



# Lipocalin Blc is a potential heme-binding protein

Nina G. Bozhanova<sup>1</sup> , M. Wade Calcutt<sup>2</sup>, William N. Beavers<sup>3</sup>, Benjamin P. Brown<sup>1</sup> ,  
Eric P. Skaar<sup>3</sup> and Jens Meiler<sup>1,4</sup>

<sup>1</sup> Department of Chemistry, Center for Structural Biology, Vanderbilt University, Nashville, TN, USA

<sup>2</sup> Mass Spectrometry Research Center, Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN, USA

<sup>3</sup> Department of Pathology, Microbiology, and Immunology, Vanderbilt University Medical Center, Nashville, TN, USA

<sup>4</sup> Institute for Drug Discovery, Medical School, Leipzig University, Germany

## Correspondence

J. Meiler, Institute for Drug Discovery,  
Medical School, Leipzig University, Leipzig,  
SAC 04103, Germany  
Tel: +49 (341) 97 15941  
E-mail: jens@meilerlab.org

(Received 9 July 2020, revised 5 November  
2020, accepted 6 November 2020, available  
online 3 December 2020)

doi:10.1002/1873-3468.14001

Edited by Stuart Ferguson

**Lipocalins are a superfamily of functionally diverse proteins defined by a well-conserved tertiary structure despite variation in sequence. Lipocalins bind and transport small hydrophobic molecules in organisms of all kingdoms. However, there is still uncertainty regarding the function of some members of the family, including bacterial lipocalin Blc from *Escherichia coli*. Here, we present evidence that lipocalin Blc may be involved in heme binding, trans-periplasmic transport, or heme storage. This conclusion is supported by a cocrystal structure, mass-spectrometric data, absorption titration, and *in silico* analysis. Binding of heme is observed at low micromolar range with one-to-one ligand-to-protein stoichiometry. However, the absence of classical coordination to the iron atom leaves the possibility that the primary ligand of Blc is another tetrapyrrole.**

**Keywords:** Blc; crystal structure; hematoporphyrin; heme-binding protein; lipocalins

Heme is an essential molecule for aerobic organisms. Heme-containing proteins are crucial for many vital biological processes. This includes classical oxygen storage and transport, electron transfer, and catalysis, as well as more recently discovered roles in transcriptional gene regulation and ion channel regulation, among others [1].

Lipocalins are a superfamily of functionally diverse proteins with remarkably conserved tertiary structure despite significant sequence variation. These ~ 20 kDa proteins form an eight-stranded antiparallel  $\beta$ -barrel enclosing an internal ligand-binding pocket. Lipocalins bind and transport small hydrophobic molecules in organisms of all three kingdoms of life. These include, but are not restricted to, fatty acids [2], retinoic acid and retinol [3], heme [4,5], and open tetrapyrroles [6–8].

Heme-binding lipocalins include members from nitrophorins family and  $\alpha_1$ -microglobulin. Nitrophorins are salivary proteins found in several blood-sucking arthropods. Nitrophorins use their prosthetic heme molecule to bind and transport nitric oxide (NO). NO induces vasodilation when released in a host organism upon biting. Nitrophorins are also shown to bind histamine, reducing inflammation and immune response elicited by the insects' salivary proteins [9].  $\alpha_1$ -microglobulin is a small globular protein found in a number of vertebrate species. It is known to play a role in clearing cytosols and extravascular fluids of heme groups and free radicals released from hemoglobin [10]. Contrary to classical  $\alpha$ -helical globins with deeply buried ferrous heme, heme-binding lipocalins have shallow binding pockets. This leads to high solvent exposure of the heme and allows only for its binding in the ferric form. These

## Abbreviations

CGD, conjugate gradient descent; ESP, electrostatic potential; HMM, hidden Markov state model; HP, hematoporphyrin; MD, molecular dynamics; NO, nitric oxide; PDB, Protein Data Bank; wtBlc, wild-type Blc.

differences in structural organization and heme redox properties cause functional diversity between these groups of proteins [11].

The precise biological function of some lipocalin family proteins is not yet clear, even for those discovered more than two decades ago. This includes the first bacterial lipocalin protein identified, *Escherichia coli* lipocalin Blc. Based on monitoring of  $\beta$ -galactosidase activity from a *blc::lacZ* translational fusion reporter plasmid, it was proposed that Blc is involved in cell stress response (starvation, high osmolarity) [12,13]. Later, ligand-binding studies using tryptophan fluorescence quenching proposed that Blc plays a role in the metabolism of lysophospholipids based on its nanomolar affinity for oleoyl-*sn*-2-lysophosphatidylethanolamine and palmitoyl-*sn*-2-lyso-phosphatidylethanolamine [14]. That study, as well as the previous structural analysis of Blc in *apo* state [15], suggests that the protein exists as a functional homodimer. However, biochemical and structural characterization of Blc expressed using a different construct design questions the formation of the homodimer and, correspondingly, the correctness of the previous findings [16]. Recently, an *E. coli* single-gene knockout library screening showed involvement of Blc in the regulation of host intestine immune response against microbes [17]. Particularly, the complete absence of the protein or the mutation of glycine to glutamic acid at position 84 of the protein (a mutation frequently observed in pathogenic *E. coli* strains) is reported to cause activation of local inflammation in the host gut. The G84E Blc mutation has been associated with the higher *E. coli* abundance in the gut and the decrease in lysophosphatidylethanolamine in both the bacteria and the host intestine lumen.

Here, based on a Blc crystal structure and supported by mass-spectrometric data, titration analysis, and molecular dynamics (MD) simulations, we present evidence that lipocalin Blc may also be involved in ferric heme binding, transport, or storage. The absence of the classical proximal coordination of the heme-Fe atom and relatively low binding affinity for heme ( $K_d = 1.4 \mu\text{M}$ ) suggests that the Blc protein may not be specific to heme and could participate in handling of other metal-containing or metal-free tetrapyrroles.

## Materials and methods

### Molecular cloning

The vectors containing wild-type Blc (wtBlc)-split-Zip and wtBlc-split (split fragments without leucine zipper peptides) were created using plasmids pMRBad-Z-CspGFP

(Addgene, Watertown, MA, USA plasmid #40730), pET11a-Z-NspGFP (Addgene plasmid #40729) [18], which were a gift from Brian McNaughton, and wtBlc [19] as described before [20]. Correctness of the obtained constructs was confirmed by sequencing.

### Protein expression and purification

All proteins were expressed in XJb(DE3) Autolysis (Zymo Research, Irvine, CA, USA) *E. coli* strain. Cells were grown in LB or M9 media supplemented with  $100 \mu\text{g}\cdot\text{mL}^{-1}$  ampicillin (full-length wtBlc protein) or  $100 \mu\text{g}\cdot\text{mL}^{-1}$  ampicillin and  $50 \mu\text{g}\cdot\text{mL}^{-1}$  kanamycin (wtBlc-split-Zip and wtBlc-split proteins) at  $37^\circ\text{C}$ . Expression was induced by addition of 0.04% L-arabinose (full-length wtBlc protein) or 0.2% L-arabinose and  $10 \mu\text{M}$  IPTG (wtBlc-split-Zip and wtBlc-split proteins) at 0.8 OD. Cells were harvested after 3 h of expression at  $37^\circ\text{C}$  if grown in LB or after overnight expression if grown in M9 and were resuspended in PBS buffer, pH 7.4. Suspensions were frozen at  $-80^\circ\text{C}$  and thawed at room temperature three times. DNA was destroyed by short sonication, and the lysates were centrifuged to obtain cell-free extracts. The proteins were first purified using gravity flow columns with TALON metal affinity resin (Clontech, Mountain View, CA, USA) and further purified by size-exclusion chromatography on a HiLoad 16/600 Superdex 75 pg or Superdex 200 pg 10/300 GL column (GE Healthcare, Marlborough, MA, USA) pre-equilibrated with 50 mM sodium phosphate buffer, pH 6.0.

### Protein concentration calculation

Protein concentrations were estimated using the Bradford dye-binding method-based [21] colorimetric assay (Bio-Rad, Hercules, CA, USA) and bovine serum albumin standard. Single point absorption measurements (595 nm) were performed using a FlexStation 3 microplate reader (Molecular Devices, San Jose, CA, USA). All measurements were performed in triplicate.

