Forum Original Research Communication

Structural and functional characterization of the globin-coupled sensors of *Azotobacter vinelandii* and *Bordetella pertussis*

Francesca Germani[1], Marco Nardini[2], Amy De Schutter[3], Bert Cuypers[3], Herald Berghmans[1], Marie-Louise Van Hauwaert[1], Stefano Bruno[4], Andrea Mozzarelli[4], Luc Moens[1], Sabine Van Doorslaer[3], Martino Bolognesi[2], Alessandra Pesce[5,*], and Sylvia Dewilde[1,*]

[1] Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium

[2] Department of Biosciences, University of Milano, Via Celoria 26, I-20133 Milan, Italy

[3] Department of Physics, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium

[4] Department of Pharmacy, University of Parma, Parco Area delle Scienze 23/A, 43124 Parma, Italy

[5] Department of Physics, University of Genova, Via Dodecaneso 33, I-16146 Genova, Italy

[*]Authors to whom correspondence should be addressed; *e-mail*: sylvia.dewilde@uantwerpen.be, *tel*: +32-03-265-2323; *e-mail*: pesce@fisica.unige.it, *tel*: +39-010-353-6243

Running title: Structure and function of globin-coupled sensors.

Keywords: heme-based sensor, oxygen affinity, c-di-GMP and enzyme specificity, biofilm, crystal structure.

word count (excluding references and figure legends): 8288

reference numbers: 75

number of greyscale illustrations: 1 Figure and 4 Tables

number of color illustrations: 6 Figures

Abstract

Aims: Structural and functional characterization of the globin coupled sensors (GCS) of *Azotobacter vinelandii* (*Av*GReg) and *Bordetella pertussis* (*Bpe*GReg).

Results: UV/Vis and RR spectroscopies confirm the presence in *Av*GReg and *Bpe*GReg of a globin domain capable of reversible gaseous ligand binding. In *Av*GReg an influence of the transmitter domain on the heme proximal region of the globin domain can be seen, and k'_{CO} is higher than for other GCSs. The O_2 binding kinetics suggest the presence of an open and a closed conformation. As for *Bpe*GReg, the fully oxygenated *Av*GReg show a very high DGC activity. The CO rebinding to *Bpe*GReg indicates that intra- and intermolecular interactions influence the ligand binding. The globin domains of both proteins (*Av*GReg-Gb and *Bpe*GReg-Gb*) share the same GCS fold, a similar proximal but a different distal side structure. They homodimerize through a G-H helical bundle as in other GCS. However, *Bpe*GReg-Gb* shows also a second dimerization mode.

***Innovation*:** This paper extends our knowledge on the GCS proteins and contributes to a better understanding of the GCSs' role in the formation of bacterial biofilms.

Conclusions: *Av*GReg and *Bpe*GReg conform to the GCS family, share a similar overall structure but they have different properties in terms of the ligand binding. In particular, *Av*GReg shows an open and closed conformation that in the latter form will very tightly bind oxygen. *Bpe*GReg has only one closed conformation. In both proteins, it is the fully oxygenated GCS form that catalyzes the production of the second messenger.

Introduction

Globin coupled sensors (GCSs) are heme-based molecules widespread in Bacteria and Archaea, consisting of chimeric proteins characterized by an N-terminal globin sensor domain coupled to a variety of C-terminal transmitter domains. Classification based on the function of the latter distinguishes: diguanylate cyclase (DGC), phosphodiesterase (PDE), histidine kinase, methyl-accepting chemotaxis and a transmembrane function (20,33).

The globin domain exhibits an efficient sensor function based on the reactivity of the heme iron atom (56,69). Its structure is a variant of the 3-over-3 helical fold with an extra N-terminal Z-helix and no D-helix (44,63,70,75). Binding of diatomic gaseous ligands causes a conformational change of the sensor domain that is transferred *via* a linker region to the transmitter domain, regulating its activity. A comprehensive model for a full-length GCS has been proposed for *E. coli* diguanylate cyclase DosC (*Ec*DosC) (63), although the precise mechanism of signal transduction within the molecule has still to be elucidated further.

Most known GCS have a transmitter domain mainly consisting of a GGDEF domain with DGC activity. GCS with DGC (or PDE) activity produce (or degrade) cyclic-di-(3',5')-GMP (c-di-GMP), a key second messenger involved in the conversion from a free-living into a sessile-living bacterial population through biofilm formation (23,32,57) and modulation of pathogenic factors (53). Bacteria in a biofilm show resistance to antibiotics and the host immune system and often result in persistent and chronic diseases (16). As such, biofilms are considered an important medical and industrial issue and, consequently, GCS with DGC activity are becoming of increasing scientific interest. Two examples of such biofilm-forming bacteria with a GCS are *Azotobacter vinelandii* and *Bordetella pertussis*.

Azotobacter vinelandii is a Gram-negative, alginate-biofilm-producing bacterium, capable of diazotrophic growth and N-fixation in the presence of O_2 (48). As such it has economic importance and it is widely used as a research model system. Genomic analysis revealed the presence of a GCS (*Av*GReg) (55,64). *Av*GReg is a 472 residues soluble protein consisting of a 178 residues N-terminal

globin sensor domain (*Av*GReg-Gb) and a 170 residues C-terminal GGDEF transmitter domain endowed with a DGC activity.

