Structural modeling of a novel membrane-bound globin-coupled sensor in Geobacter sulfurreducens

Dietmar Hammerschmid, Francesca Germani, Salvador Drusin, Charline Fagnen, Claudio Schuster, David Hoogewijs, Marcelo Marti, Catherine Venien-Bryan, Luc Moens, Sabine van Doorslaer, et al.

Computational and Structural Biotechnology Journal 19 (2021) 1874–1888

COMPUTATIONAL AND STRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Structural modeling of a novel membrane-bound globin-coupled sensor in *Geobacter sulfurreducens*

Dietmar Hammerschmid [a,b,2], Francesca Germani [a,3], Salvador I. Drusin [c,1], Charline Fagnen [d,1], Claudio D. Schuster [c,1], David Hoogewijs [e], Marcelo A. Marti [c], Catherine Venien-Bryan [d], Luc Moens [a], Sabine Van Doorslaer [f,*], Frank Sobott [b,g,h,*], Sylvia Dewilde [a]

[a] *Proteinchemistry, Proteomics and Epigenetic Signalling, Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium*
[b] *Biomolecular & Analytical Mass Spectrometry, Department of Chemistry, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium*
[c] *Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires (FCEyN-UBA) e Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN) CONICET, Pabellòn 2 de Ciudad Universitaria, Ciudad de Buenos Aires C1428EHA, Argentina*
[d] *Sorbonne Université, UMR 7590, CNRS, Muséum National d'Histoire Naturelle, Institut de Minéralogie, Physique des Matériaux et Cosmochimie, IMPMC, 75005 Paris, France*
[e] *Section of Medicine, Department of Endocrinology, Metabolism and Cardiovascular System, University of Fribourg, Switzerland*
[f] *Biophysics and Biomedical Physics, Department of Chemistry, University of Antwerp, Universiteitsplein 1, 2610 Wilrijk, Belgium*
[g] *Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom*
[h] *School of Molecular and Cellular Biology, University of Leeds, LS2 9JT, United Kingdom*

ARTICLE INFO

Article history:
Received 19 December 2020
Received in revised form 24 March 2021
Accepted 24 March 2021
Available online 26 March 2021

Dedication: We would like to dedicate this work to the memory of our friend and colleague Prof. Sylvia Dewilde who passed away at the age of 47 during the writing of this paper. She was an excellent scientist. We will miss her all.

Keywords:
Globin-coupled sensor
Geobacter sulfurreducens
Transmembrane domain
Transmembrane-coupled globins

ABSTRACT

Globin-coupled sensors (GCS) usually consist of three domains: a sensor/globin, a linker, and a transmitter domain. The globin domain (GD), activated by ligand binding and/or redox change, induces an intramolecular signal transduction resulting in a response of the transmitter domain. Depending on the nature of the transmitter domain, GCSs can have different activities and functions, including adenylate and di-guanylate cyclase, histidine kinase activity, aerotaxis and/or oxygen sensing function.

The gram-negative delta-proteobacterium *Geobacter sulfurreducens* expresses a protein with a GD covalently linked to a four transmembrane domain, classified, by sequence similarity, as GCS (*Gs*GCS). While its GD is fully characterized, not so its transmembrane domain, which is rarely found in the globin superfamily.

In the present work, *Gs*GCS was characterized spectroscopically and by native ion mobility-mass spectrometry in combination with cryo-electron microscopy. Although lacking high resolution, the oligomeric state and the electron density map were valuable for further rational modeling of the full-length *Gs*GCS structure. This model demonstrates that *Gs*GCS forms a transmembrane domain-driven tetramer with minimal contact between the GDs and with the heme groups oriented outward. This organization makes an intramolecular signal transduction less likely.

Our results, including the auto-oxidation rate and redox potential, suggest a potential role for *Gs*GCS as redox sensor or in a membrane-bound e^-/H^+ transfer. As such, *Gs*GCS might act as a player in connecting energy production to the oxidation of organic compounds and metal reduction.

Abbreviations: CCS, collision cross section; CMC, critical micelle concentration; CIU, collision-induced unfolding; CV, cyclic voltammetry; DDM, n-dodecyl-β-D-maltoside; DPV, differential pulse voltammetry; FMRF, H-Phe-Met-Arg-Phe-NH2 neuropeptide; Gb, globin; GCS, globin-coupled sensor; GD, globin domain; GGDEF, Gly-Gly-Asp-Glu-Phe motive; IM-MS, ion mobility-mass spectrometry; NH_4OAc, ammonium acetate; OG, n-octyl-β-D-glucopyranoside; PDE, phosphodiesterase; PsiE, phosphate-starvation-inducible E; RR, resonance Raman; SCE, saturated calomel electrode; SHE, standard hydrogen electrode; TD, Transmitter domain; TmD, Transmembrane domain; *Af*GcHK, *Anaeromyxobacter* sp. Fw109-5 GcHK; *As*FRMF, *Ascaris suum* FRMF-amide receptor; *Av*GReg, *Azotobacter vinilandii* Greg; *Bp*GReg, *Bordetella pertussis* Greg; *Bs*HemAT, *Bacillus subtilis* HemAT; *Ce*GLB6, *Caenorhabditis elegans* globin 6; *Ce*GLB26, *Caenorhabditis elegans* globin 26; *Ce*GLB33, *Caenorhabditis elegans* globin 33; *Ec*DosC, *Escherichia coli* Dos with DGC activity; *Gint*Hb, hemoglobin from *Gasterophilus intestinalis*; *Gs*GCS, *Geobacter sulfurreducens* GCS; *Gs*GCS^{162}, GD of *Gs*GCS; *Lm*HemAC, *Leishmania major* HemAC; *Ma*Pgb, *Methanosarcina acetivorans* protoglobin; mNgb, mouse neuroglobin; *Mt*TrHbO, *Mycobacterium tuberculosis* truncated hemoglobin O; *Pcc*GCS, *Pectobacterium carotivorum* GCS; *Pc*Mb, *Physether catodon* myoglobin; *Sa*ktrHb, *Streptomyces avermitilis* truncated hemoglobin-antibiotic monooxygenase; SwMb, myoglobin from sperm whale.

* Corresponding authors.
E-mail addresses: sabine.vandoorslaer@uantwerpen.be (S. Van Doorslaer), f.sobott@leeds.ac.uk (F. Sobott).
[1] Contributed equally to this work.
[2] Present address: Department of Chemistry, King's College London, Britannia House, 7 Trinity Street, London SE1 1DB, United Kingdom.
[3] Present address: Akott Evolution, S.r.l., Via Rosso di San Secondo 1, 20134 Milan, Italy

https://doi.org/10.1016/j.csbj.2021.03.031

Database searches indicate that GDs linked to a four or seven helices transmembrane domain occur more frequently than expected.

1. Introduction

Globin-coupled sensors (GCS) are multi-domain hemo-proteins capable of sensing and responding to environmental stimuli [1–4]. From its first introduction for heme-based aerotactic transducer (HemAT) in 2001 [5], the number of identified GCS increased steadily in Archaea and Bacteria [2,3]. Currently, GCS are understood as a subgroup of heme-based sensors [2] akin to a two-component regulatory system [6] consisting of (i) a sensor globin domain (GD), (ii) a middle domain (MD), and (iii) an accompanying transmitter domain (TD) (see detailed information on GCS in supplementary information, e.g. Figs. S1 and S2 as well as Tables S1 and S2). The GD is an all α-helical bundle shaped in a specific manner, i.e. the globin fold, capable of noncovalent heme incorporation shielding its active site, the heme iron atom, from the solvent and as such determining its reactivity. The protein moiety itself consists of 6–8 α-helices (A–H) folded either as a 3-over-3 or 2-over-2 helical sandwich (A-E-F/B-G-H or B-G/E-H) [7,8]. In GCS however the GD is slightly longer and exhibits some other specific adaptations relative to the classic globin fold: (i) the presence of a pre-A (Z-)helix, (ii) the absence of the D-helix, and (iii) an extended CE- and FG-region [9]. Moreover, additional to the proximal histidine (HisF8), which is essential to all globins (Gbs), a Phe and Tyr, at the end of the B-helix (B9/B10) have been identified as crucial in the GD of GCS (Fig. S2). These adaptations along with phylogenetic analysis define GCS as S-type members of the globin superfamily which branches into three lineages, i.e. the myoglobin-like (M-family), the sensor-like (S-family) and the truncated hemoglobin family (T-family), belonging to two structural classes (3/3 and 2/2 fold) [10,11].

The GD as a key component in GCS is well-known for its capability of binding diatomic gases (O_2, CO, NO) and/or allowing changes of the redox state of the heme iron atom. These ligand interactions are normally accompanied by a series of subtle conformational changes in both heme group and the surrounding protein moiety. This conformational flexibility makes the GD an ideal sensor for incoming environmental signals which can, based on MD and TD, potentially be converted via intramolecular signal transduction, provoking a response. To that end, GD and TD are connected via the MD which is variable in length and structure. In spite of its importance in signal transduction, current information is limited to the MD of *Ec*DosC [12]. The TD displays a great diversity in structure, i.e. domain type and organization, as well as function. Depending on its activity, four functional groups are classified, i.e. (i) aerotactic activity having a MCP-domain (type: *Bs*HemAT), (ii) gene regulation via protein–protein interaction with histidine kinase activity (type: *Af*GcHK), second-messenger production (type: *Ec*DosC) or degradation (PDE) [13], (iii) enzymatic activity (type: *Sa*-ktrHb), and (iv) unknown function [2,13,14].

Despite some progress over the last decade, our knowledge of GCS (Table S1) is predominantly limited to individual domains (Table S2), while a more comprehensive view of the structure and the intramolecular signal transduction of the full-length protein is lacking, unsurprisingly, as there is no structure of a full-length GCS available today. Crystal structures are largely limited to the globin domain only [12,15–17] or to each domain individually, e.g. for (*Ec*DosC) [12]. Therefore, sensors with diguanylate cyclase (DGC) and histidine kinase (HK) activity represent the best-characterized GCS.

*Ec*DosC consist of an S-type GD that connects via a 5 α-helix MD to a standard cyclase domain [18,19] (TD) with a catalytic GGDEF-motif. Functional *Ec*DosC reflects a highly symmetrical homodimer which gets distorted upon ligand binding or redox change of the heme iron atom. The altered symmetry, along with other conformational rearrangements in CE- and EF-regions of the GD, causes a rearrangement and/or change in flexibility of the C-terminal B-helix of the MD involving a π-helix segment. Consequently, the TD of *Ec*DosC gets enabled to relocate its active side, the Gly^{374}-Gly-Asp-Glu-Phe^{378} (GGDEF) motif, catalyzing di-cyclic-GMP production [12]. *Af*GcHK consists also of an S-type GD, structurally similar to *Gs*GCS, and a histidine kinase domain (TD), both connected via a 15 residue long flexible loop (MD). Both GD and TD form a dimer in the active state stabilized by Tyr15 and a four-helix bundle formed by two α-helices (H8, H9) of the dimerization subdomain, respectively [20,21]. Similar to *Ec*DosC, *Af*GcHK is inactive in its unligated Fe(II) form and activation requires ligand binding (Fe(III)-OH, Fe(III)-CN, Fe(II)-O_2 and Fe(II)-CO), which results in an auto-phosphorylation of His183 in the H8 helix of the TD. Ligand dissociation or heme reduction, i.e. inactivation, triggers intramolecular signal transduction which is proposed to (i) cause a conformational change in the GD, resulting in widening and splitting of the dimeric interface, (ii) signal transfer through the MD to the kinase domain, and consequently (iii) separation of the ATP-binding site from His183, hindering auto-phosphorylation of the latter [21–23]. Although *Ec*DosC and *Af*GcHK are substantially different in structure and function, their intramolecular signal transduction displays remarkable similarities, e.g. (i) a conformational shift of the GD upon activation which leads to (ii) the TD activation requiring a dimeric structure of both the GD and the overall GCS molecule.

More intriguing, because still categorized with unknown function, are GCS/receptors consisting of a transmembrane-coupled GD. Two such proteins have been identified in the nematodes *Caenorhabditis elegans* and *Ascaris suum*. Although having a eukaryotic origin, these proteins coincide with the domain architecture of GCS. Both contain an N-terminally located seven transmembrane domain (TmD) to a Phe-Met-Arg-Phe (FMRF)-amide receptor and a C-terminal GD which represents either a classic or shortened globin fold for *C. elegans* and *A. suum*, respectively [24]. Another GCS containing a TmD, and the subject of this study, has been identified in the genome of the bacterium *Geobacter sulfurreducens* (*Gs*).

