded-hinge. Ligand binding site is found at the domain interface. Hinge bending motion drives open-to-closed conformational transition upon ligand binding. The extent of this movement varies among different PBPs; C_α RMSD between the crystal structures of holo and apo forms of *E.coli* MurG glycosyltransferase is 1.5 Å, while this value goes up to 3.3 Å for *E.Coli* maltodextrin binding protein. A conformationally active region, usually a hinge, is where the FP is attached.

For this work, we have chosen *H. influenza* FBP for biosensor design which undergoes a moderate degree of ligand induced conformational change with an apo-holo C_α RMSD of 2.4 Å (Fig 1.b). In the holo form, iron is coordinated by residues from both N and C domains (Y195, Y196, E57, H9), as well as a synergistic phosphate ion and a water molecule. In contrast, apo form has an open conformation where phosphate ion is accommodated on the C domain. Two tyrosines in the binding site have downshifted pK_a values in the presence of iron and were previously modelled in deprotonated form in MD simulations of the apo form. This achieves charge neutrality of the octahedral coordination shell around iron. Perturbation response scanning (PRS) studies have revealed that iron binding and re lease dynamics are allosterically controlled by a distant loop spanning residues 44-52. PRS method is based on systematic perturbation of all the residues with a variety of forces and predicting the conformational change using linear response theory. We considered 44-52 loop to be a promising site for FP insertion since it is composed of mostly charged residues, which provides an opportunity to control the pKa of the chromophore. Moreover, it was previously shown that D52 is an allosteric controller of the conformations of FBP in both the apo and the holo forms [11]–[13]. This protein was also used previously to create an iron probe in nanopipette format [13].

3. Methods

Primary sequence of the biosensor was designed using the sequence of a previously reported cpGFP [6] This construct was originally derived from GFP from *Aequorea Victoria* by removing the two-residue bulge region and creating new termini. The original termini were connected by a flexible linker (GGSMV). cpGFP sequence was inserted between T49 and P50 of FBP. Histidine and phenylalanine were chosen as gate post residues based on their usage frequencies in reported biosensor [5]. No linker was used on either side (Fig. 2).

Colabfold which is an optimized version of AF2 was used to predict the structure of designed biosensors [14]. To obtain ligand-bound and ligand-free models, crystal structures of apo (PDB:1D9V) and holo (PDB:1MRP) FBP were used as homology templates. Due to the fact that chromophore is a non-standard residue that is not recognized by AF2, predictions were initially run with the chromophore sequence TYG and models were then superimposed with circularly permuted GFP (PDB: 3EVP). Coordinates of the cyclized chromophore were taken from superimposed cpGFP structure and integrated into the models. Highest scoring models were used as initial structures for MD simulations. As a reference for hydrogen

bond network around the chromophore, we have also simulated the intact GFP alone with its chromophore in neutral and in deprotonated forms. Crystal structure of GFP (PDB: 1EMA) was used as initial structure. The chromophore was modelled in neutral form in both apo and holo biosensor models. Additonally, we have removed iron from the holo model manually and performed a separate run in order to probe the holo-to-apo transition within the time frame of the simulation.

Green Fluorescent Protein from *Aequorea Victoria*

SKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGK
LPVPWPTLVTTFYVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGN
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADK
QKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSAL
SKDPNEKRDHMVLLEFVTAAGITHGMDELYK

1. Circular permutation of GFP

HNVYIMADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDN
HYLSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKG
EELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPV
PWPTLVTTLVYQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNF

2. Insertion of cpGFP between T49-P50 of FbP

DITVYNGQHKEAATAVAKAFEQETGIKVTLNSGKSEQLAGQLKEEGDKTHN
VYIMADKQKNGIKANFKIRHNIEDGGVQLAYHYQQNTPIGDGPVLLPDNHY
LSTQSKLSKDPNEKRDHMVLLEFVTAAGITLGMDELYKGGTGGSMVSKGEE
LFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPW
PTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYIQERTIFFKDDGNYKT
RAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFPADVFYTEQTATFA
DLSEAGLLAPISEQTIQQTAQKGVPLAPKKDWIALSGRSRVVVYDHTKLSE
KDMEKSVLDYATPKWKGKIGYVSTSGAFLEQVVALSKMKGDKVALNWLKGL
KENGKLYAKNSVALQAVENGEVPAALINNYYWYNLAKEKGVENLKSRLYFV
RHQDPGALVSYSGAAVLKASKNQAEAQKFVDFLASKKGQEALVAARAEYPL
RADVVSPFNLEPYEKLEAPVVSATTAQDKEHAIKLIEEAGLK

Figure 2. Construction of biosensor amino acid sequence. Bulge region and cpGFP are colored gray and green respectively.

VMD and NAMD packages were used to prepare the simulation box and simulate the dynamics of biosensor models in water. General simulation protocol for all models was as follows: Protein was placed in a cubic box with at least 10 Å layer of water in each direction from any atom in the system. TIP3P model was used to define water molecules Simulation box was neutralized with K^+ and Cl^+ ions while maintaining an ionic strength of 150 mM. Protein atoms were modelled using CHARMM36 all atom force field. Chromophore atoms were modelled using a combination of CGenFF and CHARMM36 parameters. Force field parameters for Fe^{3+} was adopted from literature [15]. Periodic boundary conditions were used and trajectories were calculated using velocity Verlet algorithm with a timestep of 2 fs. Long-range electrostatic interactions were calculated by Particle Mesh Ewald method with a cutoff distance of 12 Å. All systems were subjected to energy minimization before running in NPT ensemble of 310 K and 1 atm for at least 800 ns. Coordinates were saved every 2 ps for analysis. Hydrogen bond occupancies of equilibrated trajectories were calculated using the hydrogen bond plugin of VMD with default settings (donor-acceptor distance: 3 Å, angle cutoff: 20°).

4. Results

As a measure of per-residue accuracy of the prediction, AF2 produces a metric called predicted local distance difference test (pLDDT). As a rule of thumb, regions with pLDDT score higher than 90 are expected to be high accuracy including side chain conformations. A pLDDT between 70 and 90 indicates good backbone prediction. In Fig 3. apo templated model is colored based on pLDDT score using Alphafold interface of UCSF ChimeraX-1.4. Overall the prediction has acceptable backbone accuracy of the apo template model. Positions with lower accuracy correspond to the circular permutation linker of GFP and connection points of GFP and FBP.

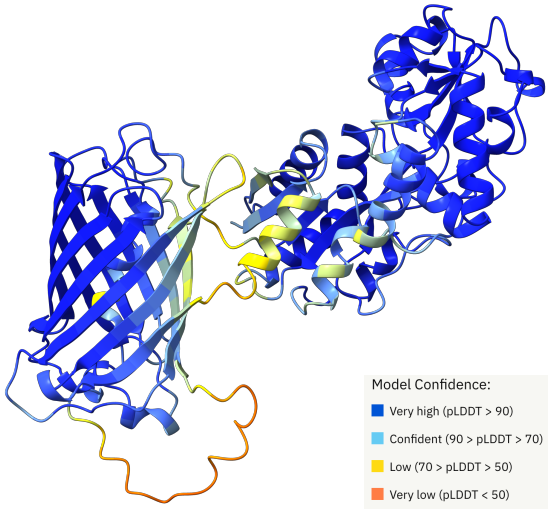

Figure 3. AF2 apo model colored according to pLDDT score.

Predicted structures of both cpGFP and FbP domains of the sensor are superimposable with their free templates (Table 1) which indicates that no linker may be required for the proper folding of this sensor. Binding site residues align well with their template (Fig. 4). When no template was provided, AF2 predicted FBP to be more similar to the apo form (model not shown). This observation is consistent with a previous report which found that AF2 predominantly predicts the binding site in apo form, ready to accommodate the ligand [16].

