Research Article

Exploring Performance Parameters of Artificial Allosteric Protein Switches

Cagla Ergun Ayva [1,2,3], Maria M. Fiorito [1,2,3], Zhong Guo [1,2,3], Selvakumar Edwardraja [5], Joe A. Kaczmarski [1,4], Dejan Gagoski [6,7], Patricia Walden [2,3], Wayne A. Johnston [2,3], Colin J. Jackson [1,4,5,8,9], Tom Nebl [10] and Kirill Alexandrov [1,2,3,11,12]*

*1 - **ARC Centre of Excellence in Synthetic Biology,** Australia*
*2 - **Centre for Agriculture and the Bioeconomy,** Queensland University of Technology, Brisbane, QLD 4001, Australia*
*3 - **School of Biology and Environmental Science,** Queensland University of Technology, Brisbane, QLD 4001, Australia*
*4 - **Research School of Biology,** Australian National University, Canberra, ACT 2601, Australia*
*5 - **Australian Institute for Bioengineering and Nanotechnology,** The University of Queensland, Brisbane, QLD 4072, Australia*
*6 - **Department of Biological Sciences,** Columbia University, New York, NY 10027, USA*
*7 - **Department of Chemistry,** Columbia University, New York, NY 10027, USA*
*8 - **Research School of Chemistry,** Australian National University, Canberra, ACT 2601, Australia*
*9 - **Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science,** Australian National University, Canberra, ACT 2601, Australia*
*10 - **Biology Group,** Biomedical Manufacturing Program, CSIRO, Bayview Ave/Research Way, Clayton, VIC 3168, Australia*
*11 - **CSIRO-QUT Synthetic Biology Alliance,** Brisbane, QLD 4001, Australia*
*12 - **Centre for Genomics and Personalised Health,** Queensland University of Technology, Brisbane, QLD 4001, Australia*

Correspondence to Kirill Alexandrov: Centre for Agriculture and the Bioeconomy, Queensland University of Technology, Brisbane, QLD 4001, Australia. *kirill.alexandrov@qut.edu.au (K. Alexandrov) @Jackson_Lab (C.J. Jackson)*
https://doi.org/10.1016/j.jmb.2022.167678
Edited by Igor Berezovsky

Abstract

Biological information processing networks rely on allosteric protein switches that dynamically interconvert biological signals. Construction of their artificial analogues is a central goal of synthetic biology and bioengineering. Receptor domain insertion is one of the leading methods for constructing chimeric protein switches. Here we present an *in vitro* expression-based platform for the analysis of chimeric protein libraries for which traditional cell survival or cytometric high throughput assays are not applicable. We utilise this platform to screen a focused library of chimeras between PQQ-glucose dehydrogenase and calmodulin. Using this approach, we identified 50 chimeras (approximately 23% of the library) that were activated by calmodulin-binding peptides. We analysed performance parameters of the active chimeras and demonstrated that their dynamic range and response times are anticorrelated, pointing to the existence of an inherent thermodynamic trade-off. We show that the structure of the ligand peptide affects both the response and activation kinetics of the biosensors suggesting that the structure of a ligand:receptor complex can influence the chimera's activation pathway. In order to understand the extent of structural changes in the reporter protein induced by the receptor domains, we have analysed one of the chimeric molecules by CD spectroscopy and hydrogen–deuterium exchange mass spectrometry. We concluded that subtle ligand-induced changes in the receptor domain propagated into the GDH domain and affected residues important for substrate and

cofactor binding. Finally, we used one of the identified chimeras to construct a two-component rapamycin biosensor and demonstrated that core switch optimisation translated into improved biosensor performance.

Introduction

The ability to quantitatively interconvert different types of biochemical and physical signals is a hallmark of all biological processes. Allosterically regulated proteins play a key role in such interconversion. Recent years have seen increased efforts to design artificial allosteric protein systems for applications such as molecular probes, bioengineering modules, and diagnostic reagents.[1] While to the best of our knowledge, no artificial protein switches have been commercially exploited as diagnostics tools, the allosteric protein-based probes, often referred to as protein biosensors, have revolutionised many areas of cell biology and neurobiology.[2] The majority of these biosensors are based on chimeras composed of a ligand binding and a reporter domain, with the latter frequently being a fluorescent protein or an enzyme.[3]

A recent review of the fluorescent biosensors proposed an empirical rule that practically useful biosensors should have a dynamic range of at least 10-fold and retain at least 60% of the parental reporter domain activity.[4] While this may not be a universal rule for all biosensors, it points to two critical parameters of biosensor performance to which we would also like to add a third: the response time. The latter should be significantly faster than the biological event one is trying to quantify or the time available for the assay (for instance, the length of doctor's appointment). It is also expected that longer assays are more susceptible to changes in environmental conditions increasing system's noise.

Given that domain insertion is responsible for the evolutionary emergence of nearly 1/10th of all multidomain proteins, it is not surprising that a similar strategy can be successfully utilised in protein engineering.[5] Here, receptor domain insertion converts constitutively active reporter domains into ligand-operated allosteric protein switches.[6] Apo- and holo- form of the inserted receptor exerts different influences on the reporter that may be a result of local or global conformational changes, reversible structural disorder, control of the oligomeric state or of structure dynamics.[1,7,8] Conformation change and reversible structural disorder have many functional similarities and hence have been extensively employed for switch construction. Introduction of global structural disorder is likely to negatively influence the overall stability of the switch. Therefore, particularly when *in vitro* applications are intended, chimeric switches with local conformation changes are preferred (Figure 1(A and B)). The success in construction of such biosensors is expected to increase when receptor domains with large conformational changes are employed. A recent bioinformatic study revealed, however, that such domains are not very common, calling for the development of more generic approaches for the construction of allosteric protein switches.[9] We have recently proposed an alternative approach where initially a reporter:receptor chimera regulated by a ligand peptide is constructed.[10,11] This allows the systems' specificity to be outsourced onto the external ligand-binding domains that control the concentration of the effector peptide in the vicinity of the chimeric reporter (Figure 1(C)). This approach enables utilisation of two-component biosensor architectures based on binders without conformational changes and is applicable to analytes ranging from small molecules to large biopolymers.[12]

Previously, we and others have identified calmodulin (CaM) as a peptide-operated receptor domain with a large conformational change that is an effective regulator of protein chimeras.[11,13] We demonstrated that calmodulin chimeras with PQQ-glucose dehydrogenase (PQQ-GDH) act as electrochemical molecular switches and can be used to build bioelectronic and biosensors applications.[12,14] As the original chimeras were created by an empirical insertion site selection, it remains unclear to what extent the obtained dynamic range, rate of response, and catalytic activity represent features of a particular design or are governed by the properties of the parental GDH domain. In this study, we devise an approach that allows biochemical analysis of GDH:calmodulin (GDH-CaM) chimera libraries quantifying their performance parameters such as dynamic range, catalytic activity and the rate of response.

Results

In vitro pipeline for production and analysis of protein chimera libraries

In order to understand the parameters governing the emergence of synthetic allosteric regulation in GDH-CaM chimeras, we sought to quantitatively analyse a significant number of such chimeras. However, GDH activity cannot be easily converted into a cellular survival assay or a fluorescence-activated sorting assay that allows the analysis of large variant libraries. Therefore, we set out to develop an approach that allows multiplexed

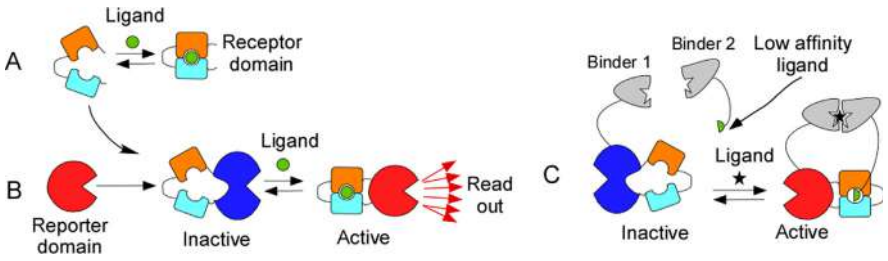

Figure 1. ***Rational design of modular protein biosensor architectures by domain insertion***. (A) A ligand-binding receptor domain (shown in orange and turquoise) that undergoes a conformational change upon ligand (green circle) binding. (B) Construction of a chimeric allosteric switch where the ligand-binding domain stabilises the reporter domain in an inactive conformation. Ligand-induced conformational changes in the receptor domain result in the structural rearrangement of the reporter domain and its activation. (C) A two-component biosensor architecture based on the artificial allosteric switch shown in B where a low-affinity ligand of the chimeric switch molecule is scaffolded by the target analyte resulting in its high local concentration that leads to activation of the reporter domain.

recombinant expression of GDH mutants and their biochemical analysis without the need for protein purification.