The facultative anaerobic ð-proteobacterium *Geobacter sulfurreducens* expresses a high proportion of proteins with environmental sensing characteristics including a 300 residue long GCS (*Gs*GCS, UniProtID: Q747F6) [25–27]. This has shown higher expression rates under Fe(III)-reducing conditions, suggesting a possible role in Fe(III) reduction [28]. The GD of *Gs*GCS (*Gs*GCS^{162}) has been characterized in detail using crystallization and spectroscopic approaches [16,29]. *Gs*GCS^{162} resembles the globin structures of *Af*GcHK, *Bs*HemAT and protoglobin (single domain S-type globin) from *Methanosarcina acetivorans,* i.e. it has an additional Z-helix, no D-helix, and an extended CE- and FG-region overall shaped in a symmetrical homo-dimeric quaternary structure (Fig. S2). The heme iron displays an unprecedented HisF8-Fe(III)-HisE11 hexa-

coordination [16] causing a rotation of the E-helix and further adaptations of the distal pocket [9,15,30].

In this study, we targeted full-length *Gs*GCS using native ion mobility-mass spectrometry (IM-MS) and cryo-electron microscopy (EM). Deducing information about the oligomeric status of the protein and the obtained electron density map by cryo-EM encouraged us to also build a structural model. Next to the EM density of the full-length construct, the X-ray structure of the globin domain [16], as well as a structural homolog of the TM domain based on sequence analysis, provided additional input for our computational modeling approach. A low-resolution structural model for full-length *Gs*GCS is now for the first time presented.

2. Materials and methods

2.1. Cloning, expression, and purification of $GsGCS^{162}$

The cloning, expression, and purification of recombinant $GsGCS^{162}$ (the first 162 amino acids of *Gs*GCS represent the globin domain only) was performed as described before [16].

2.2. Cloning, expression, and purification of full-length GsGCS

The cDNA of *Gs*GCS (codons 1–300, NCBI reference protein sequence WP_010943923) was amplified by PCR using 5′-CCGCT CGAGATGCTGACCATGCAGGAAATAAAGG-3′ and 5′-GGGTACCCTAGTGGGGAGTGTAGG-3′ as forward and reverse primer, respectively. The reaction conditions applied for PCR are described elsewhere [24]. The construct was cloned into a pBAD-a vector using XhoI and HindIII restriction sites. The expression vector was transformed into *E. coli* TOP10 cells (Invitrogen) which equips the target protein *Gs*GCS with an N-terminal His-tag (not removed after protein purification). This construct has a theoretical molecular mass of 39,198 Da including the noncovalent prosthetic heme group. Protein expression was performed as previously described elsewhere [31,32]. Briefly, cells were grown overnight at 220 rpm at 37 °C in 6 mL L-Broth (10 g tryptone, 5 g yeast extract and 0.5 g NaCl per L) containing 50 μg/mL ampicillin. The grown culture was transferred into 250 mL TB medium (1.2% bactotryptone, 2.4% yeast extract, 0.4% glycerol, 17 mM KH_2PO_4 and 72 mM K_2HPO_4) containing 50 μg/mL ampicillin and 1 mM δ-aminolevulinic acid and was incubated at 160 rpm at 25 °C. Overexpression was induced at OD_{600} = 0.8 by the addition with 0.2% L-arabinose and the culture was incubated overnight at 160 rpm at 25 °C. Cells were harvested at 3220 g (20 min) and were resuspended in 12 mL lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, and 1 mg/mL lysozyme). The resuspended cells were exposed to three freeze–thaw cycles and were sonicated (1.5 min, 70 Hz, and 3 s pulses at 4 °C) until thoroughly lysed. The cell lysate was centrifuged for 30 min at 10,700 g at 4 °C to separate the cytosolic material form the fraction which contains the membrane protein cell debris. The membrane fraction (pellet) was dissolved in an extraction buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 20 mM imidazole, 2% w/v n-octyl-β-D-glucopyranoside [OG]) and the His-tagged target protein was extracted under gentle agitation overnight at 4 °C.

The solubilized material was centrifuged for 30 min at 10,700 g at 4 °C and the supernatant was loaded onto a self-packed Ni-IDA column pre-equilibrated with buffer A (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 20 mM imidazole, 1% w/v OG (2× critical micelle concentration [CMC]). *Gs*GCS was washed with 10 column volumes buffer A and was eluted with buffer B (same as buffer A but with 200 mM imidazole). Subsequently, *Gs*GCS was buffer exchanged to a storage buffer (50 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1% w/v OG) using a PD-10 desalting column (GE Healthcare). Fractions containing *Gs*GCS were collected and concentrated using an Amicon® 100 K centrifugal filter. The concentration of *Gs*GCS was determined by optical absorbance measurement in the UV/Vis region using the Soret band (411 nm), originating from the conjugated double bond in the heme group, and the determined molar extinction coefficient (ε = 127,000 M^{-1} cm^{-1}, see Section 2.3). The purity of the protein was checked by SDS-PAGE analysis.

2.3. Determination of the molar extinction coefficient

The molar extinction coefficient ε was determined for full-length GsGCS using a Heme Assay Kit (cat. Nr.: MAK316, Sigma-Aldrich) to be 127,000 M^{-1} cm^{-1}.

2.4. Preparation of the ferrous deoxy and CO-bound form

A detergent-free storage buffer (50 mM Tris-HCl, pH 7.5, 50 mM NaCl) was flushed with N_2 for 15 min without exposure to the air. OG (1% w/v) and a saturated sodium dithionite solution (1% v/v) were added to the N_2-flushed buffer. The ferrous deoxy form was obtained by adding (1:10) the protein to this N_2-flushed buffer.

The CO-bound from was obtained in the same way but the buffer was flushed with CO instead of N_2.

2.5. UV/Vis spectroscopy

Optical absorption measurements were performed on a Varian Cary 5E UV–Vis-NIR spectrophotometer (Agilent, California, USA). Spectra were acquired in the range from 375 to 700 nm for the ferric (Fe(III), as-purified), ferrous deoxy (Fe(II)), and CO-bound (Fe(II)) form at a concentration of ~10 μM.

2.6. Resonance Raman spectroscopy

Resonance Raman (RR) spectra were acquired in low-dispersion mode on an 80-cm Dilor XY-800 Raman spectrometer (HORIBA, Kyoto, Japan) equipped with a liquid N_2-cooled CCD detector (split width = 200 μm). The excitation source was a Kr^+ laser (Spectra-Physics BeamLok 2060) operating 413.1 nm. Laser powers of 1 and 100 mW were used. The protein solution had a final concentration of ~40 μM and was stirred at 500 rpm to avoid local heating and photochemical decomposition. Six spectra (120–180 s recording time) were acquired and averaged after the removal of cosmic ray spikes using an in-house developed program.

2.7. Native ion mobility-mass spectrometry

$GsGCS^{162}$: The protein (~10 μM) was buffer-exchanged into a volatile ammonium acetate (100 mM NH_4OAc, pH 6.8) solution using P-6 micro Bio-Spin columns (Bio-Rad). A few microliters of the buffer-exchanged sample were loaded into in-house produced gold-coated borosilicate glass capillaries for infusion into a Synapt G2 HDMS (Waters, Wilmslow, UK) mass spectrometer [33]. Native IM-MS experiments were performed by using nano-electrospray ionization (nanoESI) to generate ions of the protein or complex which were drawn into the vacuum of the mass spectrometer. Crucial instrument settings in Mobility mode to retain the native, quaternary structure were as follows: 1.4 kV capillary voltage, 40 V sampling cone, 1 V extractor cone, 10 V and 2 V collision energy in the trap and transfer collision cell, respectively, and 45 V trap DC bias. IM parameters were set to 600 m/s wave velocity, 40 V wave height, and a gas flow of 180 mL/min and 90 mL/min in the helium and IM cell, respectively. Pressures throughout the instrument were adjusted to 5.0 mbar for backing, 3.0 mbar (N_2) in the ion mobility cell, and 0.03 mbar (Ar) in the trap and transfer collision cell. Calibration of ion mobility arrival times was per-

formed by native measurements of the following standard proteins: β-lactoglobulin (monomer: +7/+8; dimer: +11/+12/+13), avidin (tetramer: +15/+16), bovine serum albumin (monomer: +14/+15/+16), and concanavalin A (tetramer: +19/+20/+21) to calculate the collision cross section (CCS) [34].

Measurements were also conducted under denaturing conditions provided by mixing the protein solution 1:1 with acetonitrile containing 0.5% formic acid. Instrument parameters in Mobility mode were set to 1.4 kV capillary voltage, 25 V sampling cone, 1 V extractor cone, 5 V and 2 V collision energy in the trap and transfer collision cell, respectively, and 45 V trap DC bias. Pressures were lowered to 2.8 mbar backing pressure and 0.0235 mbar in both collision cells. IM parameters were the same as under native conditions.

*Gs*GCS: IM-MS experiment preparations for *Gs*GCS were similar to *Gs*GCS^{162}. A sample of ~50 μM was buffer exchanged into 400 mM NH_4OAc containing 1% (approx. 1.5–2.0x CMC) OG. The micelle-embedded complex was transferred into the vacuum of the mass spectrometer where the protein was liberated by gentle gas collisions from its surrounding hydrophobic environment. For this purpose, the instrument has to be fine-tuned so that the activation is strong enough for protein release but gentle enough to retain the protein's quaternary structure. Crucial instrument settings for *Gs*GCS were as follows: 1.6 kV capillary voltage, 200 V sampling cone, 5 V extractor cone, 45 V trap DC bias, 150 V and 5 V collision energy in the trap and transfer collision cell, respectively. IM parameters were set to 800 m/s wave velocity, 40 V wave height, and a gas flow of 120 mL/min and 60 mL/min in the helium and IM cell, respectively. Pressures throughout the instrument were adjusted to 7.5 mbar for backing, 2.2 mbar (N_2) in the ion mobility cell, and 0.028 mbar (Ar) in the trap and transfer collision cell. Calibration of the traveling wave IM cell was performed using the following standard proteins: concanavalin A (tetramer: +19/+20/+21), alcohol dehydrogenase (tetramer: +23/+24/+25), pyruvate kinase (tetramer: +31/+32), and glutamate dehydrogenase (hexamer: +37/+38/+39) to calculate the CCS [34].

Deconvolution of all spectra was performed manually using MassLynx software (Waters, Wilmslow, UK)

2.8. Collision-Induced unfolding

Collision-induced unfolding (CIU) experiments were performed by stepwise increase of the collision energy voltage in the trap collision cell. *Gs*GCS^{162} was measured in the energy range from 10 to 50 V using 2.5 V increments. For micelle-embedded *Gs*GCS much higher collision energies were used, as *Gs*GCS is a much bigger complex compared to *Gs*GCS^{162} and a reasonable amount of the collision energy is consumed by the detergent release process. Therefore, *Gs*GCS was measured in the range from 100 to 200 V using 10 V increments.

2.9. Electron microscopy and image analysis

Negatively stained grids were prepared applying small aliquots (3.5 μL) of *Gs*GCS (0.025 mg/mL) to glow-discharged, carbon-coated grids and allowed to absorb for 60 s. The sample was then stained with 2% uranyl acetate. Vitrified *Gs*GCS was prepared on glow-discharged carbon quantifoil grids. Specimen were imaged at a nominal magnification of 30,000× with a 2 k × 2 k Gatan CCD camera (corresponding to a pixel size of 3.57 Å at the specimen level) in a JEOL 2100, LaB6 operating at 200 kV.

For negatively stained images, a total number of 7086 particles were selected manually from 66 micrographs with a box size of 36 pixels. Defocus was determined using CTFTILT. 2D and 3D classification and the refinement of the initial model were realized with EMAN2.2 [35]. For cryo conditions, a total number of 21,414 particles were automatically picked using Gautomatch software (the reference model in this process was the model previously build from negative staining images). A 2D classification was performed using Relion 2.0 [36,37]. Bad classes were discarded and 21,037 particles were kept for 3D classification using Relion 2.0. The most convincing 3D class was used for 3D refinement. For this purpose, the initial model (class1: 8709 particles) was filtered to 40 Å and a mask of 80 Å was used. When needed, C3 and C4 symmetry was applied, during the image analysis (2D and 3D classification as well as 3D refinement).

2.10. Database search for GsGCS-like sequences

The *Gs*GCS sequence (NCBI reference: WP_010943923) was used as query to perform a search in the PFAM database [38]. Two domains were indicated, i.e. an N-terminal globin domain (residues 2–156), and a C-terminal phosphate-starvation-inducible E (PsiE) domain (residues 218–285). The PsiE Hidden Markov Model (HMM) derived from PFAM was used for an HMMsearch [39] querying the UniProtKB [40] database. Sequences retrieved from this search were further used to score against the three globin HMMs from PFAM retaining all sequences harboring also a globin domain [38]. In total, 168 sequences with a globin-PsiE domain organization (Table S3) were identified, all of which belong to the S-family (Fig. S3). The taxonomical distribution and the multiple sequence alignment (MSA) are shown in Figs. S4 and S5/S6, respectively. The same PFAM database search followed by an HMMsearch with 7TmD and globin HMMs was repeated using the sequence of full-length *Ce*GLB33 (NCBI reference: NP_001300285.1) as query for the PFAM database. All MSAs were obtained with MAFFT [41] and plotted using Jalview (https://academic.oup.com/bioinformatics/article/25/9/1189/203460).