Bordetella pertussis is a Gram-negative, aerobic, biofilm forming bacterial pathogen (9,36,73). It is the causative of pertussis or whooping cough a highly contagious human disease that still causes, yearly, many deaths among children worldwide (World Health Organization: http://www.who.int/topics/pertussis/en/). In the *Bordetella pertussis* genome, one GCS with gene regulating function has been identified, namely *Bpe*GReg (3, 4). As *Av*GReg, *Bpe*GReg is a di-domain molecule (475 residues) consisting of a globin-like sensor domain (residues:1-155) covalently attached by a linker region (residues:156-296) to a transmitter domain (residues: 297-475) with DGC activity. This enzymatic activity is regulated by ligand (O_2) binding to the sensor domain, product (c-di-GMP) binding to an inhibitory site (RxxD motive in the transmitter domain), and the oligomerization state (mono- di- or tetramer) of the full GCS protein (6,7,50,70). The participation of c-di-GMP in the process of biofilm formation of *Bpe*GReg has been demonstrated *in vivo* (70). The iron bound O_2 is stabilized by a TyrB10 and SerE11 residue of the globin sensor domain. Dimerization of this domain involves the G and H helices (7,50).

*Av*GReg and *Bpe*GReg not only merit scientific attention due to their respective economic and medical importance, but a thorough comparison of the structural and functional properties of these two molecules can contribute to a better understanding of the general role of GCSs in the signaling mechanism associated with the formation of bacterial biofilms. We have therefore combined different biochemical and biophysical techniques to unravel the structural-function relationship of these two proteins. The study involves the comparative analysis of the *Av*GReg and *Bpe*GReg full-length molecules, their respective globin domains and some mutants (*Av*GRegYB10F and Cys-mutant *Bp*eReg*). In the latter notation, asterisk indicates the Cys-Ser mutation of Cys16, Cys45, Cys114 and Cys154, needed for crystallization purposes, see Materials and Methods for details. Moreover, the full-length *Av*GReg was for the first time obtained *via in-vivo* folding, which led to a molecule

with marked differences in the ligand binding properties than reported earlier for *in-vitro* folded *Av*GReg (from now on referred to as †*Av*GReg) (64).

Results

UV/Vis spectroscopy

The absorbance maxima of the UV/VIS spectra of full-length *Av*GReg and *Bpe*GReg, in different ligation forms are reported in Table 1. The spectra of the 'as purified' forms are typical for low-spin O_2-ligated ferrous globins, while those of the reduced forms are characteristic for ferrous high-spin pentacoordination of the heme. The absorbance maxima do not change for the truncated proteins consisting only of the globin domains (*Av*GReg-Gb and *Bpe*GReg-Gb), nor for the Cys mutant (*Bpe*GReg*) and the YB10F mutant of *Av*GReg (data not shown). Overall these results confirm the presence of a globin domain capable of gaseous ligand binding and they are comparable to what was already reported for GCS in the literature (50,70).

Resonance Raman spectroscopy

The resonance Raman (RR) spectra for *Av*GReg(-Gb) and *Bpe*GReg(-Gb) were recorded in the as-purified (O_2-bound ferrous) state, and in the dithionite-reduced (deoxy ferrous) state (Fig. 1). The corresponding spectra of *Bpe*GReg* and *Bpe*GReg-Gb* are depicted in Figure S1. The high frequency region (1300-1800 cm^{-1}) of the spectrum provides insights on the oxidation (ν_4), coordination (ν_3), and spin (ν_2) states of the heme iron atom (59) (Fig. 1B,D, Table 2). The observed marker frequencies are in accordance with those reported for *Sw*Mb (60,61) and different gas sensor proteins, such as the *Bacillus subtilis* heme-based aerotaxis transducer (*Bs*HemAT) (2), the isolated globin domain of *Ec*DosC (*Ec*DosC-heme) (28), and the globin-coupled histidine kinase from *Anaeromyxobacter sp.* Fw109-5 (*Af*GcHK) (29) (Table 2). They confirm the assignment of the ferrous form of *Av*GReg(-Gb) and *Bpe*GReg to a high-spin pentacoordinated heme site. The as-purified

proteins are all in the O_2-ligated ferrous form. Interestingly, the RR spectrum of *Bpe*GReg (Fig. 1D,a) reveals partial photolysis of the O_2 ligand (two ν_4 bands at 1359 cm^{-1} and 1373 cm^{-1}, agreeing with the deoxy and oxy ferrous form). The partial photolysis hampers the analysis of the low-frequency part of the RR spectra of O_2-*Bpe*GReg, which will therefore not be considered in the below discussion.