Table 1. Ca RMSD of GFP and Fbp domains in the AF2 models against their templates

Template PDB	C_α RMSD of AF2 models / Å		
	Apo templated	Holo templated	No template
3evp (GFP)	0.17	0.16	0.30
1mrp (holo)	2.39	0.38	1.65
1d9v (apo)	0.28	2.28	1.04

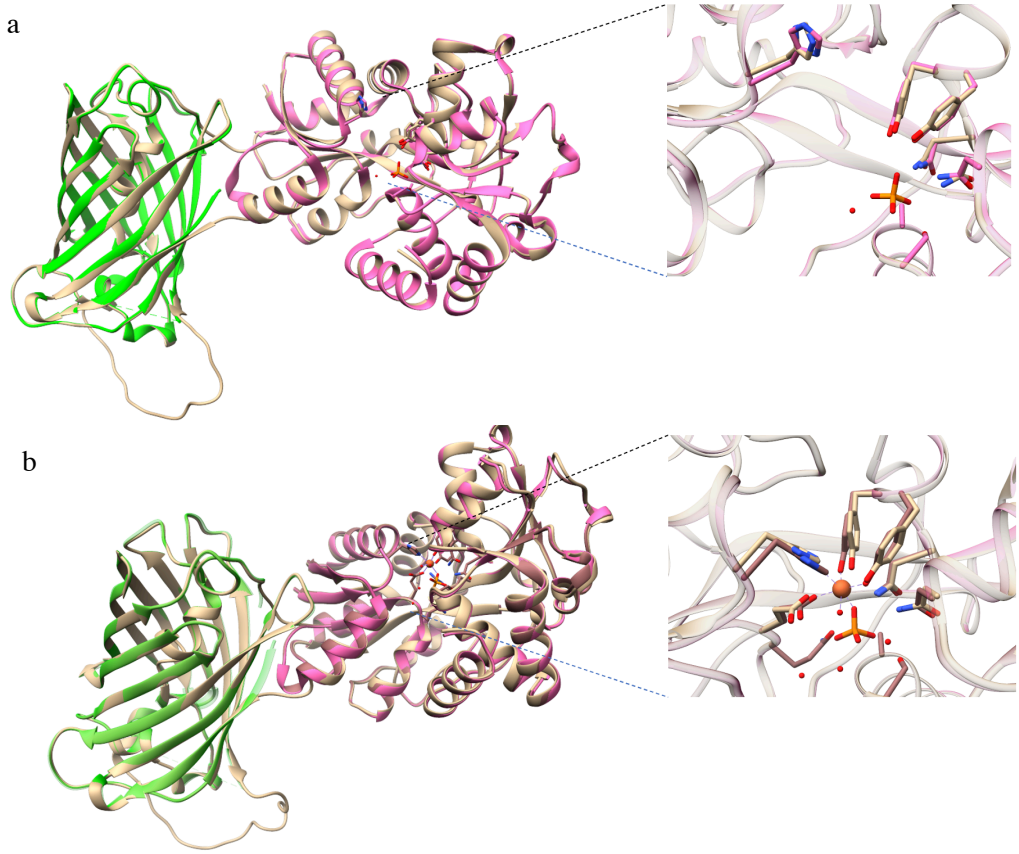

Figure 4. Apo (a) and holo (b) templated AF2 models of designed biosensors superimposed with their templates with iron binding site highlighted. Biosensor models, cpGFP template and apo/holo FBP template is colored tan, green and pink respectively.

Average backbone RMSD of the holo templated model over the 800 ns trajectory is 2.0 ±0.5 Å while that of apo form is 2.5±1.0 Å. In apo run, Fe^{3+} and $H_2PO_4^-$ stay engaged in the binding site for the entire run. When the iron is manually removed from the holo model, backbone RMSD starts to climb after 200 ns which shows that holo-to-apo transition of FbP affects the whole protein. Intact GFP alone had an average backbone RMSD of 1.1±0.2 Å in the neutral form and 1.3±0.2 Å in the deprotonated form.