2.11. Modeling of GsGCS

A topological model of the *Gs*GCS PsiE domain was built through comparative modelling using the structure of lipid II flippase MurJ (pdb 5T77) as template. A multiple sequence analysis (MSA) was performed using MAFFT-G-INS [41], and the template was manually aligned to their respective domains by adjusting the corresponding secondary structure elements, i.e. the four TM helices (Fig. S6). Secondary structure prediction of the *Gs*GCS PsiE domain was performed using I-Tasser [42]. Both the globin domain (pdb 2W31 [16]) and the PsiE domain were fit independently into one of the monomers of the cryo-EM envelope using Chimera software [43]. A simulated map of each domain at the resolution of the cryo-EM map (15 Å) was first fit manually by comparing features of the domains and the envelope, and subsequently adjusted using the volume fit tool of Chimera to optimize the correlation. The complete monomer was then built through the combination of both domains using Modeller [44] to join the C-terminal helix from the globin domain with the N-terminal helix from the PsiE domain. Finally, the tetramer was formed using Chimera to replicate the monomer applying the C4 symmetry from the cryo-EM envelope.

2.12. Auto-oxidation rate

The auto-oxidation rate was determined at 20 °C on a Sx20 stopped-flow instrument (Applied Photophysics, Leatherhead, UK). For this purpose, the as-purified ferric *Gs*GCS was reduced to the ferrous form at 20 °C as explained above. Ferrous GsGCS (100 mM potassium phosphate, pH 7.5, 150 mM NaCl, 1% w/v OG, 1 mM Na-dithionite) was loaded into the stopped-flow syringe. The second syringe and the cuvette were filled with O_2 saturated buffers (100 mM potassium phosphate pH 7.5, 150 mM NaCl, 1,250 μM O_2) (20 °C). Sample and buffer were mixed 1:1 under

high pressure (8 bar) into the stopped-flow cuvette resulting in the following final buffer composition: 100 mM potassium phosphate, pH 7.5, 150 mM NaCl, 0.5% w/v OG, 0.5 mM Na-dithionite, 625 μM O_2. The auto-oxidation rate of *Gs*GCS was monitored by detecting the decrease in absorbance at 561 nm at 20 °C over a period of 9500 s. The absorption at 561 nm was plotted against the time and the reaction rate constant was calculated by fitting the recorded trace to a single exponential function.

2.13. Reduction potential

Measurements of the reduction potential were performed under nitrogen atmosphere at room temperature (22 ± 1 °C) using a μAutolab III interface controlled by Nova 1.10 software (Metrohm-Autolab BV, Netherlands). A saturated calomel electrode (SCE, Radiometer analytical, France) and a glassy carbon rod were used as reference and auxiliary electrodes, respectively. The working electrodes (gold, 1.6 mm diameter, BASi, USA) were polished with 1 and 0.05 μm alumina slurry, washed, and sonicated in ultrapure water. Electrodes were cleaned electrochemically by cyclic voltammetry (CV) in 0.5 M NaOH at the potential range from −0.35 to −1.35 V and in 0.5 M H_2SO_4 at the potential range from 0.2 to 1.5 V versus SCE until cyclic voltammograms were obtained repeatably. Following, electrodes were incubated overnight in 8 mM 6-mercapto-1-hexanol (97%, Sigma-Aldrich) solution. Prior to measurements, the electrodes were washed thoroughly with H_2O and dried in an air stream.

The working electrode was equilibrated for 30 min in 50 μL sample buffer (50 mM Tris-HCl, pH 7.5, 1% w/v OG) under nitrogen atmosphere and blank voltammograms were recorded. 10 μL of *Gs*GCS (120 μM final concentration) were added to the cell. Due to protein sensitivity, voltammograms were measured within 2 min after injection, which did not allow a complete removal of O_2. This also hampered the evaluation of the reversibility of the electrochemical reaction. Voltammograms were recorded applying CV (20 mV/s scan rate) and differential pulse (20 mV modulation amplitude, 50 ms modulation time, and 1 s interval time) voltammetry (DPV).

3. Results

3.1. Membrane-bound globin domains – A database query

Despite the large number of globin-related proteins, a search of the literature shows that proteins consisting of a GD covalently bound to a TmD are limited to just a few examples, i.e. *Gs*GCS and *Ce*GLB33. Therefore, in order to gain more information about the structure and function of these proteins, we used the full-length sequence of *Gs*GCS (NCBI reference: WP_010943923) as query for a PFAM database search. The result of the database query indicates the presence of two domains, as expected an N-terminal globin domain (residue 2 to 156), and a C-terminal transmembrane (4TmD) phosphate-starvation-inducible E (PsiE) domain (residues 218 to 285; NCBI reference sequence: NC_000913.3) [45]. The PsiE domain was originally identified in *E. coli* K-12 during whole genome sequencing (substr. MG1655) and its function is yet uncharacterized. Quite surprisingly, further database searches using both domain HMMs identified a total number of 168 sequences with a globin-PsiE domain organization (Table S3). A more detailed analysis regarding origin and globin type family shows that: (i) all these globin domains belong to the S-family and (ii) all proteins entirely belong to bacteria (mostly proteobacteria), consistent with the prokaryotic nature of GCS (Figs. S3 and S4).

To further characterize *Gs*GCS we performed MSA of both the Gb and 4TmD PsiE domain. Results (Figs. S6 and S5) show that 4TmD evidence high overall sequence conservation, particularly in helices 2 and 3, where several hydrophobic conserved residues are present (usually Leu or Phe), and indels are only present between helices 1 and 2. Several key residues, or small motifs, can also be identified, such as charged Asp/Glu residues flanking helices 1 and 2, and a conserved Arg at the end of helix 3, which possibly facilitate proper anchoring of the transmembrane helices to the membrane. Also, key conserved glycines mark the interhelix loops. Concerning the associated Gb domain, sequences show many conserved residues which are hallmarks of the globin fold, such as HisF8, PheCD1 and key interhelix Gly and Pro residues. Positions B9 and B10 show a preference for the typical residues observed in GCS, i.e. PheB9-TyrB10 motif (with some minor variations like Ile, Met, and Leu). Interestingly, distal E7 alternates between His, as in mammalian globins or Leu as in most GCS. Extension of the analysis to globin sequences of other GCS shows similar results (Fig. S2). Furthermore, the crystal structure obtained for the GD of *Gs*GCS is comparable with those from *Ec*DosC [12] and *Bp*GReg [17,46] which further supports the GCS-like character of the globin folded structure in *Gs*GCS [16].

3.2. Structural characterization of full-length GsGCS

3.2.1. UV/Vis spectroscopy

The heme group exhibits distinct electronic absorbance bands, i.e. the Soret and the Q-bands (α and β) in the visible region, which strongly depend on its oxidation, spin state and heme–iron ligation [47]. Therefore, UV/Vis spectroscopy is a useful tool to gain insight into the configuration of the heme pocket and especially into the coordination of the central iron atom. Fig. 1 shows the spectra acquired for *Gs*GCS in the as-purified ferric (Fe(III)), ferrous deoxy (Fe(II)), and the CO-bound ferrous (Fe(II)-CO) form. The trace of as-purified *Gs*GCS displays the Soret band at 411 nm and the maxima of the α and β Q-bands at 571 and 543 nm, respectively, revealing that *Gs*GCS oxidizes spontaneously (auto-oxidation rate: 1.88 h^{-1} at 20 °C and pH 7.5; see Fig. S12 and Table S4) to the ferric state upon exposure to the air. Therefore, a role in O_2 storage can be excluded for *Gs*GCS, as the protein has to be in its ferrous state (Fe(II)) for stable dioxygen binding [48]. A rapid transient binding of O_2, to perform oxygen catalyzed reactions, may however be still possible. The spectral features are typical for a low-spin (LS) ferric heme group in line with the bis-histidine coordination observed for the globin domain of *Gs*GCS (*Gs*GCS^{162}) [16]. However, it should be noted that the Soret band reported for as-purified *Gs*GCS^{162} is narrower than the one of ferric *Gs*GCS, potentially pointing to a somewhat larger heterogeneity in the heme-bounding for the latter. However, this may also be due to a lower overall stability of the full-length protein under the measuring conditions.

Addition of dithionite to as-purified *Gs*GCS leads to the ferrous deoxy form with absorption maxima at 427, 561, and 532 nm, typical of LS ferrous Gbs in which the heme iron is ligated to two histidines (so-called hexa-coordination), e.g. neuroglobin [49], cytoglobin [50], *Ce*GLB6 [51], *Ce*GLB26 [52], and *Ce*GLB33 [24]. Of note, the Q-bands of the ferrous deoxy form appear less sharp than expected for hexa-coordination, indicating a fraction of penta-coordination which might facilitate the binding of gaseous ligands, as has been shown *in vitro* for *Gs*GCS^{162} [16]. In the latter case, the relative fraction of penta-coordination seems to be even higher. The *in vitro* binding of CO to ferrous *Gs*GCS leads to absorption maxima at 420, 561, and 545 nm similar to what has been reported for *Bs*HemAT [53], a GCS with an aerotaxis transducer domain. However, this conformation might be of subordinated role as the distal HisE11 ligand favors the binding to the heme group [16].

Overall, the comparison of UV/Vis spectra reported for the as-purified ferric, ferrous deoxy, and CO-bound ferrous form of *Gs*GCS^{162} [16] and full-length *Gs*GCS (Fig. 1) do not show strong

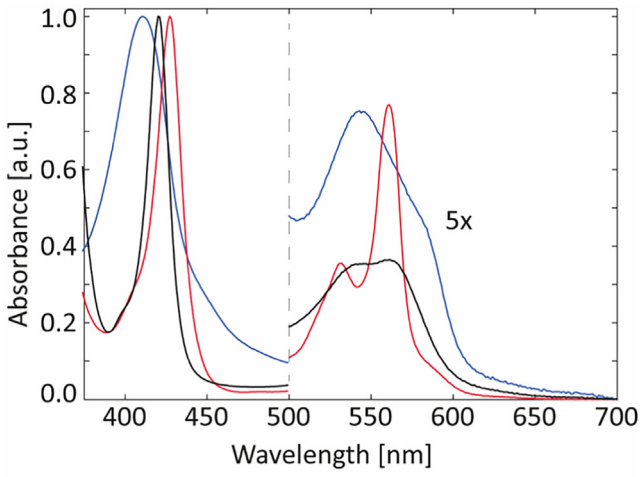

Fig. 1. UV/Vis spectra of *Gs*GCS in different heme iron oxidation and coordination states. The spectra correspond to the as-purified ferric (blue), ferrous deoxy (red), and CO-bound ferrous (black) form. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

differences, indicating only a minor to negligible impact of the TmD on pocket formation and heme coordination.

3.2.2. Resonance Raman spectroscopy

RR spectroscopy can provide a more thorough insight in oxidation, spin, and coordination state of the heme group. As for UV/Vis spectroscopy, RR spectra were also acquired in the as-purified, ferrous deoxy, and CO-bound ferrous form and largely confirm the UV/Vis results. The signal for the ν_4, ν_3, and ν_2 bands of ferric *Gs*GCS (Fig. 2a, i) lie at 1376, 1504, and 1582 cm^{-1}, respectively, typical of ferric heme proteins in their hexa-coordinated low-spin state [54]. *Gs*GCS in its ferrous deoxy form (Fig. 2a, ii), however, shows a mixture of penta- and hexa-coordination with clear marker lines at ν_4 = 1358 cm^{-1}, ν_3 = 1471 cm^{-1}, and ν_2 = 1558 cm^{-1} for penta-coordination high-spin (HS), and ν_3 = 1496 cm^{-1} for hexa-coordination low-spin (LS). This co-existence of the two forms agrees with the optical absorption spectrum in Fig. 1 and points to reduced binding affinity of the distal histidine in the ferrous compared to the ferric form. Of note, the different intensity between the ν_3 lines, i.e. 1471 and 1494 cm^{-1}, in favor of the penta-coordinated (HS) as compared to the hexa-coordinated (LS) species, does not reflect the true ratio, as this band is known to be higher in intensity for penta-coordinated Gbs in RR spectra. The appearance of two ν_4 modes i.e. 1358 and 1372 cm^{-1}, in the RR spectra of CO-bound *Gs*GCS (Fig. 2a, iii) at low laser power (1 mW) points to considerable photolysis of the bound gaseous ligand even at relatively low laser power (1 mW).

