

# Constant pH Accelerated Molecular Dynamics Investigation of the pH Regulation Mechanism of Dinoflagellate Luciferase

Patrick H. Donnan,<sup>‡</sup> Phong D. Ngo,<sup>‡</sup> and Steven O. Mansoorabadi\*

Department of Chemistry and Biochemistry, Auburn University, 179 Chemistry Building, Auburn, AL 36849, United States

## Supporting Information

**ABSTRACT:** The bioluminescence reaction in dinoflagellates involves the oxidation of an open-chain tetrapyrrole by the enzyme dinoflagellate luciferase (LCF). The activity of LCF is tightly regulated by pH, where the enzyme is essentially inactive at pH  $\sim 8$  and optimally active at pH  $\sim 6$ . Little is known about the mechanism of LCF and the structure of the active form of the enzyme, although it has been proposed that several intramolecularly conserved histidine residues in the N-terminal region are important for the pH regulation mechanism. Here, constant pH accelerated molecular dynamics was employed to gain insight into the conformational activation of LCF induced by acidification.

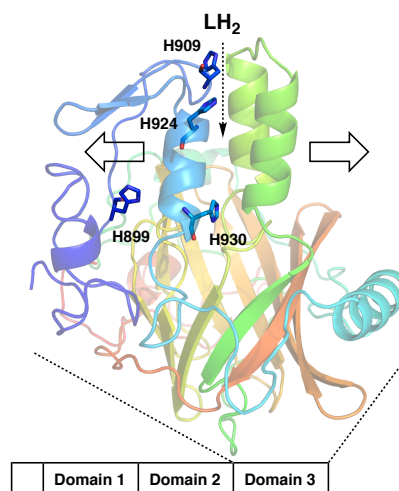
Bioluminescence, the production of light by living organisms, has arisen independently many times throughout the course of evolution.<sup>1</sup> Consequently, the biological uses of bioluminescence are varied, ranging from communication and courtship to camouflage and predation.<sup>2</sup> The chemistries of bioluminescent reactions are also quite diverse, although each involves the enzymatic, luciferase-catalyzed, oxidation of a luciferin substrate with molecular oxygen.<sup>3</sup>

Dinoflagellates are the predominant bioluminescent microorganisms in the sea.<sup>4</sup> Light production by dinoflagellates is proposed to serve as a defense mechanism that attracts predators of dinoflagellate grazers, the so-called burglar alarm hypothesis.<sup>5</sup> The bioluminescence system in dinoflagellates is comprised of an open-chain tetrapyrrolic luciferin (LH<sub>2</sub>) derived from the catabolism of chlorophyll, a luciferase (LCF) that is a divergent member of the lipochalin protein family, and in most species, a luciferin-binding protein (LPB).<sup>6-8</sup> Each of these components is housed together in small, subcellular compartments called scintillons that are contiguous with an acidic vacuole.<sup>9-10</sup>

The dinoflagellate bioluminescence reaction is unique in that it is induced by physical agitation.<sup>11</sup> Shear forces are thought to activate a G-protein coupled receptor on the surface of the cell, which initiates a signal transduction cascade and leads to an increase in cytoplasmic cal-

cium levels.<sup>12,13</sup> The increase in calcium concentration depolarizes the vacuolar membrane, generating an action potential that opens voltage-gated ion channels and allows the influx of protons into the scintillons.<sup>14</sup> Acidification of the scintillons is then proposed to trigger the release of LH<sub>2</sub> from LPB and the activation of LCF, which leads to the bright flashes of blue light characteristic of dinoflagellate bioluminescence.<sup>15</sup>

The LCFs from most bioluminescent dinoflagellates contain three homologous catalytic domains in a single polypeptide (Figure 1).<sup>16</sup> Each individual domain exhibits a sharp pH-rate profile, wherein activity is essentially zero at pH  $\sim 8$  and optimal at pH  $\sim 6$ .<sup>17</sup> Four intramolecularly conserved histidine residues in the N-terminal region of each domain have been implicated in the loss of activity in the alkaline region.<sup>17</sup> Mutation of any of these histidine residues to alanine results in an increase in the relative activity of the domains at pH 8.<sup>17</sup>



**Figure 1.** Crystal structure of LCF domain III from *Lingulodinium polyedrum* (PDB ID: 1VPR) showing the location of the intramolecularly conserved histidine residues and the proposed pH-induced conformational change.