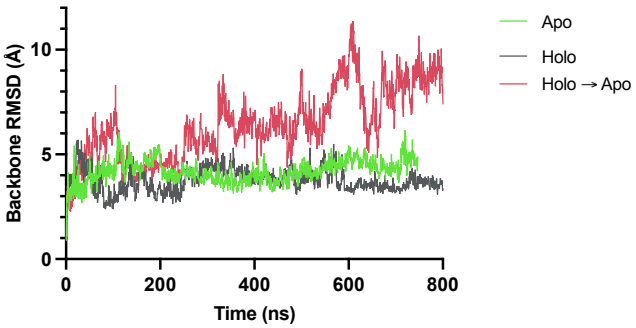

Figure 5. Backbone RMSD of designed biosensor over 800 ns MD trajectory

Hydrogen bond occupancies around the chromophore differ in the deprotonated and neutral form of intact GFP, as well as in apo and holo forms of the designed sensor. 3 specific residues in intact GFP that are involved in hydrogen bonds with the chromophore, namely Q94, Q183 and T203 show marked differences in their percent occupancies. These residues are structurally aligned with Q241, Q85 and T105 in our biosensor model (Fig 6), which again are involved in hydrogen bonds with the chromophore and show similar differences in their hydrogen bond occupancies (Fig 7). T203 has been shown to be important in maintaining the anionic state of the chromophore [17-19]. This data suggests that the holo form of the biosensor is likely to have a neutral chromophore and therefore will be dark, whereas the apo state resembles more to the deprotonated state of GFP and will be bright.

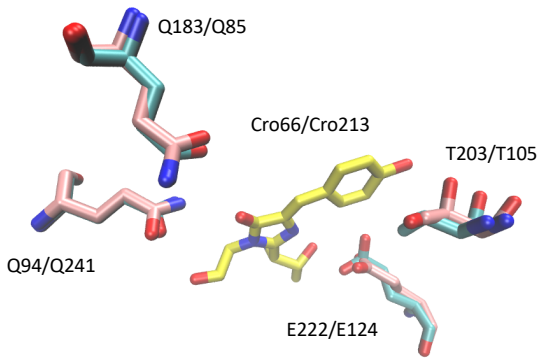

Figure 6. Structural alignment of chromophore enviroment in intact GFP (pink) and sensor model (cyan). Chromophore is shown in yellow.

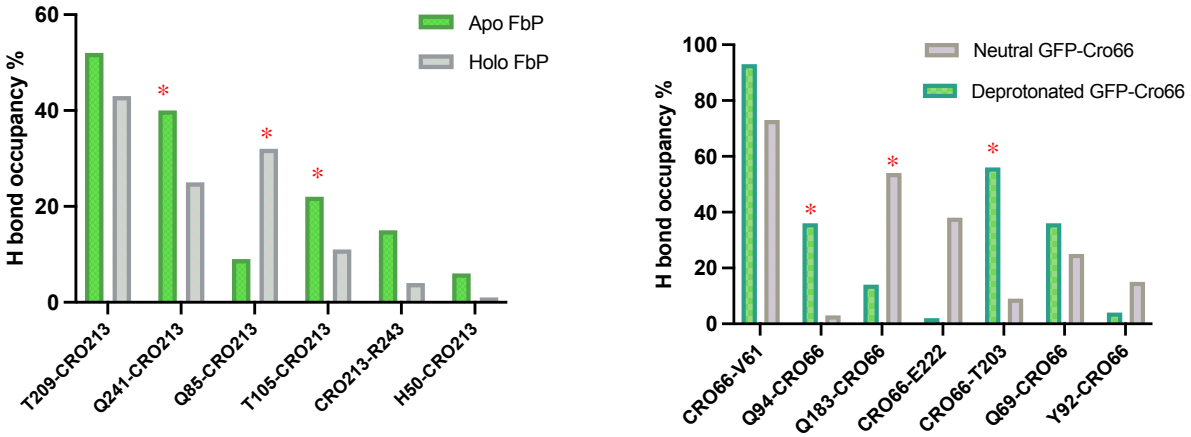

Figure 7. Occupancies of hydrogen bonds where chromophore is either acceptor or donor in apo/holo sensor (a) and in different charge states of intact GFP alone (b)