The low wavenumber area allows for a more in-depth view of the heme stabilization by the protein surrounding. Overall, these spectra are again very similar to what has been reported previously for *Gs*GCS^{162} [16]. The heme propionate bending mode $\delta(C_\beta C_c C_d)$ observed at 376 cm^{-1} in the ferric form (Fig. 2b, i) shifts to higher wavenumber for the ferrous deoxy (382 cm^{-1}) and CO-bound ferrous form (380 cm^{-1}) (Fig. 2b, ii-iii). All values indicate a strong hydrogen bond between the propionate and the surrounding amino acids, with the highest wavenumber indicating the strongest interaction between the propionate group and the protein matrix [55,56]. In this context, the observation of a clear shoulder at 694 cm^{-1} on the ν_8 peak (680 cm^{-1}) for ferric *Gs*GCS is remarkable (Fig. 2b, i). A peak at this position is not common in heme proteins and has, to our knowledge, only been reported for peroxidases [57,58]. It was tentatively assigned to an out-of-plane CH_2 asymmetric wag of the porphyrin vinyl groups [57] and points to substantial interaction of the protein matrix with heme group. Interestingly, the mode is not observed in the ferrous deoxy or CO-bound ferrous form of *Gs*GCS (Fig. 2b, ii-iii), but was observed for ferric *Gs*GCS^{162} [16]. It points to an altered stabilization of the heme group by the protein matrix in the ferrous forms as compared to the ferric form. The broad peaks at 439 cm^{-1} (ferric *Gs*GCS), 438 cm^{-1} (ferrous deoxy *Gs*GCS) and 432 cm^{-1} (Fe(II)-CO *Gs*GCS) is due to the vinyl bending mode $\delta(C_\beta C_a C_b)$ and has been observed in other heme proteins in this region [59]. The peak at 398 cm^{-1} that is only observed in ferric *Gs*GCS (Fig. 2b, i) and ferric *Gs*GCS^{162} [16] is intriguing. It most likely is due to the vinyl bending modes $\delta(C_\beta C_a C_b)_4$, although it is lower than what is usually reported for this mode in globins and other heme proteins. It may be linked to the appearance of the 694 cm^{-1} as is also substantiated by the fact that this mode is not visible in the (CO-bound) ferrous form of *Gs*GCS (Fig. 2b, ii-iii). Moreover, a clear shift of the $\nu(C_a = C_b)$ modes from 1632 cm^{-1} to 1620 cm^{-1} is found upon reduction of the heme iron. Overall the results point to a marked effect of the oxidation state of the heme iron on the interaction of the heme group with the surrounding protein matrix, mainly the interaction of the vinyl substituents with the surrounding amino acids and the strength of distal His binding. The effect is observed both in the presence and absence of the TmD.

Finally, the appearance of a peak at 492 cm^{-1} for the Fe-CO stretching mode of CO-bound ferrous *Gs*GCS (Fig. 2b, iii) points to an open heme pocket structure in which no amino-acid group is sufficiently close to the CO to affect the electronic structure of the Fe-C-O unit as also observed for other globins [60,61].

3.2.3. Native mass spectrometry

Spectroscopic techniques are valuable methods to obtain information on both pocket formation and heme coordination; yet, a detailed structural view of oligomerization and/or the overall structure remains inaccessible by these techniques. Native mass spectrometry (MS), however, has become an established technique in structural biology enabling the characterization of the proteins' architecture, oligomerization, ligand binding, and conformational dynamics [62–65]. Unlike in bottom-up or top-down proteomics [66], proteins are still in their native, folded state in solution [67] before being transferred into the vacuum of the mass spectrometer by soft ionization techniques, e.g. nano-electrospray ionization [68]. As such, it is possible to retain the proteins' quaternary structure [63] enabling to show co-existing oligomeric and/or conformational states, though the technique does not provide atomistic resolution [64,69]. Recent advances made native MS amenable to membrane proteins [70,71]. In such an approach, a surrounding hydrophobic environment, e.g. detergent micelle, protects the membrane protein upon release from solution into gas phase of the mass spectrometer [72,73] where proteins are liberated in a collision-activated manner [71]. Therefore, it seems obvious that native MS would help to gain structural insights into *Gs*GCS.

To determine how the different subdomains are involved in the oligomerization, we conducted native MS experiments on both *Gs*GCS^{162} and full-length *Gs*GCS. As *Gs*GCS^{162} is entirely soluble (no TmD), this construct was studied without the need for a detergent-supplied hydrophobic environment and its inherent fine-tuning process of protein liberation in the gas phase. Fig. 3 shows the MS spectra of *Gs*GCS^{162} under native (a) and denaturing (b) conditions. As can be seen, *Gs*GCS^{162} assembles into monomeric and dimeric species, with dimers as the most prevalent complex (Fig. 3a). Although protein crystals cannot be compared directly with the solution phase, the dimerization of the protein was not surprising, as the published X-ray structure of *Gs*GCS^{162} (Fig. 9b) is a dimer too [16]. Both monomers and dimers reveal two narrow charge state distributions that center around charge states +8 and

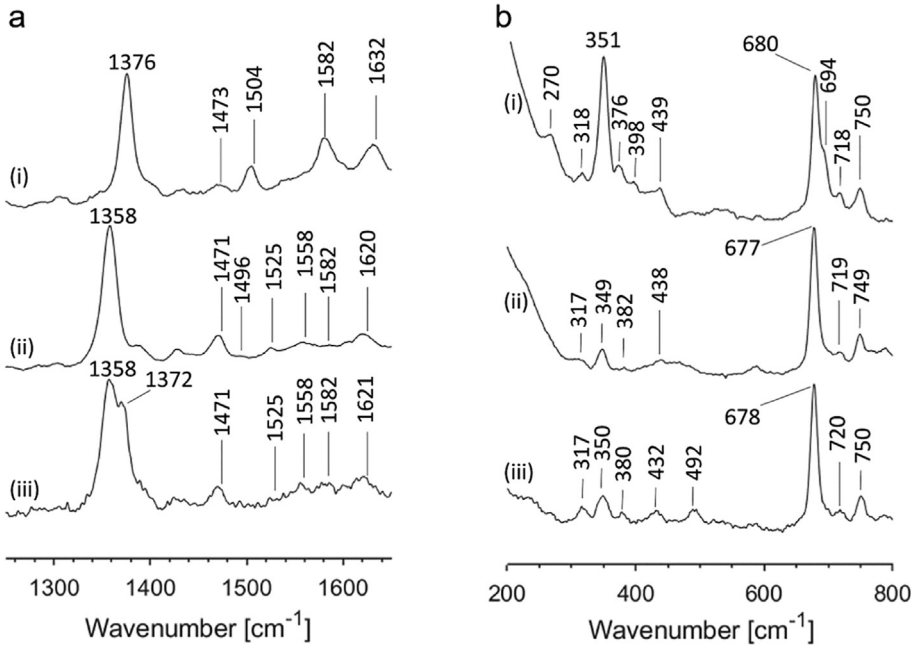

Fig. 2. Resonance Raman spectra of *Gs*GCS in different iron oxidation and coordination states. High (a) and low frequency (b) resonance Raman spectra corresponding to as-purified ferric (i), ferrous deoxy (ii), and CO-bound ferrous (iii) forms of *Gs*GCS. The applied laser power was 1 mW (iii) and 100 mW (i and ii).

+12, respectively. Noteworthy, experimental conditions were gentle enough to preserve the incorporation of the noncovalent prosthetic heme group, depicted by a small black circle in Fig. 3. Narrow distributions that are low in charge indicate a native/folded, compact character of the protein, with the extent of charging reflecting the extent of the solvent-exposed surface area, in agreement with the proteins' structure of the solution phase [74]. Unfolded proteins have a relatively higher surface accessible during ionization, and therefore, elongated proteins result in higher charged species that are relatively broader distributed in the spectrum [74]. As expected, the denaturation of the protein, i.e. through addition of acetonitrile, reveals such a distribution for monomeric *Gs*GCS^{162}, whereas the dimeric signal disappeared entirely (Fig. 3b). Of note, the monomeric signal distributes over three Gaussian-like distributions indicating a stepwise unfolding that involves three species, i.e. native, intermediate, and unfolded conformations. Moreover, denaturing conditions (pH 4.0) also lead, unsurprisingly, to the release of the noncovalently-bound heme group, illustrated by a purple circle with a white dot. The detected monomeric mass (19,166 Da) of GsGCS^{162} closely matches the theoretical mass of 19,164.75 Da based on the protein sequence. The fact that dimers are no longer present under denaturing conditions emphasizes a non-covalent character of the *Gs*GCS^{162} dimerization.

Native MS analysis of full-length *Gs*GCS was more challenging, as the C-terminal TmD requires a hydrophobic environment in order to keep the protein soluble [75]. We used native MS to find a detergent suitable for both the solubilization of *Gs*GCS and compatibility with detection in the mass spectrometer [71–73]. To that end, we tested the detergents n-dodecyl-β-D-maltoside (DDM), triton X-100, and OG. Both DDM and triton X-100 required high collision energies which exclusively led to highly charged, monomeric species, i.e. non-native, unfolded *Gs*GCS (Fig. S7). Contrary to that, OG reveals trimeric and tetrameric *Gs*GCS complexes obtained under gentle detergent release (Fig. 4). Comparing *Gs*GCS^{162}, which assembles into monomers and dimers (Fig. 3), with full-length *Gs*GCS clearly suggests the TmD as driving force behind this higher-order oligomerization. Of note, both oligomeric forms retain the full complement of noncovalently bound heme groups, indicating that they originate from solution. Excess activation of the complex during micelle release on the other hand would cause dissociation of heme groups as well as ejection of a highly charged, unfolded monomeric subunit which is not observed here, suggesting that both tri- and tetramer exist in solution.

3.2.4. Collision-induced unfolding

Together with native MS, ion mobility (IM) can be used to interrogate the gas-phase mobility of analytes providing a global view of their rotationally averaged size and shape. The theory behind this is simple, the ions are propelled through an ion mobility cell filled with inert gas, e.g. He or N_2, and the arrival time is measured for each ion [76,77]. The mobility of the ions thereby depends on their charge and their collision cross section, i.e. shape and size. Exploiting ion mobility in combination with a stepwise increase in ion activation in MS, termed collision-induced unfolding (CIU) [78], enables to study the proteins' fold stability and domain architecture [79,80]. Such CIU assays are particularly useful to compare the intrinsic stability of different protein variants and/or conformations in a solvent-free environment. As the heme group and its coordination are highly likely to play an important role, we performed CIU experiments for both *Gs*GCS^{162} and full-length *Gs*GCS to probe subunit stability as well as inter-subunit interactions in different coordination and oxidation states, i.e. the ferric, ferrous deoxy, and CO-bound ferrous form (Fig. 5).

In the case of *Gs*GCS^{162}, the collision energy was ramped from 10 to 50 V using steps of 2.5 V. Fig. 5a shows the unfolding trajectories of the three different coordination states of the dimeric +12 species of *Gs*GCS^{162} (selected m/z range of the holo complex: 3297–3302). Higher collision energies led to heme loss, however, the holo complex remained the dominant species throughout. All GD dimer forms display a clear unfolding transition at approximately 32.5 V. A similar unfolding transition for the different species suggests that the heme iron oxidation and coordination state has no discernible effect on the conformational stability and subunit interactions of *Gs*GCS^{162} upon activation. Therefore, the competing effect between the CO ligand and His(66)E11 on the distal site, as

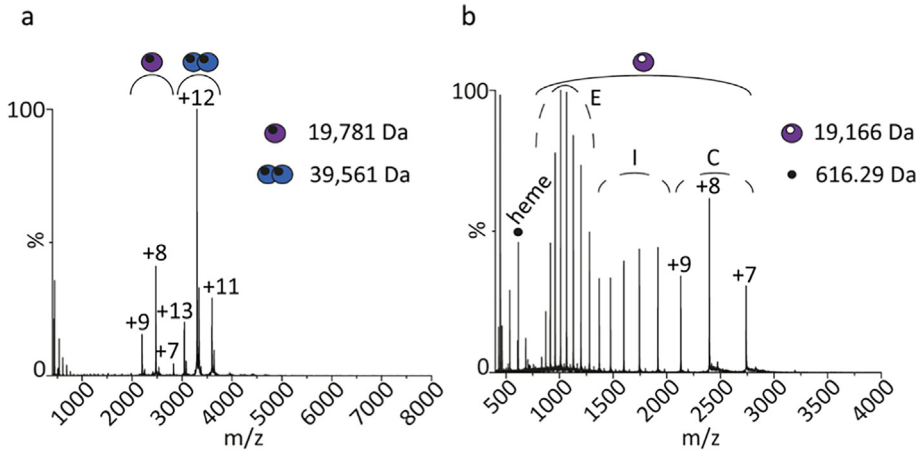

Fig. 3. Mass spectra of *Gs*GCS[162] under native and denaturing conditions. The native spectrum (a) shows two species corresponding to monomeric and dimeric *Gs*GCS[162]. The detected mass reflects the protein species including the noncovalent prosthetic heme group (black dot). The denaturing spectrum (b) shows only monomeric *Gs*GCS[162] upon denaturation, indicating the non-covalent character of the dimerization. The extended charge state distribution relative to the native spectrum highlights the stepwise unfolding of the protein from compact (C) to intermediate (I) and extended (E) structures, displayed by three Gaussian-like peak distributions and ion mobility. Denaturation also led to the release of the noncovalently-bound heme group (616.29 *m/z*).