The crystal structure of domain III from *Lingulodinium polyedrum* was solved in the inactive form at pH 8 (Figure 1).<sup>7</sup> The structure consists of a  $\beta$ -barrel that houses the presumed active site and a three-helix bundle

cap.<sup>7</sup> The four conserved histidine residues (H899, H909, H924, and H930) are positioned at the interface of the helices in the bundle, which suggests that this structural motif plays an important role in the pH regulation mechanism of LCF.<sup>7</sup>

Disruption of the hydrogen bonding interactions formed by the conserved histidine residues, either by protonation or mutation to alanine, is proposed to induce a conformational change in the three-helix bundle that activates the enzyme.<sup>7</sup> Consistent with this hypothesis, preliminary molecular dynamics (MD) simulations of the quadruple alanine variant showed an expansion of the three-helix bundle, creating a solvent channel that would allow LH<sub>2</sub> access to the active site (Figure 1).<sup>7</sup> However, these simulations were only performed for 10 ns, which is likely of insufficient length to allow for a large scale conformational change, and it is unclear whether the observed changes will be reflective of those induced by acidification.

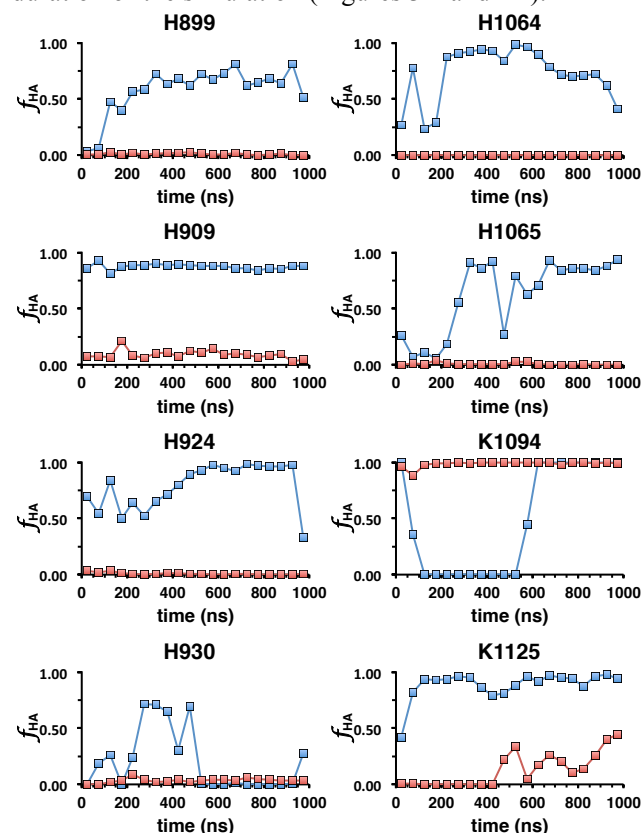
To appropriately investigate the pH regulation mechanism of LCF computationally, the protonation state of titratable residues must be allowed to vary in response to proton concentration and the calculations must be performed on sufficiently long time scales (or with enhanced sampling) so as to observe the conformational transition between the active and inactive forms. One promising approach that addresses both criteria is to couple constant pH molecular dynamics (CpHMD) with accelerated molecular dynamics (aMD).<sup>18-20</sup> The CpHMD method, as implemented in Amber 16, employs generalized Born solvated molecular dynamics with periodic Monte Carlo sampling of protonation states.<sup>18,21</sup> In aMD, a continuous bias potential is applied when the true potential falls below a chosen threshold energy, which enhances the escape rate from potential basins, improves sampling of the conformational landscape, and converges to the correct canonical distribution.<sup>19</sup>

Constant pH accelerated molecular dynamics (CpHaMD) was therefore used to investigate the conformational dynamics of LCF domain III at pH 8 and pH 6. Given the proximity of their free amino acid pK<sub>a</sub> values to the pH range of the simulation, the protonation states of all cysteine and histidine residues were made titratable. In addition, aspartate, glutamate, lysine, and tyrosine residues having less than 15% solvent accessibility (and thus most likely to have microenvironments that deviate significantly from that of solution) were also allowed to vary during the simulation. After structure minimization, heating, equilibration, and a 10 ns conventional MD simulation to obtain boost parameters, CpHaMD was performed for 1  $\mu$ s of aMD time (see Supporting Information for full computational details and a list of calculated pK<sub>a</sub> values).

