Biochemistry

pubs.acs.org/biochemistry Article

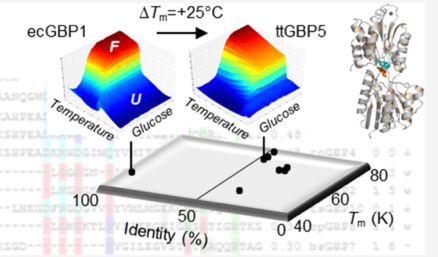
Discovery of Thermostable, Fluorescently Responsive Glucose Biosensors by Structure-Assisted Function Extrapolation

Malin J. Allert and Homme W. Hellinga*

Cite This: *Biochemistry* 2022, 61, 276−293 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Accurate assignment of protein function from sequence remains a fascinating and difficult challenge. The periplasmic-binding protein (PBP) superfamily present an interesting case of function prediction because they are both ubiquitous in prokaryotes and tend to diversify through gene duplication "explosions" that can lead to large numbers of paralogs in a genome. An engineered version of the moderately thermostable glucose-binding PBP from *Escherichia coli* has been used successfully as a reagentless fluorescent biosensor both *in vitro* and *in vivo*. To develop more robust sensors that meet the challenges of real-world applications, we report the discovery of thermostable homologues that retain a glucose-mediated conformationally coupled fluorescence response. Accurately identifying a glucose-binding PBP homologue among closely related paralogs is challenging. We demonstrate that a structure-based method that filters sequences by residues that bind glucose in an archetype structure is highly effective. Using fully sequenced bacterial genomes, we found that this filter reduced high paralog numbers to single hits in a genome, consistent with the accurate separation of glucose binding from other functions. We expressed engineered proteins for eight homologues, chosen to represent different degrees of sequence identity, and tested their glucose-mediated fluorescence responses. We accurately predicted the presence of glucose binding down to 31% sequence identity. We have also successfully identified suitable candidates for next-generation robust, fluorescent glucose sensors.



INTRODUCTION

Engineered *Escherichia coli* glucose−galactose-binding proteins (ecGBP) can be used as fluorescent biosensors for glucose.[1−19] Unlike enzyme-based systems, these glucose sensors are reagentless in the sense that they do not consume substrates to report on glucose concentrations. Several such sensors have been incorporated into optodes and show significant promise as next-generation continuous glucose monitors for the management of diabetes mellitus,[8−10,15] a disease that afflicts nearly a half-billion people worldwide with ever-increasing numbers.[20−23] ecGBP is a member of the periplasmic solute-binding protein (PBP) superfamily.[24−27] The protein comprises two α/β domains connected by a three-stranded β-hinge.[28−30] Glucose binds within the interface between the two domains (Figure 1). A Ca^{2+}-binding site confers additional thermostability.[31−34] Glucose binding shifts the equilibrium between open, glucose-free and closed, glucose-bound conformations.[28,29,35−38] These two conformations represent the extremes of a "bending" motion of the β-hinge, which is a common conformational coupling event in the PBP superfamily. Fluorescently responsive ecGBP-based glucose sensors exploit this mechanism by engineering fluorescent conjugates that link the emission properties of covalently attached fluorophores to the ligand-mediated protein conformational changes.[39,40]

Two general classes of ecGBP-based fluorescent glucose sensors have been developed using either single, covalently coupled organic dyes[1−12,14,15] or dual genetic fusions of green fluorescent protein variants at both N- and C-termini of the protein.[18,19] The fluorescent organic dye conjugates typically are attached site-specifically to a single cysteine that has been engineered into the protein. *Ab initio* prediction of the combination of fluorophores and attachment sites at which coupling between fluorescence emission and ligand-dependent conformational changes can be established has proven to be elusive, although a general strategy of introducing cysteine mutations at locations of residues that are located adjacent to the binding site (peristeric), or that contact the ligand (endosteric), or that are located in distal regions that move in concert with the hinge-bending motions (allosteric) improves the likelihood of success compared to selecting a random residue.[6,42]

An acrylodan conjugate endosterically coupled at W183C switches between blue and green emissive states corresponding to glucose-free and -bound forms of the protein conjugate, respectively.[43] This glucose-mediated spectral shift enables

Received: November 10, 2021
Revised: January 13, 2022
Published: January 27, 2022

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

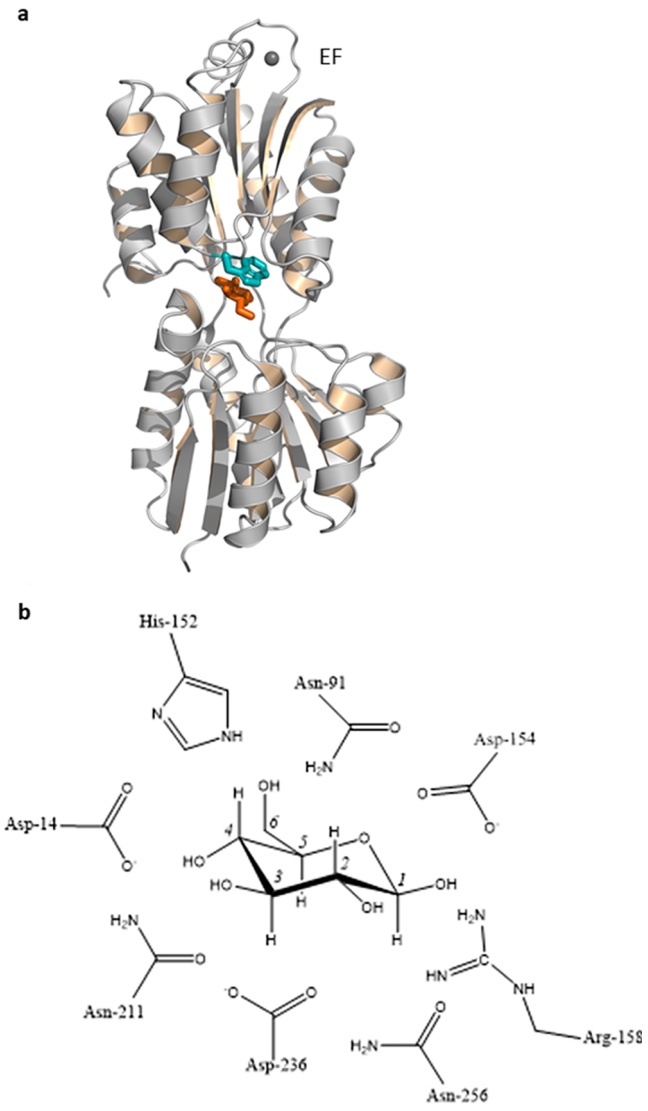


Figure 1. Structure of the *E. coli* glucose−galactose-binding protein (ecGBP, PDB identifier 2gbp[41]). (a) Glucose (orange) binds at the interface between N- and C-terminal domains. Calcium (gray) binds at a surface loop resembling an EF hand. Glucose is intercalated by van der Waals contacts with the rings of Phe16 and Trp183 (teal). Trp183 is replaced by acrylodan in the W183C conjugate. (b) Residues that form hydrogen bonds with the ring of hydroxyls in the bound glucose. The automatically identified primary complementary surface (PCS) comprises these residues and the aromatic sandwich.

glucose concentration to be measured as the ratio of the emission intensities of blue and green bands.[6,44,45] Such ratiometric sensing takes advantage of the simplicity of intensity-based measurements while avoiding problems associated with absolute measurements.[9,44,45] Acrylodan conjugates enable temperature-, glucose-, and Ca^{2+}-dependent multidimensional fluorescent "landscapes" to be collected. Analysis of these landscapes using statistical thermodynamics provides a rich source of information on thermostability, glucose or Ca^{2+} binding, osmolyte effects, and conformational coupling interactions.[32,43]

The response midpoint of ecGBP183C·acrylodan corresponds approximately to the euglycemic glucose concentration without additional mutations, unlike other fluorescent sensors.[7,46,47] Thus far, optodes that incorporate ecGBP183C·acrylodan are the only disclosed fluorescent glucose sensors that have been deployed successfully in animal and human trials for continuous glucose monitoring.[9,10,48,49] The midpoint thermal denaturation temperatures, T_m values, of ecGBP183C·acrylodan conjugate and its variants vary between 40 and 50 °C.[32,43] This modest thermostability may affect the deployment of these sensors in real-world applications by potentially limiting longevity or complicating device incorporation, manufacturing processes, and distribution logistics. We therefore set out to identify ecGBP homologues that both increase thermostability and preserve the conformational coupling mechanism of their acrylodan conjugates attached site-specifically at an endosteric single-cysteine mutant equivalent to ecGBP183C.

The analysis of enzyme families has shown that overall sequence identity below ∼60% is a weak predictor of function conservation,[50−52] and that below ∼35% (the "twilight" zone), strict conservation of structure as measured by the deviation of backbone atomic positions becomes increasingly uncertain.[53,54] Furthermore, functional assignments based on sequence homology alone can be challenging in the PBP superfamily because accurate detection of binding specificity is a particularly stringent constraint.[55−61] For instance, PBPs that, by overall sequence identity, are predicted to bind oligopeptides were found to bind oligosaccharides.[62,63] Furthermore, analysis of fully sequenced bacterial genomes has revealed that some PBPs have undergone massive gene duplications;[64,65] the resulting paralogs are likely to encode a diverse set of functions.[66] Enzyme functional assignments are improved greatly if a sequence selection filter based on conservation of catalytic residues identified from protein structures is included.[67]

We developed a structure-assisted functional extrapolation (SAFE) method to measure the degree of conservation of residues that contact bound glucose in ecGBP (Figure 1), the primary complementary surface (PCS). Predicted glucose-binding homologues were selected from a multiple sequence alignment (MSA) by matching residue positions and identities to the seed PCS. Conservation of glucose binding does not contain explicit information about retention of coupling between ligand binding and response of the conjugated fluorophore. Homologues that retain both properties therefore are likely to exhibit a significant degree of local structural conservation. We identified three homologues from thermophilic organisms with 48−51% sequence identities and conserved PCS residues that fully retained the glucose-mediated response of their acrylodan conjugate and improved thermostability by 25−31 °C. The homologue from *Thermoanaerobacter thermosaccharolyticum* (ttGBP5) was selected as the candidate for constructing robust, fluorescently responsive, ratiometric glucose sensors.

MATERIALS AND METHODS

Bioinformatics. All calculations were executed using the Biomolecular Computing (BMC) package developed by us. Original data, scripts used for analysis, and the calculation results can be downloaded from https:/doi.org/10.5061/dryad.ttdz08m0f. Fully sequenced 39 781 bacterial genomes comprising 139 034 346 protein open reading frames (ORF) were downloaded from the National Center for Biotechnology Information (June 3, 2020; www.ncbi.nlm.nih.gov) and assimilated into the Bioinformatics Environment component of the BMC. The BMC environment constructed an internal

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

database from the downloaded GenBank flat files to separate out DNA and protein sequences, their annotations, and positional information. Homology searches were carried out using BLAST[68] in our multiprocessor environment (Parallel Machine Queue, PMQ). Hits were collated and a multiple sequence alignment (MSA) constructed with ClustalO[69] divided into "chunks" that were executed on the PMQ for large data sets, from which a single alignment was collated. BMC scripts were used to calculate a positional filter encoding the primary complementary surface (PCS) using the structure of the ecGBP−glucose complex (2gbp).[41] BMC scripts were used to select hits from the MSA, which satisfied the PCS alignment criteria. The degree of alignment is measured as the number of mismatches between amino acids in the seed and homologue sequence at the subset of (discontinuous) locations defined by the PCS in the MSA. This "Hamming distance," H, can be based on 20 amino acids, identity Hamming distance, H_I, or on a reduced alphabet that combines groups of functionally similar amino acids, functional Hamming distance, H_F (see main text). Both alphabets include insertion−deletions ("gaps") as a member. A "hit" is scored if the Hamming distance is less than or equal to an upper threshold ($H_I = 0$ and $H_F = 0$ are the most stringent conditions: the two PCS sequences are identical or functionally equivalent, respectively).

The resulting hit list was used to calculate paralogs. Hit statistics were calculated from the MSA to determine fractional sequence identities relative to the entire ecGBP seed sequence, or its PCS, using full or functional alphabets. We used the average diversity of a hit set as a descriptive statistic to evaluate the efficacy of our filters. At any residue position, i, the normalized Shannon sequence entropy, S_i, is

$$S_i = -\frac{\ln M}{M} \sum_{a=1}^{M} f_{a,i} \ln f_{a,i} \quad (1)$$

where M is the number of amino acids (21 for identity; 8 for reduced alphabet, counting insertion−deletions as a mutation) and $f_{a,i}$ is the frequency of residue a in M at position i. The average diversity of N residues is the average normalized Shannon entropy

$$D = \frac{1}{N} \sum_{i=1}^{N} S_i \quad (2)$$

Gene Synthesis and Mutagenesis. The amino acid sequences for the GBP homologues identified in the bioinformatic search were edited using BMC scripts to construct a mature polypeptide with a single cysteine that replaces the equivalent of W183 in ecGBP for site-specific labeling with acrylodan; all other cysteines were changed to alanine. A hexahistidine tag was placed behind a GGS linker at the C-terminus of the mature protein to enable metal-mediated affinity purification.[70] The final amino acid sequence was back-translated into a DNA sequence encoding the open reading frame (ORF), which was placed in a construct behind an efficient Shine−Dalgarno ribosome-binding site, and flanked by a T7 promoter and terminator at the 5′ and 3′ ends, respectively.[71] The resulting ORF sequence was optimized in context by OrfOpt,[72] a program designed to predict highly expressed mRNA sequences in *E. coli*. This program identifies isocodon combinations that minimize mRNA secondary structure, set average AT nucleotide compositions to correspond to that of *E. coli* (48%), and also introduce high(er) AT levels within the first 100 bases of the 5′ end of the ORF while preferentially using codons that are used at high frequency in *E. coli* gene. The optimized DNA sequences were synthesized by oligonucleotide assembly and cloned into pUC57 by GeneWiz, Inc. (South Plainfield, New Jersey). Subsequent single- and multiple point mutations were constructed by preparing mutant sequences of the synthetic ORF using the most prevalent codon in *E. coli* for an amino acid, followed by total gene synthesis.