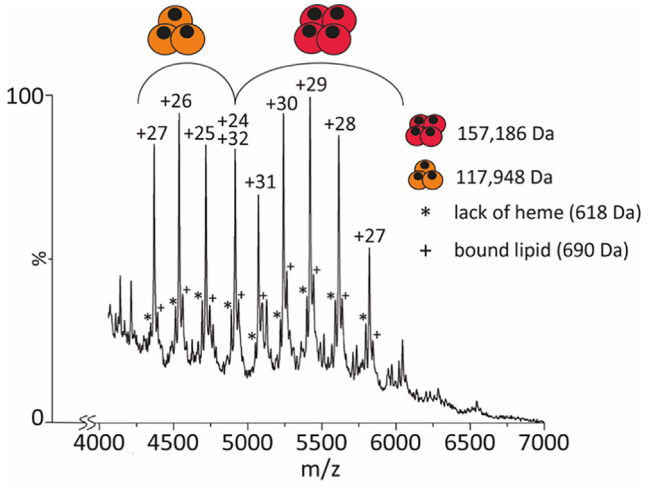

Fig. 4. Native MS spectrum of *Gs*GCS solubilized in OG micelles. The spectrum shows two charge state distributions corresponding to trimeric (117,948 Da) and tetrameric (157,186 Da) *Gs*GCS. The detected mass reflects the protein including three and four noncovalent prosthetic heme groups for trimers and tetramers, respectively. Minor satellite peaks indicate the loss of one heme group (*, 616 Da) and likely a lipid bound to the complex (+, 690 Da).

indicated by RR in a previous study [16], has no unfavorable impact on the strength of the dimeric interaction between the GDs.

As outlined earlier, native MS of detergent-solubilized membrane proteins requires the release of the protein from micelles in a collisional activation process [71,72]. Thus, only the additional excess energy is available for unfolding of the released protein, and CIU experiments performed on *Gs*GCS required much higher collision energies relative to soluble *Gs*GCS[162]. Moreover, *Gs*GCS as tetramer forms a much bigger complex relative compared to the smaller soluble construct *Gs*GCS[162], which also clearly affects the collision energy necessary for protein unfolding. We increased the collision energy in the trap cell from 100 V (onset of *Gs*GCS liberation) to 200 V, using steps of 10 V. Fig. 5b reveals CIU plots obtained for tetrameric +28 *Gs*GCS (selected *m/z* range of the holo complex: 5612–5650). Most notably, the unfolding energy is significantly higher in the ferrous deoxy (160 V) compared to the ferric (120 V) form, indicating a stabilization of the complex upon the reduction of the heme iron (Fe(III) to Fe(II)). The CO-bound *Gs*GCS complex, on the other side, unfolds at 110 V and therefore does not show such a stabilization effect. The fact that the unfolding trajectories for full-length *Gs*GCS (Fig. 5b) clearly pronounce variations in the stability between the ferric and ferrous form, which is obviously not present in *Gs*GCS[162] (Fig. 5a), suggests a state-dependent (de)stabilization effect of the heme iron on inter- and intra-subunit interactions of the GDs in the membrane-bound tetramers.

3.2.5. 3D structure of full-length GsGCS using cryo-electron microscopy and image analysis

All attempts to crystallize full-length *Gs*GCS remained unsuccessful so far. Over the last couple of years, however, cryo-electron microscopy (EM) has increasingly become successful in characterizing membrane protein structures [81,82], and we employed the technique to investigate the structure of full-length *Gs*GCS. Fig. 6a shows a characteristic image of negative-stain full-length ferric *Gs*GCS. Although the images clearly display heterogeneity, some of the particles look similar and exhibit a lack of density in their center. 2D classification performed on 21,414 automatically picked particles from 90 cryo images is shown in Fig. 6b. For further data processing, two non-representative classes with a total number of 377 particles were removed. Thus, the 3D classification was performed on 21,307 frozen particles using RELION 2.0 [36,37] with no symmetry applied. Class 1 shows a lack of density in the center, a characteristic feature also visible in isolated full-length *Gs*GCS single particles (Fig. S8a). We therefore choose to refine this 3D preliminary model (Fig. S8b).

Since native MS of *Gs*GCS suggested an assembly into trimeric and tetrameric states, we applied two different symmetries, i.e. C3 and C4. Fig. 7a illustrates the 3D structures from three different views. The 3D model obtained by C3 symmetry (yellow) displays three twisted arms with a hollow center, particularly visible in the top view. The second 3D model (cyan), calculated with C4 symmetry, shows a cubic structure with four arms protruding from the corners on the same plane. The resolution for these structures is about 15 Å. On the way towards a realistic structure we must consider that the structure of *Gs*GCS must enable a spatial separation of the soluble domain and the membrane domain, as the hydrophilic and hydrophobic parts cannot be in close proximity to each other. In both models, C3 and C4, we can clearly distinguish the membrane domain and the globin domains (Fig. 7b), and in both models, the TmD is the driving force behind the oligomerization, since the Gb domains are not in contact. Therefore, it is rather dif-

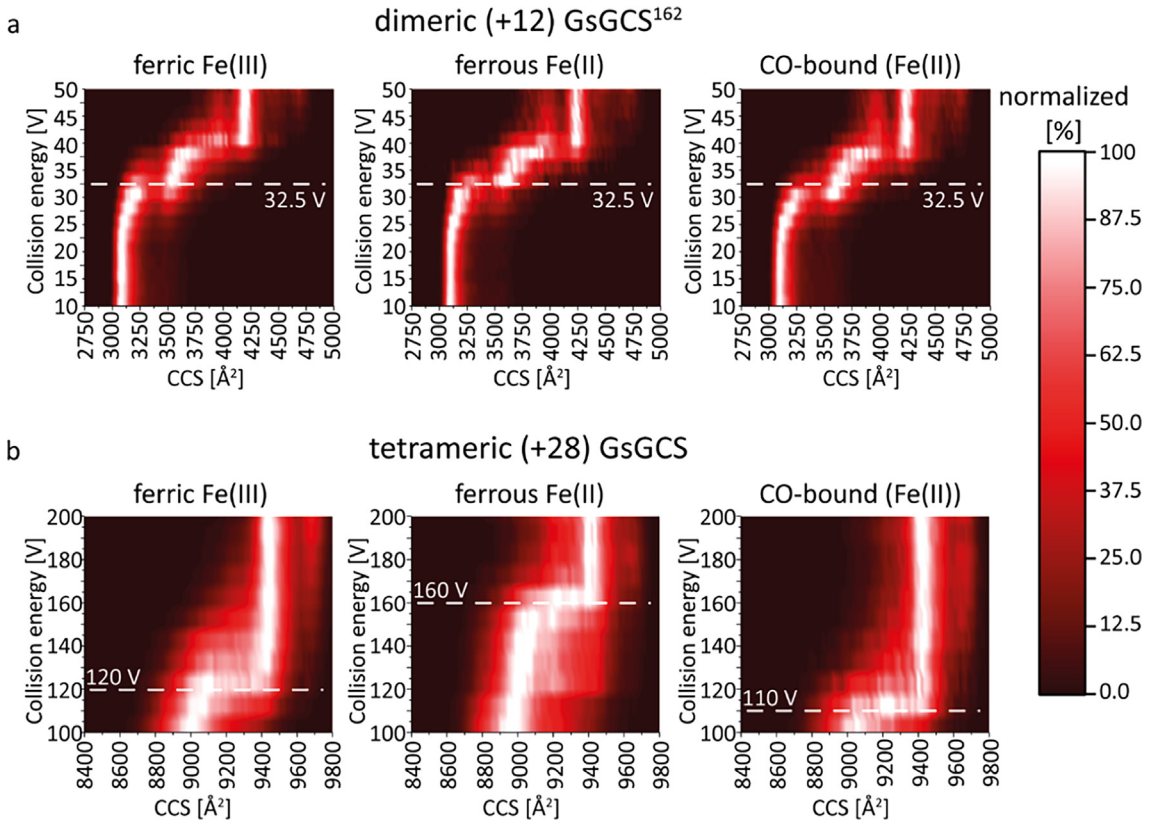

Fig. 5. CIU of dimeric (+12) *Gs*GCS[162] and tetrameric (+28) *Gs*GCS in different iron oxidation/coordination states. CIU plots of dimeric (+12) *Gs*GCS[162] (a) show the unfolding transition (32.5 V collision energy voltage) for each species with almost no difference between them. CIU plots of tetrameric (+28) *Gs*GCS (b) however show large differences among the different states. The as-purified Fe(III) ferric and the Fe(II) CO-bound form unfold at 120 and 110 V collision energy voltage, respectively. Contrary, the ferrous deoxy species show a much higher unfolding collision energy voltage (160 V), indicating a more stable form of this complex.

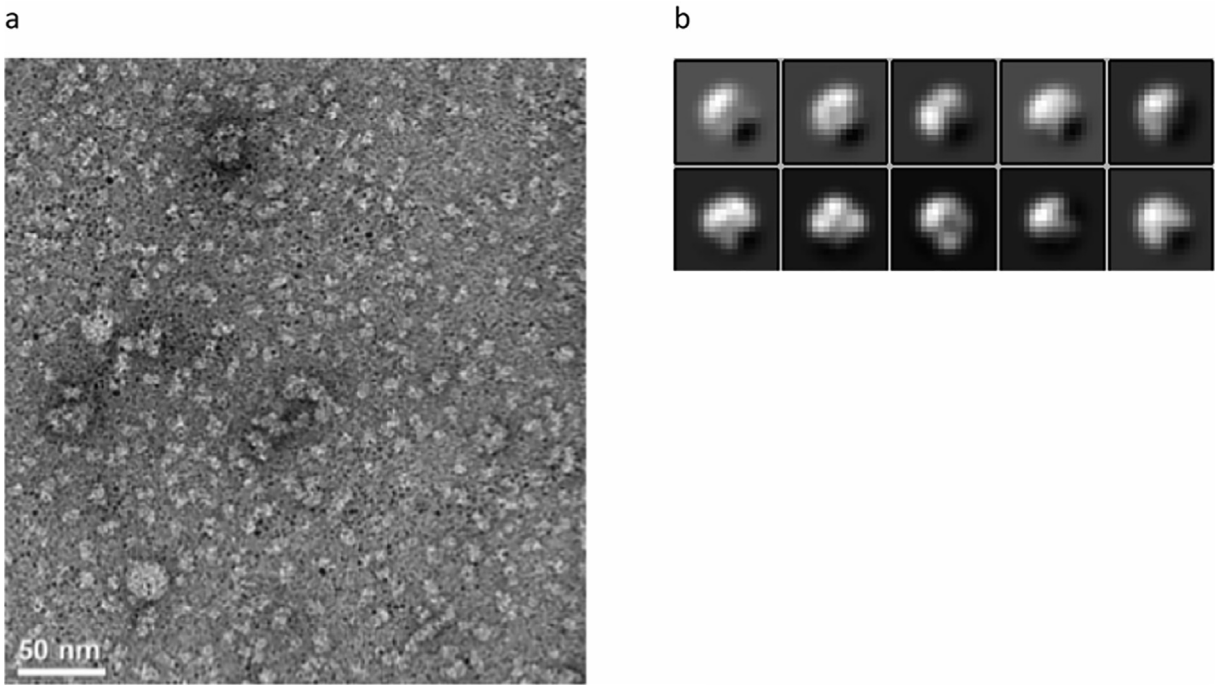

Fig. 6. Electron Micrograph and 2D classification of full-length *Gs*GCS. (a) Electron micrograph of negatively stained full-length *Gs*GCS in OG detergent. Protein is bright against a darker background. The scale bar equals 50 nm. (b) 2D classification: class averaged from 21,414 cryo-particles, the box size equals 135 × 135 Å^2.

ficult at this stage to validate which model, C3 or C4, is the more representative.

3.2.6. Modeling of full-length GsGCS

Since the resolution of the structure obtained by cryo-EM was not sufficiently high, we aimed to build a model of full-length *Gs*GCS using the tetrameric cryo-EM envelope. Such a computational approach requires a realistic structure homolog of the TmD. The identified PsiE domain is predicted to adopt a 4TmD. Therefore, to obtain a structural model of full-length ferric *Gs*GCS, we first built a PsiE domain structure using comparative modelling. For this purpose, we selected lipid II flippase MurJ (pdb 5T77), a protein consisting of an all α-helical transmembrane bundle, as template and manually aligned the corresponding sequences, preserving the predicted 4 transmembrane helix boundaries and key residues (Figs. S5 and S6). In consideration of given uncertainties, i.e. low template similarity, this PsiE model represents a first topological approximation rather than a detailed structure. Towards a reliable tetrameric model, we first aimed to develop a monomeric structure of full-length *Gs*GCS by fitting both domains into the cryo-EM envelope and joining the C-terminal helix from the globin domain with the N-terminal helix from the PsiE domain. To obtain a model for tetrameric *Gs*GCS, we replicated the monomer using C4 symmetry (Fig. 8).

