

RESEARCH ARTICLE

Computational design of Periplasmic binding protein biosensors guided by molecular dynamics

Jack M. O'Shea^{1,2*}, Peter Doerner¹, Annis Richardson¹, Christopher W. Wood^{1*}

1 School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, **2** School of Natural Sciences, Technical University of Munich, Center for Functional Protein Assemblies (CPA), Garching, Germany

* jack.oshea@tum.de (JMOS); chris.wood@ed.ac.uk (CWW)



Abstract

Periplasmic binding proteins (PBPs) are bacterial proteins commonly used as scaffolds for substrate-detecting biosensors. In these biosensors, effector proteins (for example fluorescent proteins) are inserted into a PBP such that the effector protein's output changes upon PBP-substrate binding. The insertion site is often determined by comparison of PBP *apo/* *holo* crystal structures, but random insertion libraries have shown that this can miss the best sites. Here, we present a PBP biosensor design method based on residue contact analysis from molecular dynamics. This computational method identifies the best previously known insertion sites in the maltose binding PBP, and suggests further previously unknown sites. We experimentally characterise fluorescent protein insertions at these new sites, finding they too give functional biosensors. Furthermore, our method is sufficiently flexible to both suggest insertion sites compatible with a variety of effector proteins, and be applied to binding proteins beyond PBPs.

OPEN ACCESS

Citation: O'Shea JM, Doerner P, Richardson A, Wood CW (2024) Computational design of Periplasmic binding protein biosensors guided by molecular dynamics. *PLoS Comput Biol* 20(6): e1012212. <https://doi.org/10.1371/journal.pcbi.1012212>

Editor: Alexander MacKerell, University of Maryland School of Pharmacy, UNITED STATES

Received: January 21, 2024

Accepted: May 30, 2024

Published: June 17, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pcbi.1012212>

Copyright: © 2024 O'Shea et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All code and data required to recreate all the analysis in this paper is available on GitHub under a permissive MIT

Author summary

“Biosensors” are microscopic tools that can detect specific molecules of interest and are made of biological building blocks, such as proteins. Upon coming into contact with their target molecule, such as a marker of a specific disease, biosensors change shape and their properties are altered. For example, upon contact with a disease marker, a biosensor could glow brightly to work as a diagnostic, or perhaps it could produce a chemical signal that can be detected by the immune system. Understanding how biosensors change shape in response to their target is key to developing more biosensors with complex responses. In this paper, we use computer simulations of a biosensor component to understand what parameters are important for biosensor design. We use these parameters to generate new biosensors that respond to their target molecule by producing light, but put forward a case that the lesson learned are generalisable enough to inform more complex sensor outputs as well.

License: <https://github.com/wells-wood-research/oshea-j-wood-c-pbp-design-2023>.

Funding: This work was supported by a BBSRC Engineering Biology Breakthrough Award (BB/W013320/1 to CWW). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Introduction

Protein-based biosensors are important tools for research in biochemistry and medicine. Being genetically encodable, they can monitor real-time *in vivo* conditions and can be optimised by directed evolution. The minimum requirement for a biosensor is that it possess an input-sensing "detector" domain and an output-generating "effector" domain. Modularity of detector and effector domains makes design of new sensors much simpler, which is why the periplasmic binding protein scheme of protein-based biosensors has been so successful.

Periplasmic binding proteins (PBPs) are a large family of bacterial proteins that scavenge and sense nutrients, and so are very well suited to detecting small molecules [1,2]. They have evolved to bind a wide range of metabolically important ligands with high affinity and specificity, and new ligand specificities have been engineered by directed evolution [3,4] and computational methods [5,6]. Upon substrate binding they undergo significant conformational change (Fig 1A). For this reason their *apo* (without ligand) and *holo* (ligand-bound) states to be distinguished by receptors and thereby regulate many processes downstream in bacterial cells [7,8]. Structurally, PBPs possess two distinct folded lobes that are spanned by flexible inter-lobe loops, and upon substrate binding these lobes close around the substrate with a Venus flytrap- or hinge-like movement. In other words, when PBPs are *apo* (without ligand) they are most likely in an "open" conformation, and when they are *holo* (ligand-bound) they are in "closed" conformation (Fig 1A).

In PBP biosensors, the PBP's structural changes upon substrate binding are physically transmitted to a fused effector protein so as to measurably change the effector's output. Sensors with outputs such as fluorescence [9–16], transcriptional repression [17,18], and enzymatic activity [19–22] have been generated by inserting different effector proteins into PBP sequences, demonstrating the modularity of this approach. Typically, effectors are circularly permuted: their sequence is rearranged such that their N and C termini are very close to each