Protein Expression, Purification, and Fluorescent Conjugate Preparation. Plasmids carrying the expression constructs (see above) were transformed into KRX competent cells (Promega) and grown overnight at 37 °C on Luria broth (LB) agar plates (100 mg/mL ampicillin). A single colony was picked and grown overnight at 37 °C in Terrific Broth (TB; Research Products International). The overnight cultures were diluted 1:20 in 500 mL of TB (100 mg/mL ampicillin, 1 mM $CaCl_2$), grown to an optical density of $A_{600} = 0.5$ at 37 °C in vigorously aerated shaker flasks, induced by the addition of 2.5 mL of rhamnose (20% w/v), and grown for a further 3−4 h. The cells were harvested by centrifugation (5000 rpm, 10 min). After decanting the supernatant, the cell pellets were stored at −80 °C. The cell pellets were thawed, resuspended in 8 mL of binding buffer (10 mM imadozole, 20 mM 3-(*N*-morpholino)propanesulfonic acid (MOPS), 500 mM NaCl, 1 mM $CaCl_2$, pH 7.8). Following resuspension, 3 mL of BugBuster HT (EMD Millipore) was added. After incubation (20 min, 25 °C), the cells were lysed on ice by sonication (2 min of 1 s on/off pulses, 20−30% power). A clarified lysate was prepared by centrifugation (15 000 rpm, 20 min, 4 °C), from which recombinant protein was purified by batch immobilized metal-affinity chromatography (IMAC). Resuspended IMAC agarose beads (5 mL; Sigma-Aldrich, P6611) were added to the lysate. After incubation at 4 °C in a Mini LabRoller (Labnet International) for 1 h, the beads were washed at least five times with binding buffer. The immobilized protein beads were resuspended in labeling buffer (20 mM MOPS, 100 mM NaCl, 1 mM $CaCl_2$, pH 6.9) and labeled overnight (4 °C, rotating end-over-end) with the thiol-reactive fluorophore acrylodan (5-fold stoichiometric excess over protein). Following two rinses with labeling buffer to remove unincorporated label, the proteins were eluted from the beads. To elute labeled protein from the IMAC beads, 6 mL of elution buffer (400 mM imidazole, 500 mM NaCl, 1 mM $CaCl_2$, 20 mM MOPS, pH 7.8) was added, incubated for 30 min (4 °C, rotating end-over-end), and the beads removed by centrifugation. Following dialysis of the eluate against three changes of assay buffer (20 mM MOPS, 20 mM KCl, 1 mM $CaCl_2$, pH 7.4), using 10 kDa semipermeable membrane (Snakeskin tubing, Thermo Scientific), fluorescent conjugates were concentrated in a 10 kDa cutoff spin concentrator (Vivaspin, GE Healthcare). Protein purity was assessed by sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE). Protein concentrations were determined by a Nanodrop1000 at 280 nm (using extinction coefficients calculated from their sequence[73,74]) or at the fluorophore absorbance peak (acrylodan, 391 nm).

Emission Spectra. Acrylodan fluorescence emission spectra were recorded at room temperature on a Nanodrop3300 spectrofluorimeter using the UV excitation diode (365 nm), which matches the acrylodan 391 nm absorbance peak.[75] We define the barychrome wavelength, λ_B, for each emission spectrum as

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

$$\lambda_B = \frac{\sum_i I_i \lambda_i}{\sum_i I_i} \tag{3}$$

where I_i is the emission intensity recorded at wavelength λ_i.

A ratiometric signal at a given point t in a titration series, $R_{12,t}(t)$, is given by the ratio of intensities at two wavelengths, I_1 and I_2, measured at that point

$$R_{12,t} = \frac{a_t I_{1,t}}{a_t I_{2,t}} \tag{4}$$

where a_t is an attenuation factor that describes the effect of variations in sample size (*i.e.*, the amount of observable fluorophore) in sample t on the wavelength-independent intensity of the entire emission spectrum. Ratiometry removes wavelength-independent emission intensity attenuation effects due to variations in conjugate concentration, photobleaching, fluctuations in excitation source intensities, and detection efficiency.[44,45] It is a key aspect for high-precision sensing using the reagentless fluorescently responsive sensors described here. Ratiometric signal at wavelengths λ_1 and λ_2, R_{12}, was obtained as the ratio of the integrated intensities over the integration band with half-width δ

$$R_{12} = \frac{\sum_{\lambda=\lambda_1-\delta}^{\lambda=\lambda_1+\delta} I_\lambda}{\sum_{\lambda=\lambda_2-\delta}^{\lambda=\lambda_2+\delta} I_\lambda} \tag{5}$$

Fluorescence Landscapes. Fluorescence emission intensities were recorded as a function of temperature and ligand concentrations (glucose or Ca^{2+}) in 384-well plates using a Roche LightCycler, as described.[43] The acrylodan conjugate was excited at 440 ± 35 nm within the tail of the absorbance spectrum. Emission intensities were recorded using 488 ± 20 and 510 ± 20 nm band-pass filters from which the ratiometric signal, $R_{488,510}$, was calculated as the ratio of the blue divided by the green emission intensities. Landscapes were plotted and analyzed with scripts that use data-handling components of the BMC package.

Isothermal Ligand-Binding Curves. Glucose-dependent signals, S (single-wavelength intensities, barychromes, or ratiometric), obtained at constant temperature were fit to a single-site Langmuir binding curve[43]

$$S = \alpha(1 - \bar{y}) + \beta\bar{y} \tag{6}$$

where $\bar{y}$ is the binding polynomial for a single ligand-binding site

$$\bar{y} = \frac{K_a L}{1 + K_a L} \tag{7}$$

with K_a is the association constant (reporting disassociation constants $K_d = 1/K_a$), L is the ligand concentration, α is a constant corresponding to the signal in the absence of the ligand (*i.e.*, the apoprotein), and β is the signal of the ligand complex, which can be constant or a baseline that is linearly dependent on ligand concentration

$$\beta = a + bL \tag{8}$$

Fits used least-squares minimization with simplex and conjugate gradient algorithms in succession, as described.[43]

We developed a dual-wavelength method in which the ratiometric signal, the two single-wavelength intensities, and a vector, **a**, describing attenuation values at each titration point (see eq 4) were fit to the data. We minimized the following least-squares differences at all of the points t in a titration series

$$\Delta_t^2 = (a_t {}^{obs}I_{1,t} - {}^{calc}I_{1,t})^2 + (a_t {}^{obs}I_{2,t} - {}^{calc}I_{2,t})^2 + \left(\frac{{}^{obs}I_{1,t}}{{}^{obs}I_{2t}} - \frac{{}^{calc}I_{1,t}}{{}^{calc}I_{1,t}}\right)^2 \tag{9}$$

where $obs_{I_{i,t}}$ and $calc_{I_{i,t}}$ (i = 1, 2) were the observed and calculated intensities at the two-wavelength intervals, respectively. The third term in the sum contains the ratiometric information. The $calc_{I_{1,t}}$ and $calc_{I_{2,t}}$ intensities were calculated with the same K_a value, and four individual baselines (two constants for the apo-state baseline, and two linear baselines, eq 8, for the glucose complex at each wavelength), using eq 6. The K_a and six baseline parameters were fit by least-squares minimization using constant values for the individual a_t elements in **a**. Upon completion of a least-squares fit, a_t values were updated by taking the average comparison of the observed and calculated intensities at the two wavelengths

$$a_t = \frac{1}{2}\left(\frac{{}^{calc}I_{1,t}}{{}^{obs}I_{1,t}} + \frac{{}^{calc}I_{2,t}}{{}^{obs}I_{2,t}}\right) \tag{10}$$

The least-squares minimization was then repeated, keeping the updated **a** constant. The minimization/update cycles were iterated until **a** converged. Following convergence, the a_t value was applied to all wavelengths λ of the emission spectra to construct corrected spectra at each titration point

$${}^{corr}I_{\lambda,t} = a_t {}^{obs}I_{\lambda,t} \tag{11}$$

The corrected spectra were used to identify isosbestic points.

Thermal Stabilities. The apoprotein thermal melts were obtained from the fluorescent landscapes of the acrylodan conjugates by extracting the temperature dependence of the signal in the absence of glucose. The resulting ratiometric signals, $R_{488,510}$, were fit with a quadratic Savitzky–Golay filter from which the first derivative of the signal with respect to temperature, $dR_{488,510}/dT$, was calculated directly.[76] The largest peak was identified automatically in this first derivative and 11 data points comprising the maximum and five flanking values were fit using a least-squares conjugate gradient minimization to the van't Hoff equation for a two-state thermal equilibrium, K_u, to obtain the midpoint transition temperature, T_m, and enthalpy of unfolding, ΔH_u, at that temperature with[43]

$$\frac{dR_{488,510}}{dT} = f_u T^2 (1 - f_u) A \tag{12}$$

with the fraction of unfolded protein, f_u

$$f_u = \frac{K_u}{1 + K_u} \tag{13}$$

and

$$K_u = e^{-(\Delta H_u(1/T - 1/T_m))/RT} \tag{14}$$

SYPRO Analysis. Thermostability was also measured by the differential binding of the fluorescent SYPRO dye to the unfolded state.[77,78] Glucose affinity was determined by measuring the effect of glucose on the T_m values as described.[43]

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

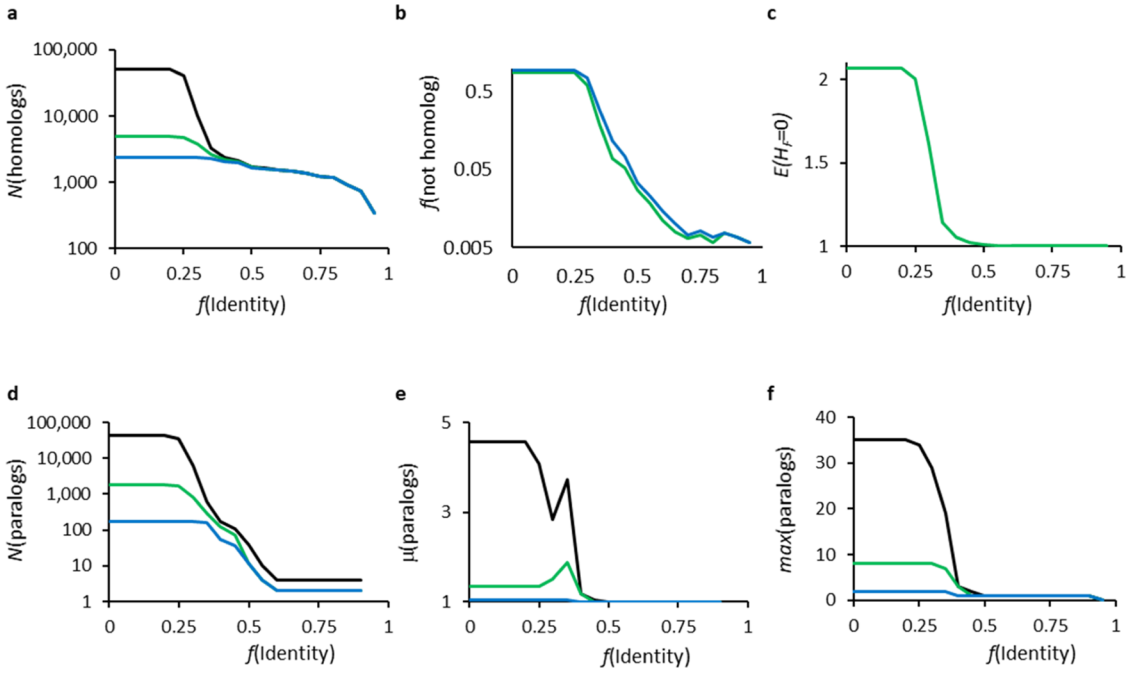


Figure 2. Identification of sequence homologues by bioinformatics. (a) Cumulative counts of sequence homologues (black: total number of hits, 50 696, N_T) and subsets of glucose-binding proteins predicted using identity, $H_I = 0$ (blue: total number of homologues, 2382), or functional similarity, $H_F = 0$ (green: total number of homologues, 4925) filters. f(Identity), f_I, is the fraction of identical residues relative to the ecGBP sequence measured in the full multiple sequence alignment. Each data set was binned into 20 intervals ($f_{I,i} = [0,0.05]_1$, ..., $[0.95,1]_{20}$), and the subtotal number of homologues whose f_I value exceeded the lower limit of the $f_{I,i}$ was recorded, N_i, and plotted for each bin i. (b) Cumulative frequency distribution of the incidence of sequence homologues that are predicted not to bind glucose, f(not homologue), identified as sequences with $H_I > 0$ (blue) or $H_F > 0$ (green). For each bin i, f_i(not homologue) = N_i(not homologue)/N_T = (N_i(sequence homologue) − $N_i(H = 0)$)/N_T. (c) Cumulative distribution of the excess of PCS sequences identified by the functional filter, $E(H_F = 0)$ relative to the identity filter, as used in the SAFE method. For each bin i, $E_i(H_F = 0) = N_i(H_F = 0)/N_i(H_I = 0)$. (d) Cumulative frequency distribution of the number of paralogs. (Black, all sequence homologues; blue, PCS identity homologues with $H_I = 0$; green, PCS functional similarity homologues with $H_F = 0$.) (e) Cumulative frequency distribution of the average number of paralogs per genome (coloring as in (d)). By definition, the minimum value of μ(paralogs) = 1. (f) Cumulative frequency distribution of the maximum number of ecGBP paralogs detected in the collection of genomes.

RESULTS AND DISCUSSION

SAFE mining of sequence databases for novel ecGBP homologues consisted of seven steps: (1) identification of an ecGBP sequence homologue set (SHS) within a large collection of sequences, (2) multiple sequence alignment (MSA) of the SHS, (3) identification of the PCS sequence using the ecGBP structure, (4) prediction of true homologues within the SHS by applying the PCS conservation filter to the MSA, (5) selection of representative homologues for experimental verification, (6) construction of synthetic genes of single-cysteine mutants with mRNA sequences optimized for heterologous expression in *E. coli*, (7) experimental testing for conservation of the coupling between protein and fluorophore by determination of fluorescence and glucose-binding properties of their acrylodan conjugates. The numerical analysis methods used to execute this strategy are described in the Materials and Methods section. Equations are numbered in the order of their development.