Our structural model allows performing an analysis of the relative subunit orientations. Subunits are positioned with the heme group facing outwards from the tetramer edge (Fig. 9a), and thus quite solvent exposed which enables easy access for potential ligand binding and/or electron transfer. Unlike in the observed X-ray structure of the Gb domain dimer (Fig. 9b), G/H helices are not establishing tight interactions, and are instead facing the E/F helices and heme edge of the following monomer (going clockwise) (Fig. 9a). Indeed, no major direct interaction is observed between the GDs, which seem to be maintained in place by their link to the 4TmD domains, confirming native MS observations of a membrane domain driven higher-order oligomerization. The fact of loosely connected globin domains suggests that they may function independently from each other, i.e. working non-cooperatively. Statements about the role of the PsiE (4TmD) domain, in this context, remain difficult. Likely, its main function is related to cellular localization on the membrane and consequently an increase in *Gs*GCS effective local concentration.

4. Discussion

At a first glance, the association of Gbs with cellular membranes seems to be rather uncommon. Yet, membrane-associated Gbs are relatively widespread as suggested by membrane-related functions of some prokaryotic Gbs [2], the occurrence of acylated Gbs in plants, brown algae and fungi [83] as well as the N-terminal acylation in globin proteins from a broad range of phyla [83–85]. This diversity in membrane-associated Gbs even led to the hypothesis that animal respiratory Gbs might originate from a membrane-bound ancestor [83]. Therefore, the coupling of a GD with a TmD is probably less exceptional than initially thought. Consistently, our *in silico* searches for Gbs with 4TmD (*Gs*GCS) and 7TmD (*Ce*GLB33) domains (see below) identified 168 and 14 potential Gbs linked with a TmD, respectively.

The identification of an unexpectedly high number of globin-PsiE proteins prompted us to perform ortholog searches for GDs fused to a TmD using also the sequence of full-length *Ce*GLB33 (NCBI reference: NP_001300285.1) as query. In addition to UniProt we also interrogated the WormBase (version WS277) database, as preliminary searches indicated a nematode origin for most of the retrieved proteins. Database searches revealed, in addition to the expected nematode sequences, two mammalian sequences (UniProt IDs: S7NGW8 and A0A061I7G5) consisting of a GDx fused with a 7TmD. The first belongs to a bat species (*Myotis brandtii*) and displays an N-terminal 7TmD and a C-terminal GD, whereas the second belongs to a rodent (*Cricetulus griseus*) and carries two N-terminal GD and a 7TmD. It is important to note that both sequences are not reviewed, as indicated in the UniProt database, and have been derived from whole genome sequencing experiments. Therefore, although these findings should be taken strictly as hypothetical awaiting confirmation, they would potentially represent the first two mammalian sequences to display both a transmembrane domain and a GD. Notably, *M. brandtii* displays the typical 7TmD-GD organization found in nematodes (like *Ce*GLB33), whereas *C. griseus* has an unprecedented GD-GD-7TmD organization. As shown in Fig. S9, these three GDs display the traditional mammalian globin hallmark HisE7 and hydrophobic residues in B9 and B10. The nematode globins, on the contrary, although highly similar among them, diverge significantly from the mammalian sequences, displaying instead a hydrophobic residue at E7 (with two exceptions which display a Glu at that position: *A. simplex* and *As*FMRF). The 7TmD MSA clearly shows conservation of all helix lengths with gaps clustered in the inter-helix regions (Fig. S10). Some minor hallmarks are a proline after helix 3 and a Cys after helix 4. Interestingly, the same pattern of two groups, mammals and nematodes, is observed. The difference is such that they are assigned to different PFAM subfamilies, with those of nematodes belonging to family 7TmD_1, and those from mammals to 7TmD_4.

As shown by the above analysis, the two variations of transmembrane-associated Gbs (globin-PsiE(4TmD) and globin-7TmD) exhibit differences in both globin fold and in occurrence. It is therefore unlikely that both types reflect the same class of proteins. Instead, it makes sense to clearly distinguish between these two variants based on GD and host organism. Thus, globin-7TmD sequences (M-type and eukaryotic) with evident similarities to a receptor must be classified as receptor-like Gbs, whereas globin-PsiE sequences (S-type and prokaryotic) must be considered as proper GCS similar to *Ec*DosC and *Bp*GReg.

The fact that our searches resulted in a non-negligible number of GD-TmD structures indicates that many more unknown globin-containing structures, included GCS, may lie dormant in databases - unsurprisingly, as the last major ‘globin census’ dates back to 2013 [86,87]. Meanwhile the databases have grown exponentially, thanks to genome-sequencing efforts along with an increasing computing power. Indeed, a more recent database interrogation analysis for trHbs (2/2) by Hade et al. (2017) [88] shows a bewildering extension of 2/2 Gbs, including chimeric multi-domain containing structures. Without doubt, the ongoing update of the globin superfamily census including their phylogenetic analysis will reveal a similar extension for GCS sequences (Schuster, C. Personal Communication). Thus, GD-TmD structures represent only one piece of the globin superfamily puzzle.

The presence of a TmD in *Gs*GCS (4TmD) made the structural characterization a challenging endeavor. Membrane protein solubilization requires a hydrophobic environment capable of mimicking the phospholipid bilayer. To that end, we tested different detergents, including OG, DDM, and triton X-100, for *Gs*GCS stabilization. Gel filtration and native MS experiments were performed on all these samples. The gel filtration experiment indicated some heterogeneity by showing a peak with an additional shoulder for all detergents, but did not allow a clear separation. However, among the tested detergents, OG performed best in terms of solubilization and stabilization, showing an assembly of *Gs*GCS trimeric and tetrameric complexes (Fig. 4). This heterogeneity substantially hampered attempts at high resolution structures in cryo-EM. We find a mixture of trimers and tetramers, possibly due to a still

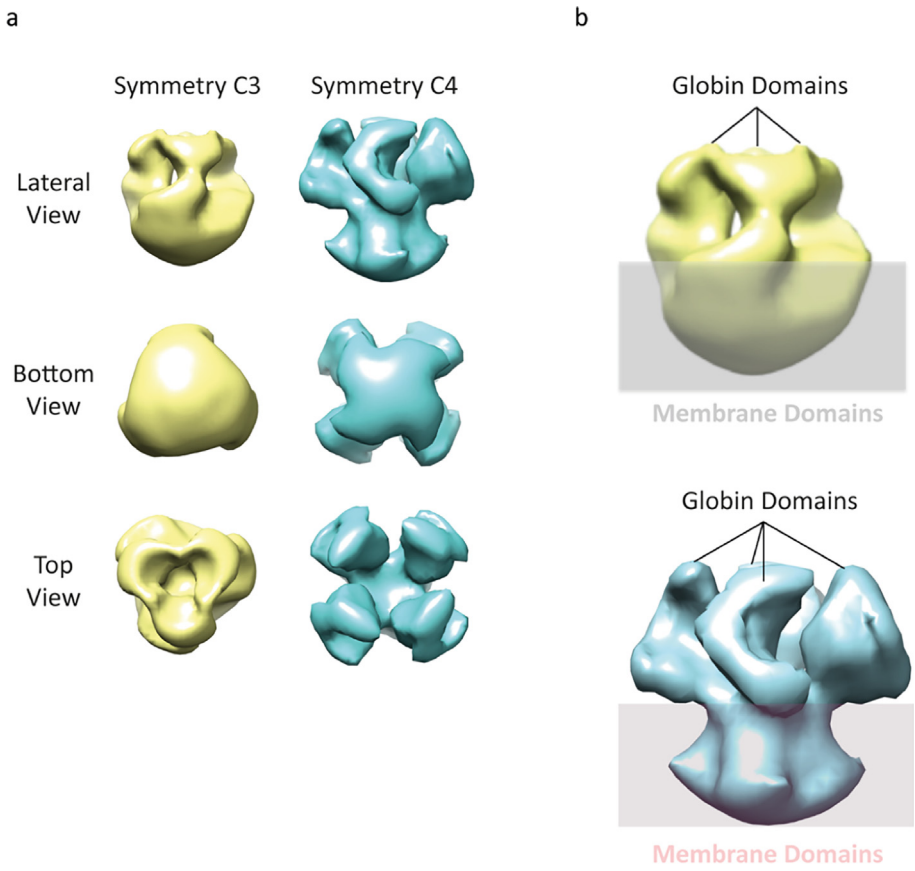

Fig. 7. 3D models of full-length *Gs*GCS in OG. (a) Structures obtained by applying of C3 (yellow) and C4 (cyan) symmetry are illustrated from different angles. (b) Description of the various domains. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

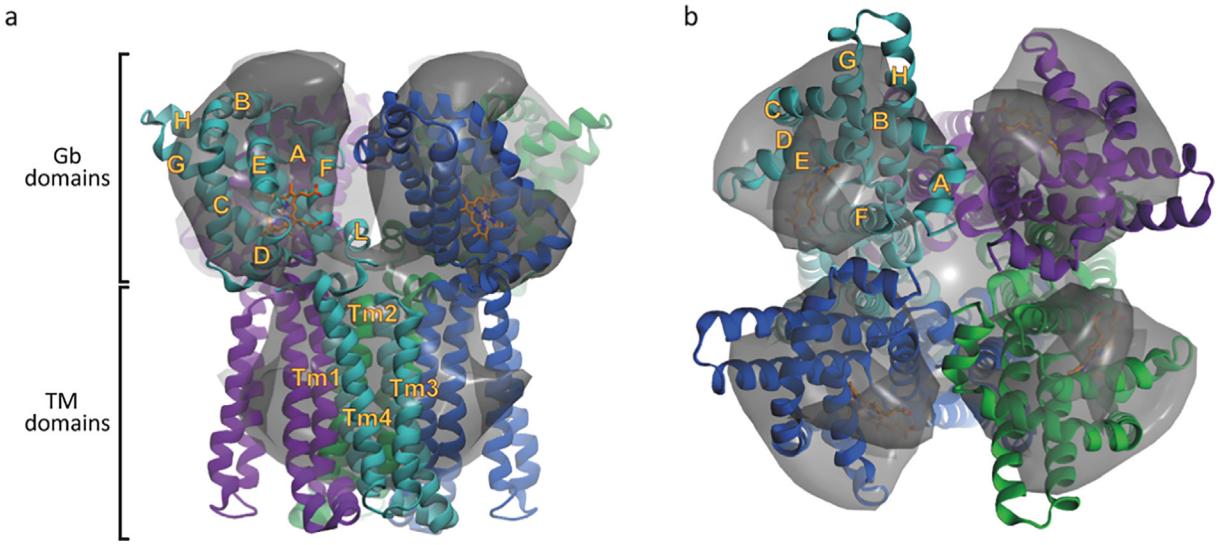

Fig. 8. Structural model of *Gs*GCS-PsiE tetramer. Side (a) and top (b) view of the tetrameric model obtained after fitting four monomeric *Gs*GCS-PsiE subunits into the cryo-EM envelope. The model shows four loosely connected globin domains with outward-oriented heme groups, providing easy access. Each monomeric subunit is shown in a different color. Helices of the globin domain (A–H), the linker (L) and the transmembrane domain (Tm1–Tm4) are labeled accordingly in one subunit.

not sufficiently "native" detergent environment, that we cannot split into two distinct classes. Applying C3 or C4 symmetry for 3D particle reconstruction generated two different low-resolution models corresponding to trimer and tetramer, respectively, and in both models we can discriminate the membrane domain from the rest of the protein (Fig. 7). In order to validate the oligomeric state of the molecule in OG (trimeric or tetrameric), we took the 3D model, with both symmetry C3 and C4 imposed, and re-projected them in various 2D orientations (25 projections for each 3D model; Fig. S11) for comparison with 2D classifications obtained in cryo conditions (Fig. 6b). Re-projections from the model with imposed C4 symmetry often show a horseshoe-like

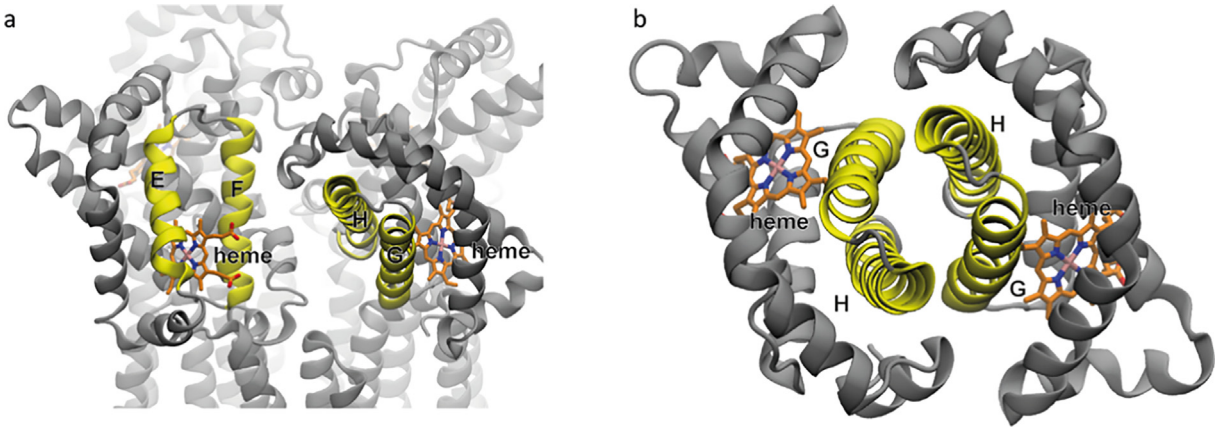

Fig. 9. Close-up view of the globin domain. Helix arrangement and heme group of the globin domains of both the model (a) and the X-ray structure *Gs*GCS[162] (pdb 2W31) (b). Comparing the X-ray structure with the model clearly highlight differences in the interactions between their globin domains. The model of the full-length protein, including the TmD, no longer forms a tightly connected 4-α-helical bundle between G/H helices. Instead, the G/H helices of one monomer are facing the E/F helices of the next subunit by going clockwise. The heme groups (orange) are orientated to the outer edge of the molecule enabling access for potential ligands. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

shape, e.g. class number 11, 16, 20 and 23. Of note, this characteristic shape is also present in the 2D classification (Fig. 6b bottom row, second from the left). Such characteristics have not clearly been noted on re-projections from the model with C3 symmetry. Thus, this observation might favor the tetrameric over the trimeric form of full-length *Gs*GCS. The tetramer is also preferred for two other reasons: (i) the application of a non-ideal detergent for membrane protein stabilization could lead to the decomposition of a bigger to a smaller complex, i.e. tetramer to trimer, and not the other way around; and (ii) trimeric assembly of Gb representatives have not been identified yet, while dimers and tetramers are quite common. Future characterizations would probably benefit from the application of membrane-like environments such as nanodisc [89] or SMALPs [90] which would provide a more native lipid environment.