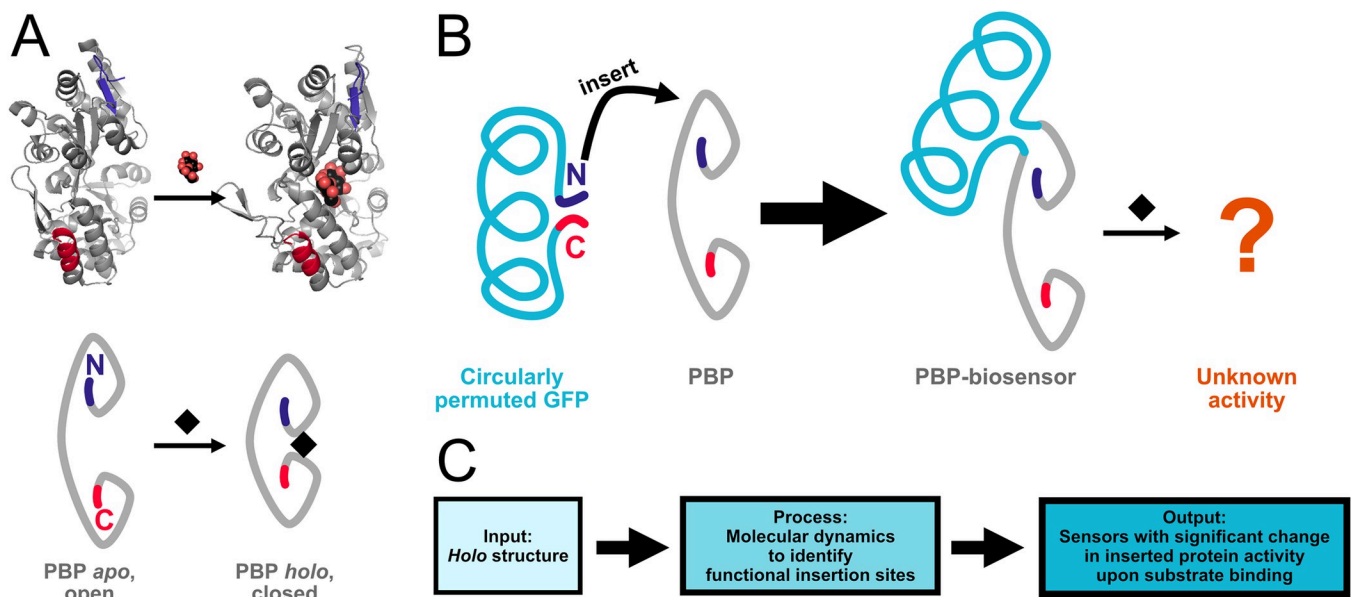


Fig 1. Schematics explaining design of PBP-biosensors. A) Crystal structures of *apo* (1ANF) and *holo* (1OMP) states of periplasmic binding proteins (PBPs), with schematic representations underneath. N and C termini are highlighted in blue and red respectively. B) Circularly permuted GFP has its N and C termini close together, so it can be inserted into PBPs to generate biosensors, but it is difficult to predict sensor activity from insertion site. C) Our method takes *holo* PBP structure as input, and using molecular dynamics simulations can provide insertion sites to generate functional biosensors.

<https://doi.org/10.1371/journal.pcbi.1012212.g001>

other, ideal for insertion into the PBP sequence [2,23,24]. However, despite many examples of PBP biosensors, their design is impeded by our limited ability to identify sites in PBPs where effectors can be inserted to produce active sensors (Fig 1B), which we term “the identification of functional insertion sites”.

Functional insertion sites were first rationally identified by the comparison of *apo* and *holo* crystal structures [10]. Marvin *et al.* showed that regions of greatest change in backbone dihedral torsion angles between the *apo*-open and *holo*-closed states of the maltose binding protein were associated with production of viable sensors upon insertion of circularly permuted GFP (cpGFP). This method has been successfully applied to other PBPs, either by referring to *apo/holo* crystal structures or by homology modelling of known successful insertion sites onto the new PBP [9,11,25]. However, change in crystal structure backbone torsion angle can be an unreliable indicator of good insertion sites: sites of high change can be nonfunctional [10] and sites of low change can be excellent [12,18]. Clearly crystal structure backbone torsion does not explain all factors necessary for functional insertion into PBPs. Alternatively, functional insertion sites have been identified by screens of random insertion libraries [12,19,21], but these methods are labour intensive, difficult to get full insertion coverage with, and do not explain why the identified insertion sites are functional.

In order to address these current shortcomings in PBP-based biosensor design, we reasoned that molecular dynamics (MD) could contain data that discern functional insertion sites better than crystal structure backbone torsion. We applied structural modelling, simulation, and analysis to the task (Fig 1C) using the well-researched maltose binding protein (MBP) as a model subject. In simulations, requiring only the *holo* structure as a starting point, we captured MBP's transition from closed to open state. We found that regions with substantial changes in residue contacts were correlated with the previously identified functional insertion sites, but the analysis also identified additional sites. From these additional sites, we generated new viable MBP sensors with a range of properties. We therefore propose that “change in residue contacts” is a more reliable indicator for functional insertion sites than “crystal structure backbone torsion”. This represents a significant advance in our understanding of the design parameters of PBP-based biosensors and demonstrates that computational analysis can greatly reduce barriers in the novel design of such sensors.

Results and discussion

Previous successful design of PBP biosensors has relied on inserting effector proteins into PBPs at regions of putative substrate-induced conformational change. We reasoned that functional insertion sites could be identified by using molecular dynamics simulations that capture such conformational change. As a test subject for this approach, we chose the maltose binding protein (MBP). Two previous studies provide screening data of large libraries of MBP with randomly inserted effector proteins. Nadler *et al.* randomly inserted a cpGFP, observing the greatest maltose-induced fluorescence changes from insertions at residues 169–171, followed by residues 355–348 [12]. Younger *et al.* inserted a zinc-finger transcription factor finding that insertion at position 335 produced the greatest maltose-induced change in transcription [18]. Agreement in 335 as a good insertion site supports 355's significance and shows that some insertion sites can be used with many different effectors. Therefore, the two regions of interest for our study are 169–171 and 335–348, and our aim was to find properties in molecular dynamics that distinguish these regions from all others.

We performed simulations of the *holo* / “closed” state using a crystal structure (PDB: 1ANF) as a starting point [26]. 10 x 100 ns simulations were conducted (Fig 2A and 2B). Simulations to sample the open state were also performed from a starting point of 1ANF, except