Steps 1−4: Identification of a Family of Periplasmic Glucose-Binding Protein Homologues Using SAFE. The initial SHS was identified (step 1) using 39 781 publicly available fully sequenced bacterial genomes, comprising 139 034 346 open reading frames (ORFs). The BLAST sequence alignment tool[79] with the ecGBP sequence as a seed identified 50 697 sequence homologues with sequence identities >20 and 70% matching of the seed length (Figure 2a). An MSA of this SHS was constructed (step 2) using ClustalO.[69]

The primary complementary surface (PCS) was identified automatically (step 3) using the structure of the ecGBP−glucose complex (2gbp).[41] The PCS was defined as the set of residues with side chains that contact the bound ligand. Such contacts comprised both hydrogen bonds and van der Waals contacts between "heavy" atoms using implicit protons. A hydrogen bond was present if the pairwise distance between nitrogen or oxygen atoms located on the residue and the ligand was less than a maximum (set to 3.5 Å). A van der Waals contact between atoms i in the residue and j in the ligand was present if the interatomic distance between them, r_{ij}, was less than the sum of the van der Waals radii of each atom, R_i and R_j, and an adjustable layer distance, $R_{contact}$ (set to 0.4 Å). A van der Waals contact between residue i and the ligand was present if the number of interatomic van der Waals contacts exceeded a minimum, N_{vdw} (set to 3). For residues that satisfied these contact constraints, an angular constraint was applied such that the angle between the C_α and C_β atoms on the residue and all atoms L_i in the ligand, C_α−C_β−L_i, exceeded a minimum, θ_{min} (set to 90°). This procedure identified 10 residues in the PCS, comprising eight residues forming a ring that binds the sugar hydroxyls (Figure 1b) and two aromatic residues that intercalate the pyranose ring. The secondary contact surface (SCS) was identified as the residues not in the PCS that contact one or more PCS residues. Contact criteria were

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

identical to interactions described above, substituting a PCS residue for ligand.

The MSA was used to identify glucose-binding homologues by filtering on conservation of residue identity at positions that correspond to the discontinuous PCS positions (step 4). Conservation of the glucose-binding function was inferred for each sequence homologue by measuring the Hamming distance, H (number of nonidentical residues), of the PCS sequence relative to that of ecGBP. Two types of Hamming distances were calculated. The first, H_I, used the identities of all 20 amino acids and is a measure of strict functional conservation. The second, H_F, used a reduced alphabet of seven letters by grouping functionally similar amino acids into one letter (Table 1) and is based on the premise that

Table 1. Relationship between Functional and Identity Alphabets

	alphabets	
name	functional	identity[a]
structural	s	A, V, I, L, M
flexibility	f	G, P
acid/base	a	E, D, Q, N, H
cationic	c	K,R
aromatic	r	W, F, Y
hydroxyl	h	S, T
thiol	t	C
insertion/deletion		

[a]Single-letter amino acids.

chemically similar residues are likely to retain functional equivalence. Glucose-binding homologues (Figure 2a) were extracted from the 50 697-membered SHS by identifying hits with $H_I = 0$ (identity conservation, 2382 homologues) or $H_F = 0$ (functional similarity conservation, 4925 homologues). The fraction of nonhomologous hits ($H_I > 0$ or $H_F > 0$) increased below ~60% fractional sequence identity, f_I, of the entire sequence relative to ecGBP, exceeding >10% for $f_I < 40\%$, and >50% for $f_I < 35\%$ (Figure 2b). For sequences $f_I < 50\%$, the less stringent functional similarity filter identified more predicted homologues than the identity filter (Figure 2c), as would be expected.

The unfiltered SHS had a high degree of diversity at the PCS positions for both functional (Figure 3a) and identity alphabets (Figure 3d), as indicated by their average diversities, $\bar{D}$ (eq 2). As intended, filtering removed PCS functional alphabet diversity of the homologue set, $\bar{D} = 0$ (Figure 3b) but retained some level of identity alphabet diversity in their PCS, $\bar{D} = 0.15$ (Figure 3c). We also examined the diversity of the residue layer that contacts the PCS residues, but not the ligand, the secondary complementary surface (SCS, Figure 3e–h). This residue layer retained some degree of diversity in the selected hits for both alphabets (Figure 3f,g) but at a lower level than the unfiltered hits, indicating that the SCS also contains some degree of information for ligand binding.

The number of paralogs that were identified within each set assesses how likely it is that true homologues were selected by SAFE. A sequence is paralogous if one or more additional homologues are identified within the same genome (*i.e.*, two or more homologues in a genome). It is postulated that paralogs typically encode different but related functions.[66] Consequently, it would be expected that the application of the PCS filter limits the number of paralogs by eliminating all sequence

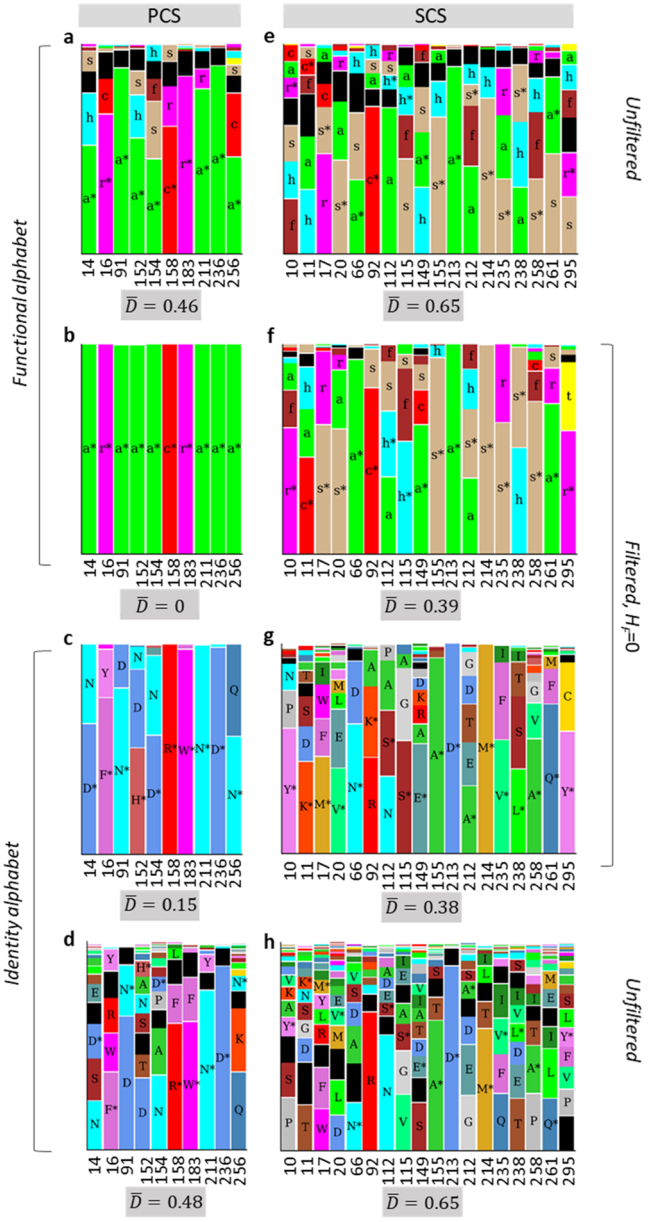


Figure 3. Effect of SAFE filtering on the diversity of the residues in the primary (PCS) and secondary (SCS) complementary surfaces. In each panel, the fraction of different amino acids is shown at each position of the complementary surface (left column: PCS; right column: SCS), with sequence numbers along the abscissa referring to ecGBP. Within each, the fractional representation of an amino acid within all of the sequences in a collection (top and bottom rows: unfiltered SHS; middle two rows: filtered glucose-binding homologues) at that position is given by the height of the colored bar (all bars sum to 1). Color choice is arbitrary but consistent (top two rows: functional alphabet; bottom two rows: identity alphabet) with letters showing the amino acid identity (see Table 1 for functional alphabet); asterisk indicates amino acid found in ecGBP at that position. At each position, the amino acids are displayed in ranked order; each column sums to 1, but amino acids with frequencies less than 0.01 are not shown. The average diversity, $\bar{D}$ (eq 2), is shown for each panel.

homologues with different ligand-binding specificities. The number of paralogs in the unfiltered SHS was high, whereas PCS filtering reduced this count by 2–3 orders of magnitude (Figure 2c). The average number of glucose-binding paralogs per genome was limited to 1 for $H_I = 0$ and slightly above 1 for

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276–293

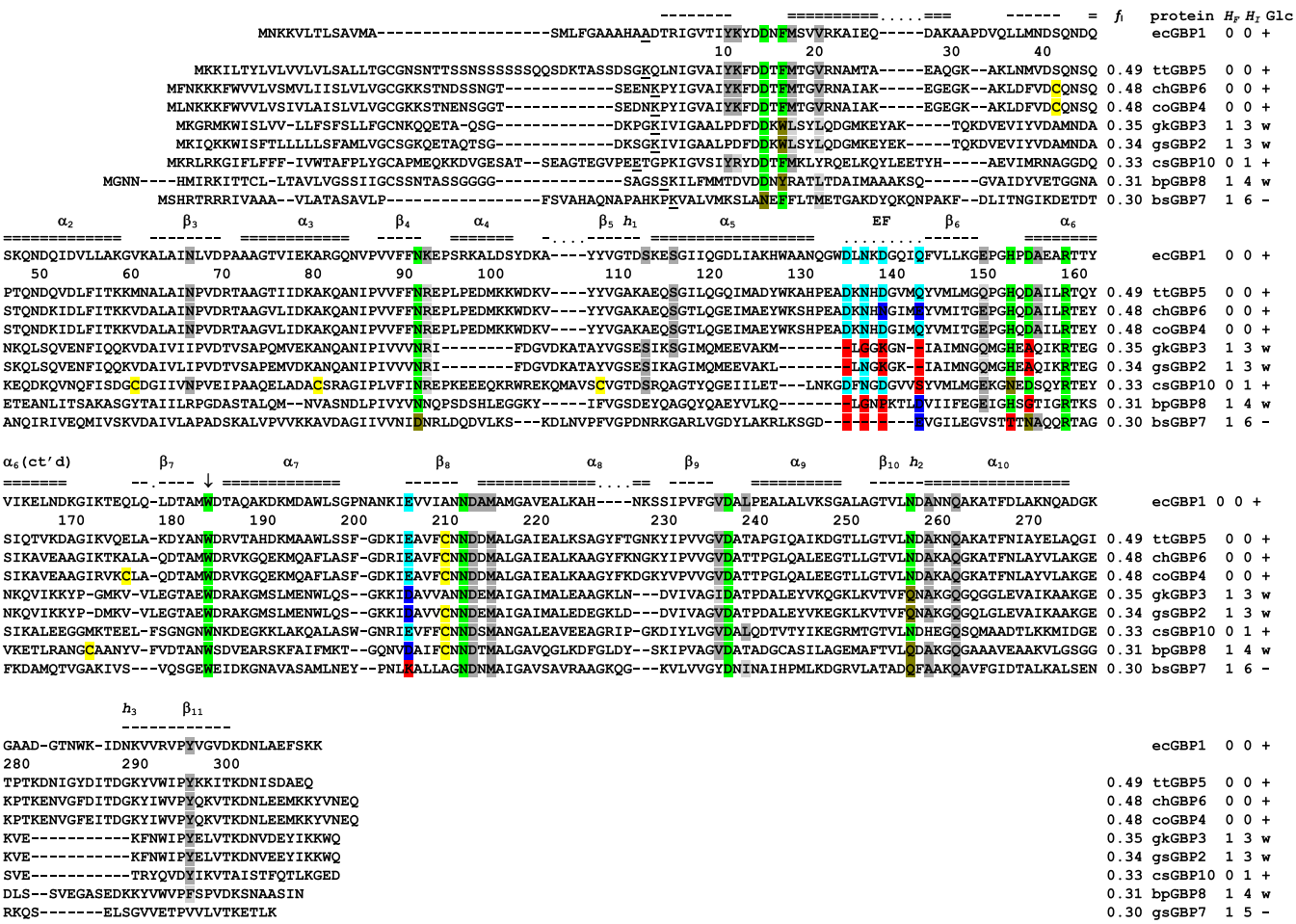


Figure 4. Alignment of the selected sequence homologues. Residue positions are numbered according to the ecGBP sequence. The secondary structure assignment is indicated above the ecGBP sequence (α, α-helix; β, β-strand; h, hinge strand; EF, EF-hand). Underlined, start of the heterologous cytoplasmic expression constructs (predicted site of leader peptide cleavage). Green, conserved positions of glucose-binding PCS residues (dark green, functionally equivalent). Blue, conserved positions of Ca^{2+}-binding PCS residues (dark blue, functionally equivalent). Red, PCS position (glucose or Ca^{2+}) not conserved with functional similarity. Gray, conserved SCS positions (dark, identical; light, functionally equivalent). Yellow, endogenous cysteines mutated to alanine in expression constructs. Vertical arrow, the position of endosteric acrylodan placement (mutated to cysteine in expression constructs). Annotation columns, protein, identities, and sources of the proteins are detailed in Table 2; f_I, fractional identity of the aligned sequence relative to ecGBP (excluding leader peptide); H_F, the Hamming distance according to the functionally similar alphabet; H_I, identity Hamming distance; and Glc, glucose-mediated fluorescence response of the protein acrylodan conjugate (+, similar to ecGBP; w, measurable but with weaker affinity; −, no binding).