Oligomerization varies among different types of GCS (Table S2). However, despite some differences it seems that (i) GCS must be minimally a dimer to obtain an active TD, and (ii) the dimeric symmetry of both GD and full molecule play a role in intramolecular signal transduction. The activity of the GCS dimer was shown to be modulated by different factors such as ligand (+) or product (−) binding and oligomerization (+/−). Clear examples of an oligomerization-driven activity regulation (DGC-group) exist. However, a tendency for aggregation *in vitro*, probably by low solubility, as seen in gel filtration and ultracentrifugation experiments may not be excluded [91–93]. Further analysis of the protein structure including its oligomerization state can be performed based on the atomic resolution tetramer model using the X-ray structure of *Gs*GCS[162] [16] (Fig. 9b) along with lipid II flippase MurJ (pdb 5T77) as transmembrane structure homolog. Our model for the Fe(III) state nicely fits the cryo-EM tetramer envelope, both in each GD as well as in the TmD tetramer (fitting to the trimer is less satisfactory, further supporting our choice). The 4 transmembrane helices are highly hydrophobic and the tetramer is tightly packed within the lipid bilayer, while the GDs are not touching each other in this model. This loose connection of the GDs in *Gs*GCS is substantially different to well characterized GCS, e.g. *Ec*DosC and *Af*GcHK (Table S2), which have shown to form a proper dimeric structure with the resulting symmetry to be most likely important for intramolecular signal transduction. Therefore, these faintly connected GDs might have functional consequences raising doubt whether there is an intramolecular signal transduction in *Gs*GCS such as in conventional GCS, i.e. *Ec*DosC and *Af*GcHK. The linker, which has not shown any significant homology to other known sequences, is predicted to adopt a helical conformation with high confidence, forming a semi-rigid backbone and possibly limiting motion of the GD. The heme group facing the tetramer outer edge in the Fe(III) state makes it fairly accessible to potential ligands or electron transfer partners. The RR of the full-length *Gs*GCS are very similar to those of the GD. The interaction of the heme vinyl groups and protein matrix is strongly depending on the oxidation state of the heme iron, an effect which is independent of the presence of the TmD. From the ion-mobility studies of the GD dimer and full-length *Gs*GCS tetramer (Fig. 5), it follows that these oxidation and spin state related changes affect the unfolding stability of the tetramer, possibly due to changes in the local heme environment which may destabilize the globin domains, or affect the interactions between them. Dimeric GsGCS[162] does not show such changes in stability among the heme iron oxidation states which might be explained by the different GD oligomerization interface (G/H helix bundle) observed in the X-ray structure.

The tetrameric structure exhibits loosely connected globin domains in the Fe(III) state, indicating that they may function independently from each other, i.e. working non-cooperatively. Notably, the GD dimer observed in the corresponding X-ray structure is not compatible with C4 symmetry and at odds with the tetrameric structure of full-length *Gs*GCS which has been clearly shown to be driven by the TmD (Figs. 3 and 4). Interestingly, the closely packed GD dimer exhibited non-cooperative carbonylation [16]. Statements about the role of this PsiE (4TmD) domain, however, remain difficult due to its unknown function. It is however apparent that its main role, in the current context, is related to oligomerization and membrane localization, which might consequently lead to an increase in effective local concentration to possibly support the electron transport chain in Geobacter sulfurreducens [94,95]. A functional role, for example the transport of H^+ across the membrane, which might even be coupled to an electron transfer [96], is also conceivable and would suggest conformational changes in the TmD upon a sensing event in the GD and/or the reduction of the heme iron.

Our model is based on the ferric form of both *Gs*GCS[162] (X-ray structure) and full-length *Gs*GCS (cryo-EM) raising the question about possible changes upon the reduction of the heme iron atom. However, the resolution of the presented cryo-EM structure is too low, which makes further assumptions about the functional role of *Gs*GCS, difficult. The lack of observation of a clear contact between

the Gb domains, at least in the ferric form, seems to differ from other genuine GCS, where the dimeric globin structure has been shown to be important for signal transduction. Additionally, the auto-oxidation rate (1.88 h^{-1} at pH 7.5 and 20 °C; Fig. S12 and Table S4) and the reduction potential (-283 mV vs. SCE, or −37 mV vs. SHE; Fig. S13 and Table S5) highlight spontaneous oxidation of *Gs*GCS and a high efficiency of donating electrons, which is similar to *Ce*GLB6 and *Ce*GLB26. Of note, these globins have shown low binding affinity for O_2 due to a deviating globin structure. Major differences compared to standard Gbs are the absence of a D-helix connecting the C-helix directly to an elongated E-helix, and strong bis-His hexa-coordination (HisF8-Fe-HisE11) [51,52,97–99]. These characteristics are very similar to those of *Gs*GCS where the His(66)E11 coordination results in an anticlockwise rotation of about 30° towards the heme of the E-helix, positioning the E-helix almost parallel to the heme and orthogonal to the C-helix [16]. Due to these similarities with *Ce*GLB6 and *Ce*GLB26, i.e. unfavorable O_2 binding, we scrutinize whether *Gs*GCS functions as an O_2 sensor as proposed by Pesce et al. [16] In our opinion, an alternative redox sensing mechanism or a role in e^-/H^+ transfer are equally likely, and we therefore propose further investigations in this direction.

In summary, the presented work exhibits a full-length oligomeric state of *Gs*GCS showing a novel and uncommon membrane-associated structure for a Gb which resembles main features of GCS.

5. Conclusion

Proteins with a GD covalently linked to a TmD (4- or 7TmD) seem to be much more prevalent than initially thought and thus form a non-negligible subgroup of the globin superfamily. Here, we present a low-resolution model of full-length *Gs*GCS (GD-4TmD) which shows a TmD-driven tetramerization of the protein responsible for membrane localization. The heme groups are oriented outward in the ferric (Fe(III)) state, and other than the TmD, the GDs display no tight interactions, which makes an intramolecular signal transduction, as occurring in genuine GCS, less likely. Collectively, the loosely connected GDs, the fairly accessible heme groups, the auto-oxidation and the redox potential characteristics might suggest a potential role of *Gs*GCS in a membrane-bound e^-/H^+ transfer or as a redox sensor.

CRediT authorship contribution statement

DHa: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **FG:** Conceptualization, Methodology, Writing – review & editing. **SID:** Formal analysis, Visualization, Writing – review & editing. **CF:** Formal analysis, Visualization, Writing – review & editing. **CDS:** Data curation, Formal analysis, Writing – review & editing. **DHo:** Data curation, Formal analysis, Writing – review & editing. **MM:** Supervision, Writing – original draft, Writing – review & editing. **CV-B:** Supervision, Writing – review & editing. **LM:** Project administration, Writing – original draft, Writing – review & editing. **SVD:** Supervision, Formal analysis, Visualization, Writing – review & editing. **FS:** Conceptualization, Funding acquisition, Supervision, Writing – review & editing. **SD:** Conceptualization, Funding acquisition, Methodology, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Fund of Scientific Research-Flanders (FWO) (Grant number G.0247.09N) and the University of Antwerp through the GOA biofilm project 25624 (GOA BOF UA 2011-2014) both to SD. DHa, SD, and FS acknowledge the Antwerp University Research Fund for the Concerted Research Actions grant (BOF-GOA 4D protein structure). DHo acknowledges funding from the Swiss National Science Foundation (Grant 31003A 173000). FG, LM, SVD, and SD would like to thank Prof. Dr. Karolien De Wael (University of Antwerp), Prof. M. Bolognesi and Prof. M. Nardini (University of Milano) and Prof. A. Pesce (University of Genova) for their contribution to the initial part of this work. Furthermore, we also want to acknowledge Herald Berghmans (University of Antwerp) for repeating the auto-oxidation experiment and determining the molar extinction coefficient of full-length *Gs*GCS. We also thank the reviewers for their detailed and constructive scrutiny which allowed us to correct a mistake in the original manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.03.031.

References

[1] Freitas TAK, Saito JA, Wan X, Hou S, Alam M. Protoglobin and globin-coupled sensors. *Smallest Biomol Diatomics Interact Heme Proteins* 2008;175–202. https://doi.org/10.1016/B978-044452839-1.50008-5.
[2] Freitas TAK, Hou S, Alam M. The diversity of globin-coupled sensors. FEBS Lett 2003;552:99–104.
[3] Hou S et al. Globin-coupled sensors: a class of heme-containing sensors in archaea and bacteria. Proc Natl Acad Sci 2001;98:9353–8.
[4] Walker JA, Rivera S, Weinert EE. Mechanism and role of globin-coupled sensor signalling. Adv Microb Physiol 2017;71.
[5] Hou S et al. A globin-coupled oxygen sensor from the facultatively alkaliphilic Bacillus halodurans C-125. Extremophiles 2001;5:351–4.
[6] Stock AM, Robinson VL, Goudreau PN. Two-component signal transduction. Annu Rev Biochem 2000;69:183–215.
[7] Gell DA. Structure and function of haemoglobins. Blood Cells Mol Dis 2018;70:13–42.
[8] Storz JF. Hemoglobin: Insights into protein structure, function, and evolution. Hemoglobin: Insights into Protein Structure, Function, and Evolution (Oxford University Press, 2018). doi:10.1093/oso/9780198810681.001.0001.
[9] Pesce A, Bolognesi M, Nardini M. Protoglobin: structure and ligand-binding properties. Adv Microb Physiol 2013;63:79–96.
[10] Vinogradov SN et al. Three globin lineages belonging to two structural classes in genomes from the three kingdoms of life. Proc Natl Acad Sci USA 2005;102:11385–9.
[11] Vinogradov SN et al. A phylogenomic profile of globins. BMC Evol Biol 2006;6:31.
[12] Tarnawski M, Barends TRM, Schlichting I. Structural analysis of an oxygen-regulated diguanylate cyclase. Acta Crystallogr D Biol Crystallogr 2015;71:2158–77.
[13] Martínková M, Kitanishi K, Shimizu T. Heme-based globin-coupled oxygen sensors: Linking oxygen binding to functional regulation of diguanylate cyclase, histidine kinase, and methyl-accepting chemotaxis. J Biol Chem 2013;288:27702–11.
[14] Germani F, Moens L, Dewilde S. Haem-based sensors: a still growing old superfamily. Adv Microb Physiol 2013;63:1–47.
[15] Zhang W, Phillips Jr GN. Structure of the oxygen sensor in Bacillus subtilis: Signal transduction of chemotaxis by control of symmetry. Structure 2003;11:1097–110.
[16] Pesce A et al. HisE11 and HisF8 provide bis-histidyl heme hexa-coordination in the globin domain of geobacter sulfurreducens globin-coupled sensor. J Mol Biol 2009;386:246–60.
[17] Germani F et al. Structural and functional characterization of the globin-coupled sensors of Azotobacter vinelandii and Bordetella pertussis. Antioxidants Redox Signal 2020;32:378–95.
[18] Dahlstrom KM, O'Toole GA. A symphony of cyclases: specificity in diguanylate cyclase signaling. Annu Rev Microbiol 2017;71:179–95.
[19] Halls ML, Cooper DMF. Adenylyl cyclase signalling complexes – Pharmacological challenges and opportunities. Pharmacol Ther 2017;172:171–80.