Cell-free protein expression systems are powerful tools for rapid and high-throughput production of large polypeptide libraries. They can be coupled with biochemical assays allowing rapid analysis of gene products within hours.[15] For the analysis of the GDH-CaM insertion library, we chose a *Leishmania tarentolae*-based eukaryotic cell-free expression system, LTE.[16,17] The LTE system has several advantages over commonly used *E. coli*-based cell-free systems, such as low protein aggregation propensity and high protein integrity that is particularly pronounced in the case of larger proteins.[18] To test the suitability of the LTE system for GDH-CaM expression and activity analysis, we expressed the N-terminally EGFP-tagged version of wild-type *Acinetobacter calcoaceticus* PQQ-GDH and GDH-CaM-403S chimera that was developed in an earlier study.[11] As shown in Figure 2(A), *in vitro* translated PQQ-GDH and GDH-CaM-403S chimera could be identified on an SDS-PAGE gel as a single fluorescent band indicating integrity of the translation product. Biochemical analysis demonstrated that GDH activity could be reliably quantified in as little as 1–2 μL of translation reaction. Using this assay, we were able to clearly distinguish activity of GDH-CaM-403S chimera in the presence and absence of the calmodulin-binding peptide (CaM-BP) (Figure 2(B)). To obtain a more quantitative activity measure of *in vitro* produced proteins we sought to obtain a reliable reference by performing kinetic analysis of purified wild-type GDH and the ligand-saturated GDH-CaM-403S chimera (Figure S1). The obtained k_{cat} values were 2444 s^{-1} and 785 s^{-1} respectively and the activity of the wild-type GDH enzyme matched closely the previously published data.[19,20] We then compared the observed reaction rates of the identical concentrations *in vitro* produced wild-type GDH (k_{obs}: 0.51 min^{-1}) to GDH-CaM-403S chimera (k_{obs}: 0.16 min^{-1}) and concluded that it retained ~30% catalytic activity. Therefore, both methods gave us comparable results.

With the availability of a simple, quantitative and multiplexable protein production and biochemical analysis platform, we sought to find an approach for the equally efficient production of DNA templates. To this end, we chose the rolling circle DNA amplification (RCA) that can be used to amplify plasmid DNA rapidly and directly from *E. coli* colonies and use it directly to prime *in vitro* translation reactions.[21–23] This provides a work-around of the time-consuming conventional plasmid DNA isolation methods that rely on liquid–solid phase separation and represent a bottleneck for high-throughput protein expression. This is particularly important for automated protocols where the ability to perform template preparation using only solution-based methods enables significant protocol simplification.

To assess the performance of RCA-products as *in vitro* translation templates, we used a single colony of *E. coli* transformed with EGPF-GDH-CaM-403S construct for template amplification. When used in LTE cell-free expression, both colony RCA (cRCA) and plasmid RCA (pRCA)-amplified templates produced similar yield and quality of protein compared to mini-prep purified plasmid (Figure S2(A)). Moreover, the GDH activity assay confirmed that reaction kinetics and affinity of CaM binding M13 peptide for GDH-CaM produced by RCA-primed cell-free reactions were similar to that of purified recombinant protein (Figure S2(B and C)). Therefore, we concluded that we developed a pipeline of sufficient throughput to be able to screen libraries in the order of 10^3 (Figure 2(E)).

Comparative analysis of the GDH-CaM chimeras

The developed assay platform allowed us to biochemically characterise a library of GDH-CaM chimeras in an effort to understand the

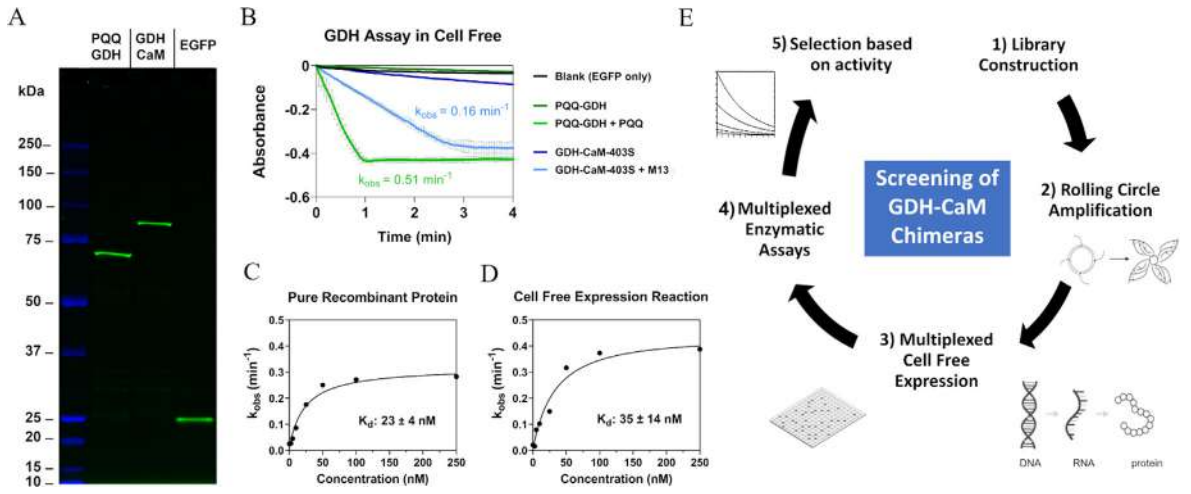

Figure 2. ***Development of cell-free biosensor expression and testing platform.*** **(A)** SDS-PAGE analysis of EGFP-GDH (molecular weight: ~78 kDa) and EGFP-GDH-CaM-403S (molecular weight: ~96 kDa) fusions expressed in *Leishmania tarentolae* cell-free system. The *in vitro* translation reactions primed with colony-templated RCA (cRCA) were resolved on the PAGE gel without boiling and photographed in UV light. **(B)** Time-resolved changes in absorbance of electron-accepting dye DCPIP in the presence of cell-free expressed PQQ-GDH protein with or without PQQ and GDH-CaM-403S protein with or without M13 calmodulin-binding peptide (CaM-BP). The observed reaction rates (k_{obs}) were calculated from three independent experiments and are shown as mean ± SD (dashed lines). **(C)** and **(D)** represent plots of initial reaction rates of 10 nM purified recombinant GDH-CaM-403S protein in the presence of increasing concentrations of M13 CaM-BP **(C)** or of protein expressed by RCA-primed LTE cell-free *in vitro* translation **(D)**. Fitting of the data to the quadratic equation (Supplementary Equation 1) led to the K_d values of 23 ± 4 nM and 35 ± 14 nM, respectively (Figure S2(B and C)). **(E)** Schematic representation of the chimera library construction and screening process.

relationship between the calmodulin insertion site and the emergence of synthetic allostery. In particular, we wanted to understand whether the switch characteristics such as the catalytic rate in fully activated state, the dynamic range, and its response rate are independent or correlated.

To this end, we designed and constructed a mixed gene library containing 221 GDH gene variants with CaM domain coding sequence placed into all solvent-exposed loop residues (Figure 3(A)). The library was cloned into pCellFree vector and transformed into DH5α *E. coli* strain. The previously characterised GDH-CaM-403S chimera was used as an internal activity control.[11] The library was *in vitro* expressed and its activity individual GDH-CaM chimeras was analysed in the presence and absence of CaM-BP using the workflow depicted in Figure 2(E). A typical output of such a screening campaign is shown in the Figure S3. In order to archive 99% library coverage 1018 clones needed to be analysed.[24] Out of nearly 1000 *in vitro* expressed and activity-tested clones, approximately 50 clones (ca. 5% of the screened clones) displayed CaM-BP-dependent activity. This corresponds to 23% of the total GDH-CaM library consisting of 221 chimeras. Clones that displayed CaM-BP-dependent activity were sequenced and further analysed for dynamic range, response time and observed catalytic rate and the extrapolated k_{cat} (Table S1). The data are summarised in Figure 3(B), demonstrating that more than half of the analysed clones displayed similar or faster k_{obs} compared to the control GDH-CaM-403S chimera. For mutants 180H, 191D, 204L, 216N, 217G, 237P, 387D, 402K, and 414P, both the dynamic range and the maximal catalytic rate were close to or greater than that of the control. Chimeras GDH-CaM-216N and 387D displayed half response time to around 5 minutes which is significantly shorter than that of the control chimera GDH-CaM-403S that had a longer half response time of 15 minutes. Considering all these three parameters of biosensor performance, the chimeras 216N, 387D and 414P stood out as switch modules with superior properties (Figure 3(C)).

