Structure-based design of versatile biosensors for small molecules based on the PAS domain of a thermophilic histidine kinase

Kai U. Cormann, Meike Baumgart, Michael Bott[#]

IBG-1: Biotechnology, Institute of Bio- and Geosciences, Forschungszentrum Jülich, 52425 Jülich, Germany

#Address for correspondence: m.bott@fz-juelich.de, phone +49 2461 613294, fax +49 2461 612710

KUC: orcid.org/0000-0003-0772-6456

MBa: orcid.org/0000-0002-9874-1151

MBo: orcid.org/0000-0002-4701-8254

ABSTRACT

The development of biosensors for *in vitro* quantification of small molecules such as metabolites or man-made chemicals is still a major challenge. Here we show that engineered variants of the sensory PAS domain of the histidine kinase CitA of the thermophilic bacterium *Geobacillus thermoleovorans* represent promising alternatives to established biorecognition elements. By combining binding site grafting and rational design we constructed protein variants binding L-malate, ethylmalonate or the aromatic compound phthalate instead of the native ligand citrate. Due to more favorable entropy contributions, the wild-type protein and its engineered variants exhibited increased (nano- to micromolar) affinities and improved enantioselectivity compared to CitA homologs of mesophilic organisms. Ligand binding was directly converted into an optical signal which was preserved after immobilization of the protein. A fluorescently labeled variant was used to quantify ethylmalonate, an urinary biomarker for ethylmalonic encephalopathy, in synthetic urine, thereby demonstrating the applicability of the sensor in complex samples.

KEYWORDS

rational design; protein engineering; PAS domain; conformational entropy; switch-based biosensor; thermophilic bacteria

Owing to their almost infinite combinatorial sequence space combined with established natural or nature-inspired selection processes, biopolymers such as proteins and nucleic acids offer unmatched affinity and selectivity for chemosensing of low molecular weight (LMW) compounds. Antibodies are one of the earliest and still the most widespread class of biorecognition elements as they can readily be obtained by high-throughput screening approaches [1] and show affinities in the subnanomolar range. However, despite some remarkable successes [2], raising antibodies against small molecules (haptens) is still a major challenge because it requires covalent linkage of the hapten to a larger carrier molecule prior to immunization. In consequence these antibodies show reduced affinities for the unmodified target compared to their counterparts raised against larger peptides or proteins. Aptamers, a second class of biorecognition elements with increasing relevance, are nucleic acids molecules binding specific ligands [3]. In contrast to antibodies, aptamers undergo major conformational changes upon ligand binding, thereby enabling direct detection of binding events with optical or electrochemical readout [4, 5]. Even though a few aptamers with nanomolar affinities for compounds <200 Da are available, there seems to be a positive correlation between the molecular weight of the analyte and its affinity to the aptamer [6].

To circumvent these limitations, naturally occurring ligand binding domains (LBDs) offer a promising alternative for small molecule sensing because over billions of years these proteins have already been selected by evolution to bind a vast set of different compounds with high affinity and selectivity. Similar to aptamers, LBDs frequently couple ligand binding to structural switching making them an ideal tool for the construction of reagentless single molecule sensors [7]. Although different classes of LBDs such as transcription factors [8] were exploited for the detection of LMW compounds, up to now most studies have focused on bacterial periplasmic binding proteins [9]. These efforts resulted in an electrochemical sensor for maltose quantification [10], single molecule FRET sensors for *in vivo* imaging [11], as well as variants modified with fluorescent dyes for the detection of various sugars, ions and amino

acids [12, 13]. Fluorescent periplasmic binding proteins were integrated into devices for continuous monitoring of glucose concentrations [14] and entered clinical studies for applications inside the human body [15]. Nevertheless, the application range of LBD-based biosensors is so far limited to analytes for which evolutionary pressure has generated a corresponding protein variant. Hellinga and co-workers addressed this drawback by rational protein engineering and reported periplasmic binding proteins with nanomolar affinities for the explosive trinitrotoluene and a degradation product of the nerve agent soman [16, 17], but a further study showed that these variants were severely destabilized and did not bind their respective ligands in isothermal titration calorimetry (ITC) experiments [18]. More recent breakthroughs in rational protein design finally succeeded in building semi-artificial LBDs with nanomolar affinities for digoxigenin [19] and the opioid fentanyl [20]. So far the scaffolds used in these designs have not been employed for biosensing *in vitro*, and furthermore they harbor prearranged binding sites and thus lack a conformational change which can be coupled to an optical signal.

Here we report a set of semi-artificial LBDs based on the extracytoplasmic Per-Arnt-Sim (PAS) domain of the citrate-sensing histidine kinase CitA of the thermophilic bacterium *Geobacillus thermoleovorans*, which bind their non-native ligands with micromolar affinities. Due to the increased affinity compared to homologs of mesophilic organisms, the stability towards mutations, chemical modifications and immobilization as well as the functionality in complex sample matrices, we propose that *Gt*CitAP represents an ideal scaffold for the design of semi-artificial LBDs in future biosensor applications.

RESULTS AND DISCUSSION

Wild-type GtCitAP is a high-affinity citrate binder. The purpose of our study was to explore the potential of the PAS domain scaffold for *in vitro* biosensing as this protein family is one of the most widespread sensory domains in bacterial signaling proteins [21], and thus

there are ligand-binding PAS domains for a huge variety of biologically relevant molecules available. In previous studies we identified the periplasmic PAS domain of the histidine kinases CitA from *Klebsiella pneumoniae* and *Escherichia coli* as citrate-binding proteins [22, 23] and identified amino acid residues important for binding [24]. For the *K. pneumoniae* protein high-resolution crystal structures are available [25, 26] and it has already been converted into a fluorescent single molecule sensor in previous studies [27, 28]. As *in vitro* biosensors should exhibit long-term stability and our initial mutational studies with the *K. pneumoniae* protein failed due to instability, we decided to search for CitA homologs in the genomes of thermophilic bacteria, because these organisms are known to be an excellent source of highly stable protein variants[29]. A BLAST search [30] identified CitA homologs in several *Geobacillus* species out of which the protein of *G. thermoleovorans* (Fig. S1) was chosen for further characterization. In the course of this study a highly similar histidine kinase of *Geobacillus thermodentrificans* was reported to function as citrate sensor [31], thereby confirming that the chosen PAS domain was most likely specific for citrate.

The DNA sequence encoding residues 31 - 161 of *Gt*CitA was cloned into plasmid pIVEX2.4IN [32] to express the PAS domain N-terminally fused to His-tagged immunity protein 7 of *Escherichia coli* (Im7; see Fig. S2). The Im7 tag allows stable and oriented immobilization on surfaces coated with the DNase domain of colicin E7 (DNaseE7) [32, 33]. The protein termed *Gt*CitAP was purified to apparent homogeneity by Ni-NTA chromatography and showed a band at the expected size of ~27 kDa in SDS-PAGE (data not shown). As initial ITC experiments did not provide any evidence for citrate binding to *Gt*CitAP, we speculated that this result was caused by co-purification of the protein with citrate. To test our hypothesis we released putative ligands by boiling of purified *Gt*CitAP and determined the citrate concentration in the supernatant by an enzymatic assay, which showed that the protein was indeed loaded with an equimolar amount of citrate. To prepare a citrate-free *Gt*CitAP, we dialyzed the purified protein against 6 M urea followed by extensive dialysis against PBS for

refolding. Approximately 20 % of *Gt*CitAP were correctly refolded and showed high affinity citrate-binding as judged by the step-shaped binding isotherm of subsequent ITC experiments (Table 1, Fig. S3). Although this high affinity prevents an accurate determination, the method's lower detection limit gives a value of ≤ 10 nM for the K_D [34]. Taking the determined enthalpy change of -67 kJ·mol^{-1} into account, this corresponds to a change in Gibbs enthalpy of ≤ -45 kJ·mol^{-1} and an entropy contribution of ≤ 22 kJ·mol^{-1} at 298 K. These thermodynamics differ remarkably from the CitA domains of the mesophilic organisms *K. pneumoniae* and *E. coli*, which exhibit slightly more favorable enthalpy changes of -76 and -86 kJ·mol^{-1}, respectively, but show higher entropy costs of 46 and 49 kJ·mol^{-1} at 298 K, respectively [22, 23]. This results in lower affinities with a K_D of 5.5 µM for *K. pneumoniae* CitA and 470 nM for *E. coli* CitA.

Given the fact that the residues in the first coordination shell of the citrate ligand in modeled *Gt*CitAP match almost exactly the crystallized domain of *K. pneumoniae* (Fig. 1A, B), we propose that the extraordinary affinity of *Gt*CitAP for citrate is not a special feature of its binding site configuration, but rather of the thermostable scaffold as a whole. While the domain of *K. pneumoniae* exhibits a high level of flexibility in the citrate-free state, which is lost upon the citrate-induced conformational change [26], proteins of thermophilic organisms tend to be rigid at room temperature [35, 36] and thus cannot lose much flexibility. Combined with our ITC data, these considerations imply that the loss of conformational entropy caused by ligand binding is lower for *Gt*CitAP than for its mesophilic homologs, and consequently *Gt*CitAP is better suited for the design of high-affinity LBD-based biosensors.