The low frequency region (200-900 cm^{-1}) reveals in-plane and out-of-plane vibrational modes of the heme (59). The $\nu_{Fe\text{-}His}$ is observed at 227 cm^{-1} for the ferrous unligated *Av*GReg (Fig. 1A,b). This value is similar to that of ferrous *Bs*HemAT (225 cm^{-1}) (2) and *Ec*DosC-heme (227 cm^{-1}) (28). The $\nu_{Fe\text{-}His}$ peak broadens and shifts to ~214 cm^{-1} for *Av*GReg-Gb (Fig. 1A,d). For all deoxy ferrous *Bpe*GReg* and *Bpe*GReg-Gb* forms, the $\nu_{Fe\text{-}His}$ mode is observed as a broad peak around 229-233 cm^{-1} (Fig. 1C,b,d, Fig. S1Ab,d). The $\nu_{Fe\text{-}His}$ stretching mode is known to depend on different parameters (2,39,43). The higher $\nu_{Fe\text{-}His}$ mode found in ferrous unligated *Av*GReg and *Bpe*GReg can indicate a stronger H-bonding to the proximal His N^{δ} nitrogen, less strain on the Fe-His bond, and/or a smaller dihedral angle ϕ between imidazole plane and the nearest N(pyrrole)-Fe-N(pyrrole) axis. Thus, in keeping with the sensor functionality, the $\nu_{Fe\text{-}His}$ difference observed between *Av*GReg and *Av*GReg-Gb points to an influence of the signaling domain on the heme proximal region of the globin domain. Similarly, the γ_7 pyrrole bending mode, observed usually between 288 - 313 cm^{-1} in heme proteins and indicative of an out-of-plane distortion of the heme, changes slightly from a sharp peak at 301 cm^{-1} for deoxy *Av*GReg to a broadened peak at 299 cm^{-1} for deoxy ferrous *Av*GReg-Gb (Fig. 1A). This is also found for the deoxy form of horse heart Mb (302 cm^{-1}) (25). In both cases, O_2 ligation causes a reduction of the out-of-plane distortion (less pronounced γ_7 mode). A similar trend is observed for the γ_7 bending modes of *Bpe*GReg(-Gb) (Fig. 1C).

The heme propionate and vinyl modes reveal important information about the stabilization of the heme by the surrounding protein. The vinyl $\delta(C_\beta C_a C_{b2,4})$ bending modes more or less coincide and give rise to broadened peaks around 421, 414 and 417 cm^{-1} for O_2-*Bpe*GReg-Gb, O_2-*Bpe*GReg* and O_2-*Bpe*GReg-Gb*, respectively (Fig. 1C,c, Fig. S3A,a,c). This indicates that the out-of-plane

distortions influence each of the pyrrole units in a similar way (68). The increase in wavenumber of this mode for the globin domain compared to the full length domain, indicates an effect of the transmitter domain on the globin domain. This effect was not observed for the *Av*GReg proteins (415 cm^{-1} (Fig. 1Aa,c)). Deoxygenation leads for *Bpe*GReg* and *Bpe*GReg-Gb* to $\delta(C_\beta C_a C_{b2,4})$ bending modes at 412 cm^{-1}, whereby the peaks are clearly less broad than in the case of the oxygenated proteins (Fig. 1C, Fig. S1). This means that O_2 ligation influences the stabilization of the two vinyl groups in a different way. Again, this effect is far less pronounced in the *Av*GReg case (Fig. 1A).

O_2-*Av*GReg and O_2-*Av*GReg-Gb display the $\delta(C_\beta C_c C_d)$ propionate bending mode at 381 cm^{-1} and 377 cm^{-1}, respectively (Fig. 1Aa,c), while those of O_2-*Bpe*GReg-Gb, O_2-*Bpe*GReg* and O_2-*Bpe*GReg-Gb* are found at 382, 375 and 378 cm^{-1}, respectively (Fig. 2C,c, Fig. S1A,a,c). This suggests a moderate to strong electrostatic interaction of the heme propionate groups with the surrounding amino-acid side chains. Interestingly, the $\delta(C_\beta C_c C_d)$ mode is found at 363 cm^{-1} for ferrous *Av*GReg, and 380 cm^{-1} for ferrous *Av*GReg-Gb, respectively. The former low value reflects the pronounced effect of the second domain on the heme surrounding upon ligand binding. While this mode is still clearly detectable for the deoxy ferrous state of *A*vGReg(-Gb) (Fig. 1A,b,d), the $\delta(C_\beta C_c C_d)$ mode can only be weakly observed for deoxy ferrous *Bpe*GReg-Gb (Fig. 1Cd, small peak at 379 cm^{-1}).

For the *in vivo* folded O_2-*Av*GReg and O_2-*Av*GReg-Gb, two $\nu_{Fe\text{-}O2}$ modes were found (Fig. 1Ab,d, Table 2). The first mode is located at 563 cm^{-1} (561 cm^{-1}), while the second mode is located at 573 cm^{-1} (see Fig. S2, for a details of this spectral region in the RR spectrum of O_2-*Av*GReg). The 573 cm^{-1} mode is similar to the one observed for other globins, such as Mb which show $\nu_{Fe\text{-}O2}$ around 570 cm^{-1} (12,24,27,58). Lower Fe-O_2 stretching modes of 557-560 cm^{-1} have been found for the oxygenated forms of *Bs*HemAT (2) and *Af*GcHK (29) (Table 2). In this respect, previous studies on *Mycobacterium tuberculosis* (74) and *Chlamydomonas eugametos* 2/2Hbs (11) indicate that the low $\nu_{Fe\text{-}O2}$ mode is related to a strong hydrogen bond between the bound O_2 molecule and residue TyrB10.