### HPLC-high-resolution MS and MS/MS

Mass-spectrometric experiments were performed using an Acquity UPLC system (Waters, Milford MA, USA) and an LTQ Orbitrap mass spectrometer equipped with a standard electrospray ionization source (Thermo, San Jose, CA, USA). The Acquity system was equipped with a binary solvent manager, refrigerated sample manager, and a thermostated column heater. Samples were chromatographically resolved on a Waters Symmetry300 reverse phase column ( $\text{C}_{18}$ ,  $2.0 \times 100 \text{ mm}$ ,  $3.5 \mu\text{m}$ ) using the following gradient conditions: 0–1 min, B = 0%; 1–8 min, B = 0–100%; 8–10 min, B = 100%; 10–10.5 min, B = 100–0%; 10.5–15 min, B = 0%. The injection volume for all samples was  $10 \mu\text{L}$ . The autosampler injection valve and syringe needle were flushed

(1 mL) and rinsed (1 mL) sequentially with mobile phases B and A between each injection. The mass spectrometer was operated in positive ionization mode; the electrospray source parameters were as follows: N<sub>2</sub> sheath gas, 40; N<sub>2</sub> aux gas, 5; spray voltage, 4.0 kV; capillary temp, 300 °C; tube lens voltage, 130 V at  $m/z = 1150$ ; capillary voltage, 31 V; skimmer offset, 10 V. Full scan spectra of 300–2000  $m/z$  were acquired in FTMS mode at a resolving power of 30 000 FWHM with the following AGC parameters: FT target, 1e6; one microscan, maximum inject time, 50 ms. Collision-induced dissociation of the precursor ion 616.17  $m/z$  was performed in the linear ion trap using the following activation parameters: IT target, 1e4; one microscan, maximum inject time, 25 ms; activ Q, 0.25; activ time 30 ms; norm collision energy 15%; isolation width 2.0  $m/z$ . Data acquisition and analysis were carried out using XCALIBUR v. 2.0.7 (Thermo Fisher Scientific, Waltham, MA, USA).

### Spectrophotometric hemin titration

A fresh 2 mM hemin stock solution was prepared by adding 5 mL 0.1 M NaOH to 6.5 mg hemin chloride (Sigma-Aldrich, Darmstadt, Germany). The required amount of hemin stock solution (0–6  $\mu$ L) supplemented with 0.1 M NaOH to a final volume of 6  $\mu$ L was mixed either with 994  $\mu$ L of the 50 mM sodium phosphate buffer, pH 6.0 or with the appropriate amounts of the full-length wtBlc or the wtBlc-split protein and 50 mM sodium phosphate buffer, pH 6.0 to obtain a 5  $\mu$ M final protein concentration. Samples were incubated for 30 min or 2 h, and absorbance spectra between 200 and 600 nm were detected using a double-beam Shimadzu (Kyoto, Japan) UV-1800 UV/Vis spectrophotometer.

### Spectrophotometric HP titration

A fresh 1 mM HP solution was prepared by diluting 10 mM HP stock solution (MedChem Express, Monmouth Junction, NJ, USA) in 100% DMSO. The required amount of HP 1 mM solution (0–10  $\mu$ L) supplemented with 100% DMSO to a final volume of 10  $\mu$ L was mixed either with 990  $\mu$ L of the 50 mM sodium phosphate buffer, pH 6.0 or with the appropriate amounts of the full-length wtBlc protein and 50 mM sodium phosphate buffer, pH 6.0 to obtain a 5  $\mu$ M final protein concentration. Samples were mixed, and absorbance spectra between 200 and 600 nm were detected using a double-beam Shimadzu UV-1800 UV/Vis spectrophotometer. Sample preparation was performed with all possible precautions to minimize light exposure of the sample and therefore prevent photoinduced changes.

### Tryptophan fluorescence quenching assay

Fresh HP solutions (100, 12.5, 10, and 1.6  $\mu$ M) were prepared by diluting 10 mM HP stock solution (MedChem Express) in 100% DMSO. The required amount of HP

solution (0.5–3  $\mu$ L) supplemented with 100% DMSO, if needed, to a final volume of 3  $\mu$ L was mixed with the appropriate amounts of the full-length wtBlc protein and 50 mM sodium phosphate buffer, pH 6.0 to obtain 10 nM final protein concentration. Fluorescence emission spectra were acquired using a Horiba Jobin Yvon Fluoromax-3 fluorometer in the 300–550 nm range with excitation at 290 nm and a step size of 0.25 nm. Sample preparation was performed with all possible precautions to minimize light exposure of the sample and therefore prevent photoinduced changes.

### $K_d$ calculation

Assuming 1 : 1 protein:ligand binding stoichiometry, the protein–ligand complex concentration in equilibrium ( $[PL]_{eq}$ ) can be expressed as following:

$$[PL]_{eq} = \frac{([L]_{init} + [P]_{init} + K_d) - \sqrt{([L]_{init} + [P]_{init} + K_d)^2 - 4[L]_{init}[P]_{init}}}{2} \quad (1)$$

where  $[L]_{init}$  and  $[P]_{init}$  are initial concentrations of a ligand and a protein, and  $K_d$  is the dissociation constant. Then, the total absorption at 411 nm ( $A_{411}$ ) of the protein–ligand solution in equilibrium can be calculated as follows:

$$A_{411} = [PL]_{eq} \times \epsilon_{[PL]} + ([L]_{init} - [PL]_{eq}) \times \epsilon_{[L]} \quad (2)$$

where  $\epsilon_{[PL]}$  and  $\epsilon_{[L]}$  are extinction coefficients of, correspondingly, the protein–ligand complex and the free ligand at 411 nm at that concentration.

The total fluorescence intensity at 354 nm ( $I_{354}$ ) of the protein–ligand solution in equilibrium can be calculated as follows:

$$I_{354} = [PL]_{eq} \times I_{[PL]} + ([P]_{init} - [PL]_{eq}) \times I_{[P]} \quad (3)$$

where  $I_{[PL]}$  and  $I_{[P]}$  are fluorescence intensities of, correspondingly, the protein–ligand complex and the free protein at 354 nm when excited at 290 nm.

To determine the apparent  $K_d$ , the experimentally obtained data were fitted to the Equation (2) or (3) by the *leastsq* method from the Scipy optimize package. Extinction coefficients of the free ligands were obtained by fitting experimental data of the free ligand absorbance spectra to the polynomial (in the case of the heme titration; Fig. S1A) or linear (in the case of the HP titration; Fig. S1B) function.

### Crystallization, data collection, and structure determination

WtBlc-split (12 mg·mL<sup>−1</sup> in 50 mM sodium phosphate buffer, pH 6.0) was crystallized at 21 °C in 1.6 M ammonium

sulfate, 0.1 M MES, pH 4.5 with protein to precipitant volume ratio of 1 : 1 or in 1.6 M ammonium sulfate, 0.1 M MES, pH 4.5 supplemented with 0.1 M Iron(III) chloride hexahydrate or 5% w/v n-Dodecyl- $\beta$ -D-maltoside according to the Hampton Research Additive Screen protocol using the sitting drop vapor diffusion technique. Crystals grew within 1 week and were flash frozen in liquid nitrogen using Parabar 10312 oil as cryoprotectant.

Diffraction data were collected at the Life Sciences Collaborative Access Team beamline 21-ID-F at the Advanced Photon Source, Argonne National Laboratory. The diffraction data were processed using the XIA2 software suite [22]. The crystal structures were solved by molecular replacement with MOLREP [23] using the DiB2-split structure as a search model (PDB ID 6UKL). Model building and iterative refinement were performed with COOT [24] and REFMAC [25], respectively. The final statistics of the structure are shown in Table S1. The model has been deposited into the Protein Data Bank (PDB ID 6VRI). Structure figures were prepared using PYMOL (v.2.2.3; Schrödinger, LLC, New York, NY, USA).

## Molecular modeling

### Model building and protein–ligand docking

Starting coordinates for hematoporphyrin (HP) are based on the common atom coordinates in the crystallographically determined pose of heme in complex with wtB1c-split fit with a maximum common substructure alignment [26]. Split site residues were remodeled with Rosetta comparative modeling (RosettaCM) using a previously solved crystallographic structure of wtB1c (PDB ID 1QWD) to obtain a model of the intact wtB1c. Subsequently, complexes of each ligand with wtB1c were relaxed into the Rosetta Talaris2014 score function with constant starting coordinate constraints.

Following relaxation, we performed local docking-based refinement of wtB1c:HP with RosettaLigand. Small molecule conformers were pregenerated with the BioChemical Library (BCL) cheminformatics toolkit [27], and 2000 rounds of local docking (conformer swap, high-resolution docking, and minimization) were performed starting from the crystallographically determined pose of wtB1c-split: heme.

### Molecular dynamics (MD) simulations

The best scoring pose of wtB1c:HP refined in RosettaLigand was used to initiate all MD simulations. Amino acid protonation states of wtB1c were assigned by H++ (<http://biophysics.cs.vt.edu/H++>) [28] at pH 6.0 to match the conditions of the HP titration experiments. HP geometry optimization and electrostatic potential (ESP) calculations

were performed in Gaussian [29] at the B3LYP/6-31G\* level of theory. Restrained ESP fitting was subsequently performed with AmberTools18 [30] to parameterize HP for classical MD simulations. Full system parameterization was completed using Leap in AmberTools18 [30]. Simulations were performed in Amber18 [30] with the ff14SB force field [31] and the generalized Amber force field 2 (GAFF2) [32]. Structures were solvated in a rectangular box of TIP3P explicit solvent neutralized with Joung–Cheatham monovalent ions [33,34]. wtB1c:HP was buffered on all sides with 12.0 Å solvent. Hydrogen mass repartitioning was performed on solute atoms to allow a simulation timestep of 4.0 fs [35].