Ligand binding of GtCitAP can be coupled to an optical output. Most LBD-based biosensors rely on optical signal outputs, which are usually based on different fluorescent properties of the free and the ligand-bound protein. As the concentration dependence of ligand binding is described by a 1:1 Langmuir binding isotherm, 90% of the sensor's total signal

change occurs over a 100-fold concentration range from 0.1 to 10 x K_D. We tested whether ligand binding to *Gt*CitAP can be monitored by attachment of environmentally sensitive fluorophors. Based on the expression plasmid for wild-type *Gt*CitAP, we introduced the mutation D100C (numbering according to full-length *Gt*CitA) for site-specific attachment of thiol-reactive dyes close to the binding site. Moreover, we replaced R133 by alanine, a substitution known to decrease the affinity of CitA [37], to allow reversible binding and to adjust the affinity to physiological citrate levels. The purified protein variant *Gt*CitAP-D100C-R133A was labeled with DACM or Dansyl and incubated with different concentrations of citrate in PBS. We observed a concentration-dependent fluorescence decrease of ~28 % for the DACM-labeled protein and of ~21 % for the Dansyl-labeled protein, which could be fitted to a 1:1 binding isotherm yielding an apparent K_D of 406 ± 36 µM for the DACM-labeled protein and 520 ± 120 µM for the Dansyl-labeled protein (Table 2, Fig. S4). Even though the total change in fluorescence is relatively small, the observed K_D represents a suitable working range for analysis of biological samples. For instance, the entries in the human metabolome database [38] for the physiological concentrations of citrate in human blood vary between 30 and 400 µM and thereby closely match the working range of the DACM-labeled *Gt*CitAP-D100C-R133A protein.

To check if the observed change in fluorescence was indeed related to ligand binding, we incubated the DACM-labeled protein with various putative ligands at a high concentration of 64 mM to detect even very weak interactions (Fig. S5). It should be noted that the affinity of ligands causing saturation of the protein can still vary drastically, as shown by the comparison of the affinities for citrate and isocitrate (see below). The result clearly demonstrated a close relation between the structure of the tested compounds and the fluorescence of the protein suggesting that a decrease of fluorescence is coupled to an interaction with the respective compound: While the tricarboxylate DL-isocitrate induced a change which is comparable to citrate, most dicarboxylates caused a decrease of 10 - 20 % compared to the control with

buffer. Interestingly, compounds harboring a positively charged amino group (L- and D-aspartate, L-glutamate) did not affect the fluorescence at all. This observation can be explained by the excess of positive charges in the binding pocket, which allow more or less specific interactions with negatively charged ligands, but prevent compounds with positively charged groups from entering.

For further characterization of the DACM-labeled *Gt*CitAP-D100C-R133A protein, its affinity for DL-isocitrate was determined using the same approach as described above for citrate (Table 2). Compared to citrate, binding of DL-isocitrate was 20 times weaker (K_D = 11.0 ± 6.5 mM). This indicates that the labeled *Gt*CitAP variant is still highly specific for citrate with a physiologically suitable working range, although applications in complex samples may require an improvement of the total change in fluorescence between the citrate-free and the citrate-bound domain.

The specificity of GtCitAP can be altered by binding pocket grafting. The application range of a biosensor can potentially be expanded by changing the ligand specificity of its biological recognition element. While it is known that ligand recognition by sensory CitA domains involves multiple residues [25, 26, 39], the complex inter-residue interactions in this conformational change make it difficult to predict the effect of mutations outside of the binding pocket. In contrast, recent studies achieved considerable progress in the construction of artificial LBDs by designing an artificial coordination sphere for the intended ligand, which was subsequently transferred to a fitting protein scaffold [19, 20]. To test if the scaffold of *Gt*CitA is suitable for the integration of foreign binding sites, we tried to transfer the critical amino acid residues from the binding pocket of a homologous dicarboxylate-binding protein into *Gt*CitA, a process referred to as "binding pocket grafting" [40]. The crystallized sensory domain of the C4-dicarboxylate-specific kinase DcuS of *E. coli* (*Ec*DcuSP) in complex with L-malate [41] shares a high degree of similarity with the crystal structure of CitA (Fig. 1A, C),

and residues determining the specificity of both proteins have also been reported [42]. Based on this study and the crystal structures, we changed the five amino acid residues interacting with carbon atoms C4 and C5 of the citrate molecule in CitA to the corresponding amino acids of *Ec*DcuSP (Fig. 1B-D). This resulted in the protein variant *Gt*CitAP-G87T-M106F-N111I-K135F-S150A (hereafter *Gt*MalAP).

The *Gt*MalAP protein was produced in *E. coli* and purified in similar amounts as *Gt*CitAP. ITC experiments revealed an enthalpically driven interaction between *Gt*MalAP and L-malate (ΔH = -35.1 ± 5.3 kJ·mol^{-1}) with a K_D of 83 ± 20 µM (Table 1). Compared to the interaction between citrate and *Gt*CitAP, the lower affinity of *Gt*MalAP for L-malate is caused by a less favorable enthalpic term which probably originates from the loss of the electrostatic interactions between the carboxyl group at carbon atom C5 and the side chains of K135 and S150. In contrast, the entropic contribution nearly matches the estimate for *Gt*CitAP and citrate. The resulting micromolar affinity differs remarkably from the apparent millimolar K_D values observed for the interactions between the template protein *Ec*DcuSP and several C4-dicarboxylates [43]. The same study reported indications that ligand binding by *Ec*DcuSP is connected to rigidification of the protein backbone and concomitant entropy costs. In analogy to the increased affinity of wild-type *Gt*CitAP compared to the domains of mesophilic organisms, this argues in favor of the assumption that the affinity of a PAS domain-binding site can be increased by transferring its binding motif to the less flexible scaffold of *Gt*CitAP.

To observe binding of L-malate by optical detection methods we also introduced the D100C substitution to *Gt*MalAP and labeled the protein with DACM and Dansyl. Addition of L-malate caused a fluorescence decrease of 68 % for DACM-labeled *Gt*MalAP and of 71 % for Dansyl-labeled *Gt*MalAP (Fig. 2). The respective K_D values of 1.74 ± 0.17 mM (DACM-labeled variant) and 0.943 ± 0.053 mM (Dansyl-labeled variant) for *Gt*MalAP-D100C are 11- to 21-fold higher than the one determined for the unmodified protein. The most likely

explanation for this effect is an interaction of both labels with the ligand binding site, which necessitates displacement of the dye prior to ligand binding. This hypothesis is supported by the ligand-induced decrease of fluorescence and a red-shifted emission in the ligand-bound state. The latter effect was especially prominent for the DACM-labeled protein, which showed two emission maxima at 452 and 470 nm in the unbound state out of which the maximum at 452 nm disappeared almost completely upon ligand addition. Both phenomena are typical for the displacement of solvatochromic dyes from a protein environment into a polar solvent [44]. Noteworthy, the fluorescence decrease triggered by binding of L-malate to *Gt*MalA was much stronger than the one observed for citrate binding to *Gt*CitAP-D100C-R133A and outperformed the signal change of LBDs that were used as glucose sensors in clinical trials [15, 45], opening an avenue for application of *Gt*CitA derivatives in complex samples.

As the template protein *Ec*DcuSP shows a broad ligand specificity, we tested binding of a set of di- and tricarboxylates to DACM-labeled *Gt*MalAP-D100C and determined the affinity for compounds causing a fluorescence decrease of at least 25 % (Fig. S6, Table 2). Besides L-malate, we detected considerable binding of citrate, maleate, 2-methylmaleate, phthalate, ethylmalonate, and 2-oxoglutarate, but still the affinity for L-malate exceeded the affinities for other ligands by a factor of 10 or more. The preference for L-malate and the selectivity between *cis*- and *trans*-dicarboxylates is a distinct difference to the template protein *Ec*DcuSP, which binds saturated, *cis*- and *trans*-C4-dicarboxylates with similar affinities [43]. Interestingly, the crystal structure of *Ec*DcuSP[41] shows the bound L-malate molecule in a conformation similar to *cis*-configured C4-dicarboxylates and thereby enables an interaction of the two carboxyl groups with R107 (R93 in *Gt*MalAP) and H110 (H96 in *Gt*MalAP) of the major loop (Fig. 1C, D). Although there is no *Ec*DcuSP structure with bound fumarate available, the rigid elongated structure of the *trans*-configuration requires a different arrangement of the three positively charged residues in the binding pocket (R107, H110 and

R147) for an efficient interaction with both carboxyl functions. Based on these considerations, the binding of *trans*-dicarboxylates by *Ec*DcuSP suggests that its scaffold provides sufficient flexibility for larger rearrangements in the binding pocket, whereas the scaffold of *Gt*CitAP/*Gt*MalAP puts harder constraints on side chain movements and by that restricts molecular recognition to ligands that resemble the conformation in which the preferred ligand L-malate is bound. This strong influence of the outer coordination shells is in line with the recent observation that other residues than those located in the binding site show strong shifts of their NMR signals upon citrate binding in the highly similar CitA domain of *G. thermodinitrificans* [39].