5. Discussion

Development of new GEFBs requires significant experimental optimization and in most cases the outcome can not be predicted due to randomized nature of interdomain linkers and gate post residues. Our AF2-predicted biosensor models are expected to fold properly without the need for an interdomain linker. Prediction of domain folds with high accuracy may reduce the time needed for linker screening. Without the linker, chromophore and the sensing domain interact in a more direct manner and the conformational coupling between the two can be predicted at a molecular level using the known structure of the sensing protein. Using a well characterized allosteric site in FBP, we were able to create an interaction interface where the pKa of the chromophore can be manipulated. All-atom MD simulations showed that the hydrogen bonding pattern around the chromophore changes upon iron binding similar to the difference observed between deprotonated and neutral intact GFP. Since the protonation state of the chromophore ultimately determines the brightness of the sensor, we can speculate that this sensor will potentially respond to changes in iron concentration. Our results suggest that hydrogen bond occupancies around chromophore can be good predictor of biosensor efficiency. As a further work, the fluorescence spectra of the proposed models will be experimentally measured in the presence and absence of iron. Iron binding affinity of FBP is quite high (Ka=10^{18} M^{-1})[12] and has to be fine-tuned by mutations in and around the binding site to increase the sensitivity of this sensor. With the high number of available structures in the PDB and Alphafold database of predicted structures, our rational design strategy can help accelerate development of a broader range of fluorescent biosensors.

6. Acknowledments

This study was supported by Tubitak 121Z329 project.

REFERENCES

[1] J. S. Marvin, E. R. Schreiter, I. M. Echevarría, and L. L. Looger, "A genetically encoded, high-signal-to-noise maltose sensor," *Proteins: Structure, Function and Bioinformatics*, vol. 79, no. 11, pp. 3025–3036, 2011, doi: 10.1002/prot.23118.

[2] A. A. Pakhomov and V. I. Martynov, "GFP Family: Structural Insights into Spectral Tuning," *Chemistry and Biology*, vol. 15, no. 8. pp. 755–764, Aug. 25, 2008. doi: 10.1016/j.chembiol.2008.07.009.

[3] G. S. Baird, D. A. Zacharias, and R. Y. Tsien, "Circular permutation and receptor insertion within green fluorescent proteins," *Proc Natl Acad Sci U S A*, vol. 96, no. 20, pp. 11241–11246, 1999, doi: 10.1073/pnas.96.20.11241.

[4] O. v Stepanenko, V. v Verkhusha, I. M. Kuznetsova, V. N. Uversky, and K. K. Turoverov, "Fluorescent Proteins as Biomarkers and Biosensors: Throwing Color Lights on Molecular and Cellular Processes," 2008.

[5] Y. Nasu, Y. Shen, L. Kramer, and R. E. Campbell, "Structure- and mechanism-guided design of single fluorescent protein-based biosensors," *Nature Chemical Biology*, vol. 17, no. 5, pp. 509–518, 2021, doi: 10.1038/s41589-020-00718-x.

[6] Q. Wang, B. Shui, M. I. Kotlikoff, and H. Sondermann, "Structural Basis for Calcium Sensing by GCaMP2," *Structure*, vol. 16, no. 12, pp. 1817–1827, 2008, doi: 10.1016/j.str.2008.10.008.

[7] J. Akerboom *et al.*, "Crystal structures of the GCaMP calcium sensor reveal the mechanism of fluorescence signal change and aid rational design," *Journal of Biological Chemistry*, vol. 284, no. 10, pp. 6455–6464, 2009, doi: 10.1074/jbc.M807657200.