[20] Kitanishi K et al. Identification and functional and spectral characterization of a globin-coupled histidine kinase from Anaeromyxobacter sp. Fw109-5. J Biol Chem 2011;286:35522–34.
[21] Stranava M et al. Coordination and redox state-dependent structural changes of the heme-based oxygen sensor AfGcHK associated with intraprotein signal transduction. J Biol Chem 2017;292:20921–35.
[22] Stranava M et al. Structural characterization of the heme-based oxygen sensor, AfGcHK, its interactions with the cognate response regulator, and their combined mechanism of action in a bacterial two-component signaling system. Proteins Struct Funct Bioinf 2016;84:1375–89.
[23] Skalova T et al. Disruption of the dimerization interface of the sensing domain in the dimeric heme-based oxygen sensor AfGcHK abolishes bacterial signal transduction. J Biol Chem 2020;295:1587–97.
[24] Tilleman L et al. A globin domain in a neuronal transmembrane receptor of Caenorhabditis elegans and Ascaris suum: molecular modeling and functional properties. J Biol Chem 2015;290:10336–52.
[25] Lovley DR. Cleaning up with genomics: applying molecular biology to bioremediation. Nat Rev Microbiol 2003;1:35–44.
[26] Methe BA et al. Genome of geobacter sulfurreducens: metal reduction in subsurface environments. Science (80-.) 2003;302:1967–9.
[27] Lin WC, Coppi MV, Lovley DR. Geobacter sulfurreducens can grow with oxygen as a terminal electron acceptor. Appl Environ Microbiol 2004;70:2525–8.
[28] Ding YHR et al. The proteome of dissimilatory metal-reducing microorganism Geobacter sulfurreducens under various growth conditions. Biochim Biophys Acta - Proteins Proteomics 2006;1764:1198–206.
[29] Desmet F et al. The heme pocket of the globin domain of the globin-coupled sensor of Geobacter sulfurreducens – An EPR study. J Inorg Biochem 2010;104:1022–8.
[30] Nardini M et al. Archaeal protoglobin structure indicates new ligand diffusion paths and modulation of haem-reactivity. EMBO Rep 2008;9:157–63.
[31] Dewilde S et al. Expression, purification, and crystallization of neuro- and cytoglobin. Methods Enzymol 2008;436:341–57.
[32] Pesce A et al. Neuroglobin and cytoglobin Fresh blood for the vertebrate globin family.. 1146 EMBO Rep 2002;3:1146–51.
[33] Sobott F, Hernández H, McCammon MG, Tito MA, Robinson CV. A tandem mass spectrometer for improved transmission and analysis of large macromolecular assemblies. Anal Chem 2002;74:1402–7.
[34] Bush MF et al. Collision cross sections of proteins and their complexes: a calibration framework and database for gas-phase structural biology. Anal Chem 2010;82:9557–65.
[35] Tang G et al. EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol 2007;157:38–46.
[36] Scheres SHW. A Bayesian view on cryo-EM structure determination. J Mol Biol 2012;415:406–18.
[37] Scheres SHW. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 2012;180:519–30.
[38] El-Gebali S et al. The Pfam protein families database in 2019. Nucleic Acids Res 2019;47:D427–32.
[39] Potter SC et al. HMMER web server: 2018 update. Nucleic Acids Res 2018;46:W200–4.
[40] Bateman A. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res 2019;47:D506–15.
[41] Katoh K, Rozewicki J, Yamada KD. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform 2018;20:1160–6.
[42] Yang J et al. The I-TASSER suite: protein structure and function prediction. Nat Methods 2014;12:7–8.
[43] Pettersen EF et al. UCSF Chimera – A visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–12.
[44] Šali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 1993;234:779–815.
[45] Kim SK et al. Dual transcriptional regulation of the Escherichia coli phosphate-starvation-inducible psiE gene of the phosphate regulon by PhoB and the cyclic AMP (cAMP)-cAMP receptor protein complex. J Bacteriol 2000;182:5596–9.
[46] Rivera S et al. Structural insights into oxygen-dependent signal transduction within globin coupled sensors. Inorg Chem 2018;57:14386–95.
[47] Horie S, Hasumi H, Takizawa N. Heme-linked spectral changes of the protein moiety of hemoproteins in the near ultraviolet region. J Biochem 1985;97:281–93.
[48] Jensen KP, Ryde U. How O2 binds to heme. Reasons for rapid binding and spin inversion. J Biol Chem 2004;279:14561–9.
[49] Dewilde S et al. Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. J Biol Chem 2001;276:38949–55.
[50] Sawai H et al. Characterization of the heme environmental structure of cytoglobin, a fourth globin in humans. Biochemistry 2003;42:5133–42.
[51] Yoon J et al. Structure and properties of a bis-histidyl ligated globin from Caenorhabditis elegans. Biochemistry 2010;49:5662–70.
[52] Tilleman L et al. An N-myristoylated globin with a redox-sensing function that regulates the defecation cycle in Caenorhabditis elegans. PLoS ONE 2012;7:e48768.
[53] Aono S et al. Resonance Raman and ligand binding studies of the oxygen-sensing signal transducer protein HemAT from Bacillus subtilis. J Biol Chem 2002;277:13528–38.
[54] Lou BS et al. Resonance Raman studies indicate a unique heme active site in prostaglandin H synthase. Biochemistry 2000;39:12424–34.
[55] Hu S, Smith KM, Spiro TG. Assignment of protoheme Resonance Raman spectrum by heme labeling in myoglobin. J Am Chem Soc 1996;118:12638–46.
[56] Cerda-Colón JF, Silfa E, López-Garriga J. Unusual rocking freedom of the heme in the hydrogen sulfide-binding hemoglobin from Lucina pectinata. J Am Chem Soc 1998;120:9312–7.
[57] Terner J, Sitter AJ, Reczek CM. Resonance Raman spectroscopic characterizations of horseradish peroxidase. Observations of the FeIV = O stretching vibration of Compound II. Biochim Biophys Acta (BBA)/Protein Struct Mol 1985;828:73–80.
[58] Chuang WJ, Van Wart HE. Resonance Raman spectra of horseradish peroxidase and bovine liver catalase compound I species. Evidence for predominant 2A (2u) π-cation radical ground state configurations. J Biol Chem 1992;267:13293–301.
[59] Jin Y, Nagai M, Nagai Y, Nagatomo S, Kitagawa T. Heme structures of five variants of hemoglobin M probed by resonance Raman spectroscopy. Biochemistry 2004;43:8517–27.
[60] Cuypers B et al. Antarctic fish versus human cytoglobins – The same but yet so different. J Inorg Biochem 2017;173:66–78.
[61] Han S, Rousseau DL, Giacometti G, Brunori M. Metastable intermediates in myoglobin at low pH. Proc Natl Acad Sci USA 1990;87:205–9.
[62] van den Heuvel RH, Heck AJ. Native protein mass spectrometry: from intact oligomers to functional machineries. Curr Opin Chem Biol 2004;8:519–26.
[63] Ruotolo BT, Robinson CV. Aspects of native proteins are retained in vacuum. Curr Opin Chem Biol 2006;10:402–8.
[64] Heck AJR. Native mass spectrometry: a bridge between interactomics and structural biology. Nat Methods 2008;5:927–33.
[65] Lorenzen K, van Duijn E. Native mass spectrometry as a tool in structural biology. Curr. Protoc. Protein Sci. 2020;62:17.12.1-17.12.17.
[66] Aebersold R, Mann M. Mass spectrometry-based proteomics. Nature 2003;422:198–207.
[67] Leney AC, Heck AJR. Native mass spectrometry: what is in the name?. J Am Soc Mass Spectrom 2017;28:5–13.
[68] Wilm M, Mann M. Analytical properties of the nanoelectrospray ion source. Anal Chem 1996;68:1–8.
[69] Konijnenberg A, Butterer A, Sobott F. Native ion mobility-mass spectrometry and related methods in structural biology. Biochim Biophys Acta – Proteins Proteomics 2013;1834:1239–56.
[70] Barrera, N. P., Di Bartolo, N., Booth, P. J. & Robinson, C. V. Micelles protect membrane complexes from solution to vacuum. science (80-.). 321, 243–246 (2008).
[71] Laganowsky A, Reading E, Hopper JTS, Robinson CV. Mass spectrometry of intact membrane protein complexes. Nat Protoc 2013;8:639–51.
[72] Borysik AJ, Hewitt DJ, Robinson CV. Detergent release prolongs the lifetime of native-like membrane protein conformations in the gas-phase. J Am Chem Soc 2013;135:6078–83.
[73] Reading E et al. The role of the detergent micelle in preserving the structure of membrane proteins in the gas phase. Angew Chem Int Ed. 2015;54:4577–81.
[74] Testa L, Brocca S, Grandori R. Charge-surface correlation in electrospray ionization of folded and unfolded proteins. Anal Chem 2011;83:6459–63.
[75] Lin SH, Guidotti G. Chapter 35 Purification of Membrane Proteins. in *Methods in Enzymology* 2009;vol. 463:619–29.
[76] Lanucara F, Holman SW, Gray CJ, Eyers CE. The power of ion mobility-mass spectrometry for structural characterization and the study of conformational dynamics. Nat Chem 2014;6:281–94.
[77] Smith DP et al. Deciphering drift time measurements from travelling wave ion mobility spectrometry-mass spectrometry studies. Eur J Mass Spectrom 2009;15:113–30.
[78] Dixit SM, Polasky DA, Ruotolo BT. Collision induced unfolding of isolated proteins in the gas phase: past, present, and future. Curr Opin Chem Biol 2018;42:93–100.
[79] Hopper JTS, Oldham NJ. Collision induced unfolding of protein ions in the gas phase studied by ion mobility-mass spectrometry: the effect of ligand binding on conformational stability. J Am Soc Mass Spectrom 2009;20:1851–8.
[80] Niu S, Ruotolo BT. Collisional unfolding of multiprotein complexes reveals cooperative stabilization upon ligand binding. Protein Sci 2015;24:1272–81.
[81] Goldie KN et al. Cryo-electron microscopy of membrane proteins. Methods Mol Biol (Clifton N.J.) 2014;1117:325–41.
[82] Thonghin N, Kargas V, Clews J, Ford RC. Cryo-electron microscopy of membrane proteins. Methods 2018;147:176–86.
[83] Blank M, Burmester T. Widespread occurrence of N-terminal acylation in animal globins and possible origin of respiratory globins from a membrane-bound ancestor. Mol Biol Evol 2012;29:3553–61.
[84] Ertas B, Kiger L, Blank M, Marden MC, Burmester T. A membrane-bound hemoglobin from gills of the green shore crab Carcinus maenas. J Biol Chem 2011;286:3185–93.
[85] Blank M et al. A membrane-bound vertebrate globin. PLoS ONE 2011;6:e25292.
[86] Vinogradov SN, Tinajero-Trejo M, Poole RK, Hoogewijs D. Bacterial and archaeal globins — A revised perspective. Biochim Biophys Acta – Proteins Proteomics 2013;1834:1789–800.
[87] Vinogradov SN et al. Microbial eukaryote globins. *Advances in Microbial Physiology* 2013;vol. 63:391–446.

[88] Hade MD, Kaur J, Chakraborti PK, Dikshit KL. Multidomain truncated hemoglobins: new members of the globin family exhibiting tandem repeats of globin units and domain fusion. IUBMB Life 2017;69:479–88.
[89] Bayburt TH, Grinkova YV, Sligar SG. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. Nano Lett 2002;2:853–6.
[90] Knowles TJ et al. Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. J Am Chem Soc 2009;131:7484–5.
[91] Lengalova A et al. Kinetic analysis of a globin-coupled diguanylate cyclase, YddV: Effects of heme iron redox state, axial ligands, and heme distal mutations on catalysis. J Inorg Biochem 2019;201.
[92] Rivera S, Burns JL, Vansuch GE, Chica B, Weinert EE. Globin domain interactions control heme pocket conformation and oligomerization of globin coupled sensors. J Inorg Biochem 2016;164:70–6.
[93] Burns JL, Douglas Deer D, Weinert EE. Oligomeric state affects oxygen dissociation and diguanylate cyclase activity of globin coupled sensors. Mol BioSyst 2014;10:2823–6.
[94] Bond DR, Lovley DR. Electricity production by geobacter sulfurreducens attached to electrodes. Appl Environ Microbiol 2003;69:1548–55.
[95] Poddar S, Khurana S. Geobacter: The Electric Microbe! Efficient microbial fuel cells to generate clean, cheap electricity. *Indian J Microbiol* 2011;51:240–1.
[96] Morgado L, Dantas JM, Bruix M, Londer YY, Salgueiro CA. Fine tuning of redox networks on multiheme cytochromes from geobacter sulfurreducens drives physiological electron/proton energy transduction. Bioinorg Chem Appl 2012;2012:298739.
[97] Tilleman L et al. Globins in Caenorhabditis elegans. IUBMB Life 2011;63:166–74.
[98] Kiger L et al. Electron transfer function versus oxygen delivery: a comparative study for several hexacoordinated globins across the Animal Kingdom. PLoS ONE 2011;6:e20478.
[99] Geuens E et al. Globin-like proteins in Caenorhabditis elegans: in vivo localization, ligand binding and structural properties. BMC Biochem 2010;11:17.