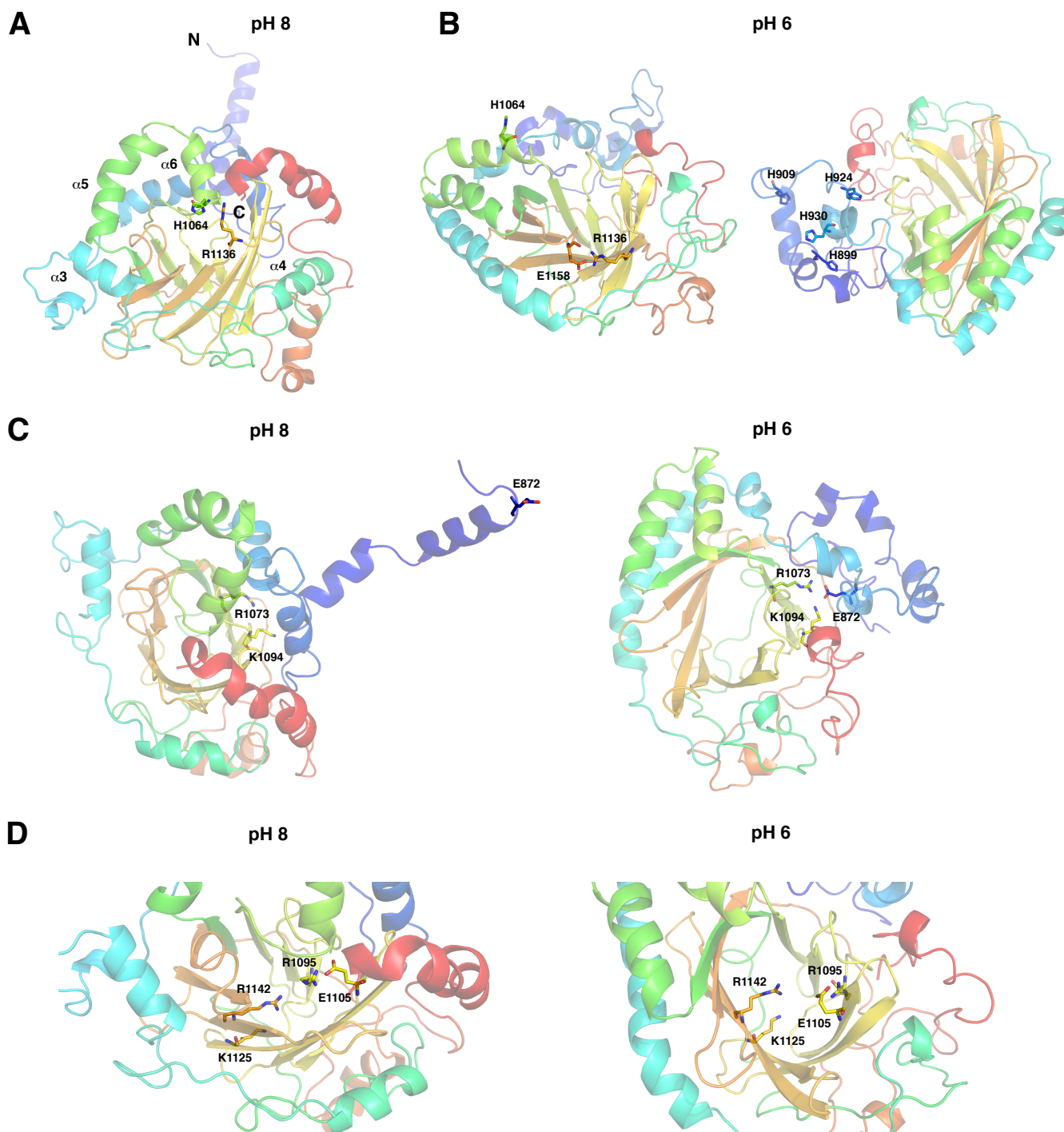
Throughout the simulation at pH 8, the four conserved histidine residues implicated in the pH regulation mechanism remain predominantly in the neutral, deprotonated form (Figure 2). Interestingly, the C-terminus adopts an