Further evidence for this interpretation is provided by the enantioselectivity of *Gt*MalAP, which did not bind C4-hydroxydicarboxylic acids with R-configured stereocenters such as D-malate or L-tartrate in our experiments. The latter compound binds to *Ec*DcuSP with a low-milimolar affinity that is similar to other C4-dicarboxylates [43]. However, if L-tartrate was bound in the same conformation as L-malate, one would expect a steric conflict between the hydroxyl group at the R-configured stereocenter and the phenyl ring of F120 (F106 in *Gt*MalAP, Fig. 1C, D). This again suggests that *Ec*DcuSP allows conformational rearrangements which are energetically unfavorable in the scaffold of *Gt*CitAP/*Gt*MalAP.

Rationally designed binding motifs enable novel ligand specificities of GtCitAP. Despite identical molecular shapes *Gt*MalAP interacts with L-malate but does not bind L-aspartate. We hypothesized that this property is caused by an electrostatic repulsion between the positively charged amino group of L-aspartate and R133. To specify the influence of R133 for the selectivity *Gt*MalAP in more depth, we inverted its positive charge by substitution with L-glutamate, expecting an interaction with L-aspartate by the introduction of the R133E exchange. Furthermore, D110 was replaced by L-asparagine to avoid repulsion between D110

and the introduced E133 residue, resulting in *Gt*CitAP-G87T-M106F-D110N-N111I-R133E-K135F-S150A (hereafter *Gt*AspAP, Fig. 1F).

Moreover, the substitution D100C was introduced into *Gt*AspAP to characterize its binding specificity after labeling with DACM and Dansyl. In accordance with our prediction, addition of L-aspartate caused the highest fluorescence decrease of DACM-labeled *Gt*AspAP-D100C (~20 %), whereas most of the other compounds did not show any effect (Fig. S7). Notably, the enantioselectivity for the L-enantiomer, which had been observed for *Gt*MalAP and L-malate, was retained in *Gt*AspAP. Although these results support our model for ligand binding by *Gt*MalAP, the affinity of labeled *Gt*AspAP turned out to be in the high millimolar range (Fig. S8), and by that is probably too low to be of practical use in sensing applications. For this reason we did not characterize this variant further and focused on other designs.

The specificity of GtCitAP can be adapted to man-made ligands. As discussed before, the characterization of *Gt*MalAP strongly suggested that its binding mode for L-malate and *cis*-configured C4-dicarboxylates closely resembles the *Ec*DcuSP crystal structure [41]. Combined with the results obtained for *Gt*AspAP, this implies that substituents at carbon atom C2 interact with R133 of *Gt*MalAP (Fig. 1D). As the fluorescence of DACM-labeled *Gt*MalAP-D100C was significantly quenched by phthalate (Fig. S6), which can be regarded as a *cis*-configured dicarboxylate with a diene substituent bridging C2 and C3, we reasoned that the affinity for this aromatic compound could be increased by introducing the mutation R133M, as methionine represents a hydrophobic amino acid of comparable size as arginine. In addition, we replaced D110 with a hydrophobic L-isoleucine residue because D110 stabilizes R133 by a conserved salt bridge and the mutation R133M alone might disrupt the connection between the helix harboring D110 and the binding pocket (Fig. 1D). The homology model obtained for the D110I substitution (Fig. 1E) showed nearly the same folding of the protein backbone as the *Ec*DcuSP structure, so that this variant was chosen for

296 further characterization (*Gt*CitAP-G87T-M106F-D110I-N111I-R133M-K135F-S150A,
297 hereafter *Gt*PhtAP).

298 After introduction of the substitution D100C into *Gt*PhtAP and fluorophore-labeling,
299 DACM-labeled *Gt*PhtAP-D100C showed by far the highest fluorescence decreases for
300 phthalate (57 %) and ethylmalonate (45 %) with similar spectral characteristics as observed
301 for *Gt*MalAP-D100C and L-malate (Fig. S9 A, B). In addition, we detected a slight decrease
302 for maleate, 2-methylmaleate, and methylsuccinate, whereas neither *trans*-dicarboxylates,
303 dicarboxylates with hydrophilic substituents nor tricarboxylates did show any effect (Fig.
304 S10). The respective affinities were 4.23 ± 0.43 mM for phthalate and 10.15 ± 0.60 for
305 ethylmalonate (Table. 2, Fig. S9 C, D). The quenching of fluorescence and the affinities of the
306 Dansyl-labeled *Gt*PhtAP-D100C protein were almost identical (Fig. S9 E - H). For
307 verification of the interactions of *Gt*PhtAP with phthalate and ethylmalonate, we performed
308 ITC experiments with the unmodified protein (Table 1). The determined K_D values of 110.0 ±
309 8.1 µM for phthalate and 323 ± 47 µM for ethylmalonate are significantly lower than the
310 values obtained by the fluorescence assay with the labeled proteins, which probably originates
311 from an interaction of the dyes with the binding site, as already discussed for *Gt*MalAP-
312 D100C. In contrast to citrate and malate binding to *Gt*CitAP and *Gt*MalAP, respectively,
313 binding of phthalate and ethylmalonate to *Gt*PhtAP was driven by favorable enthalpic and
314 entropic terms, the latter reflecting a higher contribution of hydrophobic interactions, which
315 supports our initial design strategy.

316

317 **Fluorescently labeled variants of GtCitAP are active on solid supports.** In most cases
318 the biological recognition element must be attached to solid supports prior to integration into
319 a biosensing device. This step is critical as harsh conditions of covalent immobilization
320 procedures can cause inactivation of the biomolecule, whereas non-covalent immobilization
321 frequently suffers from insufficient stability. To verify that *Gt*CitAP variants are suitable for

immobilization, we used a system which is based on the tight interaction between DNaseE7 and its cognate immunity protein Im7 and combines extremely high stability with mild immobilization conditions and defined orientation [33, 46] (Fig. 3A). The enzymatic function of DNaseE7 was inactivated by the substitution H545A to facilitate expression of the otherwise toxic protein. Purified DNaseE7-H545A was covalently attached to maleic anhydride-activated microplates. Fluorescently labeled *Gt*MalAP-D100C and *Gt*PhtAP-D100C were successfully immobilized by the interaction between their N-terminal Im7 tag and the DNaseE7-coated surface as judged by the constant fluorescence signal after removal of excess protein (Fig. S11).

Addition of the respective ligands L-malate and phthalate caused a concentration-dependent fluorescence decrease which could be fitted to a 1:1 Langmuir binding isotherm (Fig. 3B-C). Except for the interaction between Dansyl-labeled *Gt*MalAP and L-malate, the K_D values of the immobilized protein variants were increased by factors between 1.6 and 3.0 compared to the affinities determined in solution. In addition, the changes of fluorescence represented only 38 - 53 % of the quenching in solution. The latter effect could be due either to inactivation of a fraction of the immobilized protein or to an interaction of the solvatochromic dyes with the plastic surface of the microplate or other protein molecules in close proximity. In these scenarios the dye would be exposed to a more hydrophobic environment than the hydrophilic solvent water, resulting in an increased or less decreased fluorescence [44]. Although this effect and the reduced affinities suggest that the optical readout of the immobilized proteins is moderately impaired, we have demonstrated that the sensory properties of immobilized *Gt*CitAP variants were preserved to a large degree.

GtCitAP variants enable selective molecular recognition in complex sample matrices. We believe that *Gt*PthAP is the first example of a biomolecule for the detection of ethylmalonate, an urinary biomarker for ethylmalonic encephalopathy with a clinically

relevant concentration range of 0.7 - 22 mM [47]. Given that phthalate and the low affinity ligands methylsuccinate and methylmaleate do not occur in urinary levels that would cause substantial binding according the human metabolome database [38], we decided to test GtPthAP-D100C for quantification of ethylmalonate in synthetic urine. For this reason, we decided to test *Gt*PthAP-D100C for quantification of ethylmalonate in synthetic urine. For calibration of the system, we determined the affinity and the maximal fluorescence quenching of Dansyl-labeled *Gt*PhtAP upon binding of ethylmalonate in a modified assay buffer system, which included a higher buffer capacity and EDTA to avoid complex formation between the analyte and divalent cations of the matrix (Fig. 4A). In comparison to PBS, the affinity was slightly increased (K_D = 10.05 ± 0.24 mM), but the change of fluorescence was almost unaltered. To determine the influence of the synthetic urine on the fluorescence in the absence of ethylmalonate, we compared the fluorescence in buffer and in buffer supplemented with 60 % (v/v) synthetic urine. Both samples did not show any significant difference, demonstrating that the matrix does not affect the fluorescence of the sensor.