Table 2. Selected Homologues

					PCS			SCS		
		NCBI Accession codes		Identity[a]	sequence[b]	Hamming[c]		Sequence[d]	Hamming[c]	
Name	Organism	Genome	Protein	(%)		H_F	H_I		H_F	H_I
ecGGBP	*Escherichia coli* ATCC 8739	NC_010468	YP_001724487.1	100	DFNHDRWNDN	0	0	YKMVNKSSEADAMVLAQY	0	0
ttGBP5	*Thermoanaerobacterium thermosaccharolyticum* DSM571	NC_014410	YP_003852930.1	49	DFNHDRWNDN	0	0	YKMVNraSqADdMVtAQY	3	5
coGBP4	*Caldicellulosiruptor obsidiansis*	NC_014392	YP_003839461.1	48	DFNHDRWNDN	0	0	YKMVNraSEADdMVtAQY	3	4
chGBP6	*Caldicellulosiruptor hydrothermalis*	NC_014652	YP_003991244.1	48	DFNHDRWNDN	0	0	YKMVNraSEADdMVtAQY	3	4
gsGBP2	*Geobacillus* sp. Y4.1MC1	NC_014651	YP_003989526.1	34	DwNHaRWNDq	1	3	pdllirSaqqDeMVtAQY	7	11
gkGBP3	*Geobacillus kaustophilus*[e]	NZ_MQMG 01000186	WP_041467870.1	35	DwNHaRWNDq	1	3	pdllirSSqqDeMitAQY	6	11
csGBP10	*[Clostridium] saccharolyticum* WM1	NC_014376	YP_003822565.1	33	DFNnDRWNDN	0	1	YrMyNrSaEsDsMVLhQY	5	7
bpGBP8	*Butyvibrio proteoclasticus* B316	NC_014389	WP_013281503.1	31	DyNHgRWNDq	1	4	tdrlrneaEtDtMVtAQY	10	11
bsGBP7	*Burkholderia* sp.	NC_014117	YP_003608256.1	30	nFdtnRWNDQ	1	6	ksfmanngvADnMyiAQf	10	13

[a]Number of identical residues shared with the ecGGBP1 sequence according to the multiple sequence alignment (excluding leader peptide). [b]Residue identity from the ClustalO alignment at positions 14, 16, 91, 152, 154, 158, 183, 211, 236, and 256 (see Figure 3). Differences from the ecGGPB1 sequence are indicated. Functional similarity: black, conserved; red, mutated. Identity: upper case, conserved; lower case, mutated. The endosteric acrylodan placement position is indicated in bold. [c]PCS Hamming distances: H_I, identity alphabet; H_F, functional similarity alphabet. 0, identical residues; >0, number of mutations. [d]Residue identity at positions 10, 11, 17, 20, 66, 92, 112, 115, 149, 155, 212, 213, 214, 235, 238, 258, 261, and 295 (see Figure 3). [e]Also named *Geobacillus proteiniphilus*.

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

Table 3. Thermostabilities, Fluorescence Emission Maxima, and Glucose Affinities

	thermostability[a]		barychromes[b]			bands[c]		glucose affinity[b,d,e]			
protein	$^{apo}T_m$ (°C)	$\Delta^{apo}H_u$ (kcal/mol)	$^{apo}\lambda_B$ (nm)	$^{glc}\lambda_B$ (nm)	$\Delta\lambda_B$ (nm)	1 (nm)	2 (nm)	$^{B}K_d$ (mM)	$^{R}K_d$ (mM)	$^{D}K_d$ (mM)	$^{L}K_d$ (mM)
ecGBP	49.1	152	492.5	512.4	20.2	488	510	8.8	7.4	6.5	4.6
ttGBP5	74.0	241	492.3	509.6	17.3	488	510	3.7	2.7	2.3	0.54
coGBP4	78.4	192	488.5	516.0	27.5	488	510	1.1	0.78	0.72	0.41
chGBP6	80.3	206	492.9	513.8	20.9	488	510	0.93	0.70	0.71	0.63
gsGBP2	74.9	140	501.7	501.4	−0.3	460	550	17.3	82.1	90.8	131
gkGBP3	76.3	166	503.9	499.5	−4.4	460	520	63.9	68.2	68.5	102
csGBP10[f]	75.9	59	497.8	494.1	−3.7	460	520	92.3	100	24.1	1.1
bpGBP8[g]	46.0	121	490.2	479.7	−10.5	480	530	520	513	506	723

[a]Fit of van't Hoff equations (eqs 12−14) to thermal melt signal at 0 mM glucose extracted from Roche LightCycler data using the ratio of the intensities measured with 488 and 510 nm filters (Savitzky−Golay filter window of five temperatures). [b]At 25 °C. [c]Wavelengths of the integrated intensity bands (width 20 nm) extracted from the Nanodrop3300 emission spectra, used to calculate $^{R}K_d$ and $^{D}K_d$. [d]$^{B}K_d$, from fit of barychrome (eqs 3 and 6); $^{R}K_d$, from a fit of intensity ratios (eqs 4 and 6); $^{D}K_d$, from dual-wavelength fit (eqs 9 and 10); and $^{L}K_d$, from fit of isotherm extracted from Roche LightCycler landscape (eqs 4 and 6). All fits use constant baselines for the apoprotein signal and linear baselines (eqs 8) for the glucose complex signal, unless indicated otherwise. [e]The coGBP4, ttGBP5, and chGBP6 homologues also bind galactose, as expected, with $^{L}K_d$ values of 2.3, 1.5, and 7.8 mM, respectively. Other homologues were not tested. [f]Thermostability fits constrained to the [70,85] °C interval corresponding to the unfolding transition. [g]Apoprotein and glucose complex signal baselines are both constant.

$H_F = 0$ (Figure 2d). The maximum number of paralogs in at least one of the genomes reached 35 in the absence of SAFE selection (Figure 2e), consistent with massive gene duplication events observed in other PBPs, which attain functional diversity.[64,65] Up to eight paralogs could be observed in the homologues filtered with $H_F = 0$, and up to two for the more restrictively filtered $H_I = 0$.

There are at least three reasons that a genome might possess multiple copies of a glucose-binding protein. First, PBP paralogs have been found that differ in their affinities for an identical ligand, β-mannose, postulated to reflect uptake of this sugar under different environmental conditions.[80] Second, some PBPs interact with two-component signal-transduction systems rather than ABC transporters.[24] Paralogs therefore may reflect a multiplicity of protein partners rather than ligand-binding properties. Finally, the true ligand for ecGBP is not glucose, but a galactoside, glyceryl−galactopyranoside,[28] the PCS for which involves additional residues. Consequently, the paralogs identified by a PCS derived from ecGBP may retain the core glucose−galactose-binding properties but diverge in the galactoside *R*-group.

Step 5: Selection of Homologues for Experimental Characterization. Eight sequence homologues with H_F values of 0 or 1, and H_I values that varied from 0 to 6, with f_I values ranging from 30 to 49%, were selected (step 5) for experimental characterization (Figure 4 and Table 2). Four sequence homologues (gsGBP2, gkGBP3, bsGBP7, bpGBP8) lost the Ca^{2+}-binding EF-hand region that encodes Ca^{2+}-mediated thermostabilization.[32] This small collection of proteins explored the efficacy of SAFE for selecting glucose-binding homologues using the $H_F = 0$ selection criterion at moderate ($f_I \approx 50\%$: chGBP4, ttGBP5, coGBP6) and high ($f_I = 33\%$, csGBP10) degrees of sequence divergence. For these four proteins, the degree of overall sequence identity no longer guarantees glucose binding in the absence of additional information. In a high-divergence interval ($f_I = [30,35\%]$), we also explored the effects of losing functional similarity ($H_F = 0$) in the context of various degrees of identity loss ($H_I = 3$: gsGBP2, gkGBP3; $H_I = 4$: bpGBP8; $H_I = 6$, bsGBP7).

Step 6: Fluorescent Conjugate Construction. To construct sequences suitable for site-specific attachment of an endosteric acrylodan conjugate in ecGBP, all endogenous cysteines were replaced with alanine, and a single-cysteine mutation was introduced at the equivalent position of W183C in ecGBP, as predicted by the MSA (Figure 4). Cytoplasmic expression constructs were designed by removing putative leader peptides that mediate anchoring of the periplasmic-binding protein on the outside of the membrane (Gram-positive bacteria) or direct secretion into the periplasm (Gram-negative bacteria) and optimizing the resulting ORF DNA sequences for heterologous expression in *E. coli*.[81]

Step 7: Experimental Testing of Coupling between Glucose Binding and Endosterically Conjugated Acrylodan Fluorescence. The temperature-dependent glucose-binding fluorescent landscapes were determined for acrylodan conjugates of all seven proteins and compared to the ecGBP landscape by measuring the emission intensities through two filters centered at 488 nm (I_{488}) and 510 nm (I_{510}), respectively, using a Roche LightCycler (Figure 5).

The folded, glucose-free state of the ecGBP acrylodan conjugate exhibits a blue emission spectrum, the blue state (Figure 6a). Consequently, its ratiometric signal, $R_{488,510}$, is about 1.1 in the absence of glucose at 293 K (Figure 5a). Glucose binding shifts its emission into a less intense, blue-green region of the spectrum (bathochromic shift), the green state (Figure 6a), and $R_{488,510=0.67}$ in the ratiometric landscape (Figure 5a).

Thermal unfolding shifts the acrylodan conjugate to a similarly colored green state in ecGBP (Figure 5a). To a first approximation, the observed thermal landscapes therefore can be interpreted as glucose- and temperature-dependent shifts between three states: folded apo (^{apo}F, blue), folded glucose complex (^{glc}F, green), and the unfolded state (U, green), which are present at distinct combinations of glucose concentration and temperature (regions in "phase space").

Isothermal glucose affinities (Table 3) were determined by four different methods. The glucose dependence of the fluorescence emission spectra (Figure 6) was used to extract barychrometric (intensity-weighted average wavelength, λ_B; eq 3), or integrated intensities in two bands (eq 5) signals that were fit to single-site binding isotherms (eqs 6−8 for barychromes or ratios), or the dual-wavelength treatment (eqs 9−11). A ratiometric binding isotherm was also extracted from the fluorescence landscapes at 25 °C (Figure 7). The four

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

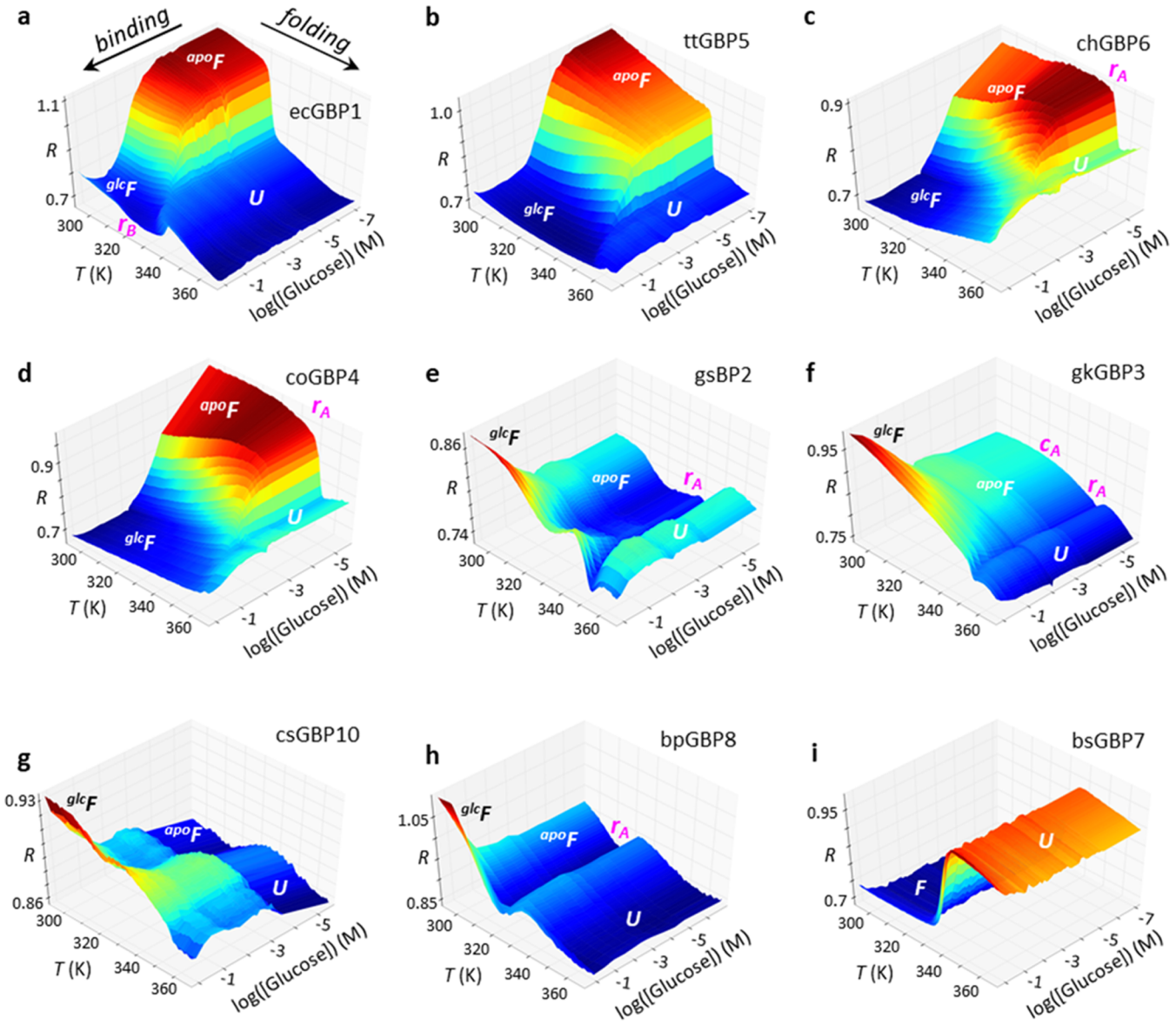


Figure 5. Temperature- and glucose-dependent fluorescent landscapes of acrylodan conjugates. Panels (a)−(i) show the effects of temperature and glucose concentration on the ratiometric fluorescence signal formed by the ratio of emission intensities measured with 488 and 510 nm filters on a Roche LightCycler for acrylodan conjugates of nine homologues. Proteins are named by their acronyms, as detailed in Table 2. Each landscape is dominated by glucose-mediated binding reactions and thermal unfolding (indicated in (a)). The regions in phase space corresponding to the folded apoprotein, ^{apo}F, the glucose complex, ^{glc}F, and the thermally unfolded protein, U, are indicated. The positions of features due to additional conformational changes are also indicated: r_B, ridge in the glucose complex; r_A, ridge in folded apoprotein; and c_A, curvature in folded apoprotein. csGBP10 (g) exhibits a small ratiometric change at these two wavelengths. bsGBP7 (i) does not bind glucose and transitions between folded, F, and unfolded, U, states.

affinity values (emission barychromes, $^{B}K_d$; ratiometric emission spectrum intensities, $^{R}K_d$; dual-wavelength emission spectrum fits, $^{D}K_d$; ratiometric fluorescent landscape, $^{L}K_d$,) are not the same because each is affected by different weights of the two end state contributions to the signal (apoprotein, and glucose complex). The ratiometric affinities ($^{L}K_d$, $^{R}K_d$) are dependent on the choice of wavelengths and the width over which each is integrated,[44,45] the effects of which vary for each emission spectrum. The barychrometric affinity, $^{B}K_d$, is not dependent on such choices and therefore is the least subjective measure of apparent glucose affinities.