The data set obtained in our screening campaign were sufficiently large (n = 41) to analyse the correlation among the performance parameters. We used Pearson's correlations as a statistical method to analyse the relationship among the developed biosensors' dynamic range, the response time, and the reaction rate. A statistically significant negative correlation (Pearson $r = -0.34$) between the dynamic range and response time was detected, meaning that chimeras with large dynamic range tend to require more time to reach full activity (Figure 3(D)). Several chimeras such as 180H and 237P displayed both large dynamic range and high reaction rate; however, these chimeras had long

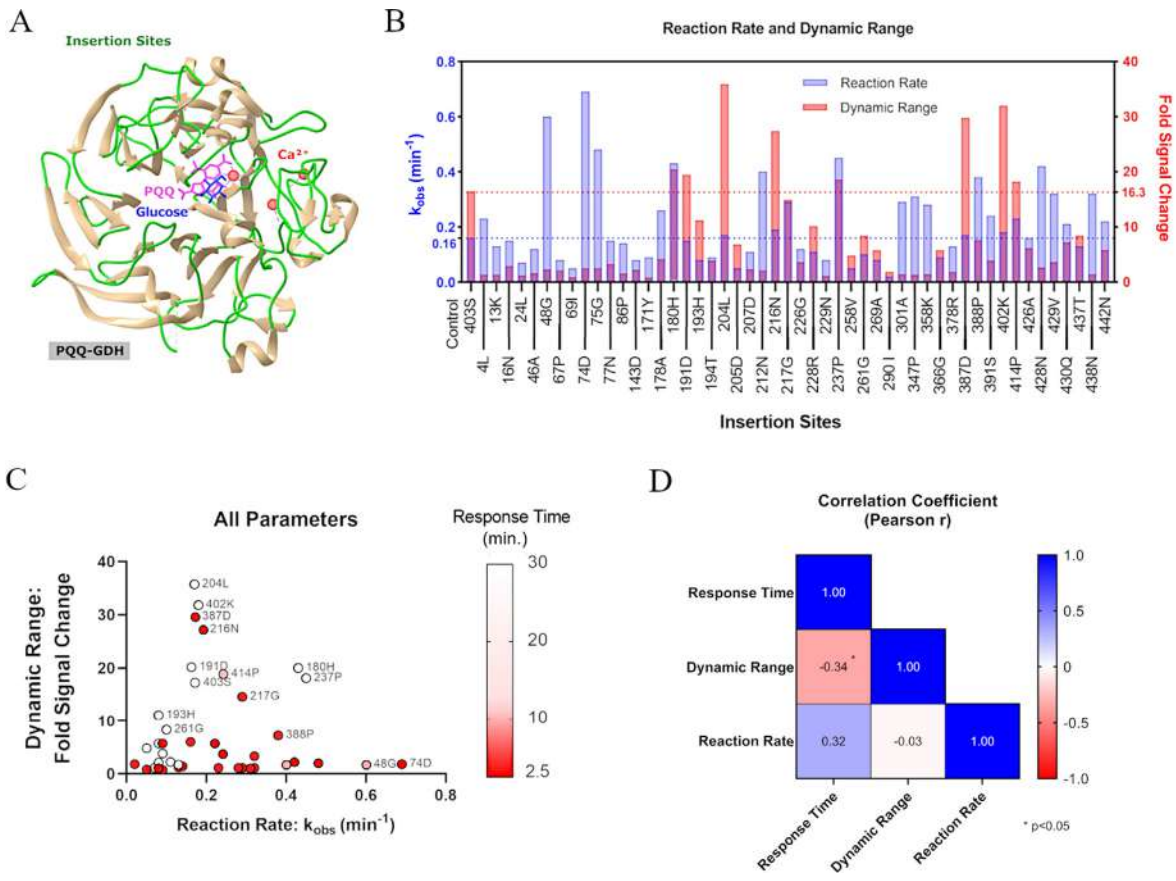

Figure 3. ***Design and analysis of the GDH-CaM chimera library.*** **(A)** Ribbon representation of PQQ-GDH structure with 221 CaM insertion sites coloured in green. Glucose (dark blue), cofactor PQQ (pink), and Ca^{2+} ions (red) are displayed in ball and stick representation. **(B)** GDH activity of the CaM-BP dependent mutants are plotted according to their observed reaction rate, k_{obs} (min^{-1}) and dynamic range (fold signal change). Horizontal dashed lines represent the activity of the internal control, GDH-CaM-403S, where the blue line is for reaction rate and the red line is for dynamic range. **(C)** A plot of reaction rate, dynamic range, and the response rate of the identified chimeras. The gradient red colour's intensity codes for response time, where red indicates the fastest response and white indicates the slowest response. **(D)** Statistical analysis of the three parameters: response time, dynamic range, and reaction rate using Pearson's correlations (sample size, n = 41). Correlation matrix with a heat map shows the Pearson r values for each pair ($p < 0.05$ labelled with an asterisk, *).

response times (Figure 3(C)). These results suggest that simultaneous optimisation of all three parameters may not be possible, at least in the case of GDH-CaM chimera.

Effect of CaM-BP sequence on the activity of the GDH-CaM chimeras

Calmodulin controls a multitude of processes in eukaryotic cells by interacting with potentially hundreds of calmodulin-binding peptides (CaM-BPs).[25] CaM-BPs often stabilise very different conformations of calmodulin and, therefore, potentially could have different effects on CaM-based chimeras.[26] While our initial screening was based on the prototypical M13 CaM-BP peptide, we were interested to explore the allosteric landscape along the axis of different CaM insertion sides and CaM-BP sequences.[27] To this end, we chose a subset of CaM-BPs that have high-resolution crystal structures solved in complex with CaM (Table S2). We tested such CaM-BPs for their ability to activate GDH-CaM chimeras at concentrations that were expected to be at least 10-fold above the K_d (Table S2 and Table S3). As shown in Figure 4 (A), BP2, BP12, BP22, BP26, and BP29 were able to stimulate GDH activity of GDH-CaM-403S to the levels comparable with M13 peptide (BP1). It should be noted that the M13 peptide (BP1) and its higher affinity version (BP2) induced the highest enzymatic activity active in most of the mutants.[28,29] In order to rationalise our findings, we analysed the available crystal structures of these peptides in complex with calmodulin and measured the distance between C-alpha atoms of N-terminal Thr 5 residue and C-terminal Thr146 residue of calmodulin in complex as a proxy of the conformational strain that such complexes

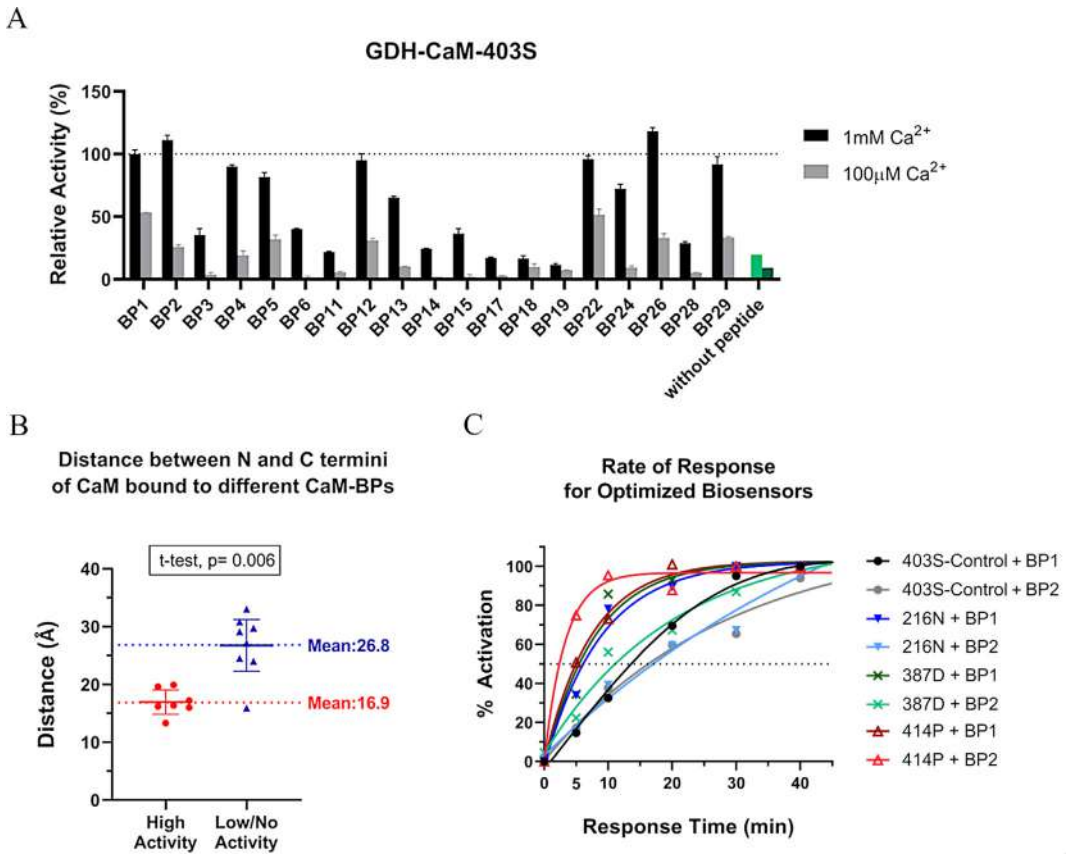

Figure 4. ***Effect of the calmodulin-binding peptides on the activity of GDH-CaM chimeras.*** **(A)** Relative activity of GDH-CaM-403S chimera activated with various calmodulin-binding peptides. The activity induced by M13 peptide (denoted here as BP1) was considered as 100% and is indicated as a horizontal dashed line. The results for the complete set of 29 peptides are provided in Figure S3. The activity for two different calcium concentrations is shown at 1 mM (black bars) and 100 μM (grey bars). Data are given as mean ± SEM (error bars) of two replicates. **(B)** Plot of the distance between C-alpha atoms of N-terminal Thr5 residue and C-terminal Thr146 residue in crystal structures of CaM bound to different CaM-BP's. Chimeras were grouped by activity levels where 'High Activity' peptides have above 50% and 'Low/No Activity' have less than 50% relative activity. Data are given as mean with a 95% confidence interval, and the two groups were compared with unpaired t-test (p = 0.006). **(C)** A plot of GDH activity for various GDH-CaM chimeras as a function of incubation time with CaM-BP. Activation is defined as percentage of the highest k_{obs} for a particular reaction after 0, 5, 10, 20, 30, and 40 minutes of incubation. The plot was fitted as a one-phase exponential.