$\alpha$ -helical conformation and remains relatively rigidly bound to the top of the  $\alpha$ -barrel adjacent to the three-helix bundle, while the N-terminus is highly mobile (Figures 3A and 4A). This is in contrast to what is observed in the crystal structure, where the C-terminus is disordered and the N-terminus forms a structured loop containing two short antiparallel  $\alpha$ -hairpins.<sup>7</sup>

In addition to the N-terminus,  $\alpha$ 5 and  $\alpha$ 6 of the three-helix bundle are also highly mobile, and the region between the glycine-rich motif and the highly conserved catalytic core (which encompasses helices  $\alpha$ 3 and  $\alpha$ 4 and wraps around the  $\alpha$ -barrel) deviates significantly from its position in the crystal structure (Figures 3A and 4A).<sup>7</sup> It is unclear whether the flexibility of the N-terminal region of domain III would be retained in the full-length enzyme (i.e., if it were tethered to domains I and II). However, despite the mobility of the N-terminus and the three-helix bundle, the latter remains tightly associated with the  $\alpha$ -barrel throughout the simulation, which is consistent with the hypothesis that this motif restricts substrate access to the active site at pH 8. In fact, after the first  $\sim$ 100 ns of the simulation, the gap between the three-helix bundle and the edge of the  $\alpha$ -barrel, as defined by the distance between the C $\alpha$ 's of H1064 and R1136, decreases from the  $\sim$ 13 Å observed in the crystal structure to  $\sim$ 6 Å and remains there for the duration of the simulation (Figures 3A and 4B).



**Figure 2.** Fraction of LCF domain III residues in the protonated form ( $f_{HA}$ ) during the CpHaMD simulations at pH 8 (red) and pH 6 (blue).

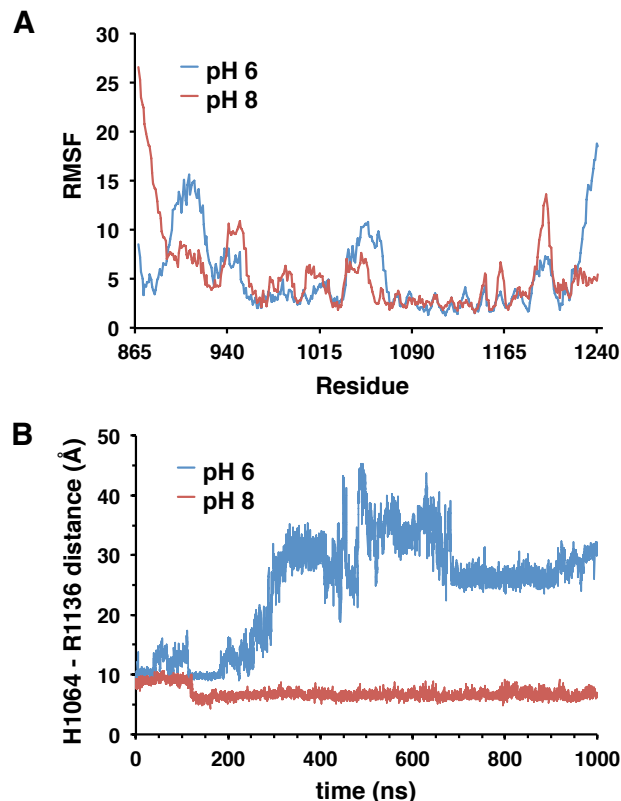


**Figure 3.** Calculated structures of LCF domain III after 1  $\mu$ s of CpHaMD simulation. A) The inactive structure at pH 8, showing the close association of H1064 and R1136. The helices  $\alpha 3$ – $\alpha 6$  and the N- and C-termini are indicated. B) Two different orientations of the activated structure at pH 6. R1136 forms a salt bridge with E1158, while the N-terminal domain containing the intramolecularly conserved histidine residues reorganizes. C) Comparison of the N-terminal domain structure at pH 8 and pH 6. K1094 participates in a hydrogen bonding network that stabilizes the N-terminal helical bundle at pH 6. D) Comparison of the active site structure at pH 8 and pH 6, showing an expansion of the  $\beta$ -barrel and repositioning of the putative catalytic base E1105.