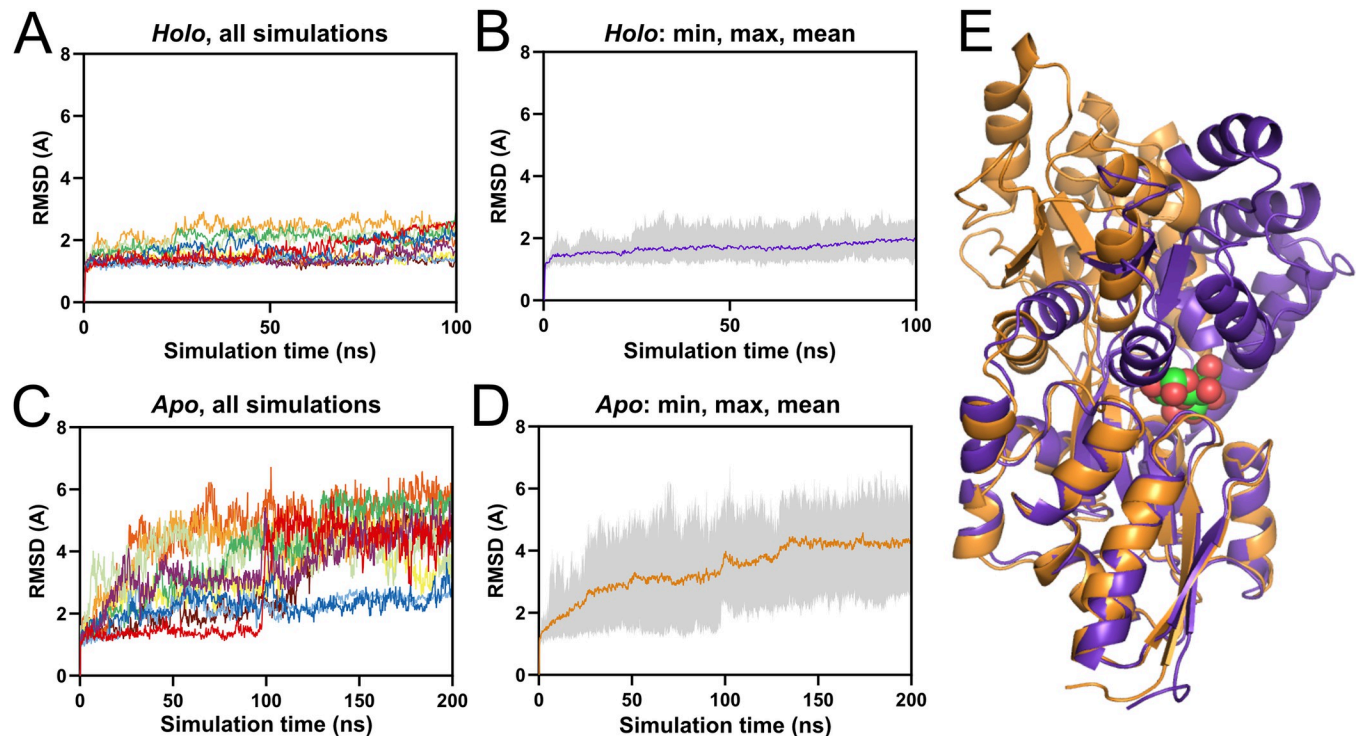


Fig 2. RMSD of molecular dynamics simulations of Maltose Binding Protein (MBP) in *apo* and *holo* forms. A, B: *Holo* simulations were generated from the starting point of the MBP/maltose complex crystal structure. A) Individual traces for 10x100ns MBP *holo* simulations. B) The mean, minimum, and maximum RMSD values for the MBP *holo* simulations. C, D: Apo simulations were generated from the starting point of the MBP/maltose complex crystal structure with maltose removed. C) Individual traces for 10x200ns MBP apo simulations. D) The mean, minimum, and maximum RMSD values for the MBP apo simulations. E) Final poses of MBP in representative *holo*/closed (purple) and *apo*/open (orange) simulations, chosen for RMSD close to the mean-of-means for each condition. Maltose is shown in the *holo* structure in green.

<https://doi.org/10.1371/journal.pcbi.1012212.g002>

with maltose manually removed, and then simulations were run long enough for the protein to transition from this closed starting point to the open state. This was a self-imposed restriction as we were aiming to produce a method that did not require crystal structures of both closed and open states (although both do exist for MBP). We ran 10 x 200 ns of these *apo* simulations, observing much higher RMSDs in most of the simulations (Fig 2C and 2D). Watching these simulations, we could see that the increase in RMSD came solely from a separation of the N- and C-lobes via a hinging motion at the inter-lobe loops (Fig 2E). Among the closed-state simulations we saw that none of the mean RMSDs exceed 2.5 Å from the starting conformation. We therefore chose to define simulations as “closed” when mean RMSD for the last 50 ns is < 2.5 Å and “open” when it is > 2.5 Å. By this definition, we saw the open state in 8/10 of the *apo* simulations. RMSD by comparison to the *apo* crystal structure (PDB: 1OMP) shows that the *apo* simulations evolve to be closer to the *apo* crystal structure, settling at a mean RMSD of ~3 Å (S1 Fig).

The first metrics we investigated were changes in backbone dihedral torsion (Δ Dihe-MD) and root mean squared fluctuation (Δ RMSF-MD) between closed and open simulations. For Δ Dihe at residue i , the dihedral angle between alpha-carbon atoms $c-\alpha_{i-1}$, $c-\alpha_i$, $c-\alpha_{i+1}$, and $c-\alpha_{i+2}$ is taken [10]. Between the MBP 1OMP (*holo*) and 1ANF (*apo*) crystal structures, residue 175 has the greatest Δ Dihe [10], although Nadler and colleagues’ random insertion at residue 171 was found to produce a better sensor [12]. Δ RMSF was chosen to investigate regions of flexibility and conformational change within the protein. Δ Dihe and Δ RMSF were calculated

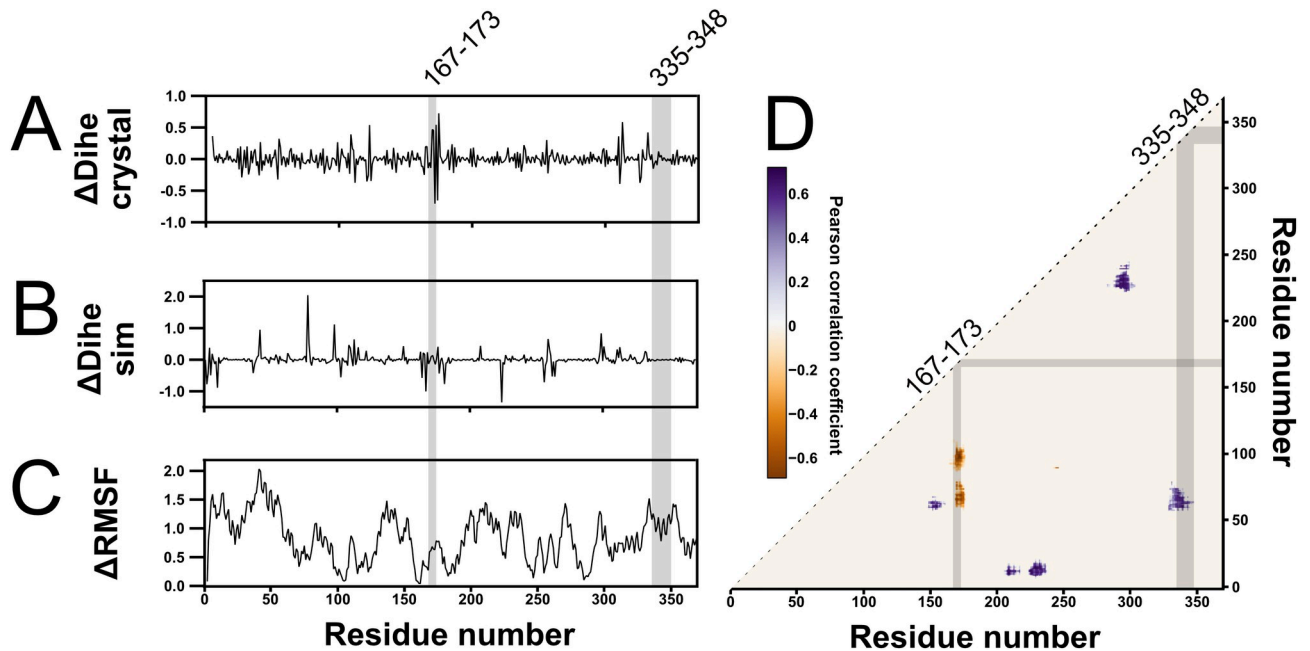


Fig 3. A comparison of metrics attempting to identify regions of MBP primary sequence known to be receptive to effector protein insertion with highlighted regions of interest. Across all figures, the residue ranges 167–173 and 335–348 are highlighted. 167–173 contains the region of interest 169–171, with flanking residues also highlighted for ease of visualisation. **A)** The change in backbone dihedral angle between *holo* and *apo* crystal structures (PDB 1ANF, 1OMP respectively). **B)** Change in backbone dihedral angle between *holo* and *apo* molecular dynamics simulations. Neither regions 167–173 nor 335–348 contain significant values for this metric. **C)** Difference in RMSF of each residue between *holo* and *apo* molecular dynamics simulations. **D)** Pearson correlation coefficient metric for all *apo* MBP simulations where the closed-to-open state transition was observed, filtered to only show peaks in the top percentile for mean, median, and maximum pixel value. Negative Pearson correlation coefficient indicates residues that become closer to each other throughout the closed-to-open transition, and positive values closer.