Minimization proceeded in three stages: (a) The system was minimized with 500 cycles of steepest descent followed by 1000 steps of conjugate gradient descent (CGD) while protein atoms were restrained with a force constant of 10.0 kcal·mol<sup>−1</sup>·Å<sup>−2</sup>; (b) the system then underwent 500 cycles of steepest descent followed by 1000 steps CGD minimization in buffer restrained with a force constant of 5.0 kcal·mol<sup>−1</sup>·Å<sup>−2</sup>; (c) finally, restraints were removed from the system for 500 steps steepest descent followed by 1000 steps of CGD minimization.

Postminimization, SHAKE [36] and SETTLE [37] (for rigid water molecules) were implemented to constrain covalent bonds to hydrogen atoms. Systems were slowly heated in the canonical (NVT; constant number of particles, temperature, and volume) ensemble to 100 K over 50 ps with a 1 fs timestep. Subsequently, systems were heated in NPT ensemble at 1 bar with isotropic position scaling from 100 to 300 K over 500 ps and 1 fs timestep. Equilibration/production simulations were run in the isothermal–isobaric (NPT) ensemble at 300 K for 2.0–4.0  $\mu$ s with a Monte Carlo barostat and integrated every 4 fs. Temperature was controlled using Langevin dynamics with a collision frequency of 1 ps<sup>−1</sup>. Periodic boundary conditions were imposed on the system throughout heating and equilibration/production. Electrostatics were evaluated using the Particle Mesh Ewald method and a distance cutoff of 8.0 Å. Trajectory RMSDs were calculated using CPPTRAJ [38].

### Markov model analysis of molecular dynamics simulations

We constructed a hidden Markov state model (HMM) distinguishing five metastable states of wtB1c:HP using PYEMMA version 2.5.7 [39]. We combined all 12.0  $\mu$ s of MD simulations and computed pairwise distances between all heavy atoms of HP and all heavy atoms of wtB1c residues within 4.0 Å of any HP atom for each trajectory frame to generate our feature trajectory. Subsequently, we performed time-lagged independent component analysis of the feature trajectory with a lag time of 10 ps. We then constructed an HMM and computed the stationary distribution of each metastable state.



## Binding free energy calculations

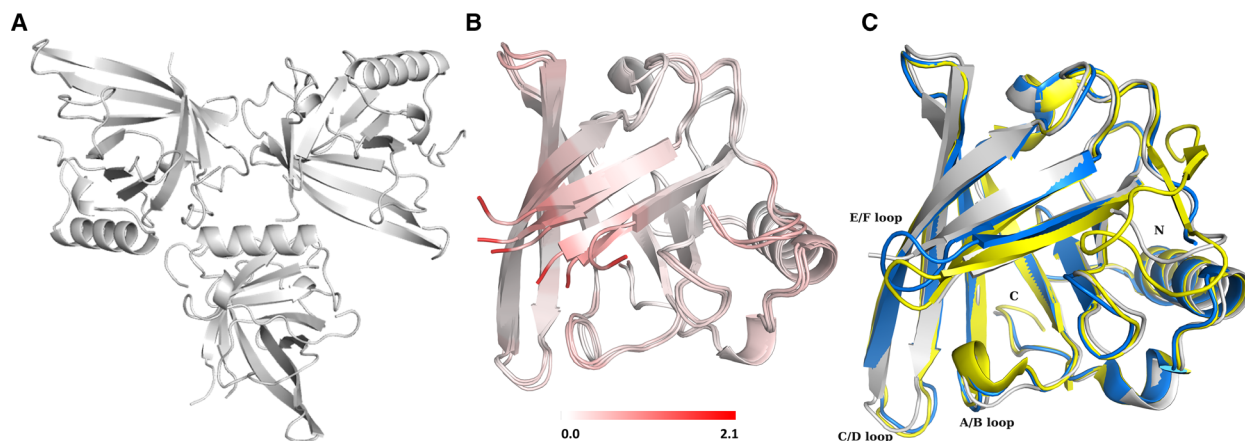
The estimated binding free energy between each of the wtBlc:HP states was computed with the molecular mechanics Poisson–Boltzmann surface area (MM-PBSA) and interaction entropy (IE) approaches using the MM-PBSA.py program [40]. From each metastable state identified in the HMM, we randomly resampled 1000 structures to use for wtBlc:HP binding free energy calculations. For the MM-PBSA calculations, the internal and external dielectric constants were set to 4.0 and 80.0, respectively. The nonpolar component of the solvation-free energy was estimated from the solvent accessible surface area with the classical method (INP = 1) using default coefficient and offset values. Atomic radii were taken from the parameter–topology file (RADIOPT = 0). The IE was computed as a *post hoc* analysis of the MM-PBSA gas phase (electrostatic and Van der Waals) contributions to the interaction energy between the protein and ligand, as described by Duan *et al.* [41].

## Results and Discussion

During our previous work on designing a new class of fluorogen-activating proteins named DiBs based on modified lipocalin Blc [19], we excluded protein variants with intensive visible coloring prior to the addition of fluorophores. The wtBlc was one of these excluded variants (Fig. 2A) despite its promising ability to bind some of the tested fluorophores because neither size-exclusion chromatography nor continuous dialysis was able to extinguish or decrease the intensity of the color. We later created a DiB-based split protein

reporter by cleaving the protein into two parts [20]. The resultant N- and C-terminal fragments were either (a) genetically fused to leucine zipper peptides and co-expressed (further referred as split-Zip proteins), or (b) co-expressed without additional modifications (further referred as split proteins) (Fig. S2). Based on the proximity of the cleavage site to the binding pocket, we hypothesized that split wtBlc could show reduced binding to the unknown colored compound. Correspondingly, we created and examined two similar constructs generated from the wtBlc: wtBlc-split-Zip and wtBlc-split. Counter to our hypothesis, both wtBlc-split-Zip and wtBlc-split retained their intense brownish color when expressed in LB medium.

Attempts to obtain crystals of the colored full-length wtBlc protein were unsuccessful. However, we were able to determine the structure of wtBlc-split protein at 1.94 Å resolution. The wtBlc-split asymmetric unit cell contained three ‘split’ molecules (Fig. 1A) that differ primarily in the area of the two β-strands adjacent to the split point (Fig. 1B). Two fragments of the split protein together fully reconstitute the typical lipocalin fold. Eight antiparallel β-strands (designated A–H) form a barrel and are accompanied by a C-terminal α-helix. Alignment of a wtBlc-split ‘monomer’ with previously determined Blc structures (PDB IDs 1QWD, 2ACO, and 3MBT) reveals no significant structural perturbation. The Cα root mean-squared deviation (rmsd) between this structure and those previously reported range from 0.98 Å (3MBT) to 1.02 Å (2ACO) over 147 shared residues. Structural deviations



**Fig. 1.** Overall wtBlc-split structure characterization. Cartoon representation of the (A) wtBlc-split crystal asymmetric unit, (B) overlaid three wtBlc-split nonsymmetric monomers, and (C) a wtBlc-split monomer (grey) aligned with previously determined full-length Blc structures (3MBT, blue; 2ACO, chain A, yellow). (B) Residues of the overlaid monomers are color-coded according to their distance to the mean position of three equivalent residues. Only residues that are present in all three nonsymmetric monomers are shown. Pairwise Cα rmsd (150 atoms) equal to 0.6, 1.2, and 1.2 Å. (C) Structure of the third full-length Blc structure, 1QWD, almost perfectly overlays with 2ACO (rmsd = 0.1 Å) and therefore is not shown in order to reduce the complexity of the figure

are primarily located in the area adjacent to the split protein cleavage point (former E/F loop). Other deviations can be observed in two other loops (A/B and C/D loops) as well as at the N-terminus (Fig. 1C). These areas colocalize with the previously identified regions where the full-length Blc structures differ from each other [16] and most likely reflect the overall flexibility of the long loops.