As such, this low ν_{Fe-O2} mode suggests a strong stabilization of the bound O_2 through hydrogen bonding in *Av*GReg, while the 573 cm^{-1} indicates the co-existence of a more open configuration in which the O_2 is not stabilized. For the oxygenated *Bpe*GReg* variants, only a ν_{Fe-O2} mode in the 556-560 cm^{-1} region is found (Table 2, Fig. 1, Fig. S1), indicating that the O_2 ligand is strongly hydrogen bonded to distal amino-acids with no open conformation present. This will be shown to correlate with the kinetics data. A very recent FTIR study on the CO-ligated ferrous *Bpe*GReg and its Tyr43Phe variant revealed that one of the CO-bound conformers is strongly hydrogen-bonded to the distal Tyr(43)B10, in line with the here observed H-bonding of the O_2 ligand (51).

In earlier work, we already performed RR analyses on †*Av*GReg(-Gb) (64). While the RR data obtained in both studies are similar for the globin domain of the protein, the RR and electron paramagnetic resonance data of †*Av*GReg showed the protein to be in the ferric form due to the *in-vitro* refolding. Figure S3 (supplementary material) shows the optical absorption and RR data for *Av*GReg oxidized with ferricyanide. It shows that (partial) oxidation of the protein can be obtained chemically, but that it also leads to partial denaturation of the protein as can be derived from the optical absorption peak at 646 nm. The spectra in Figure S3 for oxidized *Av*GReg agree with those previously obtained for †*Av*GReg, where also signs of partial denaturation were found (64).

Ligand binding kinetics on AvGReg proteins

*Av*GReg proteins and the *Av*GReg YB10F mutant were characterized in terms of the CO and O_2 binding equilibria and their kinetics as described in Materials and Methods. In the whole CO concentration range explored in the flash-photolysis experiments, the time courses of the CO binding to *Av*GReg can be described by a single exponential, allowing the use of a minimum reaction mechanism (Scheme 1, Materials and Methods (64)). The value of the kinetic parameter that defines CO binding (k'_{CO} = (0.48 ± 0.01) x 10^6 M^{-1} s^{-1}) is very similar to the value reported for *Sw*Mb (Table 3; Fig. 2). Only a minor influence is noticed of the protein truncation or YB10F mutation. However,

the other GCSs characterized to date show a very moderately (*Bs*HemAT) to significantly lower (*Ec*DosC and *Af*GcHK) CO association rate constant (Table 3).

The results of the flash-photolysis O_2/CO competition experiment can be described with a second order reaction mechanism (see Materials and methods, Scheme 2). The O_2 association/dissociation rate constants of *Av*GReg, as well as the constant for CO dissociation from *Av*GReg (Table 3; Fig. 2C,D), have been determined using Equations 1 and 2 (see Materials and methods). Interestingly, stopped-flow measurements reveal the presence of a second phase in the O_2/CO replacement mechanism (Table 3; Fig. 2B). Therefore, we can conclude that *Av*GReg is characterized by a fast O_2 binding (k'_{O2} = (12.2 ± 1.5) x 10^6 M^{-1} s^{-1}) and two O_2 dissociation constants (k_{O2} = (1732 ± 377) s^{-1} and k_{O2} = (0.53 ± 0.05) s^{-1}).

The O_2 association and dissociation rate constants we present here for the *Av*GReg differ significantly from the ones measured previously for the *in vitro* refolded protein †*Av*GReg (Table 3). In particular, †*Av*GReg appears to release dioxygen 160 times slower (high k_{O2}) and 1.3 times faster (low k_{O2}).

While two distinct k'_{O2} values were measured for *Af*GcHK (29), a biphasic O_2 dissociation was observed for *Bs*HemAT (75) as for *Av*Greg (Table 3). A difference of more than 200 fold is revealed when comparing the higher K_d values of *Av*GReg and *Af*GcHK, while the lower K_d values are in the same order of magnitude for the two GCSs. This indicates the presence of an extremely stable Fe-O_2 complex in both cases. In contrast, we find that the higher K_d value of *Av*GReg and *Bs*HemAT are in the same order of magnitude, while the lower K_d of *Av*GReg is 100 times smaller than that of *Bs*HemAT. The latter observation highlights the very high O_2 affinity of *Av*GReg, being the highest among the other globin sensors characterized to date, suggesting that it can be (and remain) activated in the presence of traces amounts of O_2.