<https://doi.org/10.1371/journal.pcbi.1012212.g003>

for each residue for each of the final 50 ns of each closed-state simulation and for each of the simulations that reached the open-state, and the mean value in the open-state was subtracted from that of the closed. We compared these to ΔDihe of the crystal structure ($\Delta\text{Dihe-Crys}$). We hoped that at least one of $\Delta\text{Dihe-MD}$ and $\Delta\text{RMSF-MD}$ would highlight our regions of interest 169–171 and 335–348.

Unfortunately, $\Delta\text{Dihe-MD}$ and ΔRMSF were worse at highlighting our regions of interest than $\Delta\text{Dihe-Crys}$ (Fig 3AC). Residues in the region of interest (ROI) 169–171 have some of the largest $\Delta\text{Dihe-Crys}$ values, with 170 and 171 ranking 10th and 11th. $\Delta\text{Dihe-MD}$ and ΔRMSF do not discern residues of this range at all. For the 335–348 range, residues did not stand apart for any measurements in MD or crystal structure. This shows that crystal structures can discern some good insertion sites (169–171) but not all (335–348). This initial analysis of the MD trajectory provided worse sites than using the static crystal structure alone, indicating that these metrics were not sensitive to the kinds of motions required to generate functional sensors.

In order to identify more subtle structural dynamics, we applied the program CONAN (CONTACT ANALYSIS) [27] to our MD data. CONAN produces many metrics on the changes in interactions of a structure throughout a simulation. We submitted our *apo*-MBP simulations to CONAN, thereby providing it dynamic data on the close-to-open transition. Of interest to us was the Pearson correlation coefficient matrix output. This matrix would show us regions that experience changes in contacts between closed and open states, so may be involved in the changes in conformation that distinguish functional insertion sites. We defined a “contact” between residues as beginning when the closest-atom distance is less than 0.5 Å. A contact is only terminated if the closest-atom distance exceeds 0.8 Å (so that momentary losses in contact

are not overly interpreted). The resulting Pearson correlation coefficient matrix has one "pixel" for each residue pair. Positive values indicate interactions that occur early in the simulation (when MBP is closed) and negative values indicate interactions that occur late in the simulation (when MBP is open). To find the most significant hotspots of the Pearson correlation coefficient matrix, first an average-Pearson matrix was produced, with each pixel value being the mean of that pixel's value across the Pearson matrices from each individual simulation (S2 Fig). Pixels were then sorted into "clusters", defined as contiguous areas of positive values (peaks) or negative values (valleys). For each cluster, mean pixel value, median pixel value, and maximum absolute pixel value were calculated. Finally, peaks and valleys were filtered for those that were in the top percentile of mean, median, and max absolute values simultaneously (peaks and valleys ranked separately).

In the resulting mean-filtered Pearson matrix (Fig 3D), it is clear that our regions of interest 169–171 and 335–348 experience some of the most substantial changes in residue contacts during MBP opening in our MD. ROI 169–171 aligns with an area of negative Pearson space, indicating contacts that only occur in the open state. ROI 335–348 aligns with an area of positive Pearson space, indicating contacts that only occur in the closed state. The Pearson matrix calculated from the comparison of the MBP closed and open crystal structures (S3 Fig) has too much noise to discern any ROIs, demonstrating the necessity of MD-generated data. These data provide the first unifying feature of the known insertion hotspots 169–171 and 335–348: changes in contacting residues between closed and open state.

It is particularly of note that this metric captures ROI 335–348, which is not identified by comparison of crystal structures (Fig 3A–3C) but is identified by all random insertion studies to date [12,18]. These studies randomly inserted cpGFP and a zinc-finger transcription factor respectively and both identified residue 335, showing that functional insertion sites can be agnostic to the effector placed there. Another rational insertion study inserted beta-lactamase at 164 to achieve maltose-dependent antibiotic resistance [19], which is at the edge of the 170-centred CONAN hotspot. Change in residue contacts seems therefore to be a consistently important factor for a variety of sensor outputs, namely fluorescence, transcription, and catalysis. From this perspective, we propose that Pearson hotspots therefore correspond to regions where an inserted effector will experience significant changes in residue environment upon substrate binding, and that these changes are likely to significantly modulate effector output.

To validate this rationale for identifying good MBP insertion sites, we used the Pearson data to guide the design of new maltose sensors. We decided to explore the areas of positive Pearson correlation highlighted in our mean-filtered matrix from residue 200 to 250 (Fig 3D). This region is sparsely sampled by random insertions with the transposon-based method, and is not distinguished by crystal structure Δ DiE, so could provide previously unknown sensors.

Sequences were generated for the insertion of cpGFP into MBP every three residues in clusters 205–217 and 224–233. Sensors will be denoted by "MBP_*i*-cpGFP", where *i* is the residue index of the insertion. For linkers, we emulated the design of Nadler *et al.* [12] (nucleotide and amino acid sequences in S1 Text). The resulting sequences were cloned into pET-28a backbone. For negative controls, cpGFP was inserted at regions of flat Pearson space:

MBP_131-cpGFP and **MBP_195-cpGFP**. For positive controls, two very good sensors from Nadler *et al.* were reproduced: **MBP_170-cpGFP** and **MBP_348-cpGFP** [12].