In contrast, during the simulation at pH 6, the four conserved histidine residues are found in the positively charged, protonated form, although the behavior of each of these residues is distinct (Figure 2). H909 is

rapidly protonated and is found almost exclusively in this form throughout the simulation. H924 is also predominantly found in the protonated form, although it is more intermittently protonated early in the simulation

and the protonated fraction increases after  $\sim 300$  ns. In contrast, H930 begins the simulation in the deprotonated form and is transiently protonated between  $\sim 250$ – $500$  ns. H899 also starts in the deprotonated form and the protonated fraction increases after a lag of  $\sim 100$  ns. Two additional histidine residues, the H1064/H1065 dyad, also display similar behavior, in which they are mostly deprotonated early in the simulation and become protonated after  $\sim 250$  ns (Figure 2).



**Figure 4.** Root mean square fluctuation (RMSF) and inter-residue distance data of LCF domain III from the CpHaMD simulations at pH 8 and pH 6. A) Per residue RMSF of LCF domain III calculated over the full 1  $\mu$ s simulation. B) Distance between the C $\alpha$ 's of H1064 and R1136 during the 1  $\mu$ s simulation.

Intriguingly, the delayed protonation of the above histidine residues correlates with a large-scale conformational change where the distance between H1064 and R1136 increases dramatically (Figures 3B and 4B). During the simulation,  $\alpha 5$  and  $\alpha 6$  from the three-helix bundle tip over and the loop containing the H1064/H1065 dyad is positioned such that the distance between H1064 and R1136 increases to  $\sim 30$  Å (Figure 3B). R1136 also moves away from its position at the top of the  $\beta$ -barrel and forms a salt bridge with E1158 on the adjacent  $\beta$ -strand (Figure 3B). The conformational change also involves a decrease in the mobility of the region surrounding the  $\beta$ -barrel containing  $\alpha 3$  and  $\alpha 4$ , an increase in the flexibility of the C-terminus,

and the reorganization of the N-terminal domain to form a helical bundle (Figures 3B and 4A).

A lysine residue (K1094) on the outside of the  $\beta$ -barrel, which is solvent exposed and in the protonated form at pH 8, rapidly becomes reverse protonated at pH 6 during the first  $\sim 600$  ns of the simulation (Figures 2 and 3C). Closer inspection shows that K1094 engages in a hydrogen bond network with R1073 and E872 (Figure 3C). R1073 is also on the outside of the  $\beta$ -barrel and moves towards K1094 as the three-helix bundle tips over, while E872 is found near the N-terminus. The interaction of E872 with K1094 and R1073 thus helps to stabilize the reorganized N-terminal domain (Figures 3B, 3C, and 4A).

Another lysine residue (K1125) found at the bottom of the  $\beta$ -barrel's interior remains predominantly in the neutral, deprotonated form throughout the simulation at pH 8, but is rapidly protonated at pH 6 (Figures 2 and 3D). This is the only residue whose protonation state changes within the presumed active site and this may be important for organizing the active site for binding of the dianionic LH<sub>2</sub> substrate. In particular, there is a notable expansion of the  $\beta$ -barrel and E1105 moves  $\sim 6$  Å from the periphery to the center of the active site, where it is positioned near R1095 and R1142 (Figure 3D). As it is likely that these two arginine residues in the active site will interact with the carboxylate moieties of the LH<sub>2</sub> chromophore, this conformational change optimally positions E1105 to serve as the catalytic base. The involvement of E1105 in the catalytic mechanism of LCF was proposed in a recent theoretical investigation of dinoflagellate bioluminescence that identified the potential bioluminophore of the reaction.<sup>22</sup>

In summary, CpHaMD was applied to investigate the conformational changes associated with the activation of LCF upon acidification. The protonation of several residues, including the previously identified intramolecularly conserved histidines and the H1064/H1065 dyad, correlate with a large scale conformational change in which the N-terminal domain reorganizes to allow substrate access to the active site. Concomitantly, the  $\beta$ -barrel expands and a putative active site base, E1105, moves into position where it can initiate catalysis. To our knowledge, this is the first example of CpHaMD being applied to investigate the conformational activation of a pH-regulated enzyme and demonstrates the power of this methodology in gaining insight into enzyme dynamics as a function of pH.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: som@auburn.edu.

### Author Contributions

The manuscript was written through contributions of all authors. \*These authors contributed equally.