A likely explanation of these results lies in the observation of two $\nu_{Fe\text{-}O2}$ modes for O_2-ligated *Av*GReg in solution, linked to an *open* and *closed* conformation for the O_2 stabilization, responsible for the two dissociation constants, similarly as for *Bs*HemAT (75). Within this model, when the globin

domain is in the *open* conformation, O_2 is less stabilized inside the heme pocket, allowing a more efficient diffusion of CO from the outer environment, and faster O_2 displacement. On the other hand, in the *closed* conformation, O_2 is firmly bound in the heme pocket, and well protected from the solvent. The competition between O_2 and CO is impaired, thus O_2 displacement is significantly slower. The results for *Av*GRegYB10F indicate that Tyr(44)B10 plays an important role in this regulation (Table 3, Fig. S4). Not only does the fastest O_2 dissociation rate increase with an order of magnitude upon mutation, the dissociation rate from the *closed* confirmation increases with a factor of ~300.

Ligand binding kinetics on Bordetella pertussis proteins

The ligand binding parameters for CO and O_2 binding to *Bpe*GReg and *Bpe*GReg-Gb, are summarized in Table 3 and Fig. 3. Interestingly, the CO binding reactions of *Bpe*GReg and the isolated *Bpe*GReg-Gb show a different decay. The double exponential rebinding to *Bpe*GReg is defined with k'_{COfast} = 0.60 μM^{-1} s^{-1} and k'_{COslow} = 0.01 μM^{-1} s^{-1}, while the rate constant of the *Bpe*GReg-Gb single exponential reaction is k'_{CO} = 0.11 μM^{-1} s^{-1}. The explanation for this difference must be related to the presence of the transmitter domain, though we believe it goes beyond the mere structural effect of the absence/presence of this domain, as no major modification of the heme environment can be ascribed to the isolated globin domain. Considering also the dimerization process, we confirm the data by Rivera *et al*. (50) that the intra- and intermolecular interactions influence the ligand binding capabilities and, therefore, result in this CO affinity difference. Surprisingly, recent FTIR of *Bpe*GReg-Gb and *Bpe*GReg showed three CO stretches in both proteins, corresponding to an open and a closed (H-bonded) conformation and a pH-dependent conformer stabilized by structural water (51).

The O_2/CO competition studies and the O_2 association/dissociation rate constants of *Bpe*GReg have been determined as for *Av*GReg. With k'_{O2} = 24.02 μM^{-1} s^{-1} and k_{O2} = 4.0 s^{-1} (Table 3), *Bpe*GReg is capable of selective and more efficient O_2 binding when a gas mixture is present, as is also the case

for myoglobin (1). Moreover, the low K_d values for O_2 (K_d = 0.17 µM) indicate that this sensor will be saturated with O_2 in physiological conditions. This is in line with the RR data that reveal only one ν_{Fe-O2} mode agreeing with a strong hydrogen bonding and thus stabilization of the dioxygen ligand. Despite the fact that the reported k'_{O2} from Wan *et al*. (70) is lower, the results are in accordance with the reported transmitter domain activation *via* O_2 binding to the sensor domain (70). Indeed, the very high affinity of the full-length *Bpe*GReg suggests that this characteristic ensures a stable promotion of the *Bpe*GReg DGC activity, resulting in the typical chronic infection-related biofilm persistence (6,7,50,70).

Diguanylate cyclase activity of AvGReg

We measured and analyzed the DGC activity of *AvGReg* to complement the already published data on *Bpe*GReg (7,70). By means of an optimized DGC activity assay, we could collect reproducible time point series within few seconds from time zero to confidently determine the initial reaction rate. We tested numerous substrate concentrations, which revealed a high sensitivity of O_2-*Av*GReg to product inhibition, as already reported for *Bpe*GReg and *Ec*DosC (28, 70). To avoid bias due to this effect, we determined the enzyme specificity constant (calculated as k_{cat}/K_m) at GTP concentrations where no product inhibition was detected as described in Material and Methods (Fig. 4) and no cooperativity was seen.

For the oxygenated form of the *in vivo* folded *Av*GReg, the reaction reached completion after 20 minutes for all the substrate concentrations used (20 to 65 µM GTP) (data not shown), yielding a k_{cat}/K_m value of (1.13 ± 0.13) x 10^4 M^{-1} s^{-1}. O_2-*Bpe*GReg shows a 66 fold lower enzyme specificity as published by Burns *et al* (k_{cat}/K_m = 0.17 x 10^3 M^{-1} s^{-1}) (6) with completion reached after 30 minutes, while *Ec*DosC is characterized by a high autoxidation rate that hindered the estimation of the exact turnover number (28). The ferrous unligated form of *Av*GReg was also assayed for DGC activity. Comparison of the O_2-*Av*GReg and unligated *Av*GReg catalyzed reactions showed that in the latter case c-di-GMP synthesis is 50% decreased (Fig. 4), as previously reported also for *Bpe*GReg and

EcDosC (28,70). These results confirm the role of O_2 as DGC reaction trigger for this class of GCSs. Considering that *A. vinelandii* does not grow anaerobically and the very high O_2 affinity of *Av*GReg, it is likely that in physiological condition the *Av*GReg sensor is in the O_2-bound active state already in the presence of traces amounts of O_2.