exert on GDH. The distances were correlated with the results of the activity analysis. As shown in Figure 4(B), this analysis revealed that in all but one case, CaM-BPs that stabilise CaM conformations with N- and C termini located closer together resulted in high catalytic activity of GDH-CaM chimeras. To an extent, this finding is not surprising as the closely located termini are expected to promote restoration of the near-native conformation of GDH. Furthermore, the chimeras were selected for their response to the M13 peptide that induces classical closed conformation of CaM with proximal N and C -termini and hence one can expect that peptides that phenocopy this conformation will stimulate chimera's activity.

Finally, we wanted to explore if the observed pattern would be maintained in other GDH-CaM chimeras identified in our screen. To this end, we tested the CaM-BP panel against GDH-CaM-414P and GDH-CaM-261G chimeras. GDH-CaM-414P has an optimal performance when considering catalytic activity, dynamic range and the activation rate, whereas GDH-CaM-261G that perform less well could potentially be improved through the use of different CaM-BP's. The obtained activity profiles were similar to that of GDH-CaM-403S, with some variations in maximal activity (Figure S4). The combination of the CaM-BP and chimera also influenced the rate at which the latter reached its maximal catalytic activity. For instance, as shown in Figure S5, GDH-CaM insertion mutants with faster response times display lower dynamic ranges (i.e., mutants with 74D, 212 N, 347P, and 428 N insertion sites). The GDH-CaM-414P chimera combined with BP2 peptide has 2-fold higher reaction rate (SI-

Figure 3) and faster response time (Figure 4(C) and Figure S6) compared to the control GDH-CaM-403S:BP1. The fastest half response time of 2.5 minutes was observed in GDH-CaM-414P:BP2 whereas it was 5 minutes with GDH-CaM-414P:BP1 (Figure 4(C)), indicating that slight differences in ligand sequence can lead to large differences in chimera's response times.

Understanding CaM-BP-induced changes in GDH-CaM chimera structure

Given the nature of peptide-induced conformation change of CaM, where the distance between termini decreases significantly in the complex with the ligand, it is intuitively obvious how the developed chimeras might operate. However,

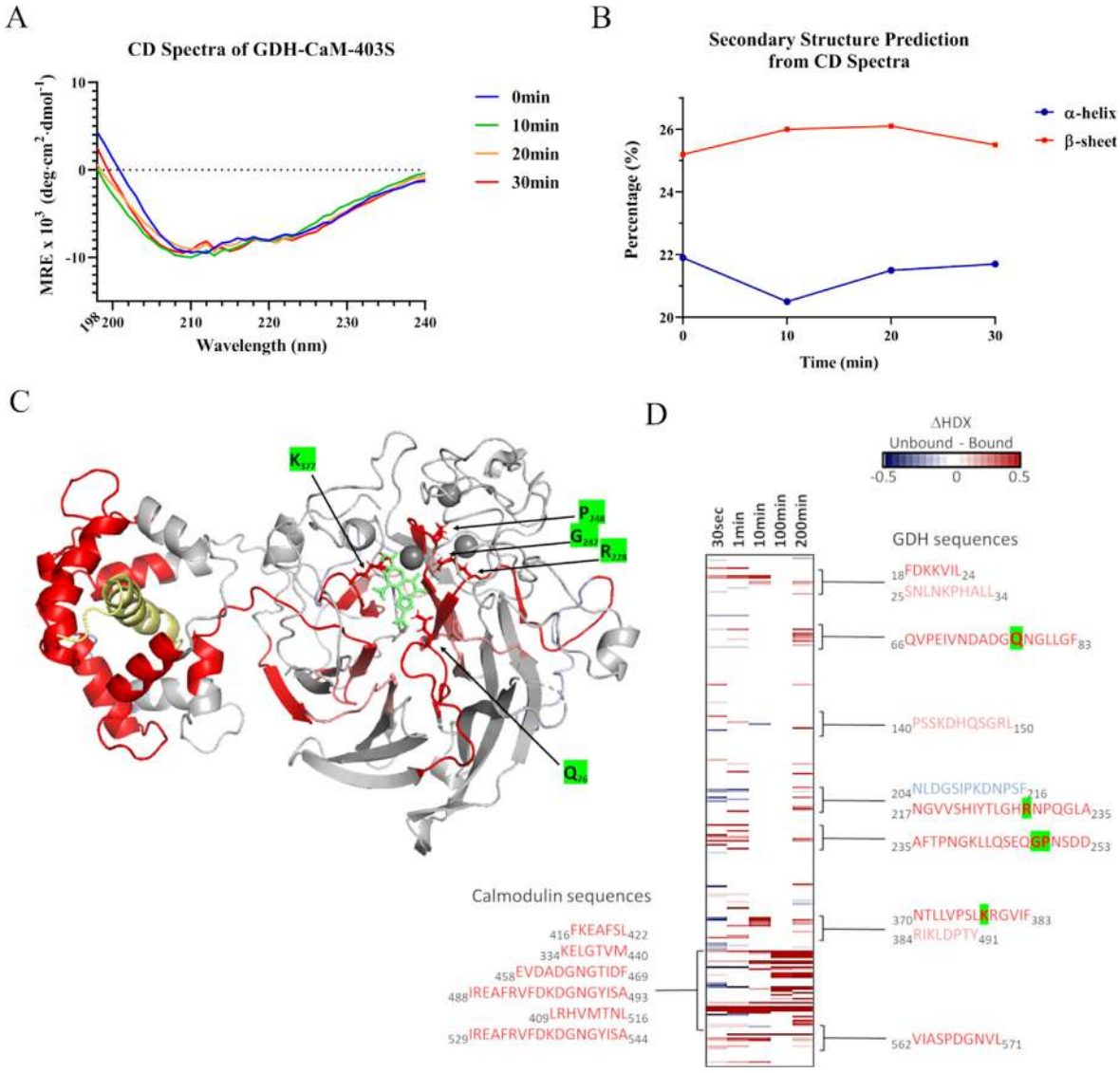

Figure 5. ***Analysis of CaM-BP induced structural changes in GDH-CaM chimera.*** **(A)** Circular dichroism spectra of recombinant GDH-CaM-403S chimera incubated with M13 peptide for different periods of time. The scan was performed after the addition of peptide and spectra were recorded at 0 min, 10 min, 20 min and 30 min. **(B)** Prediction of time-dependent change in the secondary structure content of GDH-CaM-403S after addition of M13 peptide. **(C)** Segments showing significant differences in deuterium uptake were mapped onto the 3D ribbon model of M13 bound CaM-GDH as red colouring. The co-factor interacting residues are shown as in (**D**). **(D)** The Chiclet plot shows peptide sequences as horizontal lines, and statistically significant differences in the exchange rate (i.e. delta HDX unbound–bound) are visualised from −0.5 Da (blue shades = deprotection in the bound state) to +0.5 Da (red shades = protection in the bound state) for each time point (30 sec, 1, 10, 100, and 200 min). The white colour signifies states with no statistical difference. The resulting heatmap shows peptide sequences arranged from N- (top) to C terminus (bottom), visualising 'clusters' of overlapping peptide sequences with a similar change in deuterium uptake. The residues highlighted in green (Q76, R228, G247, P248, and K377) are in close proximity to or interacting with PQQ.

what remains unclear is the extent to which CaM perturbs GDH structure and how these perturbations translate into changes in GDH activity. Therefore, we performed CD spectroscopy to investigate changes in the secondary structure of GDH-CaM chimera in response to CaM-BP binding. As can be seen in Figure 5(A and B), addition of M13 CaM-BP to the solution GDH-CaM-403S resulted in small changes in α-helical part of the spectra by the 10-minute time point with no further obvious changes occurring afterwards. Given that CaM:CaM-BP complexes are predominantly α-helical, this change may reflect their rearrangement upon complex formation. This experiment suggests that CaM-BP induced conformational changes in the calmodulin domain do not induce significant structural rearrangements in GDH reporter domains. This is in line with the behaviour of other artificial allosteric chimeras reported earlier.[30]