The thermal titration landscapes of the three moderately divergent homologues in which the identity of the PCS is conserved exhibit the closest similarities to ecGBP (Figures 5a and 6a): the ttGBP5 landscape is nearly identical (Figures 5b and 6b), and chGBP6 and coGBP4 landscapes are similar (Figures 5c,d and 6c,d). These three homologues each retain the switch from a blue ($R_{488,510} \approx 1$) to a green ($R_{488,510} \approx 0.8$) state upon glucose binding ($\Delta\lambda_B$ 17−28 nm, Table 3). The formation of green thermally unfolded states is also retained but with more variation in color: ttGBP5 is similarly green to ecGBP ($R_{488,510} \approx 0.7$), but the unfolded states of chGBP6 and coGBP4 retain more blue character ($R_{488,510} \approx 0.8$). The glucose-binding affinities of the acrylodan conjugates of the three homologues were similar to or stronger than ecGBP (Table 3). Therefore, in these three homologues, both glucose binding and conformational coupling between the fluorophore and ligand-mediated conformational changes were retained.

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

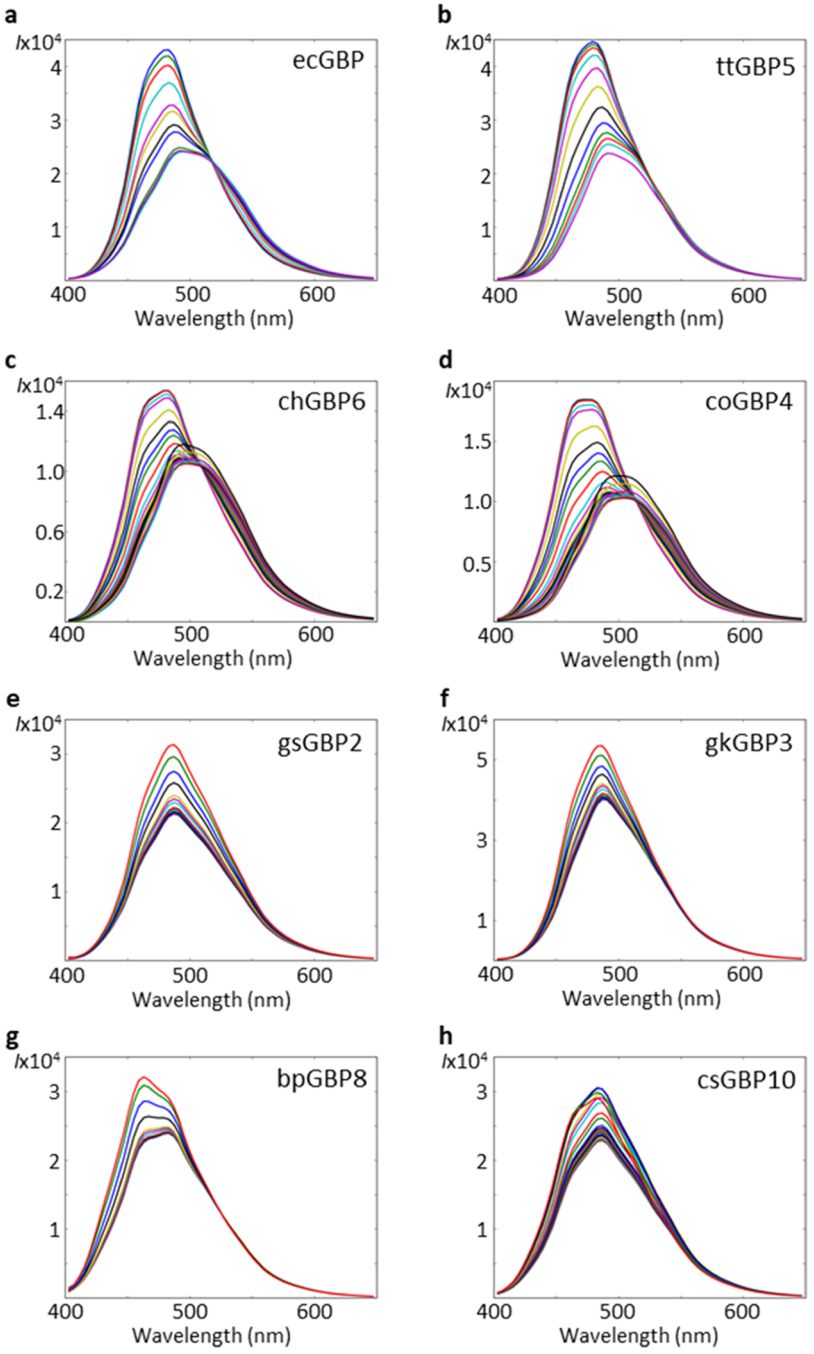


Figure 6. Glucose dependence of emission spectra. The acronym of each protein is shown for individual rows (see Table 2 for naming). Isothermal (room temperature) emission intensity spectra (arbitrary units) were recorded on a Nanodrop3300 spectrophotometer at different glucose concentrations. Dual ratiometric fits were used to calculate the corrected spectra displayed here (eqs 9−11).

The W186C·acrylodan conjugate of the low-identity csGBP10 protein, which has functionally conserved PCS (H_F = 0) with one identity change (H152N), exhibits only a small ratiometric signal using $R_{488,510}$ (Figure 5g). The magnitude of the ratiometric responses is enhanced by monitoring $R_{488,580}$ (Figure 8a). This landscape exhibits a direct glucose response that, unlike the other responsive homologues, is limited to a narrow, elevated temperature range. Extraction of glucose-binding isotherms from this landscape shows that an optimal glucose affinity of ∼9 mM is reached at 42 °C (Figure 8b), whereas at 25 °C, barely any binding is observed, consistent with the minimal shift observed in the glucose-mediated change in fluorescence emission spectra ($\Delta\lambda_B$ = −3.7 nm; Table 3) measured at room temperature (Figure 8c). Glucose-mediated thermal shifts of the wild-type protein monitored with SYPRO (Figure 8d)[77,78] reveal that the protein is a high-affinity glucose-binding homologue (Figure 8f); this affinity is diminished in the W186C·acrylodan conjugate (Figure 8e) due to either the W186C mutation or attachment of the acrylodan, or both. This protein therefore is a *bona fide* homologue of ecGBP, as would be expected from the conservation of the PCS. Despite this conservation, conformational coupling between glucose binding and the acrylodan response is largely lost, leaving the conjugate mostly "stuck" in its blue state. This

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

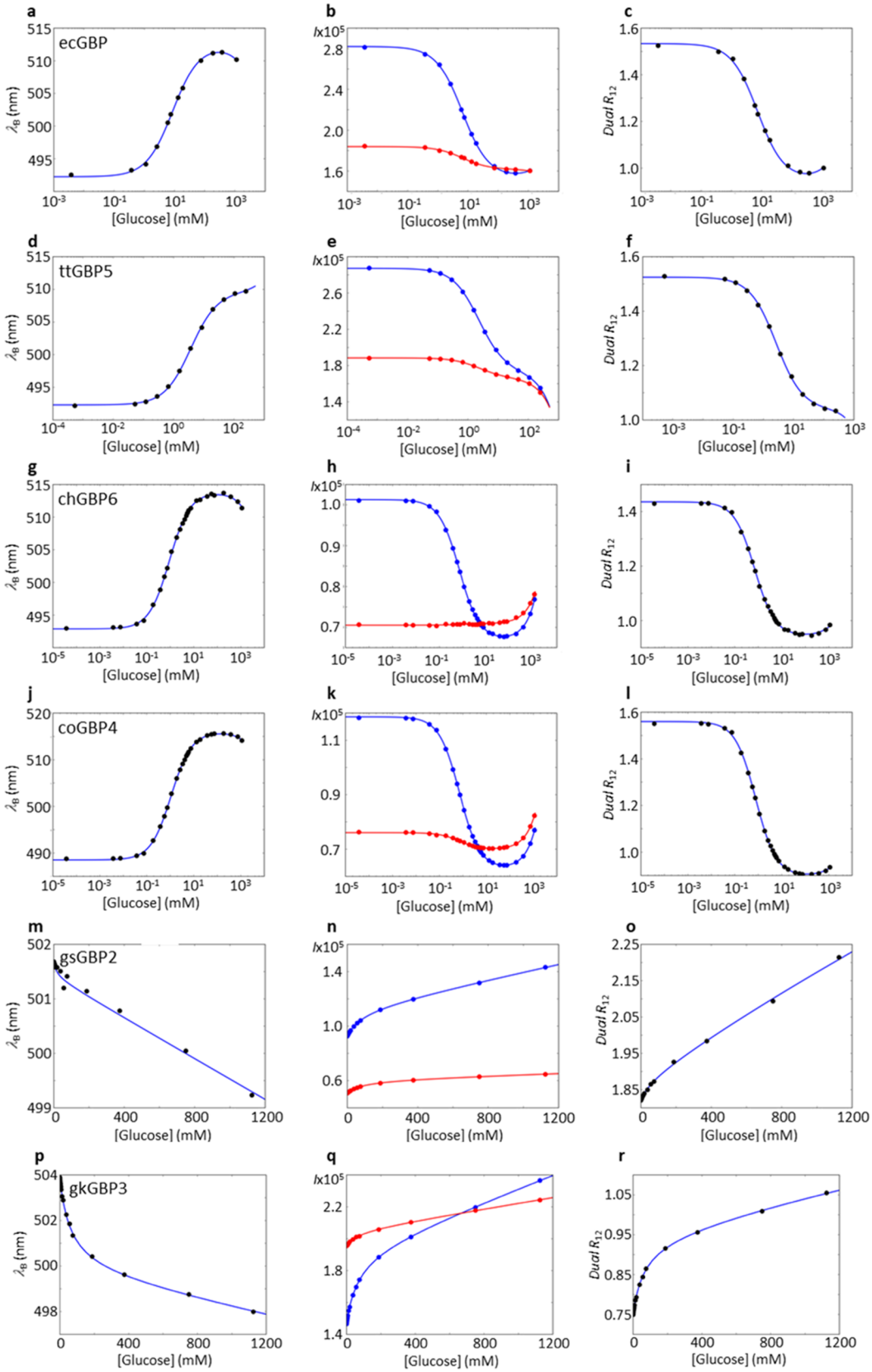


Figure 7. continued

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

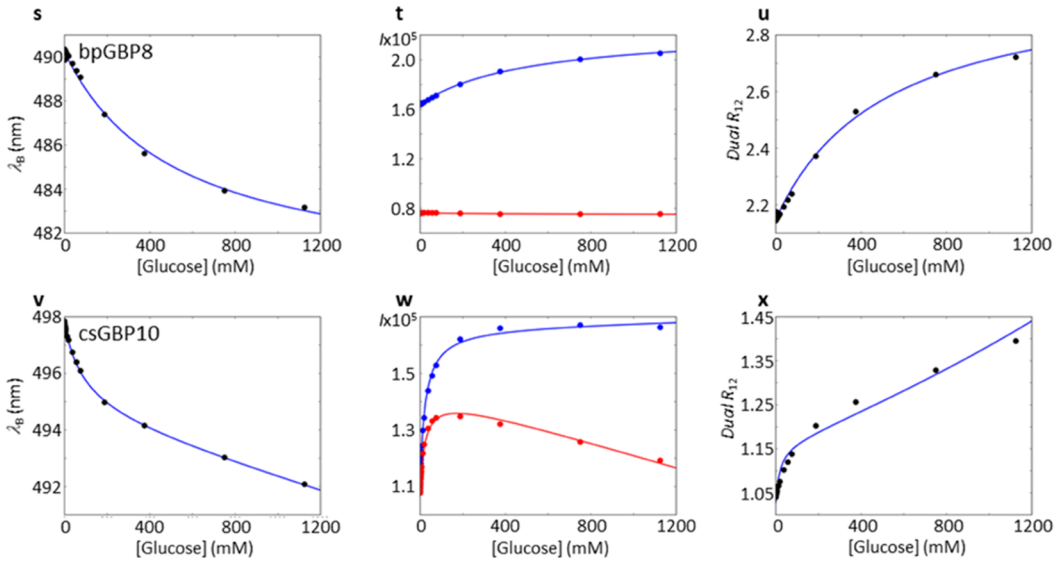


Figure 7. Glucose-binding isotherms extracted from fluorescent emission spectra. Proteins are named by their acronyms, as detailed in Table 2. Extracted emission intensity bands and calculated values for affinities and barychromes are shown in Table 3. Left column, fit of binding isotherms; lines, (eqs 6−8) to the glucose dependence of the barychromes; and circles, calculated from the emission spectra (eqs 3), using a baseline for the glucose complex that is linearly dependent on glucose concentration. ecGPB1, ttGBP5, chGBP6, and coGBP4 exhibit a bathochromic shift upon binding glucose, whereas gsGBP2, gkGBP3, bpGBP8, and csGBP10 undergo a hypsochromic shift. Middle and right columns, dual-wavelength ratiometric fit (eqs 9 and 10). Middle column, fits (lines) to the two intensity bands (circles) corrected by their attenuation values. Right column, fits (lines) to the intensity ratio (circles).