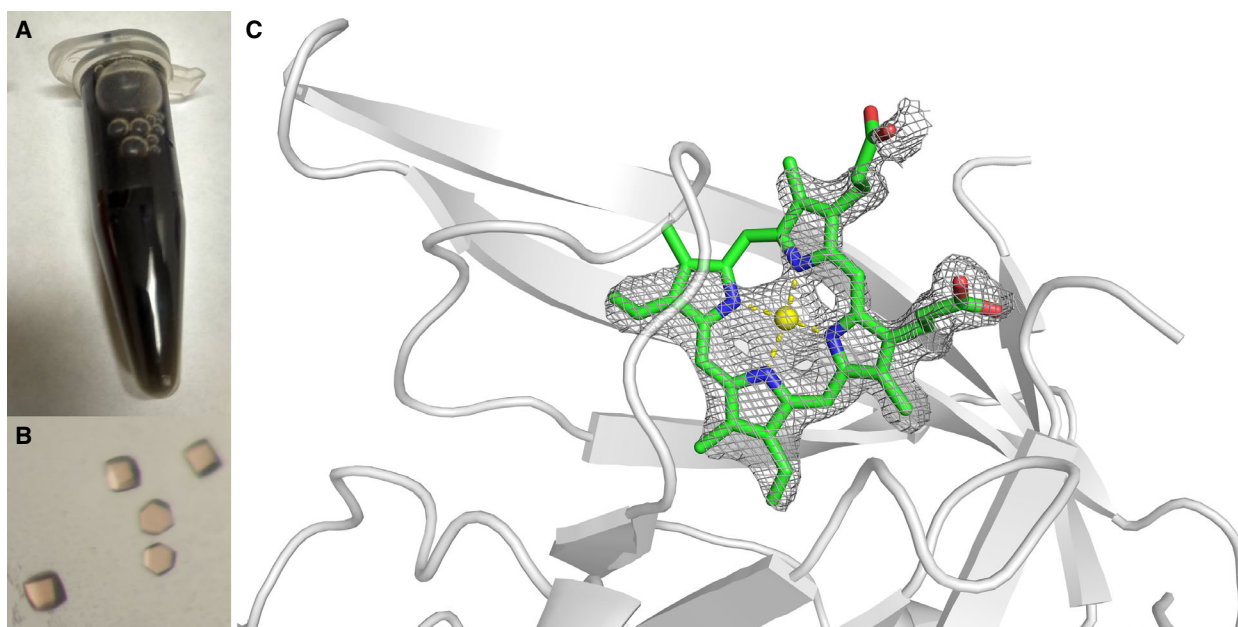
Consistent with the colored protein solution, the wtBlc-split crystals also had notable color (Fig. 2B). During structure refinement, we discovered a well-defined positive difference density map area in the binding pocket of one of the three 'split' molecules in the asymmetric unit. Based on the shape of the density as well as previous observations including color of the protein, absorbance, and fluorescence spectra, we suggested the presence of the heme B molecule in the structure (Fig. 2C). To test our hypothesis that heme B is present in the wtBlc-split structure, we performed a number of experiments.

First, we compared the wtBlc-split protein sample used for growing crystals with the commercially available hemin standard ( $\geq 97.0\%$  HPLC; Sigma-Aldrich) using HPLC-high-resolution MS and MS/MS. LC-MS/MS analysis of the wtBlc-split protein sample showed the presence of a small molecule with a  $[M + H]^+$   $m/z = 616.1774$ , having the same chromatographic

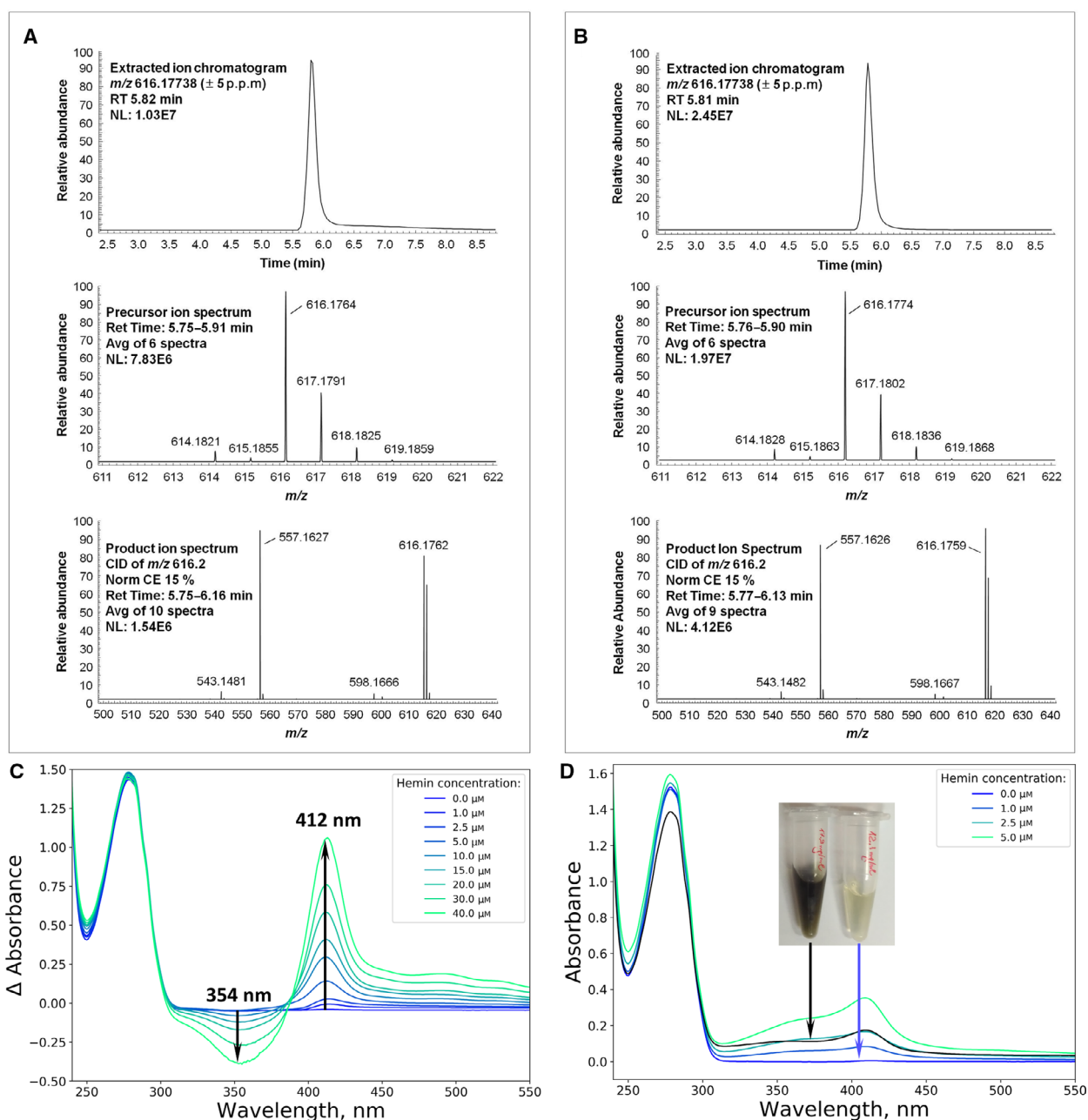
retention time, precursor ion exact mass, precursor ion isotopologue pattern, and CID product ion spectrum of the standard ( $[M + H]^+$   $m/z = 616.1764$ ; Fig. 3A,B).

Second, we used absorption spectroscopy. During our work, we discovered that the recombinant protein expression in minimal medium (M9 salts) results in much less-colored protein samples (Fig. 3D, inset). We titrated hemin into the full-length wtBlc recombinant protein expressed in minimal medium and compared these spectra with absorption of free hemin in the buffer. The resulted differential spectra demonstrated gradually decreasing absorption at  $\sim 350$  nm and increasing absorption at 412 nm (Fig. 3C). The latter absorption peak is known as the Soret peak and is characteristic for heme-binding proteins [42]. Notably, the spectrum of the protein expressed in minimal medium supplemented with hemin matches almost perfectly the absorbance of the LB-expressed wtBlc protein (Fig. 3D). However, this comparison also suggests that the LB-expressed wtBlc samples are not saturated with heme. This may explain the presence of the heme molecule only in one out of three copies of the split protein in the asymmetric unit in the obtained crystal structure.

After confirming our initial hypothesis of the presence of heme in the binding pocket of the wtBlc-split protein, we proceeded with a more detailed



**Fig. 2.** Evidence for the presence of heme in the Blc protein. (A) The concentrated ( $\sim 50$  mg·mL $^{-1}$ ) solution of the purified wtBlc protein. (B) Colored 'apo' wtBlc-split crystals. (C) Heme-like density in the wtBlc-split crystal. The  $2F_o - F_c$  electron density map from the binding pocket region of one of the wtBlc-split protein molecules after omit refinement is shown as grey mesh. The protein molecule is shown as grey cartoon. The putative heme ligand is shown as sticks.