Plasmids for expression of these sensors were transformed into BL21 *E. coli* and induced with IPTG. Pellets were recovered and lysed. Lysates were mixed with a pseudocytosol to simulate eukaryotic cellular conditions. The pseudocytosol was supplemented with either 0 mM, 1 mM, or 100 mM maltose, and the samples were tested for cpGFP fluorescence on a plate reader. Sensors were assayed by two metrics: raw fluorescence intensity (arbitrary units) and $\Delta F/F_0$ (where ΔF is the difference in raw fluorescence intensity between 0 mM maltose and the

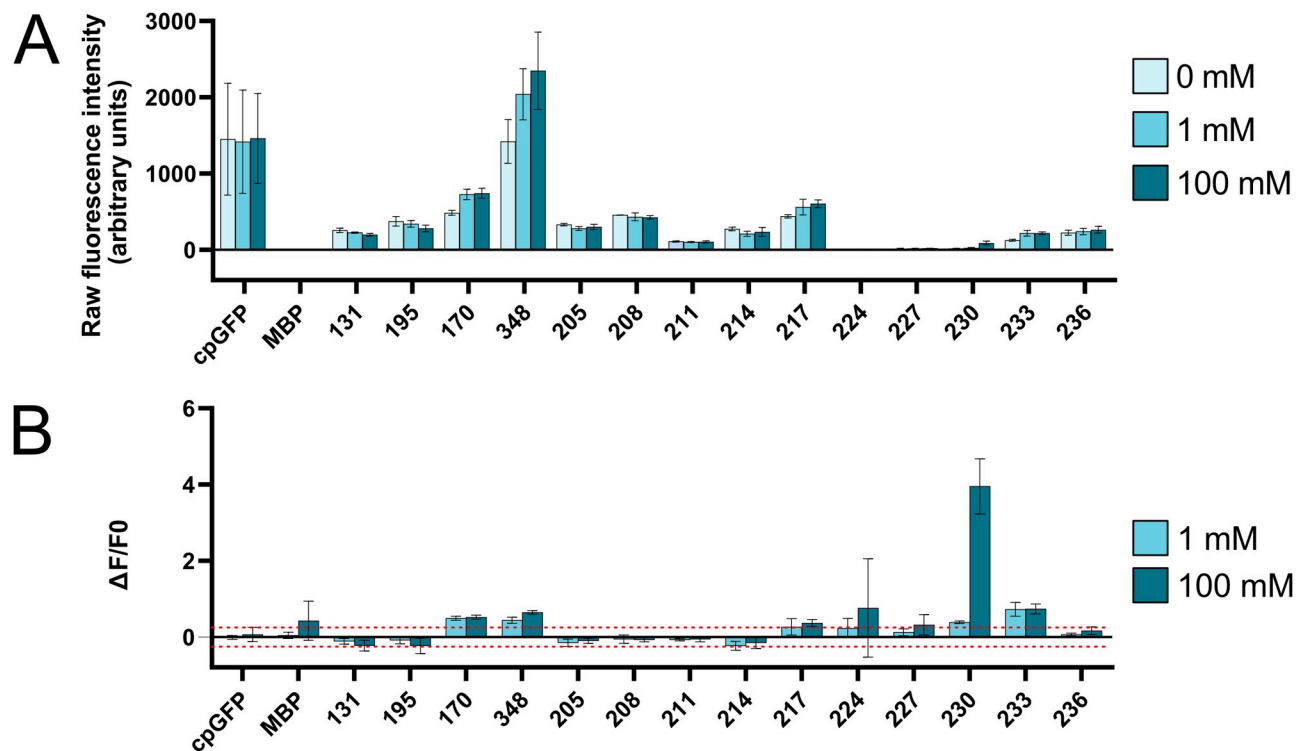


Fig 4. Comparison of fluorescence of designed sensors at varied maltose concentrations. On the left of the axes are the control conditions. cpGFP: the fluorescent protein alone. MBP: the Maltose Binding Protein alone. 131 and 195: cpGFP inserted at MBP's positions 131 and 195, which are both in areas of flat Pearson space so are predicted to generate poor quality sensors. 170 and 348: cpGFP inserted at MBP's positions 170 and 348, both very strong sensors previously identified by Nadler et al. [12]. Presented are the means and standard deviations of 3 biological repeats per condition. Tables for all values are in supplement. **A**) Raw fluorescence of sensors at 0, 1, and 100 mM maltose. **B**) Relative change in fluorescence from 0 mM maltose to 1 or 100 mM, calculated by $(F_0 - F_x) / F_0$. 0.25 and -0.25 are indicated with dashed lines.

<https://doi.org/10.1371/journal.pcbi.1012212.g004>

given concentration, and F_0 is the raw fluorescence intensity at 0 mM). $\Delta F/F_0$ is a key metric in fluorescent sensor design, with an absolute value of > 0.25 being considered a successful design [2]. The absolute $\Delta F/F_0$ values for our negative controls **MBP_131-cpGFP** and **MBP_195-cpGFP** did not exceed 0.25.

By this standard, and when compared to previously discovered sensors, our method has generated several new viable maltose sensors (Fig 4, and S1 and S2 Tables). For example, **MBP_233-cpGFP** had lower raw fluorescence than those of Nadler's **MBP_170-cpGFP** and **MBP_348-cpGFP**, but in terms of $\Delta F/F_0$ **MBP_233-cpGFP** exhibited higher values under our experimental conditions. At concentrations of 1 and 100 mM, the $\Delta F/F_0$ values of **MBP_233-cpGFP** were 0.729 and 0.741, respectively, while for **MBP_170-cpGFP** they were 0.495 and 0.523, and for **MBP_348-cpGFP** they were 0.445 and 0.650. By the " >0.25 " standard, **MBP_233-cpGFP** is certainly a viable sensor.

MBP_230-cpGFP is another sensor of note, with low raw fluorescence but by far the greatest $\Delta F/F_0$. It exhibited very little fluorescence at 0 and 1 mM and was only slightly brighter at 100 mM, with raw measured values of 22, 31, and 72 respectively. However, this small change resulted in a 100 mM $\Delta F/F_0$ of 3.954, by far the largest of those we measured. **MBP_230-cpGFP**'s fluorescence uptick at 100 mM was extremely consistent between experimental repeats, not a result of small fluctuations around a small value, making it another viable sensor, particularly for use in conditions of low background fluorescence. **MBP_227-cpGFP** also achieved $\Delta F/F_0 > 0.25$ for 100 mM, and **MBP_217-cpGFP** did so at 1 and 100 mM, totaling 4 identified viable sensors.

Not all designs were successful. **MBP_224-cpGFP** and **MBP_227-cpGFP** had no fluorescence over all conditions (the apparent large $\Delta F/F_0$ for **MBP_224-cpGFP** at 100 mM is an artefact of small fluctuations). Insertions from 205–214 did not provide viable sensors, with small if any $\Delta F/F_0$ values. Clearly yet further factors than Pearson hotspots determine whether an insertion is functional or not. That said, the aim of our work was not to generate new maltose biosensors, rather it was to develop a computational method for sensor discovery that reconciles observations of previous works in the field. The method's identification of all previously known hotspots is alone proof of its success, and the 4 new viable sensors discovered in previously unknown hotspots are only further proof. Although our discovered sensors are less bright than those of previous studies, brightness is one feature that can be relatively easily improved by directed evolution if one desired to truly apply them.