For quantitative analysis, we spiked synthetic urine with defined amounts of ethylmalonate and calculated its concentration by comparing the fluorescence quenching in relation to the blank sample with the calibration curve in buffer. The results showed a good correlation between the added amounts of ethylmalonate and the calculated concentrations (Fig. 4B), confirming that *Gt*CitA variants allow selective quantification of target analytes in complex samples.

CONCLUSIONS

Employing simple strategies of structure-guided rational protein design, our study shows that the scaffold of the PAS domain *Gt*CitAP derived from histidine kinase CitA of the thermophilic *G. thermoleovorans* can be adapted to bind non-native ligands with micromolar affinity. Importantly, the characteristics of *Gt*CitAP are significantly improved in terms of

affinity and (enantio-) selectivity compared to homologs of mesophilic organisms. Our results suggest that these benefits mainly arise from a lower flexibility of the *Gt*CitAP scaffold in the ligand-free state resulting in reduced entropy costs upon ligand binding. In consequence, we propose that these properties and the mutational stability of *Gt*CitAP as revealed by the design of functional variants with up to eight mutations make this domain an ideal subject for more sophisticated protein engineering strategies [19, 20]. Having successfully shown that fluorescently labeled *Gt*CitAP variants are functional on solid supports and suitable for applications in complex samples and given the fact that homologous CitA domains were converted into ratiometric and signal-on fluorescent biosensors [27, 28], thermophilic PAS domains provide a promising target for the construction of custom-made biosensing devices.

METHODS

Recombinant DNA work. Plasmids used or constructed in this work are listed in Table S1, oligonucleotide sequences used for PCR or assembly reactions are given in Table S2. To construct an expression plasmid for *Gt*CitAP (residues 31 - 161) fused to His-tagged immunity protein 7 (Im7), the encoding gene sequence was amplified from genomic DNA of *Geobacillus thermoleovorans* DSM 5366 (DSMZ) and cloned into plasmid pIVEX2.4IN [32] via NcoI and XhoI restriction sites yielding plasmid pIVEX2.4IN-*Gt*CitA. For generation of *Gt*CitAP variants, the designed sequences were assembled from oligonucleotide mixtures (Table S3), PCR-amplified according to a previously published protocol [48], and cloned into the backbone of plasmid pIVEX2.4IN-*Gt*CitAP via Eco105I and XhoI restriction sites. For the expression construct pIVEX2.4d-E7H545A of the catalytically inactive DNaseE7-H545A domain (residues 444 - 576 of colicin E7), the corresponding DNA sequence was amplified from plasmid pQE30DNaseE7/Im7 in two different fragments to allow introduction of the H545A mutation via primer overhangs and assembled into plasmid pIVEX2.4d (5Prime) via NdeI and SmaI restriction sites [49].

Protein expression and purification. *Escherichia coli* OverExpress™ C43(DE3) cells [50] were transformed with the respective expression plasmid and grown overnight in 5 ml LB medium supplemented with 100 µg/ml ampicillin and 10 g/l glucose at 37 °C under vigorous shaking. The culture was used to inoculate 500 ml of LB (+ 100 µg/ml ampicillin, + 10 g/l glucose) in a baffled 2 l-Erlenmeyer flask and cells were grown to an optical density at 600 nm (OD_{600}) of 0.6 at 37 °C under vigorous shaking. Isopropyl-β-D-thiogalactopyranosid was added to a final concentration of 0.5 mM to induce synthesis of the target protein and cells were grown for another 3 hours (DNaseE7) or overnight (*Gt*CitAP variants) at 37 °C. Cells were harvested by centrifugation (4,000 *g*, 20 min, 4 °C), washed with 50 ml PBS (20 mM KH_2PO_4/Na_2HPO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.8), shock-frozen in liquid nitrogen, and stored at -80 °C.

For protein purification, the cell pellet obtained from 500 ml of culture was suspended in 15 ml equilibration buffer (20 mM NaH_2PO_4/Na_2HPO_4, 20 mM imidazole, 500 mM NaCl, pH 7.4) and after addition of protease inhibitor (cOmplete™ Mini EDTA-free, Roche), cells were disrupted by sonication. After removal of cell debris by centrifugation for 20 min at 20,000 g and filtering through a filter with a pore size of 0.22 µm (Millex filter unit, Millipore), the supernatant was applied to a syringe-operated and pre-equilibrated 1 ml HisTrap FF column (GE Healthcare). After washing of the column with 15 ml of equilibration buffer, the protein was eluted with 10 ml of elution buffer (20 mM NaH_2PO_4/Na_2HPO_4, 500 mM imidazole, 500 mM NaCl, pH 7.4). Fractions containing the recombinant protein were pooled, dialyzed twice against 500 ml of PBS, shock-frozen in liquid nitrogen, and stored at -80 °C. Protein aliquots were thawed once and discarded after one day. Protein purity was verified by SDS polyacrylamide gel electrophoresis using a denaturing Tris/glycine system (Mini-PROTEAN® TGX™, Bio-Rad Laboratories) according to manufacturer's instructions. For cysteine-containing variants, 2 mM tris(2-carboxyethyl)phosphine (TCEP) was added to each buffer.

Ligand removal and refolding. For quantification of citrate bound to purified *Gt*CitAP, the protein was boiled for 8 min and after centrifugation the supernatant was analyzed with an enzymatic assay (Citrate Assay Kit, Merck) according to the manufacturer's instructions. For denaturation with subsequent refolding, purified *Gt*CitAP was dialyzed overnight against 500 ml 6 M urea. Afterwards the protein was dialyzed three times against 1 l of PBS for refolding. Precipitates formed in the refolding step were removed by centrifugation at 16,000 *g*.

Protein modification with fluorescent dyes. Purified *Gt*CitAP variants containing the mutation D100C were incubated with a 20-fold molar excess of *N*-(7-dimethylamino-4-methylcoumarin-3-yl)maleimide (DACM, Eurogentec) or *N*-[2(dansylamino)ethyl]maleimide (Dansyl, Merck) in PBS supplemented with 2 mM TCEP. The reaction was allowed to proceed overnight at 4 - 8 °C. Excess amounts of dye were removed by size exclusion chromatography of the sample using a PD 10 column (GE Healthcare) equilibrated with PBS.

Isothermal titration calorimetry. All ITC experiments were performed with a MicroCal PEAQ-ITC (Malvern Panalytical GmbH) at 298 K. The respective *Gt*CitAP variant was loaded into the cell at concentrations of 54 µM (*Gt*CitAP), 150 µM (*Gt*MalAP), or 260 µM (*Gt*PhtAP) in PBS. Ligand solutions were prepared in dialysis buffer (PBS) at concentrations of 540 or 125 µM (trisodium citrate), 1.5 mM (disodium L-malate), and 5 mM (disodium phthalate, disodium ethylmalonate). The ligand solution was injected in one 0.4 µl injection followed by 12 injections of 3.0 µl. The spacing time between the injections was 150 s. A control experiment in which the ligand solution was titrated into buffer was performed for each protein variant/ligand combination. The raw data was analyzed with MicroCal PEAQ-ITC Analysis Software (Malvern Panalytical GmbH).

Analysis of ligand binding by fluorescence. All fluorescence measurements were performed in a TECAN Infinite M1000 PRO microplate reader (Tecan). To probe for an interaction, fluorescently labeled *Gt*CitAP-D100C variants at a concentration of 1 µM were incubated with the sodium carboxylates L-malate, D-malate, L-aspartate, D-aspartate, L-tartrate, succinate, 2-methylsuccinate, maleate, 2-methylmaleate, fumarate, 2-methylfumarate, itaconate, L-glutamate, 2-oxoglutarate, phthalate, quinolinate, citrate, DL-isocitrate, L-lactate, or ethylmalonate at concentrations of 64 mM in PBS using UV-Star® microplates (Greiner Bio-One GmbH). The excitation and emission wavelengths were 380/440 nm (bandwith: 10/5 nm) for DACM-labeled proteins and 350/510 nm (bandwith: 20/5 nm) for Dansyl-labeled proteins. For affinity determination, a 1:1 dilution series of the respective compound was prepared in PBS and the impact on the fluorescence was determined accordingly. The K_D was fitted according to the following formula of a 1:1 Langmuir binding isotherm using Origin Pro 9.1 (Origin Lab):

$$F(c) = F_0 + \frac{\Delta F_{max} * c}{K_D + c}$$

where F(c) is the measured fluorescence, c is the ligand concentration, F_0 is the fluorescence in the absence of the ligand and ΔF_{max} is the fluorescence difference between the ligand-saturated and the ligand-free protein. The values for the parameters ΔF_{max} and F_0 were also determined by the fitting algorithm.