**Fig. 3.** Confirmation of the presence of heme in the Blc protein. (A–B) Detection of ferric heme by HPLC-high resolution MS and MS/MS in (A) 1 mM hemin chloride (solution in 100% DMSO) standard diluted in  $H_2O/MeOH$  (1 : 1) to a final concentration of 10  $\mu$ M, 100 pmol hemin on column. (B) 0.5 mM heterologously expressed wtBlc-split (solution in 50 mM sodium phosphate buffer, pH 6.0) diluted in  $H_2O/MeOH$  (1 : 1) to a final concentration of 5  $\mu$ M, 50 pmol protein on column. (C) Differential absorbance spectra of heme bound to M9-expressed wtBlc protein. Increasing aliquots of 1 mM hemin solution in DMSO were added to 40  $\mu$ M protein solution and measured using the mixture of the same amount of the hemin solution with the buffer as a reference read. (D) Absorbance spectra of wtBlc protein expressed in minimal medium and supplemented with heme in comparison with the spectrum of wtBlc protein expressed in LB (black solid line). Inset: difference in the color of the solutions of the purified wtBlc protein samples expressed in LB (left) and minimal medium (right).

characterization of the discovered interaction. The heme molecule occupied the same hydrophobic cavity of the lipocalin as it was previously showed for

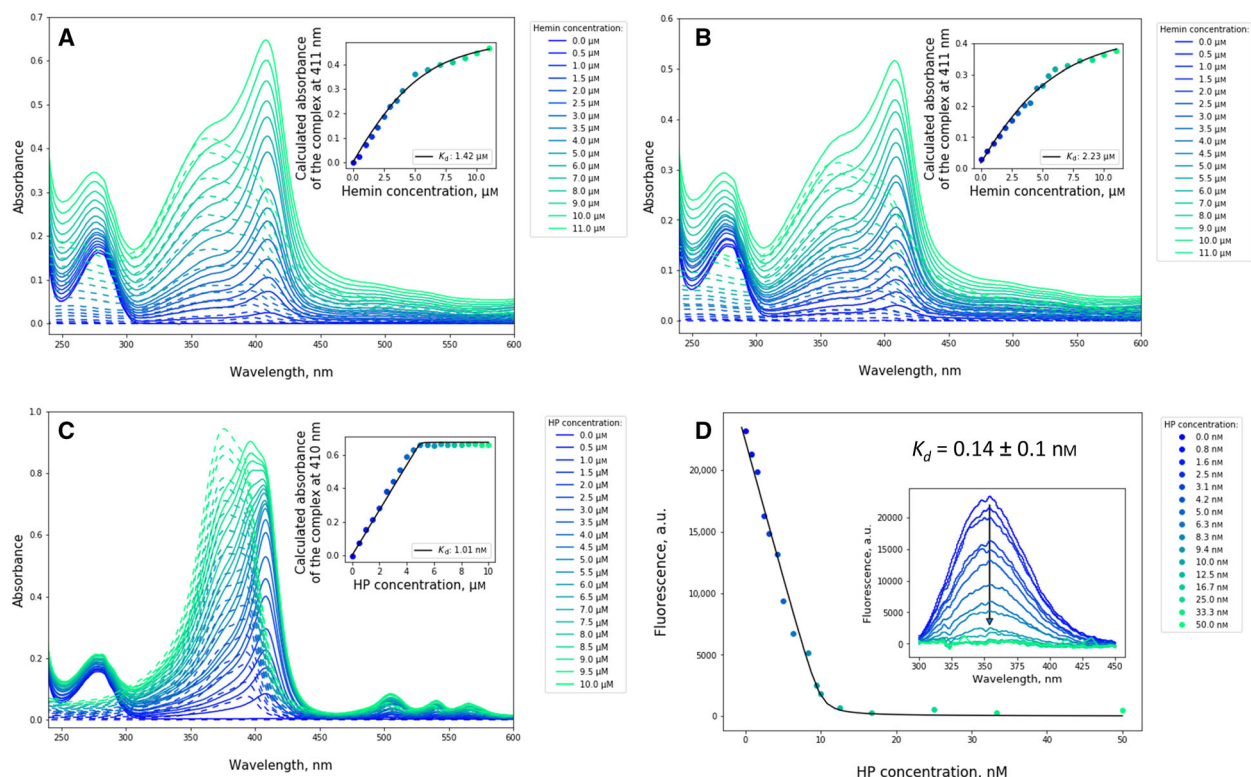
vaccenic acid [14] (Fig. S3). We further assessed the hemin-binding affinity to the wtBlc. For this, we used the full-length wtBlc protein expressed in minimal

medium. This sample failed to have any detectable absorption at 411 nm and according to the LC-MS contained negligible amounts of copurified heme. We measured absorbance spectra of the protein in the presence of different hemin concentrations (Fig. 4A). Assuming 1 : 1 protein : ligand binding stoichiometry based on the obtained crystal structure, spectrophotometric titration yielded a  $K_d$  of  $1.4 \pm 0.3 \mu\text{M}$ . This value is comparable to those determined for full-length wtB1c binding to fatty acids [14] or some other heme-binding proteins like lipocalin  $\alpha_1$ -microglobulin [43], bacterial chlorite dismutase-like protein HemQ [44], or some all-helical hemoproteins like horseradish peroxidase [45]. The affinity for heme is lower than that of classical heme-binding proteins like apohemoglobins

and apomyoglobins ( $K_d$  on the order of  $10^{-14} \text{ M}$ ) [46], some other  $\beta$ -barrel proteins like nitrobindins ( $K_d$  on the order of  $10^{-12} \text{ M}$ ) [47], or hemophores ( $K_d$  on the order of  $10^{-10} \text{ M}$ ) [48].

To further ensure that splitting of wtB1c did not fundamentally affect its ability to bind to heme and, therefore, that we can extrapolate our wtB1c-split:heme complex-based findings onto the full-length protein, we also performed the same spectrophotometric heme titration of the wtB1c-split protein. This titration yielded a similar dissociation constant of  $2.0 \pm 0.3 \mu\text{M}$  (Fig. 4B).

We then compared the heme binding and coordination by wtB1c-split with other known structures. We downloaded all available structures of lipocalins



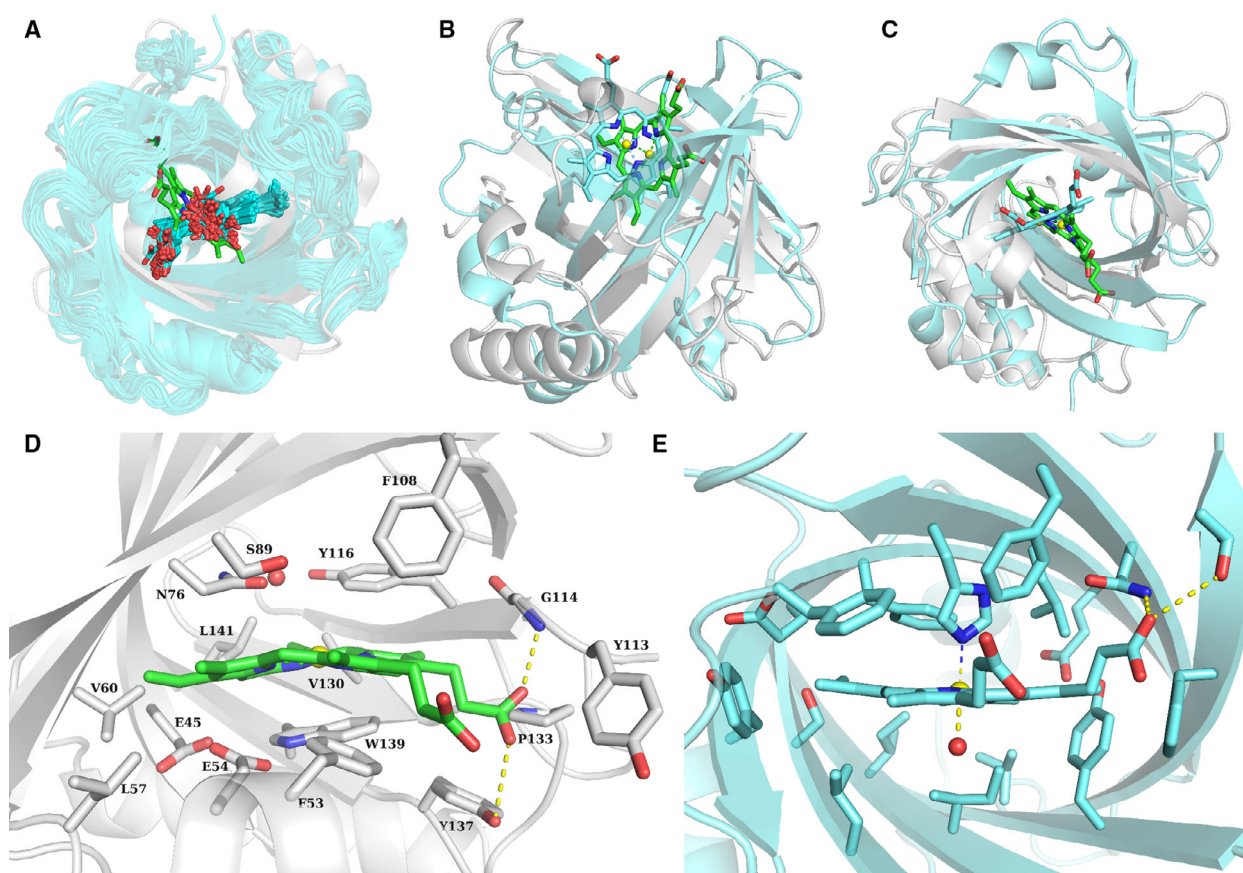
**Fig. 4.** wtB1c binding to porphyrins. Representative examples of the spectrophotometric heme titration of wtB1c (A) and wtB1c-split (B). Solid lines show absorbance spectra of 5  $\mu\text{M}$  protein solution (expressed in minimal medium) supplemented with increased concentrations (0.5–11  $\mu\text{M}$ ) of heme. Dashed lines show absorbance spectra of increased concentrations (0.5–11  $\mu\text{M}$ ) of heme in the buffer. Inset: the calculated absorbance of the formed protein:heme complex at 411 nm and the fitting curve (black line). (C) A representative example of the spectrophotometric HP titration of wtB1c. Solid lines show absorbance spectra of 5  $\mu\text{M}$  full-length minimal medium expressed wtB1c protein solution supplemented with increased concentrations (0.5–10  $\mu\text{M}$ ) of HP. Dashed lines show absorbance spectra of increased concentrations (0.5–10  $\mu\text{M}$ ) of HP in the buffer. Inset: the calculated absorbance of the formed wtB1c:HP complex at 410 nm and the fitting curve (black line). (D) A representative example of the fluorometric HP titration of wtB1c. Data points show tryptophan fluorescence intensity at 354 nm of 10 nM full-length minimal medium expressed wtB1c protein solution supplemented with increased concentrations (0.8–50 nM) of HP. The fit of the emission intensity to Equation (3) (black line) yielded a  $K_d$  of  $0.14 \pm 0.1 \text{ nM}$ . Inset: tryptophan fluorescence spectra of 10 nM full-length minimal medium expressed wtB1c protein solution supplemented with increased concentrations (0.8–50 nM) of HP. The arrow shows the direction of the changes as the concentration of HP increases.