Conclusion

Modulation of effector domain output upon substrate binding is central to biosensor design. Early PBP-biosensor studies inserted effectors at regions of high backbone dihedral angle change (high ΔDihe), and attributed output modulation to direct transmission of backbone conformational changes from detector to effector. However, random insertion studies have shown that some insertions at sites of low ΔDihe give substantial substrate-induced output changes, showing that ΔDihe the insertion sites that best modulate effector output are sometimes overlooked by this metric. Here we use Pearson correlation coefficient of residue contacts in MD simulations to show that regions of significant contact change during PBP opening correlate well with the known best insertion sites. We then designed new sensors from insertions in regions of high Pearson correlation coefficient and low ΔDihe , validating this as a new method to generate PBP-biosensors.

In this work we have demonstrated that Pearson hotspots correlate with known good insertion sites from more regions of primary sequence better than changes in dihedral angles of crystal structures or MD-derived poses. The Pearson correlation coefficient matrix can be accessible with knowledge of only the *holo* crystal structure, and could even be accessed for PBPs with no known crystal structure thanks to structural prediction algorithms such as AlphaFold2, which are known to mostly predict *holo* states of binding proteins [28]. A wider variety of candidate insertion sites will help with the successful insertions of a wider variety of effectors such as transcription factors and enzymes [18,19]. It would also be interesting to see this method applied to other sensor proteins with similar hinge-like binding dynamics, such as the *de novo* hinge proteins recently developed in the Baker lab [29]. Finally, the most immediate impact of this work is the forging of a new, computational avenue for PBP-biosensor discovery that can be performed at a much greater scale than current wet-lab methods for minimal labour and cost.

Methods

MD simulations

Simulations were run on a computer with a 36 core CPU clocked at 4.2 GHz, 64GB RAM, NVIDIA Quadro RTX 8000 48GB VRAM, 1 TB NVMe, and 2TB SSD. Simulations were set up using AmberTools and performed using the Python package OpenMM [30]. Nonbonding interactions were modelled by PME, with cut-off at 1 nm. Simulations were run at 1 bar with a Monte Carlo barostat and 300 K temperature was maintained by the Langevin integrator with frictional constant 1 ps^{-1} . Hydrogen bond length constraints were applied. A timestep of 2 fs was used. Code for running these simulations can be found at <https://github.com/wells-wood-research/oshea-j-wood-c-pbp-design-2023>.

RMSD, RMSF, dihedral torsion

RMSD and RMSF calculations were performed using the Python package MDAnalysis [31]. Dihedral torsion angles were calculated using the Python package MDTraj [32]. Code for calculating the change in RMSF and dihedral torsion angles can be found at <https://github.com/wells-wood-research/oshea-j-wood-c-pbp-design-2023>.

CONtact ANalysis (CONAN)

CONAN was applied to the 8/10 *apo* simulations where MBP opening occurred. We defined a “contact” as any residue atoms that came within 0.5 Å and then remained within 0.8 Å of each other, as this is well within the limit to exclude water (which has a radius of 1.9 Å). The input file and outputs relevant to this work can be found at <https://github.com/wells-wood-research/oshea-j-wood-c-pbp-design-2023>. For each contact in the Pearson correlation coefficient matrix, the pixel’s mean value across the 8 simulations analysed was taken. These means formed the matrix that was subjected to pixel clustering, the code for clustering was adapted from that of Galloway *et al.* [33] to differentiate between peaks and valleys. For each cluster, mean pixel value, median pixel value, and maximum absolute pixel value was calculated. Peaks and valleys were then filtered for those that were in the >99th percentile of all of these values (peaks and valleys ranked separately).

Design of constructs for expression of sensors

For a given insertion point, linker design followed that of Nadler *et al.* [12]. N-terminal linkers were res^{i+1} -Ala-Ser, where res^{i+1} is the residue after the inserted residue. C-terminal linkers were Ala-Ser- res^i , where res^i is the residue that was inserted at. The amino acid sequence of cpGFP flanked by the linkers was inserted into the amino acid sequence for MBP. The amino acid sequence was codon optimised for expression in *Escherichia coli* using the Twist Bioscience sequence input interface. Optimised sequences were cloned into the pET28a backbone at the SacI_HindIII restriction sites.

Cloning of constructs

Primers for cloning and sequencing were ordered from IDT. Fragments were amplified using Phusion polymerase (NEB), run on 1% agarose gels and recovered using GeneJET Plasmid miniprep kit (ThermoFisher). Recovered fragments were assembled using NEBuilder kit (NEB). Assembled plasmids were sequenced by the Medical Research Council Protein Phosphorylation and Ubiquitylation Unit’s DNA sequencing service at the University of Dundee.

Sensor expression

Plasmids were transformed into BL21-alpha cells (NEB) and plated overnight at 37°C on kanamycin LB plates (50 µg/mL). The next day, colonies were picked into 2 mL of TB supplemented with kanamycin (50 µg/mL), and incubated at 37°C, 190 rpm for 20 hours. 500 µL of these cultures were then used to inoculate 25 mL of LB supplemented with kanamycin (50 µg/mL), and incubated at 37°C, 240 rpm for 1.5 hours. Cultures were then supplemented with 1 mM IPTG, and further incubated at 25°C, 240 rpm for 20 hours.

Fluorescent assay

After expression, cultures were centrifuged at 4000 G, 4°C for 15 minutes. The pellets were recovered and weighed, then resuspended in 1 µL/mg of lysis buffer (1x BugBuster (Novagen), 20 mM Tris-HCl, 150 mM K-gluconate, pH 7.5 [34]). Lysis was performed with shaking at

room temperature, 1600 rpm, 30 minutes. Lysates were spun down at 16000 G for 15 minutes and the supernatant was recovered. 20 μ L of supernatant was transferred to wells in the 96 well black/clear bottom plate (ThermoFisher). 200 μ L of pseudocytosol (100 mM K-gluconate, 30 mM NaCl, 25 mM MES, 25 mM HEPES, 40% sorbitol, 1mg/mL bovine serum albumin, pH 7.5 [34–36]) supplemented with 0, 1, or 100 mM maltose. Plates were immediately imaged on an Infinite M200PRO (Tecan) plate reader, with excitation/emission of 485/515nm, with gain set to 40. In each of the three biological repeats, three technical repeats were performed for each condition and from these the mean was calculated. The mean value of the background negative control pET28a was subtracted from the positive controls and experimental conditions. From this background-subtracted mean value, $\Delta F/F_0$ values were calculated, where $\Delta F = F_X - F_0$, where $X = 1$ or 100. For the MBP negative control, 5 μ g of MBP in 5 μ L (RayBio-tech) was mixed with 15 μ L of lysis buffer and 200 μ L of pseudocytosol with 0, 1, or 100 mM maltose.