Immobilization of GtCitAP variants on DNaseE7-coated microplates. The wells of "Pierce™ Maleic Anhydride Activated Plates" (Thermo Fisher Scientific) were washed with PBS and incubated overnight with the purified catalytically inactive DNaseE7-H545A domain (residues 444 - 576 of colicin E7) at a concentration of 50 µg/ml at 4-8 °C. After washing twice with PBS remaining maleic anhydride functions were deactivated by incubation with

3% (w/v) bovine serum albumin in PBS for 1 h. The wells were washed four times with PBS, incubated with 1 µM of a purified and fluorescently labeled *Gt*CitAP-D100C variant and washed again four times with PBS. Afterwards the wells were washed ten times with PBS and the fluorescence after each washing step was recorded. Finally, a 1:1 dilution series of ligand solution in PBS was added in order of increasing concentrations, and the fluorescence for each concentration was recorded. The fluorescence measurements were performed as described in the previous paragraph.

Quantification of ethylmalonate in synthetic urine. For calibration of the system, a 1:1 dilution series of 64 mM ethylmalonate in assay buffer (100 mM MOPS/NaOH, 100 mM NaCl, 15 mM EDTA, pH 7.0) was prepared and incubated with 1 µM of Dansyl-labeled *Gt*PhtAP-D100C for fluorescence measurements. Synthetic urine (Surine^{TM}, Merck) was spiked with ethylmalonate in concentrations from 0.7 to 24.3 mM and diluted to obtain a solution of 60% (v/v) sample in assay buffer. After addition of 1 µM Dansyl-labeled *Gt*PhtAP-D100C, the fluorescence was determined. All fluorescence measurements were conducted as described in the previous paragraphs. The concentration of ethylmalonate was calculated by rearrangement of the equation for K_D determination using the values for K_D, ΔF_{max} and F_0 obtained from the calibration.

Homology modelling and ligand docking. Homology models were prepared with SWISS-MODEL [51] using the crystal structures of the sensory domains of *K. pneumoniae* CitA (pdb code: 2J80) [26] or *E. coli* DcuS (pdb code: 3BY8) [41] as template. Ligand docking was performed with the ROSETTA server [52]. Homology models of *Gt*CitAP, *Gt*MalAP, *Gt*AspAP, and *Gt*PhtAP are included as pdb files in the SI.

SUPPORTING INFORMATION

The supporting information includes Figures S1 – S11, Tables S1 – S3, and supplementary references.

Fig. S1. Sequence of the full-length histidine kinase CitA of *Geobacillus thermoleovorans* DSM 5366

Fig. S2. Sequence of His6-Im7-tagged *Gt*CitAP protein as encoded by plasmid pIVEX2.4IN-*Gt*CitAP

Fig. S3. ITC experiments with refolded *Gt*CitAP and citrate

Fig. S4. Citrate-dependent fluorescence decrease of *Gt*CitAP-D100C-R133A

Fig. S5. Analysis of the ligand selectivity of DACM-labeled *Gt*CitAP-D100C-R133A

Fig. S6. Analysis of the ligand selectivity of DACM-labeled *Gt*MalAP-D100C

Fig. S7. Analysis of the ligand selectivity of DACM-labeled *Gt*AspAP-D100C

Fig. S8. L-aspartate-dependent fluorescence decrease of *Gt*AspAP-D100C

Fig. S9. Phthalate- and ethylmalonate-dependent fluorescence decrease of *Gt*PhtAP-D100C

Fig. S10. Analysis of the ligand selectivity of DACM-labeled *Gt*PhtAP-D100C

Fig. S11. Stability of the immobilization of *Gt*CitAP variants by the interaction between DNaseE7 and Im7 fused N-terminally to the *Gt*CitAP variants

Table S1. Plasmids used or constructed in this study

Table S2. Oligonucleotides used in this study

Table S3. Assembly of expression plasmids for *Gt*CitAP variants based on pIVEX2.4IN-*Gt*CitAP

AUTHOR INFORMATION

Author contributions. KUC performed the experiments, analyzed the data and participated in the design of the research. MBo and MBa analyzed the data and designed the research. All authors participated in writing of the manuscript.

ACKNOWLEDGEMENTS

This work was funded by the German Federal Ministry of Education and Research (BMBF), funding code 031A095B, as part of the project "Molecular Interaction Engineering: From Nature's Toolbox to Hybrid Technical Systems (MIE)". We thank Dr. Marc M. Nowaczyk (Plant Biochemistry, Ruhr-University Bochum, Germany) for providing the plasmids pQE30DNaseE7/Im7 and pIVEX2.4IN.

REFERENCES

[1] Weber, J., Peng, H., and Rader, C. (2017) From rabbit antibody repertoires to rabbit monoclonal antibodies, *Exp. Mol. Med. 49*, e305.

[2] Charlton, K., Harris, W. J., and Porter, A. J. (2001) The isolation of super-sensitive anti-hapten antibodies from combinatorial antibody libraries derived from sheep, *Biosens. Bioelectron. 16*, 639-646.

[3] Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science 249*, 505-510.

[4] Ho, H. A., and Leclerc, M. (2004) Optical sensors based on hybrid aptamer/conjugated polymer complexes, *J. Am. Chem. Soc. 126*, 1384-1387.

[5] Xiao, Y., Piorek, B. D., Plaxco, K. W., and Heeger, A. J. (2005) A reagentless signal-on architecture for electronic, aptamer-based sensors via target-induced strand displacement, *J. Am. Chem. Soc. 127*, 17990-17991.

[6] Carothers, J. M., Goler, J. A., Kapoor, Y., Lara, L., and Keasling, J. D. (2010) Selecting RNA aptamers for synthetic biology: investigating magnesium dependence and predicting binding affinity, *Nucleic Acids Res. 38*, 2736-2747.

[7] Vallee-Belisle, A., and Plaxco, K. W. (2010) Structure-switching biosensors: inspired by nature, *Curr. Opin. Struct. Biol. 20*, 518-526.

[8] Turner, K., Joel, S., Feliciano, J., Feltus, A., Pasini, P., Wynn, D., Dau, P., Dikici, E., Deo, S. K., and Daunert, S. (2017) Transcriptional regulatory proteins as biosensing tools, *Chem. Commun. 53*, 6820-6823.

[9] Dwyer, M. A., and Hellinga, H. W. (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering, *Curr. Opin. Struct. Biol. 14*, 495-504.

[10] Benson, D. E., Conrad, D. W., de Lorimier, R. M., Trammell, S. A., and Hellinga, H. W. (2001) Design of bioelectronic interfaces by exploiting hinge-bending motions in proteins, *Science 293*, 1641-1644.

[11] Fehr, M., Frommer, W. B., and Lalonde, S. (2002) Visualization of maltose uptake in living yeast cells by fluorescent nanosensors, *Proc. Natl. Acad. Sci. USA 99*, 9846-9851.

[12] 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., and Hellinga, H. W. (2002) Construction of a fluorescent biosensor family, *Protein Sci. 11*, 2655-2675.