cocrystallized with heme from the PDB. All 67 of these belong to the nitrophorins family. While all proteins superimposed relatively well ( $\sim 4$  Å rmsd over backbone atoms), position of heme in the binding pocket in wtBlc-split is strikingly different from all other analyzed heme-binding lipocalins (Fig. 5A). Taking into account the overall similarity of all found nitrophorins, we continued comparison of the heme-binding modes of wtBlc-split and nitrophorins using one structure, nitrophorin 7 from *Rhodnius prolixus* (PDB ID 4XMC) [49], which has the lowest overall rmsd to wtBlc-split. In both structures, heme occupies the main cavity of the lipocalin fold and is situated similarly deep in the binding pocket (Fig. 5B). However, the plane of the heme molecule in the wtBlc-split binding

pocket is rotated almost 120 degrees relative to its position in nitrophorin 7 (Fig. 5C).

In the binding pocket formed by chains E and F of the structure, there are 19 amino acid residues containing atoms within a 4.0 Å distance from the ligand (Glu45, Phe53, Glu54, Leu57, Val60, Asn76, Lys77, Gly78, Ser89, Phe108, Tyr113, Gly114, Gly115, Tyr116, Val130, Pro133, Tyr137, Trp139, and Leu141). Phe53 and Trp139 form  $\pi$ -stacking interactions with the porphyrin rings IV and II. The side-chain hydroxyl of Tyr137 and backbone amide of Gly113 stabilize one of the propionate groups of heme by hydrogen bonds. All other listed amino acids contribute to the hydrophobicity of the binding pocket (Fig. 5D). Heme also interacts with two water



**Fig. 5.** Comparison of heme binding and coordination by wtBlc-split with other lipocalins cocrystallized with heme. (A) Overlay of 67 lipocalin structures cocrystallized with heme found in the PDB (proteins are shown as cyan cartoon, and heme molecules are shown as cyan sticks) aligned with wtBlc-split structure (protein is shown as grey cartoon, and heme is shown as green sticks). The position of heme in the binding pocket of wtBlc-split (grey cartoon, heme is shown as green sticks) and nitrophorin 7 (PDB ID 4XMC, protein is shown as cyan cartoon, and heme is shown as cyan sticks): (B) side view and (C) view from the top. (D) Heme coordination in the binding pocket of the wtBlc-split protein. The protein molecules are shown as grey cartoon. Amino acid side chains within 4 Å distance from the ligand are shown as sticks. Heme is shown as green sticks. Hydrogen bonds are shown as yellow dashed lines. (E). Classical heme coordination in the binding pocket of nitrophorin 7 (PDB ID 4XMC). Hydrogen bonds are shown as yellow dashed lines.

molecules—one at the solvent front and one buried near Y116 (Fig. S4). There are two additional amino acid residues from the adjacent ‘split’ molecule in the asymmetric unit (Ser148 and Glu150 from chain D) that contribute interactions with heme at the solvent front (Fig. S4). These latter amino acid contacts may be sterically impaired in the full-length wtB1c with the intact E/F loop. Moreover, according to the Protein Interfaces, Surfaces and Assemblies (PISA) server [50] evaluation, the interface between ‘split’ molecules in the trimer is unlikely to be stable in solution. This suggests that the observed interactions of the heme propionate group with the amino acids from chain D are most likely the result of the crystal packing. We speculate that in solution the propionate group is exposed to water.

While the presence of heme propionate anchors as well as the described above heme ‘face’ and heme edge interactions are standard interactions that were previously reported in other heme-binding proteins [51], the classical heme iron atom coordination by protein amino acid(s) (proximal ligand interaction) is missing, making the observed heme-binding mode unusual (Fig. S5E). We considered that this heme-binding mode may be a crystallographic artifact. To investigate the stability of the observed binding pose, we employed MD simulations. We performed three independent 2.0  $\mu$ s simulations of wtB1c-split:heme and the modeled full-length wtB1c:heme complexes. In all six simulations, we saw no substantial deviation from the crystallographic pose (Fig. S5A). Superimposition of the binding pockets from the final frame of both the wtB1c-split:heme (Fig. S5B) and wtB1c:heme (Fig. S5C) trajectories shows good overlap of heme with the crystallographically determined pose. These results support our experimental structural data and suggest that the observed wtB1c-split:heme interaction in the crystal structure is not a crystallographic artifact.

However, wtB1c may not have evolved specificity toward heme binding and may bind other porphyrins. To test this hypothesis, we evaluated the ability of wtB1c to interact with other chemically similar porphyrins: protoporphyrin IX, and non-natural porphyrins HP and verteporfin (Fig. S6).

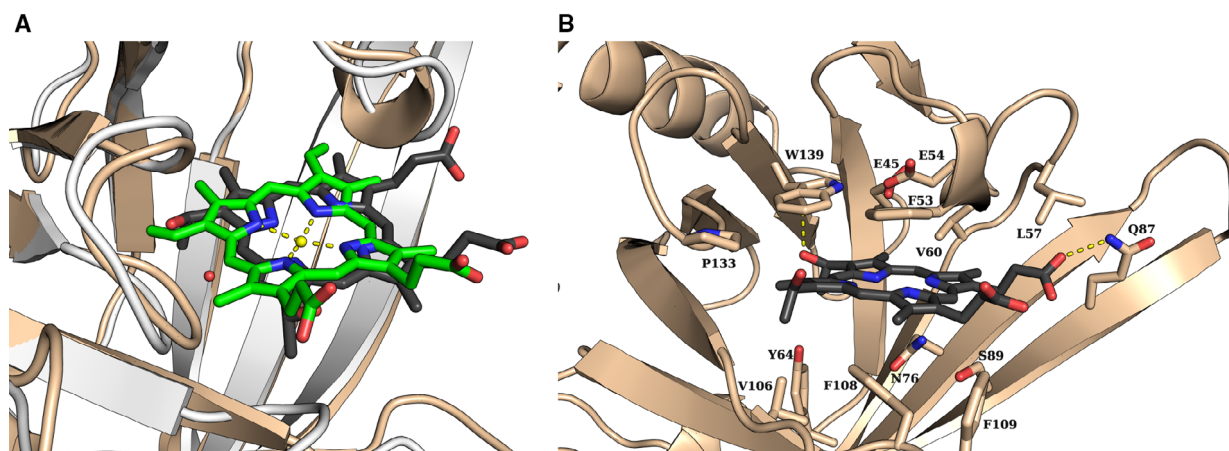
Of the three tested compounds, only HP showed significant perturbations in absorbance spectra upon adding the protein. To characterize the HP binding, we repeated an absorbance titration using a constant concentration of the full-length wtB1c protein expressed in minimal medium supplemented with different HP concentrations (Fig. 4C) analogous to the previously performed hemin titration. Based on similar size and overall shape of the heme and HP molecules, we

hypothesized the same 1 : 1 protein : ligand ratio of the wtB1c : HP interaction. Data analysis suggested a  $K_d$  in the low nM range, but the sensitivity limitations of the method did not allow for accurate  $K_d$  determination. For more precise  $K_d$  determination, we exploited the tryptophan fluorescence quenching assay that was previously used for potential B1c ligand screening [14]. The analysis yielded a  $K_d$  of  $0.14 \pm 0.1$  nM (Fig. 4D).