In order to gain insight into more subtle and local structural rearrangements of the chimera, we utilised time-resolved hydrogen–deuterium exchange mass spectrometry (HDX-MS). HDX-MS allows time-resolved quantification of dynamic changes in secondary structures which can be applied to large protein systems.[31,32] Analysis of time-dependent changes in deuterium uptake of GDH-CaM-403 chimera upon addition of M13 peptide revealed that, as expected, the calmodulin domain of the chimera underwent large conformational transition resulting in significant protection of calmodulin sequences involved in peptide binding (Figure 5(C and D)). However, we were also able to detect small, but statistically significant differences in deuterium uptake of GDH peptide sequences containing residues of the active site. For instance, there were significant differences in deuterium uptake of sequences containing residues G247, P248, and K377 located in the cofactor PQQ binding pocket that were detectable already within the first 30 seconds after peptide addition. Two former residues also play a role in the coordination of the Ca^{2+} ion in the active site, and therefore their dislocation is expected to impact the enzyme's activity. Additionally, slower structural changes occurred in sequences containing residues Q76 and R228 involved in glucose binding (Figure S8 (D)). Overall analysis reveals that the peptide binding to CaM domain leads to rearrangement of multiple elements of secondary structure, leading to changes in solvent accessibility of GDH residues Q76, R228, G247, P248, and K377 involved in coordination of PQQ and/or glucose (Figure 5(C and D) and Figure S8).

Is response kinetics of GDH-CaM affected by the choice of the GDH orthologue?

One of the caveats of the presented experimental design and its conclusions relates to the fact that chosen protein domains represent a tiny fraction of possible structure/functional space. It is unknown how the assessed properties (stability of the chimeras, catalytic activity and dynamic range) vary among homologous structures. Earlier findings suggested that closely related chimeras of identical topology do not display conservation in artificial allostery.[33] We wanted to gather first indications whether artificial switches based on the orthologous domains display similar performance parameters. To this end, we generated a sequence similarity network (SSN) that comprised sequences of 8,590 proteins containing the GDH domain clustered according to alignment scores (Figure S9). We picked sequences of ten orthologs from within the Cluster 23 that contains *Acinetobacter calcoaceticus* PQQ-GDH that was originally used for chimera library construction. The proteins were produced recombinantly in *E.coli* and several parameters were analysed using established GDH activity assay. In our assessment we considered efficiency of recombinant expression, catalytic rate, substrate specificity and activity dependence on Ca^{2+} ion (Figures S10 and S11). Based on this analysis we selected *Deinococcus ficus* PQQ-GDH homologue that displayed the highest catalytic activity in the set and had no discernible Ca^{2+} dependence (Figure S10). This ortholog displays 45% sequence identity to the *A. calcoaceticus* PQQ-GDH. We used this orthologue to construct a focused library of GDH-CaM chimeras using a design approach similar to the earlier described library but with lower domain insertion frequency. This allowed us to decrease the overall size of the library and produce it in the arrayed format that allowed direct correlation of the genotype and phenotype. We used *in vitro* translation/activity analysis assay to test 42 chimeras that led to the identification of 3 allosteric variants (Figure S10). The observed frequency of the allosteric variant recovery was below 8% and hence nearly three times lower than in the case of *A. calcoaceticus* GDH-based library. The response times of the chimeras based on the chosen ortholog GDH were similar to those based on *Acinetobacter calcoaceticus* (Figure S11). The alignment of *A. calcoaceticus* and the *D. ficus PQQ-GDH* homologues revealed that co-location of the CaM insertion did not result in coordinated emergence of allosteric chimeras (Figure S16). While a larger study will be needed to draw a definitive conclusion about frequency of allosteric switch emergence and their kinetic parameters these observations may suggest that the variation in overall fold stability and structure of the reporter domain determine properties thereon based artificial switches.

Construction of two-component GDH-CaM biosensors using developed switch modules

One of the motivations for construction of GDH-CaM chimeras relates to their use in modular biosensor systems. Here, GDH-CaM and CaM in fusion with low affinity version of CaM-BP are

connected to ligand-binding domains (Figure 6(A)). Ligand-mediated scaffolding results in an increase of local concentrations of the components leading to association of CaM-BP with the reporter chimera and its activation. While modular and sensitive, the original versions of these biosensors suffered from a slow response rate that was traced back to slow activation of the core GDH-CaM switch.[34] Having now identified CaM-GDH: CaM-BP pairs with faster response rates, we decided to test if their integration into two-component biosensor architecture would result in systems with faster response. To this end, we constructed a two-component rapamycin biosensor that is controlled by rapamycin-dependent association of FKBP and FRB domains.[35,11] We used GDH-CaM-414P and a truncated version of BP2 (Table S4) as the components of the core switch. The resulting biosensor components were produced recombinantly in *E. coli* and analysed for activity in comparison with the original rapamycin biosensor based on GDH-CaM-403S and truncated BP1. Our results demonstrated that utilisation of a faster core switch resulted in the reduction of half response time from 30 minutes to 10 minutes (Figure 6(B)). However, the overall affinity of the biosensor for rapamycin for its ligand has decreased (Figure S15), pointing to the need for further optimisation of the system. These data could not be fitted to an exponential curve due to, presumably, the kinetic complexity of two-component systems.

Discussion

Domain insertion is one of the most successful approaches for conversion of constitutively active proteins into artificial allosteric systems. This approach is more successful when utilising a receptor domain with conformational changes sufficiently large to perturb the structure of the reporter domain, thereby creating an OFF state. Ligand-mediated conformation change of the receptor results in the re-activation and emergence of the ON state. However, the paucity of receptors with sufficiently large conformational changes calls for more reliable approaches for artificial allosteric system construction. As a workaround for this problem, we previously proposed an approach for separating the binding and confirmation change functions by utilising receptors with peptide ligands that can be used in modular higher order biosensor architectures.[11] We demonstrated that calmodulin and calmodulin-binding peptide possess appropriate properties for such universal receptor:ligand pairs. In order to understand the landscape of artificial allostery created by calmodulin insertion, we developed an all-*in-vitro* protein expression and screening platform that allows identification of allosteric variants in libraries of chimeric proteins. We utilised this platform for the analysis of a library of PQQ-GDH-calmodulin chimeras and identified 50 new allosteric switches. This collection enabled us to analyse key performance parameters of the artificial allosteric molecules such as dynamic range, maximal catalytic activity, and response time. The obtained switches had dynamic ranges between 1.2- and 36-fold, maximal catalytic rates representing 4 % – 139% of parental PQQ-GDH activity (calculated k_{cat} range 99 s^{-1} to 3403 s^{-1}) and the half response time ranging from 2.5 to 30 minutes (Table S1 and Figure 3(B and C)).

Importantly we found that dynamic range and response rates that are negatively correlated. A possible mechanistic explanation is that a large dynamic range relies on more drastic structural

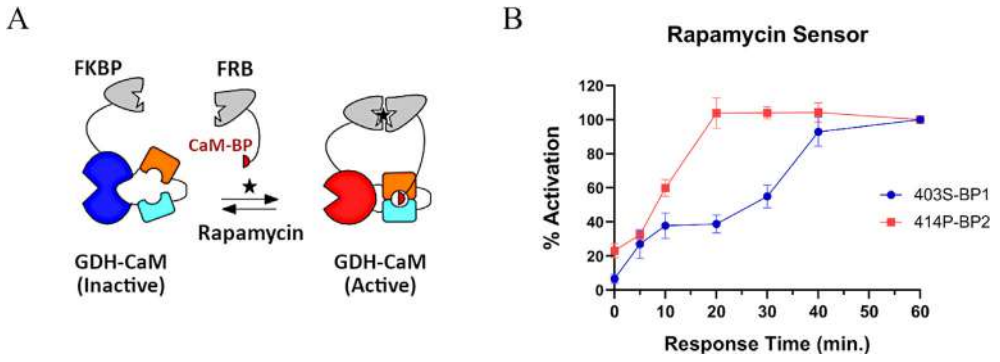

Figure 6. ***Construction of a two-component Rapamycin sensor.*** **(A)** Schematic representation of the rapamycin biosensor composed of two components: GDH-CaM fused with FKBP and CaM-BP fused with FRB domain. Rapamycin (represented by a black star) scaffolds FKBP and FRB and brings CaM-BP in close proximity to CaM, leading to the activation of GDH. **(B)** Comparison of the response rates for two-component rapamycin biosensors based on GDH-CaM-403S (blue) and GDH-CaM-414P (red) chimeras. Activation is given as a percentage of highest activity based on reaction rate, k_{obs}, after 0, 5, 10, 20, 30, 40, and 60 minutes of pre-incubation of the component with the ligand prior to the addition of glucose and reporter dyes. Data are given as mean ± SEM (error bars) of three replicates.

changes that take longer to relax and reach the active state. Given the ensemble nature of allostery, it is not entirely surprising that the kinetic separation of states plays an important role in regulating active and inactive conformations.[36] The same rationale can help explain the observation that different calmodulin-binding peptides activated the same chimera with different kinetics. We conjecture that the conformation adopted by calmodulin:peptide complex influences the isomerisation pathways proceeds with conformation-dependent rates. The extent of structural frustration in the reporter domain induced by the calmodulin is unknown and is likely vary among insertion sites. Analysis of one selected chimera by CD-spectroscopy and HDX-MS suggests that these changes are relatively minor but appear to propagate throughout the GDH molecule. This is consistent with the earlier findings on CytB-β-lactamse chimera.[30] Our analysis of GDH orthologues suggested that the frequency of artificial allosteric chimer emergence in domain insertion libraries may vary significantly even among orthologues. Systematic analysis is required to identify the structural features that determine the frequency of artificial allostery and factors that control its kinetic parameters.