[13] Donaldson, T., Iozzino, L., Deacon, L. J., Billones, H., Ausili, A., D'Auria, S., and Dattelbaum, J. D. (2017) Engineering a switch-based biosensor for arginine using a *Thermotoga maritima* periplasmic binding protein, *Anal. Biochem. 525*, 60-66.
[14] Veetil, J. V., Jin, S., and Ye, K. (2010) A glucose sensor protein for continuous glucose monitoring, *Biosens. Bioelectron. 26*, 1650-1655.
[15] Judge, K., Morrow, L., Lastovich, A. G., Kurisko, D., Keith, S. C., Hartsell, J., Roberts, B., McVey, E., Weidemaier, K., Win, K., and Hompesch, M. (2011) 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. 13*, 309-317.
[16] Allert, M., Rizk, S. S., Looger, L. L., and Hellinga, H. W. (2004) Computational design of receptors for an organophosphate surrogate of the nerve agent soman, *Proc. Natl. Acad. Sci. USA 101*, 7907-7912.
[17] Looger, L. L., Dwyer, M. A., Smith, J. J., and Hellinga, H. W. (2003) Computational design of receptor and sensor proteins with novel functions, *Nature 423*, 185-190.
[18] Schreier, B., Stumpp, C., Wiesner, S., and Höcker, B. (2009) Computational design of ligand binding is not a solved problem, *Proc. Natl. Acad. Sci. USA 106*, 18491-18496.
[19] Tinberg, C. E., Khare, S. D., Dou, J., Doyle, L., Nelson, J. W., Schena, A., Jankowski, W., Kalodimos, C. G., Johnsson, K., Stoddard, B. L., and Baker, D. (2013) Computational design of ligand-binding proteins with high affinity and selectivity, *Nature 501*, 212-216.
[20] Bick, M. J., Greisen, P. J., Morey, K. J., Antunes, M. S., La, D., Sankaran, B., Reymond, L., Johnsson, K., Medford, J. I., and Baker, D. (2017) Computational design of environmental sensors for the potent opioid fentanyl, *eLife 6*, e28909.
[21] Henry, J. T., and Crosson, S. (2011) Ligand-binding PAS domains in a genomic, cellular, and structural context, *Annu. Rev. Microbiol. 65*, 261-286.
[22] Kaspar, S., and Bott, M. (2002) The sensor kinase CitA (DpiB) of *Escherichia coli* functions as a high-affinity citrate receptor, *Arch. Microbiol. 177*, 313-321.
[23] Kaspar, S., Perozzo, R., Reinelt, S., Meyer, M., Pfister, K., Scapozza, L., and Bott, M. (1999) The periplasmic domain of the histidine autokinase CitA functions as a highly specific citrate receptor, *Mol. Microbiol. 33*, 858-872.
[24] Gerharz, T., Reinelt, S., Kaspar, S., Scapozza, L., and Bott, M. (2003) Identification of basic amino acid residues important for citrate binding by the periplasmic receptor domain of the sensor kinase CitA., *Biochemistry 42*, 5917-5924.
[25] Reinelt, S., Hofmann, E., Gerharz, T., Bott, M., and Madden, D. R. (2003) The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain, *J. Biol. Chem. 278*, 39189-39196.
[26] Sevvana, M., Vijayan, V., Zweckstetter, M., Reinelt, S., Madden, D. R., Herbst-Irmer, R., Sheldrick, G. M., Bott, M., Griesinger, C., and Becker, S. (2008) A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA, *J. Mol. Biol. 377*, 512-523.
[27] Ewald, J. C., Reich, S., Baumann, S., Frommer, W. B., and Zamboni, N. (2011) Engineering genetically encoded nanosensors for real-time in vivo measurements of citrate concentrations, *PloS one 6*, e28245.
[28] Honda, Y., and Kirimura, K. (2013) Generation of circularly permuted fluorescent-protein-based indicators for *in vitro* and *in vivo* detection of citrate, *PloS one 8*, e64597.
[29] Razvi, A., and Scholtz, J. M. (2006) Lessons in stability from thermophilic proteins, *Protein Sci. 15*, 1569-1578.

[30] Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389-3402.

[31] Graf, S., Broll, C., Wissig, J., Strecker, A., Parowatkin, M., and Unden, G. (2016) CitA (citrate) and DcuS (C_4-dicarboxylate) sensor kinases in thermophilic *Geobacillus kaustophilus* and *Geobacillus thermodenitrificans*, *Microbiology 162*, 127-137.

[32] Cormann, K. U., Bartsch, M., Rögner, M., and Nowaczyk, M. M. (2014) Localization of the CyanoP binding site on photosystem II by surface plasmon resonance spectroscopy, *Front. Plant Sci. 5*, 595.

[33] Hosse, R. J., Tay, L., Hattarki, M. K., Pontes-Braz, L., Pearce, L. A., Nuttall, S. D., and Dolezal, O. (2009) Kinetic screening of antibody-Im7 conjugates by capture on a colicin E7 DNase domain using optical biosensors, *Anal. Biochem. 385*, 346-357.

[34] Perozzo, R., Folkers, G., and Scapozza, L. (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects, *J. Recept. Signal Transduct. Res. 24*, 1-52.

[35] Radestock, S., and Gohlke, H. (2011) Protein rigidity and thermophilic adaptation, *Proteins 79*, 1089-1108.

[36] Wolf-Watz, M., Thai, V., Henzler-Wildman, K., Hadjipavlou, G., Eisenmesser, E. Z., and Kern, D. (2004) Linkage between dynamics and catalysis in a thermophilic-mesophilic enzyme pair, *Nat. Struct. Mol. Biol. 11*, 945-949.

[37] Gerharz, T., Reinelt, S., Kaspar, S., Scapozza, L., and Bott, M. (2003) Identification of basic amino acid residues important for citrate binding by the periplasmic receptor domain of the sensor kinase CitA, *Biochemistry 42*, 5917-5924.

[38] Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vazquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., Serra-Cayuela, A., Liu, Y., Mandal, R., Neveu, V., Pon, A., Knox, C., Wilson, M., Manach, C., and Scalbert, A. (2018) HMDB 4.0: the human metabolome database for 2018, *Nucleic Acids Res. 46*, D608-D617.

[39] Salvi, M., Schomburg, B., Giller, K., Graf, S., Unden, G., Becker, S., Lange, A., and Griesinger, C. (2017) Sensory domain contraction in histidine kinase CitA triggers transmembrane signaling in the membrane-bound sensor, *Proc. Natl. Acad. Sci. USA 114*, 3115-3120.

[40] Scheib, U., Shanmugaratnam, S., Farias-Rico, J. A., and Höcker, B. (2014) Change in protein-ligand specificity through binding pocket grafting, *J. Struct. Biol. 185*, 186-192.

[41] Cheung, J., and Hendrickson, W. A. (2008) Crystal structures of C_4-dicarboxylate ligand complexes with sensor domains of histidine kinases DcuS and DctB, *J. Biol. Chem. 283*, 30256-30265.

[42] Krämer, J., Fischer, J. D., Zientz, E., Vijayan, V., Griesinger, C., Lupas, A., and Unden, G. (2007) Citrate sensing by the C_4-dicarboxylate/citrate sensor kinase DcuS of *Escherichia coli*: binding site and conversion of DcuS to a C_4-dicarboxylate- or citrate-specific sensor, *J. Bacteriol. 189*, 4290-4298.

[43] Kneuper, H., Janausch, I. G., Vijayan, V., Zweckstetter, M., Bock, V., Griesinger, C., and Unden, G. (2005) The nature of the stimulus and of the fumarate binding site of the fumarate sensor DcuS of *Escherichia coli*, *J. Biol. Chem. 280*, 20596-20603.

[44] Loving, G. S., Sainlos, M., and Imperiali, B. (2010) Monitoring protein interactions and dynamics with solvatochromic fluorophores, *Trends Biotechnol. 28*, 73-83.

[45] Thomas, K. J., Sherman, D. B., Amiss, T. J., Andaluz, S. A., and Pitner, J. B. (2006) A long-wavelength fluorescent glucose biosensor based on bioconjugates of

galactose/glucose binding protein and Nile Red derivatives, *Diabetes Technol. Ther. 8*, 261-268.
[46] Ko, T. P., Liao, C. C., Ku, W. Y., Chak, K. F., and Yuan, H. S. (1999) The crystal structure of the DNase domain of colicin E7 in complex with its inhibitor Im7 protein, *Structure 7*, 91-102.
[47] Zafeiriou, D. I., Augoustides-Savvopoulou, P., Haas, D., Smet, J., Triantafyllou, P., Vargiami, E., Tamiolaki, M., Gombakis, N., van Coster, R., Sewell, A. C., Vianey-Saban, C., and Gregersen, N. (2007) Ethylmalonic encephalopathy: clinical and biochemical observations, *Neuropediatrics 38*, 78-82.
[48] Acevedo-Rocha, C. G., and Reetz, M. T. (2014) Assembly of designed oligonucleotides: a useful tool in synthetic biology for creating high-quality combinatorial DNA libraries, *Methods Mol. Biol. 1179*, 189-206.
[49] Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., 3rd, and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases, *Nat. Methods 6*, 343-345.
[50] Miroux, B., and Walker, J. E. (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, *Journal of molecular biology 260*, 289-298.
[51] Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Res. 42*, W252-258.
[52] Combs, S. A., Deluca, S. L., Deluca, S. H., Lemmon, G. H., Nannemann, D. P., Nguyen, E. D., Willis, J. R., Sheehan, J. H., and Meiler, J. (2013) Small-molecule ligand docking into comparative models with Rosetta, *Nat. Protoc. 8*, 1277-1298.

693 **Table 1.** Thermodynamic data of the interaction between *Gt*CitAP variants and their
694 respective ligands obtained by ITC at 278 K using PBS buffer pH 7.8.