HP differs from heme in two key ways: (a) HP contains no coordinated iron, and (b) the vinyl groups in heme are hydrated to alcohols in HP. Nevertheless, they are structurally similar porphyrins, and it is not immediately clear why HP has a substantially higher binding affinity to wtB1c than heme. To further investigate this observation, we performed *in silico* analysis. First, we used Rosetta to construct a model of the wtB1c:HP complex such that HP was oriented to its common substructure with heme. We then performed three independent 4.0  $\mu$ s MD simulations of wtB1c:HP to evaluate stability of HP in the binding pocket.

To our surprise, HP consistently departed from its initial heme-like binding mode to rotate  $\sim 90^\circ$ , orienting both of the propionate groups toward the solvent front. To identify the most frequently occupied HP-binding modes, we generated HMM from the simulations of the wtB1c:HP complex. Our HMM identified five HP-binding modes: one heme-like binding mode (state 1), three variants of the 90-degree rotated binding mode (states 0, 3, and 4), and one intermediate binding mode (state 2) (Fig. S7). The transition to states 0, 3, and 4 occurred between 0.8 and 2.0  $\mu$ s (Fig. S8). Based on the relative free energy profiles and stationary distributions of each state in the HMM (Fig. S7A) as well as MM-PBSA-IE binding free energy calculations of structures resampled from each HMM state (Table S2), we chose wtB1c:HP state 4 model for detailed comparison with wtB1c-split:heme crystal structure.

The porphyrin rings of heme and HP overlap, which results in involvement of generally the same set of amino acids into HP ‘face’ and HP edge interactions (Fig. 6) as in the wtB1c-split:heme complex (Fig. S5D). Interestingly, one of the HP alcohols overlaps with the heme vinyl situated next to the buried structural water molecule (Fig. 6A), potentially reducing occupancy of water in this site or providing the water molecule with additional hydrogen bonding interactions. The other HP alcohol is oriented in the same direction as one of the heme propionate groups near the ‘split’ site. Both HP propionate groups are directed outward at the solvent front. These observations suggest a possible structural explanation for the observed higher binding



**Fig. 6.** Comparison of the converged pose of wtBlc:HP MD simulations with the crystal structure of wtBlc-split:heme. (A) Overlay of a randomly resampled wtBlc:HP pose from HMM state 4 (dark gray sticks and wheat cartoon) and the co-crystallized wtBlc-split:heme (green sticks and white cartoon; ferric iron is depicted as a yellow sphere with yellow dashed lines indicating coordination to porphyrin nitrogen atoms; a structural water molecule in the binding pocket is depicted as a red sphere). (B) Interactions between HP and wtBlc in HMM state 4. Yellow dashed lines indicate hydrogen bonds.

affinity of wtBlc to HP relative to heme: The wtBlc:HP complex may benefit from a reduced desolvation penalty.

HP, as well as other porphyrins and their derivatives, is well-known for its photosensitizing properties [52]. While performing titration experiments, we discovered that the presence of the wtBlc protein in solution makes HP incredibly sensitive to light. When the sample was kept in darkness, no changes in the spectra were observed during a 1-h period. However, visible color change of the solution and dramatic absorbance spectral changes happened after the sample was outside of the spectrophotometer chamber for as little as one minute (Fig. S9). Neither of these changes were seen in the absence of the protein in the similar time scale. Further experiments in a nitrogen glove box showed that the observed color change depends on the presence of oxygen. We speculate that wtBlc accelerates the degradation of HP.

## Conclusions and Future directions

Using LC-MS/MS and absorption spectroscopy, we confirmed copurification of the heterologously expressed wtBlc and wtBlc-split proteins with heme. This allowed us to assign the observed ‘heme-like’ density in the obtained wtBlc-split crystal structure as a heme molecule and to investigate wtBlc:heme interaction in more detail.

Based on the spectrophotometric hemin titration experiments and crystal structure analysis, we assessed the wtBlc hemin-binding affinity to be in the low  $\mu\text{M}$

range with 1 : 1 ligand-to-protein stoichiometry. While most of the observed protein:heme interactions are similar to those found in other heme-binding proteins, wtBlc-split does not coordinate the iron of the heme molecule. This observation prompted us to test whether wtBlc binds any other heme-like molecules. Interestingly, while the iron-free protoporphyrin IX molecule seems not to interact with the protein according to our absorption spectroscopy experiments, additional hydration of the two vinyl groups as seen in HP not only rescues the binding, but also improves the affinity by approximately three orders of magnitude.

Blc is a membrane-associated protein anchored to the outer membrane by a covalently bound N-terminal cysteine lipid, and it faces the periplasmic space [12,13]. Based on the obtained data, we speculate that lipocalin Blc might participate in the trans-periplasmic trafficking of heme molecules. While sequestering heme from highly competitive extracellular environments requires tight heme binding, further transportation inside a protected space might be performed by a protein with much lower affinity [53]. However, the fact that the wtBlc-split:heme complex has been purified from the unnatural cytoplasmic environment together with the observed high affinity for HP leaves the possibility that the main ligand of the Blc is yet another metal-containing or metal-free tetrapyrrole. There currently exists no evidence that the protein predominantly binds to only one ligand; therefore, Blc may target a broad family of tetrapyrroles, including heme.

The observed light sensitivity of the wtBlc:HP complex was out of scope of the current research and

therefore was not closely investigated. However, further study might result in creation of new oxygen sensing indicators, genetically encoded light-induced sources of reactive oxygen species, or tags for correlative light and electron microscopy.

## Acknowledgements

We are grateful to Manuel Castro and Chuck Sanders for providing aliquots of protoporphyrin IX, HP, and verteporfin for the initial screening, to Xuan Zhang for help with protein expression and purification, and to Shannon Smith for help with editing the text of the manuscript. This work was supported by NIH NIGMS grant R01 GM099842. This research used resources of the Biomolecular Crystallography Facility in the Vanderbilt University Center for Structural Biology and the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). BPB is supported through the NIH by a Ruth L. Kirschstein NRSA fellowship (F30DK118774). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## Author contributions

NGB, MWC, WNB, and JM planned the studies. NGB, MWC, WNB, and BPB conducted experiments. NGB, MWC, WNB, BPB, EPS, and JM interpreted the studies. NGB wrote the first draft of the paper. JM and EPS obtained funding. All authors reviewed, edited, and approved the paper.

## References

- 1 Poulos TL (2007) The Janus nature of heme. *Nat Prod Rep* **24**, 504–510.
- 2 Sevvana M, Ahnström J, Egerer-Sieber C, Lange HA, Dahlbäck B and Muller YA (2009) Serendipitous fatty acid binding reveals the structural determinants for ligand recognition in apolipoprotein M. *J Mol Biol* **393**, 920–936.
- 3 Breustedt DA, Schönfeld DL and Skerra A (2006) Comparative ligand-binding analysis of ten human lipocalins. *Biochim Biophys Acta* **1764**, 161–173.
- 4 Andersen JF, Champagne DE, Weichsel A, Ribeiro JM, Balfour CA, Dress V and Montfort WR (1997) Nitric oxide binding and crystallization of recombinant nitrophorin I, a nitric oxide transport protein from the blood-sucking bug *Rhodnius prolixus*. *Biochemistry* **36**, 4423–4428.
- 5 Larsson J, Allhorn M and Kerström B (2004) The lipocalin alpha(1)-microglobulin binds heme in different species. *Arch Biochem Biophys* **432**, 196–204.
- 6 Ghosh S, Yu C-L, Ferraro DJ, Sudha S, Pal SK, Schaefer WF, Gibson DT and Ramaswamy S (2016) Blue protein with red fluorescence. *Proc Natl Acad Sci USA* **113**, 11513–11518.
- 7 Kumagai A, Ando R, Miyatake H, Greimel P, Kobayashi T, Hirabayashi Y, Shimogori T and Miyawaki A (2013) A bilirubin-inducible fluorescent protein from eel muscle. *Cell* **153**, 1602–1611.
- 8 Schultz LW, Liu L, Cegielski M and Hastings JW (2005) Crystal structure of a pH-regulated luciferase catalyzing the bioluminescent oxidation of an open tetrapyrrole. *Proc Natl Acad Sci USA* **102**, 1378–1383.
- 9 Montfort WR, Weichsel A and Andersen JF (2000) Nitrophorins and related antihemostatic lipocalins from *Rhodnius prolixus* and other blood-sucking arthropods. *Biochim Biophys Acta* **1482**, 110–118.
- 10 Olsson MG, Allhorn M, Bülow L, Hansson SR, Ley D, Olsson ML, Schmidtchen A and Akerström B (2012) Pathological conditions involving extracellular hemoglobin: molecular mechanisms, clinical significance, and novel therapeutic opportunities for  $\alpha(1)$ -microglobulin. *Antioxid Redox Signal* **17**, 813–846.
- 11 De Simone G, Ascenzi P, di Masi A and Polticelli F (2017) Nitrophorins and nitrobindins: structure and function. *Biomol Concepts* **8**, 105–118.
- 12 Bishop RE, Penfold SS, Frost LS, Hölte JV and Weiner JH (1995) Stationary phase expression of a novel *Escherichia coli* outer membrane lipoprotein and its relationship with mammalian apolipoprotein D. Implications for the origin of lipocalins. *J Biol Chem* **270**, 23097–23103.
- 13 Bishop RE (2000) The bacterial lipocalins. *Biochim Biophys Acta* **1482**, 73–83.
- 14 Campanacci V, Bishop RE, Blangy S, Tegoni M and Cambillau C (2006) The membrane bound bacterial lipocalin B1c is a functional dimer with binding preference for lysophospholipids. *FEBS Lett* **580**, 4877–4883.
- 15 Campanacci V, Nurizzo D, Spinelli S, Valencia C, Tegoni M and Cambillau C (2004) The crystal structure of the *Escherichia coli* lipocalin B1c suggests a possible role in phospholipid binding. *FEBS Lett* **562**, 183–188.
- 16 Schiefner A, Chatwell L, Breustedt DA and Skerra A (2010) Structural and biochemical analyses reveal a monomeric state of the bacterial lipocalin B1c. *Acta Crystallogr D Biol Crystallogr* **66** (Pt 12), 1308–1315.
- 17 Zou D, Pei J, Lan J, Sang H, Chen H, Yuan H, Wu D, Zhang Y, Wang Y, Wang D *et al.* (2020) A SNP of