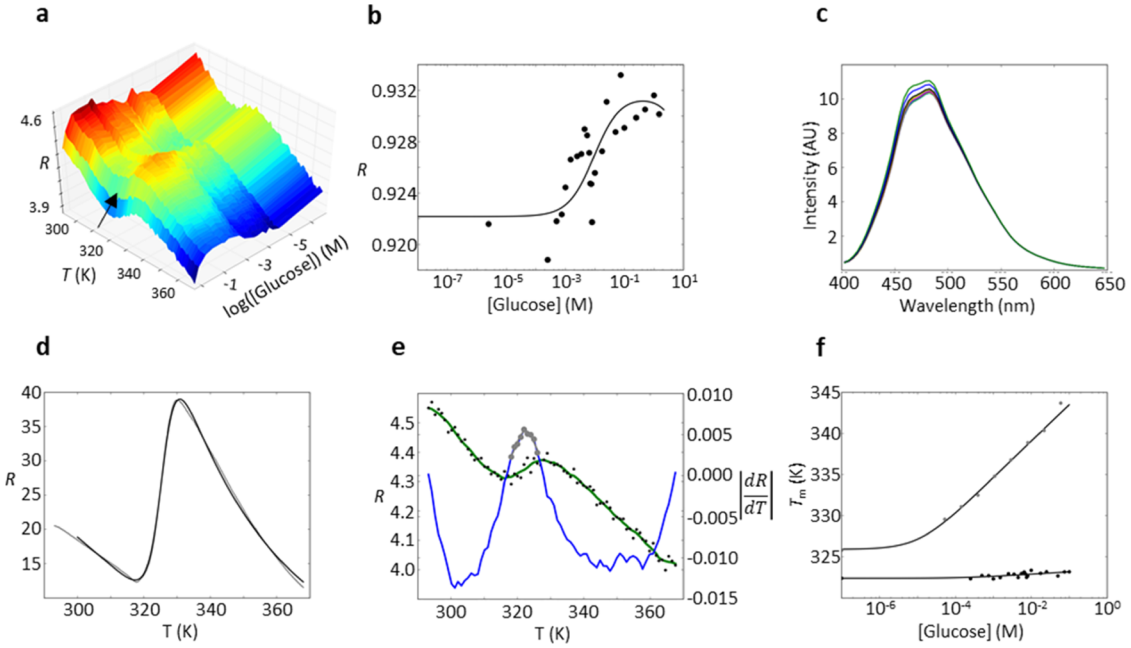


Figure 8. Glucose-binding behavior of csGBP10. (a) Thermal glucose titration landscape using the ratio of 488 and 580 nm. A weak direct response to glucose binding is observed over a narrow temperature range (arrow, $^{488,580}R$ decreases). (b) Ratiometric isotherm extracted at 315 K (42 °C) from the fluorescent landscape exhibits a glucose affinity of 8.9 mM. (c) At 25 °C, emission spectra recorded on the Nanodrop3300 barely exhibit a glucose-dependent response. (d) Thermal stability of the wild-protein (cysteine-free) apoprotein was determined by fitting a two-state unfolding equilibrium (black line) to the fluorescence of the SYPRO complex (gray line). (e) Thermal stability of the apoprotein acrylodan conjugate was determined to fitting the van't Hoff two-state approximation (gray line) to the peak (gray dots) of the first derivative (blue line) of a second-order Savitzky−Golay filter (green line) fit to the acrylodan fluorescence (black circles). (f) Fits of the glucose dependence of the thermal transition midpoint, T_m, determined for the wild-type protein (gray circles) using SYPRO (gray) and for the acrylodan conjugate (black circles) show that the wild-type protein binds glucose with a 12 μM affinity and the W186C acrylodan conjugate with 0.1 mM affinities at their respective apoprotein T_m values.

homologue therefore represents a case in which the SAFE filter accurately predicts ligand binding but does not capture the information necessary to conserve the conformational coupling mechanism.

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

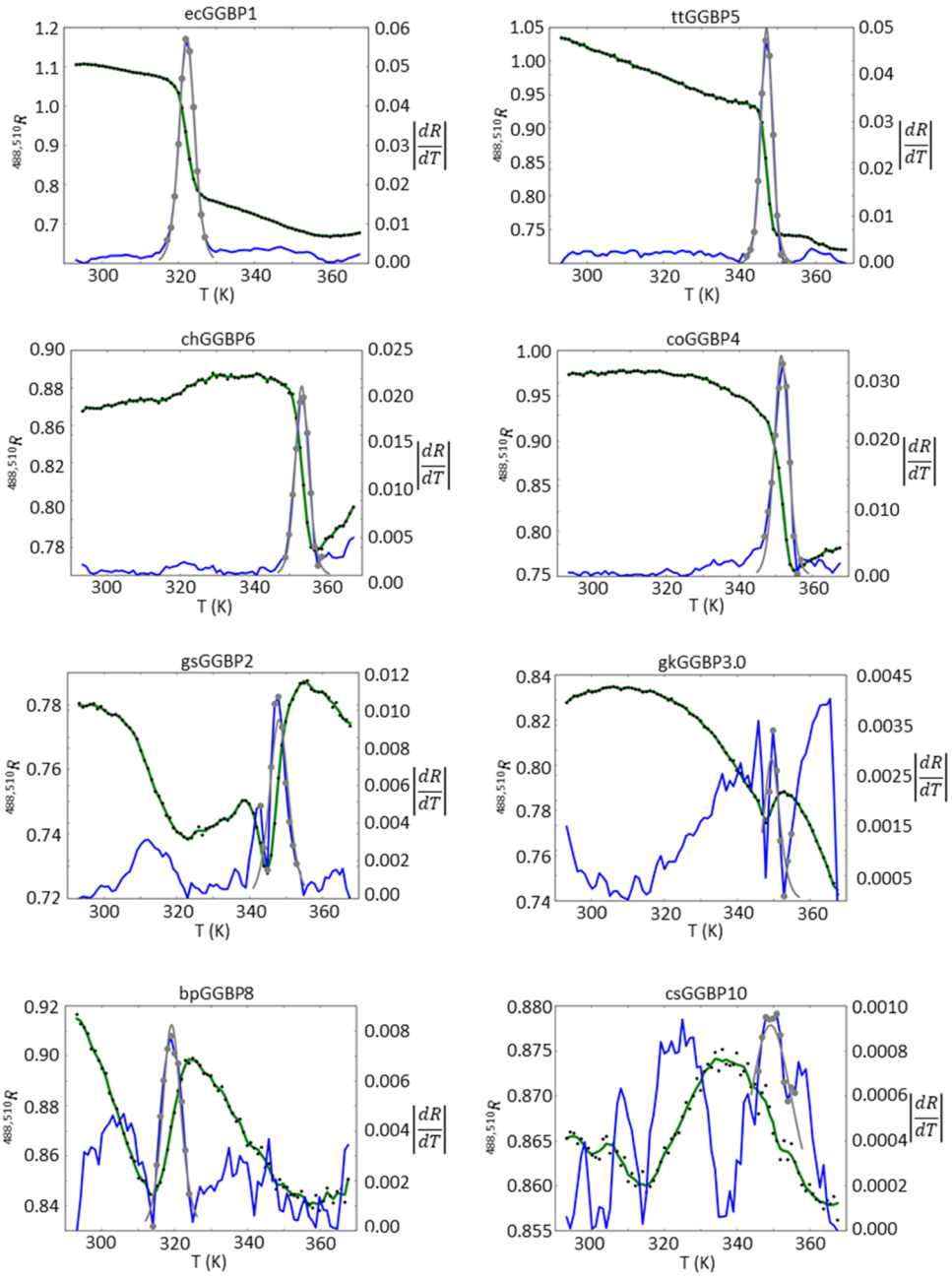


Figure 9. Thermostabilities of the apoprotein acrylodan conjugates. Proteins are named by their acronyms, as detailed in Table 2. Ratiometric intensities, $^{488,510}R$ (black circles), as a function of temperature, T, in the absence of glucose were extracted from the landscapes shown in Figure 5 and fit with a quadratic Savitzky−Golay filter (green lines) that enabled direct calculation of first derivatives, $d^{488,510}R/dT$ (blue lines), within which peaks spanning the two-state folded transition between folded F and unfolded states, U, were identified (gray circles). The van't Hoff approximation for the thermal two-state equilibrium, $F \rightleftharpoons U$, was fit to these peaks using eqs 12−14 (gray lines) to obtain values for $^{apo}T_m$ and ΔH_u (Table 3).

Three distant homologues that have partially lost both functional similarity and identity in their PCS, gsGBP2, gkGBP3, and bpGBP8, retain some degree of coupling between glucose binding and fluorophore behavior; their glucose affinities were considerably weaker than ecGBP (Table 3). In gkGBP3 glucose, binding switches colored states in the opposite direction (hypsochromic shift), from green in the apo-state to a blue glucose complex (Figures 5e,f,h and 6e−g). This effect is barely present in the closely related gsGBP2 homologue, further emphasizing that small sequence changes can affect coupling. All three proteins exhibit green thermally unfolded states as observed in ecGBP. These three homologues therefore illustrate both that partial loss of a PCS is accompanied by weakening of the ligand-binding properties associated with that PCS and that the coupling mechanism of the conjugated fluorophore with glucose binding can exhibit a reversed color pattern.

The most distant homologue, bsGBP7, has also the largest loss in PCS identity. It does not exhibit conformational coupling between glucose binding and the acrylodan response. Furthermore, the colors of the folded and unfolded states are reversed in bsGBP7 (Figure 5i). This homologue therefore illustrates that if the PCS conservation falls below a certain threshold, all semblance of ligand-binding properties is lost.

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

Table 4. Thermostabilities, Fluorescence Emission Maxima, and Glucose Affinities[a]

		thermostability		barychromes			glucose affinity			
protein	mutations	$^{apo}T_m$ (°C)	$\Delta^{apo}H_u$ (kcal/mol)	$^{apo}\lambda_B$ (nm)	$^{glc}\lambda_B$ (nm)	$\Delta\lambda_B$ (nm)	$^{B}K_d$ (mM)	$^{R}K_d$ (mM)	$^{D}K_d$ (mM)	$^{L}K_d$ (mM)
gkGBP3		76.3	166	503.9	499.5	−4.4	63.9	68.2	68.5	102
gkGBP3.6	A140D	76.2	243	503.8	496.6	−7.2	60.6	63.5	90.2	77.4
gkGBP3.10[b]	Q240N	74.4	183	507.4	498.9	−8.5	14.8	16.6	23.4	21.5
gkGBP3.14[b]	A140D, Q240N	76.7	208	507.7	490.7	−17.0	13.4	18.3	28.1	43.3
gkGBP3.18[b]	W16F	77.3	228	506.5	500.2	−4.7	40.5	45.4	52.8	47.4
gkGBP3.19[b]	W16F, A140D	76.8	194	504.9	496.9	−8.0	85.1	88.7	134	135
gkGBP3.20[b]	W16F, Q240N	76.5	180	510.0	500.7	−9.3	21.6	23.9	58.9	42.1
gkGBP3.21[b]	W16F, A140D, Q240N	79.0		507.0	496.6	−10.4	91.7	110	203	112

[a]See Table 3 footnotes. Integrated intensity bands for calculating $^{R}K_d$ and $^{D}K_d$ were extracted from the Nanodrop3300 emission spectra at 460 and 520 nm (20 nm width). [b]Thermostability fits constrained to the [70,85] °C interval corresponding to the unfolding transition.

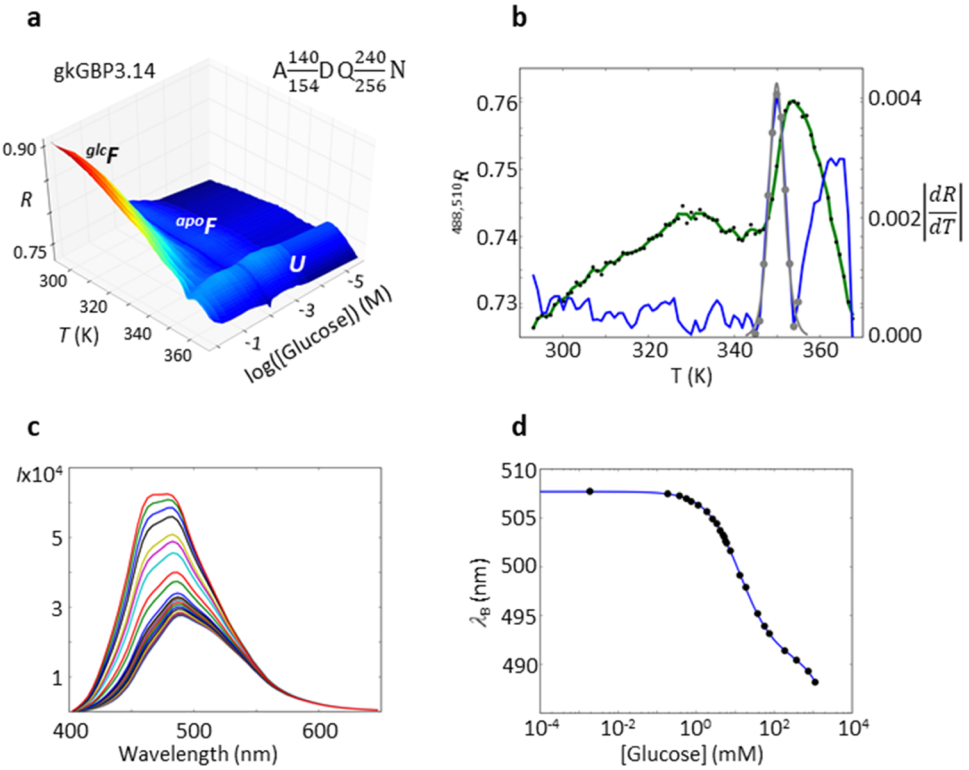


Figure 10. Effects of installing the ecGBP hydrogen-bonding PCS residues in gkGBP3. The fluorescence landscape (a), determination of the $^{apo}T_m$ in the absence of glucose (b), corrected glucose-dependent fluorescent emission spectra (c), and barychrome glucose titration (d) are shown for gkGBP3.14 (Table 4).

Furthermore, it shows that the color pattern of fluorophore states of the folded and unfolded states can also vary.

Thermostabilities. The apoprotein thermostabilities were extracted from fluorescence landscapes by fitting a two-state van't Hoff thermal unfolding equilibrium to the thermal melt determined in the absence of glucose (eqs 12−14; Figure 9). The $^{apo}T_m$ values (Table 3) revealed that six of the seven glucose-binding proteins are thermophilic, as intended.