Finally, we tested the utility of the identified GDH-CaM chimera:peptide pairs in the construction of two-component biosensor. We demonstrated that the biosensor based on the fast-acting GDH-CaM-414P:BP2 pair that has a half response time of 2.5 minutes resulted in a two-component biosensor with the half response time of 10 minutes, which compares favourably with the previously achieved half response time of 30 minutes. However, it also points to the fact that while the rate of core switch isomerisation is important, there are other parameters that influence the system's rate of activation. Future work will be required to understand the role played by the reporter domain in the performance of CaM-controlled switches. For instance, how would the performance of the chimeras differ if different reporters are used, or linkers connecting the domains are optimised?[37,38] Given that CaM binds its peptide ligand on a milliseconds scale and isomerises on a seconds scale, what is the fastest activation rate could CaM-reporter chimera display?[29,39] We provide a methodological platform and identify the key design parameters of artificial allosteric switches that will lead the way to answer these questions.

Material and Methods

Acinetobacter calcoaceticus GDH-CaM chimera library design

The GDH-CaM insertion library consists of 221 domain insertion sites covering all the surface residues in the loop regions of *Acinetobacter calcoaceticus* PQQ-GDH (Figure 3(A)) that are highlighted on the sequence in Table S4. The chimeras were named according to the residue where the CaM domain was inserted into GDH (i.e., GDH-CaM-403S denotes CaM insertion into residue 403S of GDH). Standard glycine–serine linkers (GSGG-CaM-GGSSG) were used to fuse CaM and GDH. The library was synthesized by GeneUniversal (China) and supplied as a mixed pool of linear DNA fragments that were then inserted into pCellFree_G03 vector[40] (Addgene plasmid # 67137) using Gateway cloning. The mixed clone library was transformed into *E. coli* DH5α chemically competent cells (ThermoFisher, USA), and then plated onto LB Agar plates with 100 μg/ml ampicillin to obtain approximately 100 colony forming units per plate. After overnight incubation of plates at 37, the individual colonies were randomly picked for plasmid amplification. Random screening required analysis of 4–5 times more colonies than the total number of variants (~1000 colonies for 221 variants), to achieve 99% coverage of the library. The coverage was calculated using GLUE program (http://guinevere.otago.ac.nz/cgi-bin/aef/glue.pl) based on the publication from Patrick et al.[24]

Cell-free expression of CaM-GDH chimera library

In vitro expression of the GDH-CaM insertion library in *Leishmania tarentolae*-based cell-free system (LTE) followed protocols described earlier.[17,41] The input DNA templates were amplified from colonies by the rolling circle amplification (RCA) method using the TempliPhi DNA amplification kit (Cytiva, Australia). The cell-free lysate was mixed with the feed solution (6:2 μL) as described.[42] The input DNA concentration was around 50–100 ng/μL to ensure protein expression with optimum yield and quality. For screening, the constructs were expressed in 384-well plates (PerkinElmer, Australia). All constructs of the library contained N-terminal EGFP (enhanced green fluorescent protein) fused to the GDH-CaM biosensor that allowed monitoring of the protein expression in real-time. During the expression 10 μL reaction was incubated at 26 °C for 2–3 hours, and GFP fluorescence was monitored using microplate fluorometric reader (Tecan Spark). For quality control the translation reactions were resolved on the SDS-PAGE and fluorescence of EGFP-GDH-CaM protein bands was recorded using fluorescent scanner (ChemiDoc imager, Bio-Rad).

Selection of GHD orthologues

The Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST) server[44] was used to create a sequence similarity network (SSN) representing proteins from the Glucose/Sorbosone dehydrogenase domain Pfam (PF07995) that were classified

as either having a single GSDH domain or a "GSDHx2" architecture (which the target sequence from *Acinetobacter calcoaceticus* belongs to). The databases used in this analysis were UniProt: 2020–06 and InterPro: 83. The resulting network contained 8,590 accession IDs comprising representative sequences from the UniRef50 database[45] (i.e. sequences that share ≥50% sequence identity over 80% of the sequence length are grouped together and represented by a single node ID). To generate the SSN, an all-by-all BLAST was performed to obtain the pairwise sequence identities and alignment scores. An alignment score cut-off of 100 was used to remove excess edges in the initial network. Sequences that were > 70% redundant are represented by a single node. Networks were visualised and annotated in Cytoscape,[46] using the yFiles Organic Layout. The "Cluster analysis" tool of the EFI-EST server was used to analyse individual clusters in the SSN and to produce a hidden-Markov model (HMM) profile for each of the clusters. Ten orthologs from within Cluster 23 (the same cluster as *Acinetobacter calcoaceticus* PQQ-GDH) were selected based on progressively decreasing similarity to *A. calcoaceticus GDH* (SI-Table 5). The signal sequences were removed and the open reading frames were synthesised by Gene Universal (China) and cloned into pET-28a(+).

GDH ortholog library design

The GDH Ortholog library containing 42 chimeras constructed from *Deinococcus ficus* PQQ-GDH gene and CaM domain inserted into each loop region of the former (Table S5). The *Deinococcus ficus* GDH ortholog is a PQQ dependent sugar dehydrogenase 45% sequence identity to the *A. calcoaceticus* PQQ-GDH. The preferable insertion sites replace glycine and proline residues.

Cell-free expression of GDH ortholog library

The GDH orthologue library was expressed in *E. coli* cell-free system.[43] The cell-free lysate was mixed with the feed solution and the input DNA concentration was around 50–100 ng/μL to ensure protein expression with optimum yield and quality. For screening, the constructs were expressed in 96-well plates (PerkinElmer, Australia). All constructs of the library contained N-terminal EGFP (enhanced green fluorescent protein) fused to the construct that allowed monitoring of the protein expression in real-time. During the expression, the reaction was incubated at 26 °C for 2–3 hours, and GFP fluorescence was monitored using microplate fluorometric reader (Tecan Spark).

Expression and purification of GDH orthologues and GDH-CaM chimeras

The open reading frames were cloned into a kanamycin-resistant pET-28a(+) vector. Competent *E. coli* BL21(DE3) cells were transformed with the resulting expression vectors and grown in LB broth with 50 μg/ml kanamycin (Everest Inc, Australia) with shaking at 37 °C. Protein expression was induced by adding 0.3 mM IPTG (Sigma, USA). Subsequently, cultures were incubated overnight at 18 °C. Cells were harvested at 4,000 RPM for 10 min, and the cell pellet was lysed in buffer containing 50 mM $NaHPO_4$ (pH 8.0), 300 mM NaCl, 20 mM imidazole, 1 mM AEBSF and DNAse I (all Sigma, USA). Following cell lysis and disruption using sonication, the supernatant was collected, and recombinant protein was further purified on the Ni-NTA HisTrap FF crude column driven by ÄKTA Express Purifier (Cytiva) system in buffer containing 50 mM $NaHPO_4$, 300 mM NaCl and 20 mM imidazole and a gradient of 500 mM imidazole for elution at pH 8.0. Finally, the protein was dialysed against 20 mM Tris/HCl (pH 7.0) and 100 mM NaCl for testing and storage. Protein samples were snap frozen in liquid nitrogen and stored at −80 °C.