- bacterial blc disturbs gut lysophospholipid homeostasis and induces inflammation through epithelial barrier disruption. *EBioMedicine* **52**, 102652.
- 18 Blakeley BD, Chapman AM and McNaughton BR (2012) Split-superpositive GFP reassembly is a fast, efficient, and robust method for detecting protein-protein interactions *in vivo*. *Mol Biosyst* **8**, 2036–2040.
  - 19 Bozhanova NG, Baranov MS, Klementieva NV, Sarkisyan KS, Gavrikov AS, Yampolsky IV, Zagaynova EV, Lukyanov SA, Lukyanov KA and Mishin AS (2017) Protein labeling for live cell fluorescence microscopy with a highly photostable renewable signal. *Chem Sci* **8**, 7138–7142.
  - 20 Bozhanova NG, Gavrikov AS, Mishin AS and Meiler J (2020) DiB-splits: nature-guided design of a novel fluorescent labeling split system. *Sci Rep* **10**, 11049.
  - 21 Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254.
  - 22 Winter G (2010) xia2: an expert system for macromolecular crystallography data reduction. *J Appl Crystallogr* **43**, 186–190.
  - 23 Vagin A and Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J Appl Crystallogr* **30**, 1022–1025.
  - 24 Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60** (Pt 12 Pt 1), 2126–2132.
  - 25 Murshudov GN, Vagin AA and Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53** (Pt 3), 240–255.
  - 26 Brown BP, Mendenhall J and Meiler J (2019) BCL:: MolAlign: three-dimensional small molecule alignment for pharmacophore mapping. *J Chem Inf Model* **59**, 689–701.
  - 27 Kothiwale S, Mendenhall JL and Meiler J (2015) BCL:: Conf:small molecule conformational sampling using a knowledge based rotamer library. *J Cheminform* **7**, 47.
  - 28 Anandakrishnan R, Aguilar B and Onufriev AV. H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res* 2012;**40**: W537–W541.
  - 29 Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji *Het al.* (2016) Gaussian 16 Rev. C.01, Wallingford, CT.
  - 30 Case DA, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro VWD, Darden TA, Duke RE, Ghoreishi D, Gilson MK *et al.* (2018) AMBER 2018.
  - 31 Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE and Simmerling C (2015) ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J Chem Theory Comput* **11**, 3696–3713.
  - 32 Wang J, Wolf RM, Caldwell JW, Kollman PA and Case DA (2004) Development and testing of a general amber force field. *J Comput Chem* **25**, 1157–1174.
  - 33 Joung IS and Cheatham TE (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* **112**, 9020–9041.
  - 34 Vega C, Abascal JLF and Nezbeda I (2006) Vapor-liquid equilibria from the triple point up to the critical point for the new generation of TIP4P-like models: TIP4P/Ew, TIP4P/2005, and TIP4P/ice. *J Chem Phys* **125**, 34503.
  - 35 Hopkins CW, Le Grand S, Walker RC and Roitberg AE (2015) Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput* **11**, 1864–1874.
  - 36 Ryckaert J-P, Ciccotti G and Berendsen HJC (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J Comput Phys* **23**, 327–341.
  - 37 Miyamoto S and Kollman PA (1992) Settle: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J Comput Chem* **13**, 952–962.
  - 38 Roe DR and Cheatham TE (2013) PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput* **9**, 3084–3095.
  - 39 Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H and Noé F (2015) PyEMMA 2: a software package for estimation, validation, and analysis of markov models. *J Chem Theory Comput* **11**, 5525–5542.
  - 40 Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H and Roitberg AE (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J Chem Theory Comput* **8**, 3314–3321.
  - 41 Duan L, Liu X and Zhang JZH (2016) Interaction entropy: a new paradigm for highly efficient and reliable computation of protein-ligand binding free energy. *J Am Chem Soc* **138**, 5722–5728.
  - 42 Smith DW and Williams RJP (1970). In *The Spectra of Ferric Haems and Haemoproteins*, pp 1–45. Springer, Berlin, Heidelberg.
  - 43 Karnaukhova E, Rutardottir S, Rajabi M, Wester Rosenlöf L, Alayash AI and Åkerström B (2014) Characterization of heme binding to recombinant  $\alpha$ 1-microglobulin. *Front Physiol* **5**, 465.
  - 44 Hofbauer S, Hagmüller A, Schaffner I, Mlynek G, Krutzler M, Stadlmayr G, Pirker KF, Obinger C, Daims H, Djinić-Carugo K *et al.* (2015) Structure and heme-binding properties of HemQ (chlorite dismutase-like protein) from *Listeria monocytogenes*. *Arch Biochem Biophys* **574**, 36–48.

- 45 Leung GCH, Fung SSP, Dovey NRB, Raven EL and Hudson AJ (2019) Precise determination of heme binding affinity in proteins. *Anal Biochem* **572**, 45–51.
- 46 Hargrove MS, Barrick D and Olson JS (1996) The association rate constant for heme binding to globin is independent of protein structure. *Biochemistry* **35**, 11293–11299.
- 47 Bianchetti CM, Blouin GC, Bitto E, Olson JS and Phillips GN Jr (2010) The structure and NO binding properties of the nitrophorin-like heme-binding protein from *Arabidopsis thaliana* gene locus At1g79260.1. *Proteins* **78**, 917–931.
- 48 Deniau C, Gilli R, Izadi-Pruneyre N, Letoffe S, Delepierre M, Wandersman C, Briand C and Lecroisey A (2003) Thermodynamics of heme binding to the HasA(SM) hemophore: effect of mutations at three key residues for heme uptake. *Biochemistry* **42**, 10627–10633.
- 49 Knipp M, Ogata H, Soavi G, Cerullo G, Allegri A, Abbruzzetti S, Bruno S, Viappiani C, Bidon-Chanal A and Luque FJ (2015) Structure and dynamics of the membrane attaching nitric oxide transporter nitrophorin 7. *Fl000Res* **4**, 45.
- 50 Krissinel E and Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–797.
- 51 Schneider S, Marles-Wright J, Sharp KH and Paoli M (2007) Diversity and conservation of interactions for binding heme in b-type heme proteins. *Nat Prod Rep* **24**, 621–630.
- 52 Bonnett R and Martínez G (2001) Photobleaching of sensitizers used in photodynamic therapy. *Tetrahedron* **57**, 9513–9547.

- 53 Benson DR and Rivera M (2013) Heme uptake and metabolism in bacteria. *Met Ions Life Sci* **12**, 279–332.

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Table S1.** Data Collection and Refinement Statistics

**Table S2.** Binding free energy estimates of wtBlc:HP in different HMM metastable states.

**Fig. S1.** Absorbance of solutions with different concentrations of free hemin at 411 nm (A) or free HP at 410 nm (B), and fitted curves (black dashed lines).

**Fig. S2.** Schematic representation of the design of the leucine zippers-containing (“split-Zip”) and leucine zippers-free (“split”) DiB-split proteins constructs.

**Fig. S3.** Comparison of different proposed wtBlc ligands’ binding sites.

**Fig. S4.** Heme coordination in the binding pocket of the wtBlc-split protein, side view.

**Fig. S5.** Time-dependent rmsd of heme to its crystallographically determined pose.

**Fig. S6.** Structures of (A) Heme B, (B) protoporphyrin IX, (C) hematoporphyrin, and (D) verteporfin

**Fig. S7.** Metastable states of wtBlc:HP distinguished by a hidden Markov state model (HMM).

**Fig. S8.** HP transitions from the initial heme-like binding mode to an alternative stable pose.

**Fig. S9.** Light-induced changes of the wtBlc:HP solution.