Supporting information

S1 Fig. RMSD of the *apo* and *holo* simulations with respect to the maltose binding protein's *apo* crystal structure (PDB: 1OMP).

(TIFF)

S2 Fig. Unfiltered matrix of mean Pearson Correlation Coefficient values for inter-residue contacts over all *apo* simulations where the closed-to-open transition occurred.

(TIFF)

S3 Fig. Pearson Correlation Coefficient matrix of inter-residue contacts calculated from comparison of the available crystal structures of *apo* (1ANF) and *holo* (1OMP) MBP.

(TIFF)

S1 Table. Mean and individual raw fluorescence intensity values for all sensors and controls. “Replicate” refers to “biological replicate”, each being a mean of three technical repeats.

(XLSX)

S2 Table. dF/F_0 values for all sensors and controls.

(XLSX)

S1 Text. FASTA format nucleotide and amino acid sequences used in this study. Maltose binding protein sequence (MBP) is highlighted in blue, circularly permuted GFP sequence (cpGFP) in yellow, and linker sequence between these two components in biosensors in green.

(DOC)

Acknowledgments

Thank you to Dr. Frank Machin of the Doerner lab for direction to literature and Jonathan Lecoy of the Richardson lab for help using their lab space. Thank you to the teams at the Edinburgh Genome Foundry and the Edinburgh Protein Production Facility.

Author Contributions

Conceptualization: Jack M. O’Shea, Peter Doerner, Annis Richardson, Christopher W. Wood.

Data curation: Jack M. O’Shea.

Formal analysis: Jack M. O’Shea.

Funding acquisition: Peter Doerner, Annis Richardson, Christopher W. Wood.

Methodology: Jack M. O'Shea, Christopher W. Wood.

Resources: Peter Doerner.

Software: Jack M. O'Shea, Christopher W. Wood.

Supervision: Peter Doerner, Annis Richardson, Christopher W. Wood.

Validation: Jack M. O'Shea.

Visualization: Jack M. O'Shea.

Writing – original draft: Jack M. O'Shea, Peter Doerner, Christopher W. Wood.

Writing – review & editing: Jack M. O'Shea, Peter Doerner, Annis Richardson, Christopher W. Wood.

References

1. Edwards KA. Periplasmic-binding protein-based biosensors and bioanalytical assay platforms: Advances, considerations, and strategies for optimal utility. *Talanta Open*. 2021 Aug 1; 3:100038.
2. Nasu Y, Shen Y, Kramer L, Campbell RE. Structure- and mechanism-guided design of single fluorescent protein-based biosensors. *Nat Chem Biol*. 2021 May 1; 17(5):509–18. <https://doi.org/10.1038/s41589-020-00718-x> PMID: 33558715
3. Guntas G, Mansell TJ, Kim JR, Ostermeier M. Directed evolution of protein switches and their application to the creation of ligand-binding proteins. *Proc Natl Acad Sci*. 2005 Aug 9; 102(32):11224–9. <https://doi.org/10.1073/pnas.0502673102> PMID: 16061816
4. Taylor ND, Garruss AS, Moretti R, Chan S, Arbing MA, Cascio D, et al. Engineering an allosteric transcription factor to respond to new ligands. *Nat Methods*. 2016 Feb; 13(2):177–83. <https://doi.org/10.1038/nmeth.3696> PMID: 26689263
5. Tavares D, Reimer A, Roy S, Joubin A, Sentchilo V, van der Meer JR. Computational redesign of the *Escherichia coli* ribose-binding protein ligand binding pocket for 1,3-cyclohexanediol and cyclohexanol. *Sci Rep*. 2019 Nov 15; 9(1):16940. <https://doi.org/10.1038/s41598-019-53507-5> PMID: 31729460
6. Li H, Zhang C, Chen X, You H, Lai L. Tailoring *Escherichia coli* Chemotactic Sensing towards Cadmium by Computational Redesign of Ribose-Binding Protein. *mSystems*. 2022 Jan 11; 7(1):e01084–21. <https://doi.org/10.1128/msystems.01084-21> PMID: 35014867
7. Quijcho FA, Ledvina PS. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol Microbiol*. 1996; 20(1):17–25. <https://doi.org/10.1111/j.1365-2958.1996.tb02484.x> PMID: 8861200
8. Felder CB, Graul RC, Lee AY, Merkle HP, Sadee W. The venus flytrap of periplasmic binding proteins: An ancient protein module present in multiple drug receptors. *AAPS PharmSci*. 1999 Jun 1; 1(2):2. <https://doi.org/10.1208/ps010202> PMID: 11741199
9. Alicea I, Marvin JS, Miklos AE, Ellington AD, Looger LL, Schreiter ER. Structure of the *Escherichia coli* Phosphonate Binding Protein PhnD and Rationally Optimized Phosphonate Biosensors. *J Mol Biol*. 2011 Dec 2; 414(3):356–69. <https://doi.org/10.1016/j.jmb.2011.09.047> PMID: 22019591
10. Marvin JS, Schreiter ER, Echevarría IM, Looger LL. A genetically encoded, high-signal-to-noise maltose sensor. *Proteins Struct Funct Bioinforma*. 2011; 79(11):3025–36. <https://doi.org/10.1002/prot.23118> PMID: 21989929
11. Marvin JS, Borghuis BG, Tian L, Cichon J, Harnett MT, Akerboom J, et al. An optimized fluorescent probe for visualizing glutamate neurotransmission. *Nat Methods*. 2013 Feb; 10(2):162–70. <https://doi.org/10.1038/nmeth.2333> PMID: 23314171
12. Nadler DC, Morgan SA, Flamholz A, Kortright KE, Savage DF. Rapid construction of metabolite biosensors using domain-insertion profiling. *Nat Commun*. 2016 Jul 29; 7(1):12266. <https://doi.org/10.1038/ncomms12266> PMID: 27470466
13. Hu H, Wei Y, Wang D, Su N, Chen X, Zhao Y, et al. Glucose monitoring in living cells with single fluorescent protein-based sensors. *RSC Adv*. 2018 Jan 9; 8(5):2485–9. <https://doi.org/10.1039/c7ra11347a> PMID: 35541484