Glucose dehydrogenase activity assay

The glucose dehydrogenase (GDH) activity assay was performed in a 96-well plate format (PerkinElmer, Australia) adapted from Guo et al.[11] Cell-free reaction products containing GDH-CaM protein biosensors were diluted into the reaction buffer containing 20 mM Tris-HCl (pH 7.2), 20 mM NaCl, and 1 mM $CaCl_2$. The exact amount (between 1–2 μl) of cell-free expressed protein was determined from the monitored EGFP fluorescence using final RFU readings after 2–3 hours of expression and by normalizing against the RFU reading of the control, GDH-CaM-403S chimera. All the samples were saturated with 1 μM CaM-BP and were supplemented with 50 nM PQQ (ThermoFisher, USA). When purified protein was used for the assay it was supplemented with PQQ in 1:1.5 ratio. For spectrometric measurements, 0.06 mM DCPIP redox dye (Sigma, USA) and 0.6 mM phenazine methosulfate (PMS) (Sigma, USA) as electron acceptor were added to the samples. Finally, the reaction was initiated with 20 mM D-glucose (Sigma, USA). The absorbance at 600 nm was measured kinetically at 25 with Synergy Neo2 Microplate Reader (BioTek, USA) every 10–20 seconds. The observed reaction rate constant, k_{obs}, was derived by fitting the linear phase of the curves via linear regression. To determine the rate of biosensor's response, the GDH-CaM chimeras were incubated with calmodulin-binding peptide M13 (Mimotopes, Australia) for 2.5, 10, 20, and 30 minutes and assayed for GDH activity. The initial rates were determined and plotted. The response time was selected as the pre-incubation time needed to reach 50% of the full activity (half response time). The dynamic range was defined

as the fold change in amplitude of the signal in the presence of CaM-BP compared to its absence.

Construction of two-component rapamycin biosensors

The rapamycin sensor consisted of two components: GDH-CaM fused to FKBP (GDH-CaM-FKBP) and FRB fused to a calmodulin-binding peptide (FRB-BP) (Table S4). We designed two sets of two-component rapamycin sensors. The first consisted of GDH-CaM-403S-FKBP and FRB-BP1. The second contained GDH-CaM-414P-FKBP coupled with FRB-BP2. The sequences of these components are given in Table S4. These constructs were expressed in *E. coli* BL21(DE3) and purified by Ni-NTA affinity chromatography as described above. For activity assay 10 nM of GDH-CaM-FKBP was reconstituted with 15 nM PQQ and mixed with 30 nM of FRB-BP in the reaction buffer. To determine the response rate, the GDH-CaM-FKBP constructs were incubated with FRB-BP for 5, 10, 20, 30, 40 and 60 minutes in the presence of 500 nM rapamycin.

Cd spectroscopy

Circular dichroism (CD) spectra were recorded using the J-715 Spectropolarimeter (Jasco, USA) at 0.5 nm intervals for absorbance between 198 nm and 240 nm using 1 mm quartz cells. The samples contained 1.75 μM (0.12 mg/ml) GDH-CaM-403S biosensor in 10 mM Tris/H_2SO_4 at pH 7.2, 20 nM NaCl, 1 mM $CaCl_2$, and 2 μM PQQ. The measurements were taken at 0 min, 10 min, 20 min, and 30 min immediately after the reaction was started by adding 2 μM of calmodulin-binding peptide M13. For data analysis, the CDPro Software was used to determine the percentage of secondary structures based on a reference set of proteins, SP48, which includes 5 denatured and 43 soluble proteins.[47] The predicted secondary structure of the GDH-CaM chimera contained around 25% β-sheets contributed by GDH, and 20% α-helices mostly attributed to the calmodulin domain. The remaining 45% were most likely disordered regions or loops. Overall, the CD spectra of the GDH-CaM sensor had α-helical content changing between 25–26% and β-sheet varying between 21–22% at different time points (0, 10, 20, 30 min) of incubation with the peptide.

HDX-MS analysis of GDH-CaM

To investigate allosteric changes potentially resulting in alteration of GDH function, we performed comparative HDX-MS analyses of GDH-CaM in the presence or absence of calmodulin-binding peptide M13. A LEAP HDX-2 Automation system (Trajan Scientific, Ringwood, VIC, Australia) was used to automate labelling, quenching and injection of samples into an ACQUITY UPLC M-Class HDX Manager (Waters, Milford, MA, USA). 3 μl Purified protein (15 μM GDH-CaM in PBS containing 5% (v/v) DMSO) was incubated in 57 μl PBS buffer reconstituted in D_2O (99.90%, Sigma). GDH-CaM:PQQ and GDH-CaM:PQQ:M13 complexes were prepared at a molar ratio of 1:3 and 1:3:3, and pre-incubated for 60 min to achieve binding prior to each hydrogen–deuterium exchange reaction. Deuterium labelling was performed for 30 s, 1, 10, 100 and 200 min, followed by quenching of 50 μl of the deuterium exchange reaction mixture in 50 μl pre-chilled 50 mM PBS quench solution to lower the pH to 2.5 and lower temperature to 0 °C. 80 μl Quenched samples were subjected to online digestion using an immobilised Waters Enzymate BEH pepsin column (2.1 × 30 mm), and proteolyzed peptides were separated on a Waters ACQUITY UPLC BEH C18 column (1.0 × 100 mm, 1.7 μm), as previously described.[12] A positive electrospray ionization source fitted with a low flow probe was used to ionize peptides sprayed onto SYNAPT G2-Si mass spectrometer (Waters, Milford, MA, USA). Data was acquired in MSE acquisition mode using 200 pg/μl Leucine enkephalin and 100 fmol/μl [Glu1]-fibrinopeptide B ([Glu1]-Fib) at a flow rate of 5 μl/min for lockspray correction.

ProteinLynx Global Server (PLGS) v3.0 was used to identify peptides from non-deuterated protein samples. The identified peptides were filtered in DynamX v3.0 (Waters, Milford, MA, USA) using a minimum intensity cut-off of 3000 for product and precursor ions, minimum products per amino acids of 0.3, a precursor ion mass tolerance of <5 ppm and a file threshold of 3. All deuterium exchange experiments were performed in triplicate, and reported values were not corrected for deuterium back exchange. HDX-MS data were further analyzed using the HD-eXplosion software tool and visualization of hydrogen–deuterium exchange data with statistical filtering.[48]

Statistical analysis

Statistical analysis and data fitting were performed in GraphPad Prism 9 software. Correlation analysis was performed on reaction rate, dynamic range, and response time by choosing Pearson's correlations as a parametric test and calculating two-tailed P values. The calculated correlation coefficient (Pearson's r) was plotted on a correlation matrix. In correlation analysis, 41 chimeras were selected out of total set of 50. The excluded chimeras had low expression levels, creating an obstacle for accurate response rate analysis. The difference between the means of 'High Activity' and 'Low/No Activity' groups were tested by unpaired t-test for the distance between N and C termini residues for CaM on the bound form to different CaM-BP's.

DATA AVAILABILITY

The data will be provided upon revision.

Acknowledgements

This work was supported in part by the Australian Research Council Discovery Projects DP160100973, DP150100936 as well as ARC Centre of Excellence in Synthetic Biology CE200100029 to KA. The work was also supported by NHMRC Development grants APP1113262 and APP1179001 to KA. This work was also in part supported by HFSP grant RGP0002/2018 to KA and the US Department of Defence grant W81XWH-20-1-0708 to KA. KA gratefully acknowledges the financial support of the CSIRO-QUT Synthetic Biology Alliance. We would also like to thank Waters Applications Support Specialist Caryn Hepburn and the Melbourne Mass Spectrometry and Proteomics Facility of The Bio21 Molecular Science and Biotechnology Institute at the University of Melbourne for supporting mass spectrometry analysis.

Declarations of interest

The authors declare the following competing interests: ZG and KA are named inventors on patents covering electrochemical protein biosensor technology related to this study. KA holds equity in Molecular Warehouse Ltd that owns this patent. The rest of the authors declare no competing interests.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2022.167678.

Received 22 February 2022;
Accepted 6 June 2022;
Available online 14 June 2022

Keywords:
allosteric proteins;
protein chimeras;
synthetic biology;
in vitro expression;
conformational change