Other Conformational Coupling Effects. Previously, we established that each of the three regions in the fluorescent landscape of thermal glucose titrations exhibits further subtle variations in emission color due to the presence of additional conformational states, the distribution of which can be analyzed by statistical thermodynamics.[43] For instance, the presence of a "ridge" in the glucose-bound form of ecGBP (r_B, Figure 5a) was interpreted as a thermal transition between two conformations in the glucose complex.[43] This feature appeared to be absent in the three close homologues ttGBP5, chGBP6, and coGBP4. In chGBP6 and coGBP4, a ridge was observed in the folded apoprotein, r_A, which revealed the presence of a conformational transition that is not present in ecGBP (Figure 5c,d). In the more distant homologues gsGBP2, gkGBP3, and bpGBP8, an equivalent ridge was observed in their (blue) folded apoprotein state. In gkGBP3, the plane corresponding to the apoprotein appeared more curved (c_A) than observed in ecGBP (Figure 5f). Analysis of these conformational exchange reactions by more detailed landscape fits with statistical thermodynamics[43] is beyond the scope of this paper.

Installation of the ecGBP PCS into gkGBP3. If the ligand-binding properties (specificity, affinity) of the PCS are independent of the surrounding residues (*i.e.*, the SCS), then it should be possible to convert a partial homologue into a

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

high(er) affinity glucose-binding protein by the full installation of a known PCS. To test this hypothesis, we built various degrees of ecGBP PCS identity reconstructions in gkGBP3. The predicted hydrogen-bonding interactions with bound glucose have undergone two changes in gkGBP3, removing one hydrogen bond altogether (A140, equivalent to 154D in ecGBP) and replacing another with a functionally equivalent residue (Q240, equivalent to N256 in ecGBP). F16 in the aromatic sandwich has been replaced with W16. We mutated these three residues singly and in combination and determined the effects of the mutations (Table 4) on the thermal landscape, apoprotein thermostability ($^{apo}T_m$), barychrometric glucose affinity ($^{B}K_d$), and acrylodan coupling ($\Delta\lambda_B$). The glucose affinity was improved significantly only by installation of Q240N singly (gkGBP3.10) or in combination (gkGBP3.14, gkGBP3.20, gkGBP3.21). Hypsochromic coupling to acrylodan became as effective as the bathochromic effect in ecGBP only if both A140D and Q240N were installed (gkGBP3.14, Figure 10; $\Delta\lambda_B$ = −17 nm, Table 4). The A140D Q240N double mutant exhibited a glucose response that was comparable to ecGBP (gkGBP3.14) but was adversely impacted by the W16F installation (gkGBP3.21).

Improvements in thermostability were largely observed upon installation of W16F. Installation of the complete PCS resulted in a 2.7 K increase in thermostability but at the expense of improvements in signaling and glucose affinity observed for the installation of the two hydrogen-bonding residues (gkGBP3.21). These results are consistent with the hypothesis that the PCS encodes significant functional information but that the SCS also contributes.

CONCLUSIONS

Accurate prediction of protein function from sequence remains a formidable challenge for the periplasmic-binding proteins.[55−61] We show that structure-based techniques accurately identify a subset of specific ligand-binding PBPs within a large collection of closely related sequence homologues. We used fully sequenced bacterial genomes to search for sequence homologues because the effectiveness of the filters used to predict functional conservation could be evaluated by calculating their effect on the number of paralogs. Consistent with previous studies which established that periplasmic-binding proteins can undergo massive gene duplication events,[64,65] we found that paralogs were prevalent in the unfiltered glucose-binding protein sequence family. We applied a structure-based sequence filter that selects homologues in which the residues encoding the direct interactions between protein and glucose (primary complementary surface, PCS) are conserved, using ecGBP as the structural archetype.[29] Two levels of stringency were applied: absolute and functionally equivalent conservation. Both filters drastically reduced the number of paralagous sequences, indicating that glucose-binding homologues were predicted accurately, with the few remaining paralogs representing either differences in pyranoside specificity (glucose is not the true cognate ligand of ecGBP[30]) or biological system effects such as multiple response concentration regimes[80] or multiple protein partners for different transport and signal-transduction networks.

We selected eight homologues for the experimental study of their acrylodan conjugates placed at positions equivalent to W183 in ecGBP. These eight homologues were chosen to span different degrees of predictive certainty, ranging from full to partial functional or absolute conservation of the PCS residues. Furthermore, we favored thermophilic over mesophilic homologues to meet our biotechnological goal of identifying a robust fluorescently responsive sensor that retained the excellent signaling features found in ecGBP183C·acrylodan. Within this experimentally evaluated data set, we identified three homologues (ttGBP5, chGBP6, coGBP4) from thermophilic organisms with moderate sequence conservation (48−51% sequence identities) but fully absolutely conserved PCS residues (H_I = 0). These homologues fully retained both glucose binding and the glucose-mediated response of their acrylodan conjugate. In another homologue with complete functional conservation of the PCS (H_F = 0), but low overall sequence conservation (33%) and loss of one absolutely conserved residue (H_I = 1), csGBP10, we found that glucose binding was retained, but the response of the Acrylodan conjugate was severely compromised. In four homologues with sequence identities in the 30−34% twilight range and loss of one functional residue (H_F = 1; gkGBP2, gkGBP3, bsGBP7, bpGBP8), both glucose binding and the fluorescence response were compromised. These results illustrate that the values of the SAFE filters successfully predict ligand-binding properties of a distant sequence homologue even within the twilight zone.

Function prediction is particularly challenging in the cases presented here because we also require conservation of coupling of the fluorescence properties of the covalently attached acrylodan and glucose-mediated conformational changes. A systematic study indicated that the probability of establishing linkage between ligand binding and fluorophore response in engineered fluorescent biosensors is relatively low (∼4%), even if attachment positions are placed in regions that are known to undergo local conformational changes in response to ligand binding.[6,42] It is therefore an open question to what extent equivalently positioned fluorophore conjugates to retain coupling to ligand binding in homologues. Deterioration of coupling between ligand-mediated protein motions and the acrylodan response correlates with the degree of loss of absolute PCS conservation. Furthermore, the retention of glucose and loss of fluorescent signaling in csGBP10 showed that the information required for the coupling mechanism is not fully encoded within the PCS. Even so, in the partial homologue gkGBP3, we demonstrated that "installation" of a subset of conserved residues (gkGBP3.10 and gkGBP3.14, Table 4) in a partially retained PCS largely restores the acrylodan response and improves glucose affinity but with a color change that is the reverse (green apoprotein → blue glucose complex; $\Delta\lambda_B$ < 0) of the other responsive homologues (blue → green; $\Delta\lambda_B$ > 0). These results suggest that the coupling of the fluorescence response to protein conformational changes is critically dependent on detailed local interactions involving one or more residues, the identity of which is as yet unknown. Retention of coupling therefore is uncertain even in true homologues, consistent with the challenges encountered in predicting ligand-responsive fluorescent conjugates *ab initio*.

Three homologues are excellent candidates for fulfilling the technological goal of constructing a robust version of the ecGBP183C·acrylodan conjugate: ttGBP5, coGBP4, and chGBP6. These proteins retained high-affinity glucose binding, blue → green coupling between conjugate fluorescence properties and glucose-mediated conformational change, and improved thermostability by 25−31 °C. Of these, we selected ttGBP5 because it expressed most easily and bore the closest resemblance to the ecGBP thermal glucose binding and folding

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

fluorescent landscape. We anticipate that this protein will be incorporated into the next-generation robust, fluorescently responsive, ratiometric glucose sensors.

■ ASSOCIATED CONTENT

Accession Codes

ecGBP: NC_010468, YP_001724487.1. gsGBP2: NC_014651, YP_003989526.1. gkGBP3: WP_041467870.1. coGBP4: NC_014392, YP_003839461.1. ttGBP5: NC_014410, YP_003852930.1. chGBP6: NC_014652, YP_003991244.1. bsGBP7: NC_014117, YP_003608256.1. bpGBP8: NC_014389, WP_013281503.1. csGBP10: NC_014376, YP_003822565.1.

■ AUTHOR INFORMATION

Corresponding Author

Homme W. Hellinga − *Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, United States;* orcid.org/0000-0002-4272-3800; Phone: 919-681-5885; Email: hwh@biochem.duke.edu

Author

Malin J. Allert − *Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.biochem.1c00738

Notes

The authors declare the following competing financial interest(s): The authors are co-inventors on patents for technologies based on proteins described in this manuscript. The source code for the BMC package and the experimental data can be downloaded from https:/doi.org/10.5061/dryad.ttdz08m0f.

■ ACKNOWLEDGMENTS

This work was supported in part by a Corporate Research Agreement between Duke University and SenGenix Inc. (Durham, North Carolina; SPS 201039) and by the Department of Biochemistry. The authors thank Anna Yasinovskaya for assistance with some protein preparation.

■ ABBREVIATIONS

PBP, periplasmic solute-binding protein; SAFE, structure-assisted functional extrapolation; PCS, primary complementary surface; BMC, biomolecular computing; PMQ, parallel machine queue; MSA, multiple sequence alignment; ORF, open reading frame; SHS, sequence homologue set; SCS, secondary contact surface

■ REFERENCES

(1) Marvin, J. S.; Hellinga, H. W. Engineering biosensors by introducing fluorescent allosteric signal transducers: construction of a novel glucose sensor. *J. Am. Chem. Soc.* **1998**, *120*, 7−11.

(2) Hsieh, H. V.; Pfeiffer, Z. A.; Amiss, T. J.; Sherman, D. B.; Pitner, J. B. Direct detection of glucose by surface plasmon resonance with bacterial glucose/galactose-binding protein. *Biosens. Bioelectron.* **2004**, *19*, 653−660.

(3) Khan, F.; Gnudi, L.; Pickup, J. C. Fluorescence-based sensing of glucose using engineered glucose/galactose-binding protein: a comparison of fluorescence resonance energy transfer and environmentally sensitive dye labelling strategies. *Biochem. Biophys. Res. Commun.* **2008**, *365*, 102−106.

(4) Salins, L. L.; Ware, R. A.; Ensor, C. M.; Daunert, S. A novel reagentless sensing system for measuring glucose based on the galactose/glucose-binding protein. *Anal. Biochem.* **2001**, *294*, 19−26.

(5) Scognamiglio, V.; Aurilia, V.; Cennamo, N.; Ringhieri, P.; Iozzino, L.; Tartaglia, M.; Staiano, M.; Ruggiero, G.; Orlando, P.; Labella, T.; Zeni, L.; Vitale, A.; D'Auria, S. D-galactose/D-glucose-binding Protein from *Escherichia coli* as Probe for a Non-consuming Glucose Implantable Fluorescence Biosensor. *Sensors* **2007**, *7*, 2484−2491.

(6) de Lorimier, R. M.; Smith, J. J.; Dwyer, M. A.; Looger, L. L.; Sali, K. M.; Paavola, C. D.; Rizk, S. S.; Sadigov, S.; Conrad, D. W.; Loew, L.; Hellinga, H. W. Construction of a fluorescent biosensor family. *Protein Sci.* **2002**, *11*, 2655−2675.

(7) Sakaguchi-Mikami, A.; Taneoka, A.; Yamoto, R.; Ferri, S.; Sode, K. Engineering of ligand specificity of periplasmic binding protein for glucose sensing. *Biotechnol. Lett.* **2008**, *30*, 1453−1460.

(8) Saxl, T.; Khan, F.; Ferla, M.; Birch, D.; Pickup, J. A fluorescence lifetime-based fibre-optic glucose sensor using glucose/galactose-binding protein. *Analyst* **2011**, *136*, 968−972.

(9) Weidemaier, K.; Lastovich, A.; Keith, S.; Pitner, J. B.; Sistare, M.; Jacobson, R.; Kurisko, D. Multi-day pre-clinical demonstration of glucose/galactose binding protein-based fiber optic sensor. *Biosens. Bioelectron.* **2011**, *26*, 4117−4123.

(10) Judge, K.; Morrow, L.; Lastovich, A. G.; Kurisko, D.; Keith, S. C.; Hartsell, J.; Roberts, B.; McVey, E.; Weidemaier, K.; Win, K.; Hompesch, M. Continuous glucose monitoring using a novel glucose/galactose binding protein: results of a 12-hour feasibility study with the becton dickinson glucose/galactose binding protein sensor. *Diabetes Technol. Ther.* **2011**, *13*, 309−317.

(11) Thomas, K. J.; Sherman, D. B.; Amiss, T. J.; Andaluz, S. A.; Pitner, J. B. A long-wavelength fluorescent glucose biosensor based on bioconjugates of galactose/glucose binding protein and Nile Red derivatives. *Diabetes Technol. Ther.* **2006**, *8*, 261−268.

(12) Tolosa, L.; Gryczynski, I.; Eichhorn, L. R.; Dattelbaum, J. D.; Castellano, F. N.; Rao, G.; Lakowicz, J. R. Glucose sensor for low-cost lifetime-based sensing using a genetically engineered protein. *Anal. Biochem.* **1999**, *267*, 114−120.

(13) Ge, X.; Tolosa, L.; Rao, G. Dual-labeled glucose binding protein for ratiometric measurements of glucose. *Anal. Chem.* **2004**, *76*, 1403−1410.

(14) El-Sayed, M. M.; Brown, S. R.; Mupparapu, K.; Tolosa, L. The effect of pH on the glucose response of the glucose-galactose binding protein L255C labeled with Acrylodan. *Int. J. Biol. Macromol.* **2016**, *86*, 282−287.

(15) Tiangco, C.; Fon, D.; Sardesai, N.; Kostov, Y.; Sevilla, F., III; Rao, G.; Tolosa, L. Fiber optic biosensor for transdermal glucose based on the glucose binding protein. *Sens. Actuators, B* **2017**, *242*, 569−576.

(16) Hsieh, H. V.; Sherman, D. B.; Andaluz, S. A.; Amiss, T. J.; Pitner, J. B. Fluorescence resonance energy transfer glucose sensor from site-specific dual labeling of glucose/galactose binding protein using ligand protection. *J. Diabetes Sci. Technol.* **2012**, *6*, 1286−1295.

(17) Der, B. S.; Dattelbaum, J. D. Construction of a reagentless glucose biosensor using molecular exciton luminescence. *Anal. Biochem.* **2008**, *375*, 132−140.

(18) Ye, K.; Schultz, J. S. Genetic engineering of an allosterically based glucose indicator protein for continuous glucose monitoring by fluorescence resonance energy transfer. *Anal. Chem.* **2003**, *75*, 3451−3459.

(19) Fehr, M.; Lalonde, S.; Lager, I.; Wolff, M. W.; Frommer, W. B. In vivo imaging of the dynamics of glucose uptake in the cytosol of COS-7 cells by fluorescent nanosensors. *J. Biol. Chem.* **2003**, *278*, 19127−19133.