### Notes

The authors declare no competing financial interests.

## ACKNOWLEDGMENT

This work was supported by the National Science Foundation CAREER Award CHE-1555138 and made possible in part by a grant of the high performance computing resources and technical support from the Alabama Super-computer Authority.

## ABBREVIATIONS

aMD, accelerated molecular dynamics; CpHaMD, constant pH accelerated molecular dynamics; CpHMD, constant pH molecular dynamics; LBP, luciferin-binding protein; LCF, dinoflagellate luciferase; LH<sub>2</sub>, dinoflagellate luciferin; MD, molecular dynamics; RMSD, root mean square fluctuation.

## REFERENCES

- (1) Hastings, J. W. (1983) *J. Mol. Evol.* 19, 309–321.
- (2) Wilson, T., and Hastings, J. W. (2013) *Bioluminescence: Living Lights, Lights for Living*, Harvard University Press, Cambridge, MA.
- (3) Hastings, J. W. (1996) *Gene* 173, 5–11.
- (4) Haddock, S. H. D., Moline, M. A., and Case, J. F. (2010) *Annu. Rev. Mar. Sci.* 2, 443–493
- (5) Abrahams, M., and Townsend, L. (1993) *Ecology* 74, 258–260.
- (6) Topalov, G., and Kishi, Y. (2001) *Angew. Chem. Int. Ed.* 40, 3892–3894.
- (7) Schultz, L. W., Liu, L., Cegielski, M., and Hastings, J. W. (2005) *Proc. Natl. Acad. Sci. USA* 102, 1378–1383.
- (8) Morse, D., Pappenheimer, A. M. Jr., and Hastings, J. W. (1989) *J. Biol. Chem.* 264, 11822–11826.
- (9) Desa, R., Hastings, J. W., and Vatter, A. E. (1963) *Science* 141, 1269–1270.
- (10) Nicolas, M. T., Sweeney, B. M., and Hastings, J. W. (1987) *J. Cell Sci.* 87, 189–196.
- (11) Maldonado, E. M., and Latz, M. I. (2007) *Biol. Bull.* 212, 242–249.
- (12) Chen, A. K., Latz, M. I., Sobolewski, P., and Frangos, J. A. (2007) *Am. J. Physiol. Regul. Integr. Comp. Physiol.* 292, R2020–R2027.
- (13) Von Dassow, P., and Latz, M. I. (2002) *J. Exp. Biol.* 205, 2971–2986.
- (14) Rodriguez, J. D., Haq, S., Bachvaroff, T., Nowak, K. F., Nowak, S. J., Morgan, D., Cherny, V. V., Sapp, M. M., Bernstein, S., Bolt, A., DeCoursey, T. E., Place, A. R., and Smith, S. M. (2017) *PLoS One* 12, e0171594.
- (15) Fogel, M., and Hastings, J. W. (1972) *Proc. Natl. Acad. Sci. USA* 69, 690–693.
- (16) Liu, L., Wilson, T., and Hastings, J. W. (2004) *Proc. Natl. Acad. Sci. USA* 101, 16555–16560.
- (17) Li, L., Liu, L., Hong, R., Robertson, D., and Hastings, J. W. (2001) *Biochemistry* 40, 1844–1849.
- (18) Mongan, J., Case, D. A., and McCammon, J. A. (2004) *J. Comput. Chem.* 25, 2038–2048.
- (19) Hamelberg, D., Mongan, J., and McCammon, J. A. (2004) *J. Chem. Phys.* 120, 11919–11929.
- (20) Williams, S. L., de Oliveira, C. A., and McCammon, J. A. (2010) *J. Chem. Theory. Comput.* 6, 560–568.
- (21) Case, D. A., Cerutti, D. S., Cheatham, T. E., III, Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Greene, D., Homeyer, N., Izadi, S., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Mermelstein, D., Merz, K. M., Monard, G., Nguyen, H., Omelyan, I., Onufriev, A., Pan, F., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Simmerling, C. L., Botello-Smith, W. M., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Xiao, L., York, D. M., and Kollman, P. A. (2017) *AMBER 2017*, University of California, San Francisco.
- (22) Ngo, P. D. and Mansoorabadi, S. O. (2017) *ChemPhotoChem* 1, 383–387.