References

1. Alberstein, R.G., Guo, A.B., Kortemme, T., (2022). Design principles of protein switches. *Curr. Opin. Struct. Biol.* **72**, 71–78.
2. Zhou, X., Mehta, S., Zhang, J., (2020). Genetically Encodable Fluorescent and Bioluminescent Biosensors Light Up Signaling Networks. *Trends Biochem. Sci.* **45**, 889–905.
3. Kostyuk, A.I., Demidovich, A.D., Kotova, D.A., Belousov, V., V, & Bilan D. S., (2019). Circularly permuted fluorescent protein-based indicators: History, principles, and classification. *Int. J. Mol. Sci.* **20**, 4200.
4. Nasu, Y., Shen, Y., Kramer, L., Campbell, R.E., (2021). Structure- and mechanism-guided design of single fluorescent protein-based biosensors. *Nature Chem. Biol.* **17**, 509–518.
5. Aroul-Selvam, R., Hubbard, T., Sasidharan, R., (2004). Domain insertions in protein structures. *J. Mol. Biol.* **338**, 633–641.
6. Makhlynets, O.V., Raymond, E.A., Korendovych, I.V., (2015). Design of allosterically regulated protein catalysts. *Biochemistry* **54**, 1444–1456.
7. Nicholes, N., Date, A., Beaujean, P., Hauk, P., Kanwar, M., Ostermeier, M., (2016). Modular protein switches derived from antibody mimetic proteins. *Protein Eng. Des. Sel.* **29**, 77–85.
8. McCormick, J.W., Russo, M.A.X., Thompson, S., Blevins, A., Reynolds, K.A., (2021). Structurally distributed surface sites tune allosteric regulation. Eds C.R. Landry, M.A. Marletta, G. Hochberg, A. Serohijos. *Elife* **10**, e68346.
9. Clark, J.J., Benson, M.L., Smith, R.D., Carlson, H.A., (2019). Inherent versus induced protein flexibility: Comparisons within and between apo and holo structures. *PLoS Comput. Biol.* **15**, 1–21.
10. Guo, Z., Johnston, W.A., Stein, V., Kalimuthu, P., Perez-Alcala, S., Bernhardt, P.V., Alexandrov, K., (2016). Engineering PQQ-glucose dehydrogenase into an allosteric electrochemical Ca 2+ sensor. *Chem. Commun.* **52**, 485–488.
11. Guo, Z. et al, (2019). Generalizable Protein Biosensors Based on Synthetic Switch Modules. *J. Am. Chem. Soc.* **141**, 8128–8135.
12. Guo, Z. et al, (2021). Design of a methotrexate-controlled chemical dimerization system and its use in bio-electronic devices. *Nature Commun.* **12**, 38–46.
13. Meister, G.E., Joshi, N.S., (2013). An engineered calmodulin-based allosteric switch for peptide biosensing. *ChemBioChem* **14**, 1460–1467.
14. Bollella, P., Bellare, M., Kadambar, V.K., Guo, Z., Alexandrov, K., Melman, A., Katz, E., (2020). Boolean Logic Networks Mimicked with Chimeric Enzymes Activated/Inhibited by Several Input Signals. *ChemPhysChem* **21**, 589–593.
15. Silverman, A.D., Karim, A.S., Jewett, M.C., (2020). Cell-free gene expression: an expanded repertoire of applications. *Nature Rev. Genet.* **21**, 151–170.
16. Mureev, S., Kovtun, O., Nguyen, U.T.T.U., T. T., & Alexandrov K., (2009). Species-independent translational leaders facilitate cell-free expression. *Nature Biotechnol.* **27**, 747–752.
17. Kovtun, O., Mureev, S., Johnston, W., Alexandrov, K., (2010). Towards the construction of expressed proteomes using a Leishmania tarentolae based cell-free expression system. *PLoS One* **5**, 1–11.
18. Gagoski, D., Polinkovsky, M.E., Mureev, S., Kunert, A., Johnston, W., Gambin, Y., Alexandrov, K., (2016). Performance benchmarking of four cell-free protein expression systems. *Biotechnol. Bioeng.* **113**, 292–300.

19. Igarashi, S., Ohtera, T., Yoshida, H., Witarto, A.B., Sode, K., (1999). Construction and characterization of mutant water-soluble PQQ glucose dehydrogenases with altered K (m) values - Site-directed mutagenesis studies on the putative active site. *Biochem. Biophys. Res. Commun.* **264**, 820–824.
20. Durand, F., Limoges, B., Mano, N., Mavré, F., Miranda-Castro, R., Savéant, J.-M., (2011). Effect of substrate inhibition and cooperativity on the electrochemical responses of glucose dehydrogenase. Kinetic characterization of wild and mutant types. *J. Am. Chem. Soc.* **133**, 12801–12809.
21. Dean, F.B., Nelson, J.R., Giesler, T.L., Lasken, R.S., (2001). Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res.* **11**, 1095–1099.
22. Reagin, M.J., Giesler, T.L., Merla, A.L., Resetar-Gerke, J. M., Kapolka, K.M., Mamone, J.A., (2003). TempliPhi: A sequencing template preparation procedure that eliminates overnight cultures and DNA purification. *J. Biomol. Tech.* **14** (2), 143–148.
23. Hadi, T., Nozzi, N., Melby, J.O., Gao, W., Fuerst, D.E., Kvam, E., (2020). Rolling circle amplification of synthetic DNA accelerates biocatalytic determination of enzyme activity relative to conventional methods. *Sci. Rep.* **10**, 1–8.
24. Patrick, W.M., Firth, A.E., Blackburn, J.M., (2003). User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng.* **16**, 451–457.
25. Andrews, C., Xu, Y., Kirberger, M., Yang, J.J., (2021). Structural aspects and prediction of calmodulin-binding proteins. *Int. J. Mol. Sci.* **22**, 1–26.
26. Kursula, P., (2014). The many structural faces of calmodulin: A multitasking molecular jackknife. *Amino Acids* **46**, 2295–2304.
27. Tidow, H., Nissen, P., (2013). Structural diversity of calmodulin binding to its target sites. *FEBS J.* **280**, 5551–5565.
28. Findlay, W.A., Martin, S.R., Beckingham, K., Bayley, P.M., (1995). Recovery of Native Structure by Calcium Binding Site Mutants of Calmodulin upon Binding of sk-MLCK Target Peptides. *Biochemistry* **34**, 2087–2094.
29. Montigiani, S., Neri, G., Neri, P., Neri, D., (1996). Alanine substitutions in calmodulin-binding peptides result in unexpected affinity enhancement. *J. Mol. Biol.* **258**, 6–13.
30. Edwards, W.R., Williams, A.J., Morris, J.L., Baldwin, A.J., Allemann, R.K., Jones, D.D., (2010). Regulation of beta-lactamase activity by remote binding of heme: functional coupling of unrelated proteins through domain insertion. *Biochemistry* **49**, 6541–6549.
31. Englander, J.J., Del Mar, C., Li, W., Englander, S.W., Kim, J.S., Stranz, D.D., Hamuro, Y., Woods, V.L., (2003). Protein structure change studied by hydrogen-deuterium exchange, functional labeling, and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 7057–7062.
32. Hu, W., Walters, B.T., Kan, Z.Y., Mayne, L., Rosen, L.E., Marqusee, S., Englander, S.W., (2013). Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7684–7689.
33. Tullman, J., Nicholes, N., Dumont, M.R., Ribeiro, L.F., Ostermeier, M., (2016). Enzymatic protein switches built from paralogous input domains. *Biotechnol. Bioeng.* **113**, 852–858.
34. Edwardraja, S. et al, (2020). Caged Activators of Artificial Allosteric Protein Biosensors. *ACS Synth. Biol.* **9**, 1306–1314.
35. Banaszynski, L.A., Liu, C.W., Wandless, T.J., (2005). Characterization of the FKBP-rapamycin-FRB ternary complex. *J. Am. Chem. Soc.* **127**, 4715–4721.
36. Motlagh, H.N., Wrabl, J.O., Li, J., Hilser, V.J., (2014). The ensemble nature of allostery. *Nature.* **508**, 331–339.
37. Gräwe, A., Stein, V., (2020). Linker Engineering in the Context of Synthetic Protein Switches and Sensors. *Trends Biotechnol.*, 1–14.
38. Gräwe, A., Ranglack, J., Weyrich, A., Stein, V., (2020). IFLinkC: An iterative functional linker cloning strategy for the combinatorial assembly and recombination of linker peptides with functional domains. *Nucleic Acids Res.* **48**, E24.
39. Jeon, J., Yau, W.M., Tycko, R., (2020). Millisecond Time-Resolved Solid-State NMR Reveals a Two-Stage Molecular Mechanism for Formation of Complexes between Calmodulin and a Target Peptide from Myosin Light Chain Kinase. *J. Am. Chem. Soc.* **142**, 21220–21232.
40. Gagoski, D., Mureev, S., Giles, N., Johnston, W., Dahmer-Heath, M., Škalamera, D., Gonda, T.J., Alexandrov, K., (2015). Gateway-compatible vectors for high-throughput protein expression in pro- and eukaryotic cell-free systems. *J. Biotechnol.* **195**, 1–7.
41. Kovtun, O., Mureev, S., Jung, W.R., Kubala, M.H., Johnston, W., Alexandrov, K., (2011). Leishmania cell-free protein expression system. *Methods* **55**, 58–64.
42. Johnston, W.A., Alexandrov, K., (2014). Production of eukaryotic cell-free lysate from leishmania tarentolae. *Methods Mol. Biol.* **1118**, 1–15.
43. Schwarz, D., Junge, F., Durst, F., Frölich, N., Schneider, B., Reckel, S., Sobhanifar, S., Dötsch, V., Bernhard, F., (2007). Preparative scale expression of membrane proteins in Escherichia coli-based continuous exchange cell-free systems. *Nature Protoc.* **2**, 2945–2957.
44. Zallot, R., Oberg, N., Gerlt, J.A., (2019). The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **58**, 4169–4182.
45. Suzek B. E., Wang Y., Huang H., McGarvey P. B., Wu C. H., & Consortium the U, (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* **31**, 926–932.
46. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.
47. Sreerama, N., Woody, R.W., (2000). Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287**, 252–260.
48. Zhang, N., Yu, X., Zhang, X., D'Arcy, S., (2021). HD-eXplosion: Visualization of hydrogen-deuterium exchange data as chiclet and volcano plots with statistical filtering. *Bioinformatics* **37**, 1926–1927.