[8] Z. Li, J. Zhang, and H. W. Ai, "Genetically Encoded Green Fluorescent Biosensors for Monitoring UDP-GlcNAc in Live Cells," *ACS Central Science*, vol. 7, no. 10, pp. 1763–1770, 2021, doi: 10.1021/acscentsci.1c00745.

[9] R. M. de Lorimier *et al.*, "Construction of a fluorescent biosensor family," *Protein Science*, vol. 11, no. 11, pp. 2655–2675, 2009, doi: 10.1110/ps.021860.

[10] I. Alicea, J. S. Marvin, A. E. Miklos, A. D. Ellington, L. L. Looger, and E. R. Schreiter, "Structure of the Escherichia coli phosphonate binding protein PhnD and rationally optimized phosphonate biosensors," *Journal of Molecular*

Biology, vol. 414, no. 3, pp. 356–369, 2011, doi: 10.1016/j.jmb.2011.09.047.

[11] O. Sensoy, A. R. Atilgan, and C. Atilgan, "FbpA iron storage and release are governed by periplasmic microenvironments," *Physical Chemistry Chemical Physics*, vol. 19, no. 8, pp. 6064–6075, 2017, doi: 10.1039/c6cp06961d.

[12] G. Guven, A. R. Atilgan, and C. Atilgan, "Protonation states of remote residues affect binding-release dynamics of the ligand but not the conformation of apo ferric binding protein," *Journal of Physical Chemistry B*, vol. 118, no. 40, pp. 11677–11687, 2014, doi: 10.1021/jp5079218.

[13] G. Bulbul, G. Liu, N. R. Vithalapur, C. Atilgan, Z. Sayers, and N. Pourmand, "Employment of Iron-Binding Protein from Haemophilus influenzae in Functional Nanopipettes for Iron Monitoring," *ACS Chemical Neuroscience*, vol. 10, no. 4, pp. 1970–1977, Apr. 2019, doi: 10.1021/acschemneuro.8b00263.

[14] M. Mirdita, S. Ovchinnikov, and M. Steinegger, "ColabFold - Making protein folding accessible to all," *bioRxiv*, p. 2021.08.15.456425, 2021, [Online]. Available: https://www.biorxiv.org/content/10.1101/2021.08.15.456425v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.08.15.456425v1.abstract

[15] C. Atilgan and A. R. Atilgan, "Perturbation-response scanning reveals ligand entry-exit mechanisms of ferric binding protein," *PLoS Computational Biology*, vol. 5, no. 10, 2009, doi: 10.1371/journal.pcbi.1000544.

[16] T. Saldaño *et al*., "Impact of protein conformational diversity on AlphaFold predictions," *bioRxiv*, p. 2021.10.27.466189, 2021, [Online]. Available: https://www.biorxiv.org/content/10.1101/2021.10.27.466189v1%0Ahttps://www.biorxiv.org/content/10.1101/2021.10.27.466189v1.abstract

[17] S. S. Patnaik, S. Trohalaki, and R. Pachter, "Molecular modeling of green fluorescent protein: Structural effects of chromophore deprotonation," *Biopolymers*, vol. 75, no. 6, pp. 441–452, 2004, doi: 10.1002/bip.20156.

[18] N. Reuter, H. Lin, and W. Thiel, "Green fluorescent proteins: Empirical force field for the neutral and deprotonated forms of the chromophore. Molecular dynamics simulations of the wild type and S65T mutant," *Journal of Physical Chemistry B*, vol. 106, no. 24, pp. 6310–6321, 2002, doi: 10.1021/jp014476w.

[19] J. Kaur *et al*., "Role of Ser65, His148 and Thr203 in the Organic Solvent-dependent Spectral Shift in Green Fluorescent Protein," *Photochemistry and Photobiology*, vol. 95, no. 2, pp. 543–555, Mar. 2019, doi: 10.1111/php.13018.