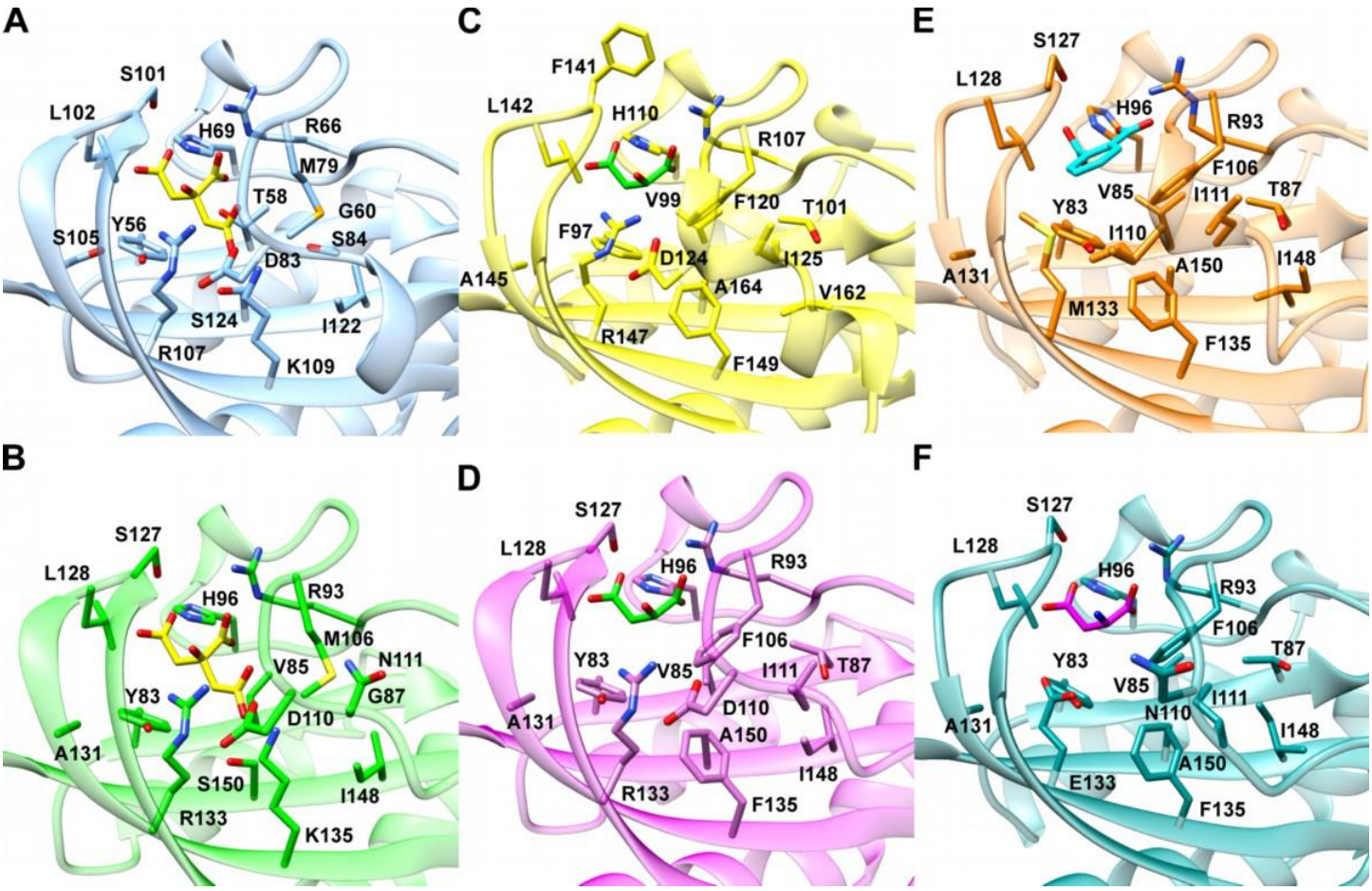
Variant	**Ligand**	**K_D (µM)[a]**	**Stochiometry**	**ΔH (kJ·mol^{-1})[a]**	**-T·ΔS (kJ·mol^{-1})[a]**	**ΔG (kJ·mol^{-1})[a]**	**n[b]**
*Gt*CitAP	Citrate	≤ 0.01	0.203 ± 0.032	-67.4 ± 2.2	≤ 22	≤ -45.	3
*Gt*MalAP	L-Malate	83 ± 20	0.682 ± 0.089	-35.1 ± 5.3	11.8 ± 5.6	-23.40 ± 0.59	6
*Gt*PhtAP	Phthalate	110.0 ± 8.1	1 (fixed)	-14.40 ± 0.17	-8.27 ± 0.35	-22.60 ± 0.21	3
*Gt*PhtAP	Ethylmalonate	323 ± 47	1 (fixed)	-11.1 ± 1.1	-8.9 ± 1.4	-20.00 ± 0.36	4

695 [a]Errors are given as standard deviations. [b]n indicates the number of ITC experiments performed.
696

Table 2. Apparent K_D values (in mM) for DACM-labeled *Gt*CitAP-D100C variants and various ligands determined by fluorescence.

Ligand	***Gt*CitAP-D100C-R133A**[a]	***Gt*MalAP-D100C**[a]	***Gt*PhtAP-D100C**[a]
Citrate	0.406 ± 0.036	22.1 ± 5.4	no binding
DL-Isocitrate	11.0 ± 6.5	weak binding	no binding
L-Malate	weak binding	1.74 ± 0.17	no binding
Maleate	weak binding	26.4 ± 5.4	weak binding
2-Methylmaleate	weak binding	20.0 ± 1.8	weak binding
2-Oxoglutarate	weak binding	39 ± 10	no binding
Phthalate	weak binding	17.6 ± 2.4	4.23 ± 0.43
Ethylmalonate	weak binding	17.9 ± 2.3	10.15 ± 0.60

[a]Errors are given as standard errors obtained from a fit of the fluorescence decrease at 440 nm to a 1:1 Langmuir binding isotherm.



703

704 **Figure 1.** Structures and models of *Gt*CitAP variants and homologs. (A) Crystal structure of
705 the sensory CitA domain of *K. pneumoniae* [26] in ribbon presentation. Residues located in the
706 binding site of the citrate ligand (yellow) are displayed as sticks. (B) Homology model of
707 *Gt*CitAP based on the crystal structure shown in (A). The position of the citrate ligand
708 (yellow) was taken from the template structure. (C) Crystal structure of *Ec*DcuSP [41]. The
709 ligand L-malate is shown in green. (D) Homology model of *Gt*MalAP based on the *Ec*DcuSP
710 structure shown in (C). The position of the ligand L-malate (green) was taken from the
711 template structure. (E) Homology model of *Gt*PhtAP based on the *Ec*DcuSP structure shown
712 in (C). The position of the ligand phthalate (cyan) was obtained by molecular docking with
713 the ROSETTA server [52] and selected due to similar positioning of the carboxyl functions of
714 phthalate compared to L-malate in the template structure. (F) Homology model of *Gt*AspAP

based on the *Ec*DcuSP structure shown in (C). The ligand L-aspartate is shown in the same position as the isosteric ligand L-malate in the template structure.