*Tertiary and quaternary structures of AvGReg-Gb and BpeGReg-Gb**

The ferric aquo-met form of *Av*GReg-Gb and *Bpe*GReg-Gb* crystal structures have been determined at 2.83 Å and 3.20 Å resolution, respectively (two molecules per asymmetric unit, referred to as A and B in both cases). The final *Av*GReg-Gb model consists of 258 amino acids; no electron density is observed for residues 48-67 and 46-69 in chains A and B, respectively (B-C-E region). The refined *Bpe*GReg-Gb* structure comprises 306 amino acids; only residues 57-63 in chains B (C-E loop) show poor electron density. Data collection and refinement statistics are reported in Table 4.

The overall fold of *Av*GReg-Gb and *Bpe*GReg-Gb* conform to the modified globin fold typically found for GCS sensor domains (44,63,75) and protoglobin (40,45,46), consisting of 8 helices which, according to the classical globin nomenclature, have been named in alphabetical order from A to H, with a Z helix preceding the A at the N-terminus and the D helix absent (Figs. 5A,B). The two *Av*GReg-Gb and *Bpe*GReg-Gb* structures superimpose well, with a root mean squared deviation (rmsd) of ~1.3 Å over 126 Cα atom pairs, with structural differences localized at the N-terminal region (including the Z helix), in the B-E region, in the EF and GH loops (Fig. 5C).

Overall, two major features, both associated to the heme pocket, uniquely characterize *Av*GReg-Gb. On the heme-proximal side *Av*GReg-Gb exhibits one of the longest F-helix seen in globins (23 residues); this is due to the fact that the N-terminus of the F helix contains one more turn at the expense of the C-terminal turn of the E helix (Fig. 5A,C). On the heme-distal side the structural disorder at the B-C-E region (in both the A and B chains) results in the unwinding of almost the whole B-helix (Fig. 5A). The B-C-E region is instead well structured in *Bpe*GReg-Gb* (Figs. 5B,C).

The *Av*GReg-Gb and *Bpe*GReg-Gb* monomers assemble in homodimers with the interface residues belonging to the Z, G, and H helices of both A and B chains (Fig. 6A,B). Thus, the dimerization interface has a four α-helical G-H bundle as the core region (including 11 H-bonds and 7 salt bridges for *Av*GReg-Gb, and 11 H-bonds and 16 salt bridges for *Bpe*GReg-Gb*), while the adjacent Z helices provide mostly van der Waals interactions. The dimeric assembly observed in *Av*GReg-Gb and *Bpe*GReg-Gb* is similar to those found in *Gs*GCS[162] (PDB code: 2W31) (44), *Bs*HemAT (PDB codes: 1OR4, 1OR6) (75), *Ec*DosC (PDB code: 4ZVA, 4ZVB) (63), and in *Ma*Pgb (PDB codes 2VEB, 2VEE) (40). However, the dimeric interface surface area can vary significantly: it is large for *Ma*Pgb (~2,100 Å^2), *Bs*HemAT (~1,800 Å^2) and *Gs*GCS[162] (~1,700 Å^2), medium for in *Ec*DosC (~1,300 Å^2) and *Av*GReg-Gb (1,304 Å^2), and only 1,042 Å^2 for *Bpe*GReg-Gb*.

Furthermore, the precise orientation of the A/B subunits in the dimer is not the same in different GCS proteins. Indeed, when the *Av*GReg-Gb A chain is superimposed to that from the other GCSs, the match with the B chain is good for *Bpe*GReg-Gb* and *Ec*DosC, while the *Bs*HemAT B chain is rotated by ~20°, and *Gs*GCS[162] and *Ma*Pgb by ~30°, around an axis approximatively orthogonal to the G-H helices. Strikingly, a similar (albeit more pronounced, ~60°) different orientation of the subunits of the dimer has been found in another *Bpe*GReg-Gb* crystal form (same spacegroup *C*2, but different unit cell), determined at 1.55 Å resolution (*Bpe*GReg-Gb*-2) (Table 4). This second quaternary structure is symmetrical around a crystal axes, covers a contact surface of 1,305 Å^2 (similar to *Av*GReg-Gb and *Ec*DosC) with 14 hydrogen bonds and 10 salt bridges, and involves the Z, G, and H helices, and the CE corner (Fig. 6C,D).

It should be noted that the different dimerization mode cannot be ascribed to a different structure of the interfacing monomers, since the *Bpe*GReg-Gb* and *Bpe*GReg-Gb*-2 structures don't show significant deviations if compared monomer by monomer (rmsd values ranging between 0.50 Å and 0.80 Å, over 138-145 Cα pairs). The different quaternary assembly is most likely responsible for the minor structural deviations observed at the N-terminus of the E helix, in the FG, G, GH, and H regions between the *Bpe*GReg-Gb* and *Bpe*GReg-Gb*-2. In particular, the C-terminal His155 residue, buried

at the dimeric interface in *Bpe*GReg-Gb *, is solvent exposed in *Bpe*GReg-Gb*-2 and interacting with an additional heme group (probably released in solution by a fraction of *Bpe*GReg-Gb*) which becomes hexacoordinated between the end of two H-helices at the top of the dimeric interface (Fig. 6E,F).