14. Shivange AV, Borden PM, Muthusamy AK, Nichols AL, Bera K, Bao H, et al. Determining the pharmacokinetics of nicotinic drugs in the endoplasmic reticulum using biosensors. *J Gen Physiol*. 2019 Jun 3; 151(6):738–57. <https://doi.org/10.1085/jgp.201812201> PMID: 30718376
15. Borden PM, Zhang P, Shivange AV, Marvin JS, Cichon J, Dan C, et al. A Fast Genetically Encoded Fluorescent Sensor for Faithful in vivo Acetylcholine Detection in Mice, Fish, Worms and Flies [Internet]. Rochester, NY; 2020 [cited 2023 Mar 18]. Available from: <https://papers.ssrn.com/abstract=3554080>
16. Jing M, Li Y, Zeng J, Huang P, Skirzewski M, Kljakic O, et al. An optimized acetylcholine sensor for monitoring in vivo cholinergic activity. *Nat Methods*. 2020 Nov; 17(11):1139–46. <https://doi.org/10.1038/s41592-020-0953-2> PMID: 32989318
17. Younger AKD, Dalvie NC, Rottinghaus AG, Leonard JN. Engineering Modular Biosensors to Confer Metabolite-Responsive Regulation of Transcription. *ACS Synth Biol*. 2017 Feb 17; 6(2):311–25. <https://doi.org/10.1021/acssynbio.6b00184> PMID: 27744683
18. Younger AKD, Su PY, Shepard AJ, Udani SV, Cybulski TR, Tyo KEJ, et al. Development of novel metabolite-responsive transcription factors via transposon-mediated protein fusion. *Protein Eng Des Sel*. 2018 Feb 1; 31(2):55–63. <https://doi.org/10.1093/protein/gzy001> PMID: 29385546
19. Guntas G, Ostermeier M. Creation of an Allosteric Enzyme by Domain Insertion. *J Mol Biol*. 2004 Feb 6; 336(1):263–73. <https://doi.org/10.1016/j.jmb.2003.12.016> PMID: 14741221
20. Ribeiro LF, Nicholes N, Tullman J, Ribeiro LFC, Fuzo CA, Vieira DS, et al. Insertion of a xylanase in xylose binding protein results in a xylose-stimulated xylanase. *Biotechnol Biofuels*. 2015 Aug 15; 8(1):118.
21. Ribeiro LF, Tullman J, Nicholes N, Silva SRB, Vieira DS, Ostermeier M, et al. A xylose-stimulated xylanase–xylose binding protein chimera created by random nonhomologous recombination. *Biotechnol Biofuels*. 2016 Jun 6; 9(1):119. <https://doi.org/10.1186/s13068-016-0529-7> PMID: 27274356
22. Choi JH, Xiong T, Ostermeier M. The interplay between effector binding and allostery in an engineered protein switch. *Protein Sci*. 2016; 25(9):1605–16. <https://doi.org/10.1002/pro.2962> PMID: 27272021
23. Yu Y, Lutz S. Circular permutation: a different way to engineer enzyme structure and function. *Trends Biotechnol*. 2011 Jan 1; 29(1):18–25. <https://doi.org/10.1016/j.tibtech.2010.10.004> PMID: 21087800
24. Ribeiro LF, Amarelle V, Ribeiro LFC, Guazzaroni ME. Converting a Periplasmic Binding Protein into a Synthetic Biosensing Switch through Domain Insertion. *BioMed Res Int*. 2019 Jan 3; 2019:4798793. <https://doi.org/10.1155/2019/4798793> PMID: 30719443
25. Marvin JS, Shimoda Y, Magloire V, Leite M, Kawashima T, Jensen TP, et al. A genetically encoded fluorescent sensor for in vivo imaging of GABA. *Nat Methods*. 2019 Aug; 16(8):763–70. <https://doi.org/10.1038/s41592-019-0471-2> PMID: 31308547
26. Quiocho FA, Spurlino JC, Rodseth LE. Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Struct Lond Engl* 1993. 1997 Aug 15; 5(8):997–1015. [https://doi.org/10.1016/s0969-2126\(97\)00253-0](https://doi.org/10.1016/s0969-2126(97)00253-0) PMID: 9309217
27. Mercadante D, Gräter F, Daday C. CONAN: A Tool to Decode Dynamical Information from Molecular Interaction Maps. *Biophys J*. 2018 Mar 27; 114(6):1267–73. <https://doi.org/10.1016/j.bpj.2018.01.033> PMID: 29590584
28. Saldaño T, Escobedo N, Marchetti J, Zea DJ, Mac Donagh J, Velez Rueda AJ, et al. Impact of protein conformational diversity on AlphaFold predictions. Valencia A, editor. *Bioinformatics*. 2022 May 13; 38(10):2742–8.
29. Praetorius F, Leung PJY, Tessmer MH, Broerman A, Demakis C, Dishman AF, et al. Design of stimulus-responsive two-state hinge proteins. *Science*. 2023 Aug 18; 381(6659):754–60. <https://doi.org/10.1126/science.adg7731> PMID: 37590357
30. Eastman P, Friedrichs MS, Chodera JD, Radmer RJ, Bruns CM, Ku JP, et al. OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J Chem Theory Comput*. 2013 Jan 8; 9(1):461–9. <https://doi.org/10.1021/ct300857j> PMID: 23316124
31. Gowers RJ, Linke M, Barnoud J, Reddy TJE, Melo MN, Seyler SL, et al. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc 15th Python Sci Conf*. 2016;98–105.
32. McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, et al. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J*. 2015 Oct 20; 109(8):1528–32. <https://doi.org/10.1016/j.bpj.2015.08.015> PMID: 26488642
33. Galloway JM, Bray HEV, Shoemark DK, Hodgson LR, Coombs J, Mantell JM, et al. De Novo Designed Peptide and Protein Hairpins Self-Assemble into Sheets and Nanoparticles. *Small*. 2021; 17(10):2100472. <https://doi.org/10.1002/sml.202100472> PMID: 33590708
34. Mukherjee P, Banerjee S, Wheeler A, Ratliff LA, Irigoyen S, Garcia LR, et al. Live imaging of inorganic phosphate in plants with cellular and subcellular resolution. *Plant Physiol*. 2015 Mar; 167(3):628–38. <https://doi.org/10.1104/pp.114.254003> PMID: 25624397

35. Messerli MA, Robinson KR. Cytoplasmic acidification and current influx follow growth pulses of *Lilium longiflorum* pollen tubes. *Plant J.* 1998; 16(1):87–91.
36. Feijó JA, Sainhas J, Hackett GR, Kunkel JG, Hepler PK. Growing pollen tubes possess a constitutive alkaline band in the clear zone and a growth-dependent acidic tip. *J Cell Biol.* 1999 Feb 8; 144(3):483–96. <https://doi.org/10.1083/jcb.144.3.483> PMID: 9971743