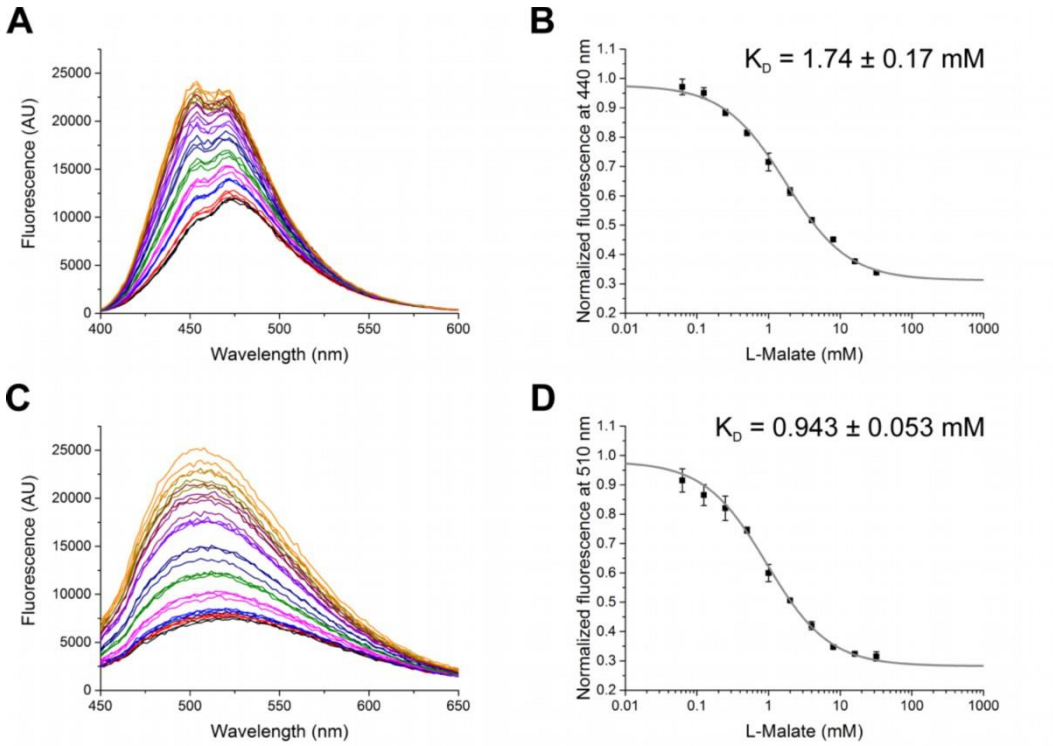
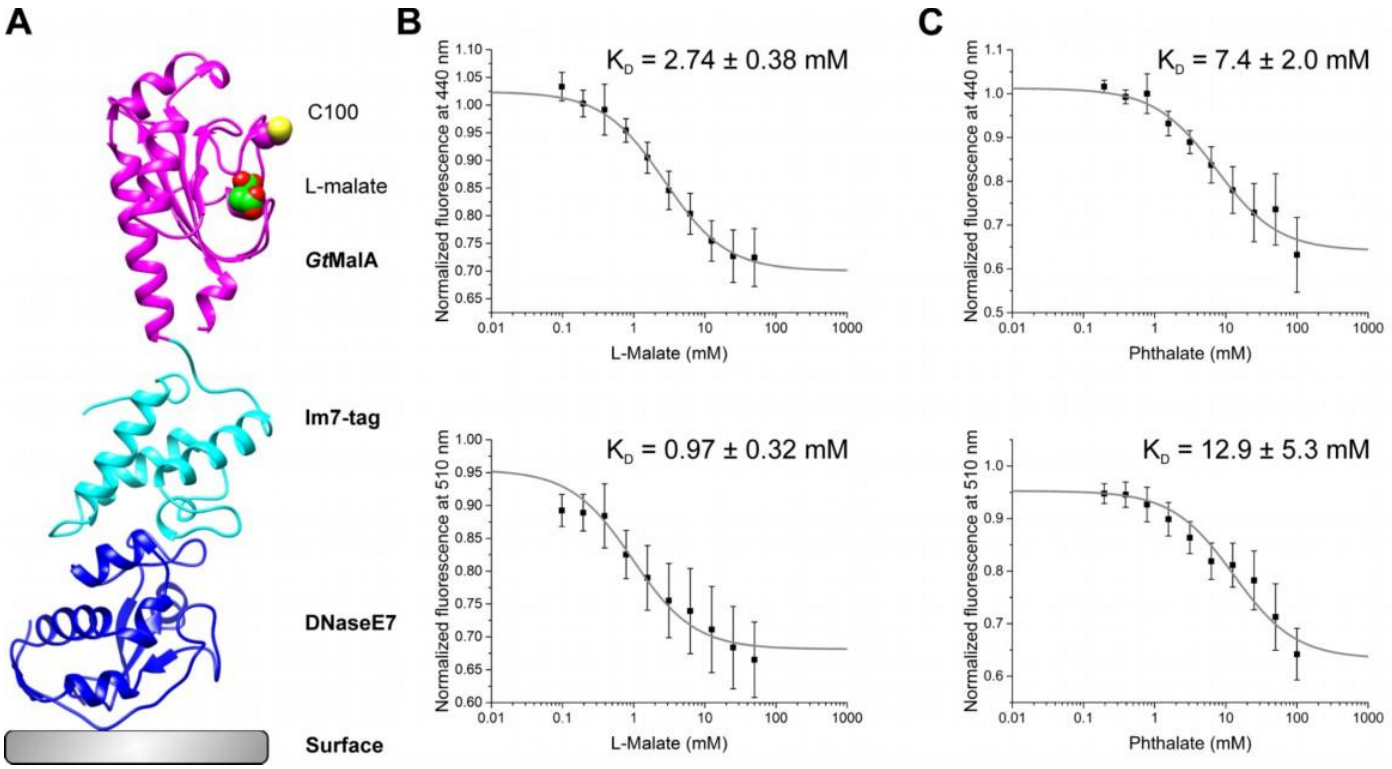


Figure 2. L-Malate-dependent fluorescence quenching of *Gt*MalAP-D100C. (A) Fluorescence spectra of DACM-labeled *Gt*MalAP-D100C in the presence of L-malate at concentrations of 32, 16, 8, 4, 2, 1, 0.5, 0.25, 0.125, and 0.0625 mM. An excitation wavelength of 380 nm was used. (B) The mean fluorescence at 440 nm was normalized to the fluorescence in PBS, plotted against the L-malate concentration, and fitted to a 1:1 Langmuir binding isotherm (grey curve) to determine the affinity. Error bars represent standard deviations of the three curves shown in (A) for each concentration. The apparent K_D is given with the standard error obtained from the fit. (C) Respective fluorescence spectra and (D) fit of the fluorescence at 510 nm of Dansyl-labeled *Gt*MalAP-D100C. A wavelength of 350 nm was used for excitation.

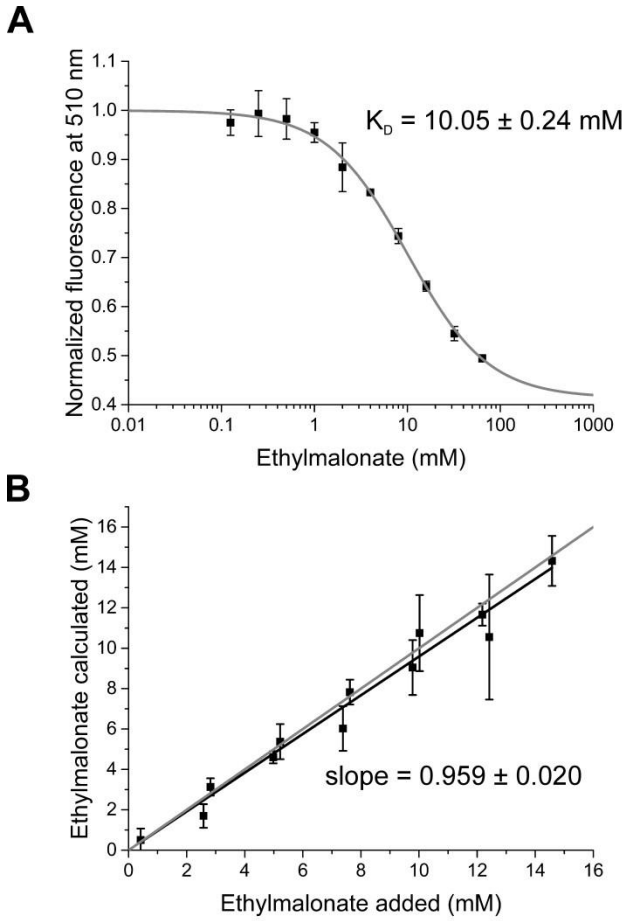


728

729 **Figure 3.** Characterization of *Gt*CitAP variants immobilized on DNaseE7-coated surfaces.
730 (A) Schematic structure of immobilized *Gt*MalAP-D100C based on the homology model
731 shown in Fig. 1 and the crystal structure of the complex between DNaseE7 and Im7 [46].
732 DNaseE7 (blue) was covalently coupled to the surface of maleic anhydride-activated
733 microplates. *Gt*MalAP-D100C (magenta) is attached to the surface via the high affinity
734 interaction between its N-terminal Im7-tag (cyan) and the immobilized catalytically inactive
735 DNaseE7-H545A domain (residues 444 - 576 of colicin E7). The ligand L-malate (green) and
736 C100, to which the label is attached, are shown as spheres. (B) The fluorescence of
737 immobilized DACM-labeled *Gt*MalAP-D100C at 440 nm (top) and the fluorescence of
738 immobilized Dansyl-labeled *Gt*MalAP-D100C at 510 nm (bottom) was fitted to a 1:1
739 Langmuir binding isotherm (grey curves) for determination of the affinity. The fluorescence
740 of each well was normalized to its fluorescence in PBS and averaged for each concentration.
741 Error bars represent standard deviations of at least four wells. Apparent K_D values are given

742 with standard errors obtained from the fit. (C) Respective fits for DACM-labeled (top) and
743 Dansyl-labeled (bottom) *Gt*PhtAP-D100C and the ligand phthalate.

744



745

746 **Figure 4.** Quantification of ethylmalonate in synthetic urine. (A) Plot of the fluorescence of
747 Dansyl-labeled *Gt*PhtAP-D100C fluorescence at 510 nm against the ethylmalonate
748 concentration. The fluorescence was normalized to the fluorescence in MOPS buffer and
749 fitted to a 1:1 Langmuir binding isotherm (grey curve), which was used as calibration curve
750 for the determination of ethylmalonate in synthetic urine. Error bars represent standard
751 deviations of three replicates. The apparent K_D is given with the standard error obtained from
752 the fit. (B) Plot of ethylmalonate concentrations which were added to synthetic urine against
753 the concentrations which were calculated based on the fluorescence quenching of Dansyl-
754 labeled *Gt*PhtAP-D100C in the respective samples. The slope of 0.959 ± 0.020 obtained by
755 linear regression (black line) reveals that the calculated concentration equals on average ~ 96

% of the added concentrations. The grey line with a slope of one represents a system with ideal response characteristics.