Heme proximal side

*Av*GReg-Gb and *Bpe*GReg-Gb* share a very similar heme pocket proximal side, due to a very high sequence similarity in this protein region. The proximal heme pocket is lined by residues Ile(97)F4, Ile(100)F7, Ile(104), Ile(106), Val(111)G4, and Met(154)H17 in *Av*GReg-Gb, and Val(95)F4, Val(98)F7, Ile(102), Ile(104), Val(109)G4, and Met(153)H17 in *Bpe*GReg-Gb*, which surround and mostly contact the porphyrin ring.

Both *Av*GReg-Gb and *Bpe*GReg-Gb* show a loose coordination bond between the Fe atom and the proximal His residue with Fe–His(101)F8 NE2 distance of 2.39 Å and 2.50 Å in the A and B monomers of *Av*GReg-Gb, and Fe–His(99)F8 distance of 2.10 Å and 2.29 Å in A and B chains of *Bpe*GReg-Gb*. In *Av*GReg-Gb and *Bpe*GReg-Gb* the F8 imidazole ring lies in a staggered azimuthal orientation relative to the heme pyrrole N-atoms in chain A, and it is eclipsed in chain B along the direction of the heme pyrrole NA-NC atoms in *Av*GReg-Gb and NB-ND in *Bpe*GReg-Gb* (Fig. 7A). In the *Bpe*GReg-Gb*-2 His(99)F8 stabilizes the heme iron atom with a distance of 2.20 Å in chain A and 2.14 Å in chain B, being always eclipsed along the direction of the heme pyrrole NB and ND atoms.

Heme distal side

The heme distal side of *Av*GReg-Gb and *Bpe*GReg-Gb* varies considerably, but both reveal a high conformational flexibility of the B-C-E region. In *Av*GReg-Gb the heme distal pocket residues Tyr(42)B8, Phe(43)B9, Tyr(44)B10, and Leu(70)E11 surround a heme-ligated water molecule (2.81 Å and 3.32 Å in the A and B chains, respectively), which is not stabilized by any polar interaction

provided by the protein (Fig. 7A). It should be noted, however, that part of the heme-distal site corresponding to the B-C-E region (residues 48-67 and 46-69 in chain A and chain B, respectively), could not be modeled due to the low connectivity of the experimental electron density map. In particular, this disordered region involves the C-terminal end of the B-helix, which in *Av*GReg-Gb appears to be almost completely unwound (Fig. 5A). As a consequence, Tyr(44)B10, a conserved residue usually involved in ligand stabilization in GCS proteins, is too far to H-bond the heme-bound water molecule (Fig. 7A). The same considerations hold for the nearby Tyr(42)B8, an amino acid which is instead specific of *Av*GReg-Gb and few others GCSs (19). However, both Tyr residues point their hydroxyl groups towards the interior of the distal site and, therefore, could be involved in ligand stabilization in the context of a structured B-C-E region.

Conformational disorder at the BE distal site region has been previously reported in GCSs and other globins, and, in some cases, associated to functional roles. For instance, in GCSs and related proteins (*i.e.* protoglobin) disorder in this region has been associated to multiple conformations of specific residues whose side-chain orientations in the heme-distal site depend on the ligation state of the heme-Fe atom (40,44,45,46,75). However, in *Av*GReg-Gb the disordered region is delimited by residues Gly(40)B6-Pro(41)B7 on one side and Gly(68)E9 on the other, which are specific for *Av*GReg-Gb and not conserved in other GCSs (19,44).

The *Bpe*GReg-Gb* heme iron atom coordinates to a water molecule (at 3.06 Å in the A chain and at 2.90 Å in the B chain) at the distal site, which is not stabilized by any interaction with the amino acids facing the distal site. In particular, Tyr(43)B10 is pointing its hydroxyl group towards the exterior of the distal site. The distal side residues are hydrophobic (Phe(42)B9, Leu56, Val(61)E3, Leu(65)E7, Met(69)E11) and well defined in the A chain of the dimer, while in the B chain the C-E region (residues 57-60) is partly disordered (Fig. 7B). Interestingly, the second crystal form, *Bpe*GReg-Gb*-2, shows an Fe-coordination state at the distal site that differs in the A and B chains and from the *Bpe*GReg-Gb* crystal form: the heme-Fe is loosely hexacoordinated to Tyr(43)B10 (3.11 Å) in chain A, (Fig. 7C) and to an imidazole molecule (2.12 Å), probably originating from the

purification protocol, in chain B, with Tyr(43)B10 swinging out from the distal site due to steric repulsion caused by the imidazole ligand (Fig. 7D). Overall, we can conclude that the different dimerization mode of *Bpe*GReg-Gb*-2 has an impact in the distal site geometry reshaping, with Tyr(43)B10 able to point inwards and outwards, in keeping with the previously proposed mechanism whereby the globin domain dimerization leads to conformation(s) of the distal pocket that adjust the position of the hydrogen bonding residues involved in O_2 dissociation kinetics (50). Interestingly, the distal site swinging movement of TyrB10 has been suggested as the mechanism of O_2 stabilization for other GCSs, as *Bs*HemAT (75) and *Ec*DosC (63).

In *Bpe*GReg-Gb* the heme group's propionates are not interacting with Lys(64)E6 at the distal site, nor with Lys(94)F3 at the proximal site (Fig. 7A), while in the *Bpe*GReg-Gb*-2 structure only a loose interaction is present with Lys(64)E6 (Fig. 7C,D).