(20) Yoo, E.-H.; Lee, S.-Y. Glucose biosensors: an overview of use in clinical practice. *Sensors* **2010**, *10*, 4558−4576.

(21) Cash, K. J.; Clark, H. A. Nanosensors and nanomaterials for monitoring glucose in diabetes. *Trends Mol. Med.* **2010**, *16*, 584−593.

(22) Mathers, C. D.; Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* **2006**, *3*, No. e442.
(23) International Diabetes Federation. *IDF Diabetes Atlas*, 8th ed.; International Diabetes Federation, 2017.
(24) Tam, R.; Saier, M. H., Jr. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* **1993**, *57*, 320−346.
(25) Berntsson, R. P.; Smits, S. H.; Schmitt, L.; Slotboom, D. J.; Poolman, B. A structural classification of substrate-binding proteins. *FEBS Lett.* **2010**, *584*, 2606−2617.
(26) Scheepers, G. H.; Lycklama, A. N. J. A.; Poolman, B. An updated structural classification of substrate-binding proteins. *FEBS Lett.* **2016**, *590*, 4393−4401.
(27) Chu, B. C.; Vogel, H. J. A structural and functional analysis of type III periplasmic and substrate binding proteins: their role in bacterial siderophore and heme transport. *Biol. Chem.* **2011**, *392*, 39−52.
(28) Flocco, M. M.; Mowbray, S. L. The 1.9 A x-ray structure of a closed unliganded form of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J. Biol. Chem.* **1994**, *269*, 8931−8936.
(29) Borrok, M. J.; Kiessling, L. L.; Forest, K. T. Conformational changes of glucose/galactose-binding protein illuminated by open, unliganded, and ultra-high-resolution ligand-bound structures. *Protein Sci.* **2007**, *16*, 1032−1041.
(30) Sooriyaarachchi, S.; Ubhayasekera, W.; Boos, W.; Mowbray, S. L. X-ray structure of glucose/galactose receptor from *Salmonella typhimurium* in complex with the physiological ligand, (2R)-glyceryl-beta-D-galactopyranoside. *FEBS J.* **2009**, *276*, 2116−2124.
(31) Vyas, N. K.; Vyas, M. N.; Quiocho, F. A. A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. *Nature* **1987**, *327*, 635−638.
(32) Allert, M. J.; Hellinga, H. W. Harnessing Environmental Ca(2+) for Extracellular Protein Thermostabilization. *Biochemistry* **2020**, *59*, 3725−3740.
(33) Herman, P.; Vecer, J.; Barvik, I., Jr.; Scognamiglio, V.; Staiano, M.; de Champdore, M.; Varriale, A.; Rossi, M.; D'Auria, S. The role of calcium in the conformational dynamics and thermal stability of the D-galactose/D-glucose-binding protein from *Escherichia coli*. *Proteins* **2005**, *61*, 184−195.
(34) D'Auria, S.; Alfieri, F.; Staiano, M.; Pelella, F.; Rossi, M.; Scire, A.; Tanfani, F.; Bertoli, E.; Grycznyski, Z.; Lakowicz, J. R. Structural and thermal stability characterization of *Escherichia coli* D-galactose/D-glucose-binding protein. *Biotechnol. Prog.* **2004**, *20*, 330−337.
(35) Luck, L. A.; Falke, J. J. Open conformation of a substrate-binding cleft: 19F NMR studies of cleft angle in the D-galactose chemosensory receptor. *Biochemistry* **1991**, *30*, 6484−6490.
(36) Ortega, G.; Castano, D.; Diercks, T.; Millet, O. Carbohydrate affinity for the glucose-galactose binding protein is regulated by allosteric domain motions. *J. Am. Chem. Soc.* **2012**, *134*, 19869−19876.
(37) Careaga, C. L.; Sutherland, J.; Sabeti, J.; Falke, J. J. Large amplitude twisting motions of an interdomain hinge: a disulfide trapping study of the galactose-glucose binding protein. *Biochemistry* **1995**, *34*, 3048−3055.
(38) Unione, L.; Ortega, G.; Mallagaray, A.; Corzana, F.; Perez-Castells, J.; Canales, A.; Jimenez-Barbero, J.; Millet, O. Unraveling the Conformational Landscape of Ligand Binding to Glucose/Galactose-Binding Protein by Paramagnetic NMR and MD Simulations. *ACS Chem. Biol.* **2016**, *11*, 2149−2157.
(39) Grünewald, F. S. Periplasmic Binding Proteins In Biosensing Applications. In *Advances in Chemical Bioanalysis*; Matysik, F. M., Ed.; Bioanalytical Reviews; Springer International: Switzerland, 2014; pp 205−235.
(40) Edwards, K. A. Periplasmic-binding protein-based biosensors and bioanalytical assay platforms: Advances, considerations, and strategies for optimal utility. *Talanta Open* **2021**, *3*, No. 100038.
(41) Vyas, N. K.; Vyas, M. N.; Quiocho, F. A. Sugar and signal-transducer binding sites of the *Escherichia coli* galactose chemoreceptor protein. *Science* **1988**, *242*, 1290−1295.
(42) Fan, C.; Plaxco, K. W.; Heeger, A. J. Biosensors based on binding-modulated donor-acceptor distances. *Trends Biotechnol.* **2005**, *23*, 186−192.
(43) Allert, M. J.; Hellinga, H. W. Describing Complex Structure-Function Relationships in Biomolecules at Equilibrium. *J. Mol. Biol.* **2020**, *432*, 1926−1951.
(44) Demchenko, A. P. The concept of lambda-ratiometry in fluorescence sensing and imaging. *J. Fluoresc.* **2010**, *20*, 1099−1128.
(45) Demchenko, A. P. Practical aspects of wavelength ratiometry in the studies of intermolecular interactions. *J. Mol. Struct.* **2014**, *1077*, 51−67.
(46) Amiss, T. J.; Sherman, D. B.; Nycz, C. M.; Andaluz, S. A.; Pitner, J. B. Engineering and rapid selection of a low-affinity glucose/galactose-binding protein for a glucose biosensor. [Erratum to document cited in CA148:073340]. *Protein Sci.* **2007**, *16*, 2775.
(47) Khan, F.; Saxl, T. E.; Pickup, J. C. Fluorescence intensity- and lifetime-based glucose sensing using an engineered high-Kd mutant of glucose/galactose-binding protein. *Anal. Biochem.* **2010**, *399*, 39−43.
(48) Hickey, J. M.; Sahni, N.; Chaudhuri, R.; D'Souza, A.; Metters, A.; Joshi, S. B.; Russell Middaugh, C.; Volkin, D. B. Effect of acrylodan conjugation and forced oxidation on the structural integrity, conformational stability, and binding activity of a glucose binding protein SM4 used in a prototype continuous glucose monitor. *Protein Sci.* **2017**, *26*, 527−535.
(49) Sahni, N.; Chaudhuri, R.; Hickey, J. M.; Manikwar, P.; D'Souza, A.; Metters, A.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B. Preformulation Characterization, Stabilization, and Formulation Design for the Acrylodan-Labeled Glucose-Binding Protein SM4-AC. *J. Pharm. Sci.* **2017**, *106*, 1197−1210.
(50) Todd, A. E.; Orengo, C. A.; Thornton, J. M. Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* **2001**, *307*, 1113−1143.
(51) Tian, W.; Skolnick, J. How well is enzyme function conserved as a function of pairwise sequence identity? *J. Mol. Biol.* **2003**, *333*, 863−882.
(52) Devos, D.; Valencia, A. Practical limits of function prediction. *Proteins* **2000**, *41*, 98−107.
(53) Chothia, C.; Lesk, A. M. The relation between the divergence of sequence and structure in protein. *EMBO J.* **1986**, *5*, 823−826.
(54) Rost, B. Twilight zone of protein sequence alignments. *Protein Eng., Des. Sel.* **1999**, *12*, 85−94.
(55) Maqbool, A.; Horler, R. S.; Muller, A.; Wilkinson, A. J.; Wilson, K. S.; Thomas, G. H. The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem. Soc. Trans.* **2015**, *43*, 1011−1017.
(56) Johnson, J. M.; Church, G. M. Predicting ligand-binding function in families of bacterial receptors. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3965−3970.
(57) Giuliani, S. E.; Frank, A. M.; Collart, F. R. Functional assignment of solute-binding proteins of ABC transporters using a fluorescence-based thermal shift assay. *Biochemistry* **2008**, *47*, 13974−13984.
(58) Boucher, N.; Noll, K. M. Ligands of thermophilic ABC transporters encoded in a newly sequenced genomic region of Thermotoga maritima MSB8 screened by differential scanning fluorimetry. *Appl. Environ. Microbiol.* **2011**, *77*, 6395−6399.
(59) Vetting, M. W.; Al-Obaidi, N.; Zhao, S.; San Francisco, B.; Kim, J.; Wichelecki, D. J.; Bouvier, J. T.; Solbiati, J. O.; Vu, H.; Zhang, X.; Rodionov, D. A.; Love, J. D.; Hillerich, B. S.; Seidel, R. D.; Quinn, R. J.; Osterman, A. L.; Cronan, J. E.; Jacobson, M. P.; Gerlt, J. A.; Almo, S. C. Experimental strategies for functional annotation and metabolism discovery: targeted screening of solute binding proteins and unbiased panning of metabolomes. *Biochemistry* **2015**, *54*, 909−931.
(60) Åstrand, M.; Cuellar, J.; Hytonen, J.; Salminen, T. A. Predicting the ligand-binding properties of *Borrelia burgdorferi* s.s. Bmp proteins in light of the conserved features of related Borrelia proteins. *J. Theor. Biol.* **2019**, *462*, 97−108.

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293

(61) Fernandez, M.; Rico-Jimenez, M.; Ortega, A.; Daddaoua, A.; Garcia Garcia, A. I.; Martin-Mora, D.; Torres, N. M.; Tajuelo, A.; Matilla, M. A.; Krell, T. Determination of Ligand Profiles for Pseudomonas aeruginosa Solute Binding Proteins. *Int. J. Mol. Sci.* **2019**, *20*, 5156−5177.
(62) Cuneo, M. J.; Beese, L. S.; Hellinga, H. W. Structural analysis of semi-specific oligosaccharide recognition by a cellulose-binding protein of *Thermotoga maritima* reveals adaptations for functional diversification of the oligopeptide periplasmic binding protein fold. *J. Biol. Chem.* **2009**, *284*, 33217−33223.
(63) Nanavati, D. M.; Thirangoon, T.; Noll, K. M. Several archaeal homologs to putative oligopeptide-binding proteins encoded by Thermotoga maritima bind sugars. *Appl. Environ. Microbiol.* **2006**, *72*, 1336−1345.
(64) Antoine, R.; Jacob-Dubuisson, F.; Drobecq, H.; Willery, E.; Lesjean, S.; Locht, C. Overrepresentation of a gene family encoding extracytoplasmic solute receptors in Bordetella. *J. Bacteriol.* **2003**, *185*, 1470−1474.
(65) Rosa, L. T.; Springthorpe, V.; Bianconi, M. E.; Thomas, G. H.; Kelly, D. J. Massive over-representation of solute-binding proteins (SBPs) from the tripartite tricarboxylate transporter (TTT) family in the genome of the alpha-proteobacterium Rhodoplanes sp. Z2-YC6860. *Microb. Genomics* **2018**, *4*, No. e000176.
(66) Koonin, E. V. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **2005**, *39*, 309−338.
(67) George, R. A.; Spriggs, R. V.; Bartlett, G. J.; Gutteridge, A.; MacArthur, M. W.; Porter, C. T.; Al-Lazikani, B.; Thornton, J. M.; Swindells, M. B. Effective function annotation through catalytic residue conservation. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 12299−12304.
(68) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.
(69) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; Thompson, J. D.; Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539.
(70) Hengen, P. N. Purification of His-Tag fusion proteins from *Escherichia coli*. *Trends Biochem. Sci.* **1995**, *20*, 285−286.
(71) Cox, J. C.; Lape, J.; Sayed, M. A.; Hellinga, H. W. Protein fabrication automation. *Protein Sci.* **2007**, *16*, 379−390.
(72) Allert, M.; Cox, J. C.; Hellinga, H. W. Multifactorial determinants of protein expression in prokaryotic open reading frames. *J. Mol. Biol.* **2010**, *402*, 905−918.
(73) Gill, S. C.; von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **1989**, *182*, 319−326.
(74) Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; Grosdidier, A.; Hernandez, C.; Ioannidis, V.; Kuznetsov, D.; Liechti, R.; Moretti, S.; Mostaguir, K.; Redaschi, N.; Rossier, G.; Xenarios, I.; Stockinger, H. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **2012**, *40*, W597−W603.
(75) Prendergast, F. G.; Meyer, M.; Carlson, G. L.; Iida, S.; Potter, J. D. Synthesis, spectral properties, and use of 6-acryloyl-2-dimethylaminonapthalene (acrylodan). *J. Biol. Chem.* **1983**, *258*, 7541−7544.
(76) Press, W. H. T.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes. The Art of Scientific Computing*, 3rd ed.; Cambridge University Press, 2007.
(77) Lo, M. C.; Aulabaugh, A.; Jin, G.; Cowling, R.; Bard, J.; Malamas, M.; Ellestad, G. Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem.* **2004**, *332*, 153−159.
(78) Layton, C. J.; Hellinga, H. W. Thermodynamic analysis of ligand-induced changes in protein thermal unfolding applied to high-throughput determination of ligand affinities with extrinsic fluorescent dyes. *Biochemistry* **2010**, *49*, 10831−10841.
(79) Chenna, R.; Sugawara, H.; Koike, T.; Lopez, R.; Gibson, T. J.; Higgins, D. G.; Thompson, J. D. Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* **2003**, *31*, 3497−3500.
(80) Ghimire-Rijal, S.; Lu, X.; Myles, D. A.; Cuneo, M. J. Duplication of genes in an ATP-binding cassette transport system increases dynamic range while maintaining ligand specificity. *J. Biol. Chem.* **2014**, *289*, 30090−30100.
(81) Petersen, T. N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785−786.

https://doi.org/10.1021/acs.biochem.1c00738
Biochemistry 2022, 61, 276−293