Discussion

The *A. vinelandii* and *B. pertussis* are highly interesting bacterial systems. Indeed, *A. vinelandii* fixes N_2 in the presence of O_2, a unique metabolic feature that renders it particularly interesting for industrial applications whereas *B. pertussis* is a highly contagious human pathogen that still causes many deaths worldwide and is, therefore, of considerable medical interest. In addition, the GCS of both systems are involved in biofilm formation. In this respect, their characterization, here presented, is significant for a more detailed understanding of the control that the sensors might exert on biofilm formation.

The combination of biochemical techniques applied in this study provides complementary pictures of *Av*GReg and *Bpe*GReg functional and structural behaviors. We performed our experiment on *in vivo* folded full-length proteins. It is important to note that in the previously published study on *Av*GReg and *Av*GReg-Gb, both proteins had been purified from inclusion bodies, followed by *in vitro* refolding (64). The high level of agreement between our data and the optical and RR spectra of the *in*

vitro refolded *Av*GReg-Gb is an indication of the reliability of the refolding method when applied to the globin domain alone. In contrast, the discrepancies detected between *in vivo* folded *Av*GReg and that purified from inclusion bodies strongly suggest that the presence of the second (transducer) domain may hamper a correct *in vitro* refolding process.

We showed that both the *in vivo* folded *Av*GReg and *Bpe*GReg sensors are oxygenated when purified and pentacoordinated when the deoxy ferrous species is produced through dithionite reduction. Additionally, based on our enzymatic *in vitro* assays and the literature, we proved that O_2 binding triggers c-di-GMP biosynthesis and that O_2-*Av*GReg and O_2-*Bpe*GReg are the catalytically competent forms. O_2-*Av*GReg and O_2-*Bpe*GReg display a high sensitivity to product inhibition confirming the literature (70). For O_2-*Av*GReg and O_2-*Bpe*GReg a k_{cat}/K_m value of $1.13 \pm 0.13 \times 10^4$ M^{-1} s^{-1} (reported here) and 0.17×10^3 M^{-1} s^{-1} (6,7) was obtained, respectively. The RR measurements indicate that the heme pocket of the fully oxygenated proteins are well structured, with moderate to strong electrostatic interactions of the heme propionate groups with the surrounding amino-acid side chains and the O_2 molecule tightly bound to the heme-Fe for all forms and an additional more open distal heme-pocket site for O_2-*Av*GReg. These data are confirmed by the recently published structure of the oxygenated form of *Bpe*GReg-Gb (structure not yet released by PDB), where O_2 is described as tightly bound to the Fe^{2+} and H-bonded to Tyr(43)B10, and Lys(64)E6 forms a hydrogen bond with heme propionate (51).

The combination of kinetics measurements and the structural data suggest mechanisms for protein activation and signal transduction. In the case of *Av*GReg the kinetics measurements reveal a fast O_2 association ($k'_{O2} = 12.2 \times 10^6$ M^{-1} s^{-1}) and a biphasic O_2 dissociation, characterized by fast and slow rates (k_{O2} = 1732 and 0.53 s^{-1}, respectively). Among GCSs a biphasic O_2 dissociation has been reported only for *Bs*HemAT (75). The higher K_d for O_2 is in the same order of magnitude for the *Av*GReg and *Bs*HemAT, while the lower K_d highlights the very high O_2 affinity of *Av*GReg, being the highest among the other globin sensors characterized to date (see Table 3), suggesting that it can be activated in the presence of traces amounts of O_2. The crucial role of Tyr(44)B10 is highlighted

by the binding kinetics study of the YB10F mutant which shows a significant increase of O_2 dissociation rate, for both the fast and slow rates (Table 3). Specific high affinity O_2 binding (needed for sensing, transport, and storage) requires significant deviation from the sliding scale rule, which predict that, although the equilibrium dissociation constants (K_d) can vary a lot, the ratios for NO:CO:O_2 binding are roughly the same, namely 1:~10^3:~10^6 when the proximal ligand is a histidine and the distal site is apolar (66). Indeed, data reported for GCSs (including *Bp*eGReg) show that, in general, CO affinity is higher than O_2, but that the $K_D(O_2)/K_D(CO)$ ratio is much smaller than that predicted by the sliding scale rule due to a preferentially lower $K_D(O_2)$ associated with hydrogen bonding stabilization of the O_2 molecule into the distal site (66).

Within this picture the biphasic behavior for *Av*GReg O_2 dissociation, measured by flash-photolysis O_2/CO competition, can be explained by the presence of two different *Av*GReg conformations (*open* and *closed*, respectively) responsible for the two dissociation rate constants. More precisely, when the globin domain is in the *closed* conformation, O_2 would be firmly bound in the heme pocket and well protected from the solvent. As a result, the competition between O_2 and CO would be hampered and O_2 displacement significantly slower. On the contrary, when the protein is in the *open* conformation, O_2 would be less stabilized within the heme pocket, thus allowing a more efficient diffusion of CO from the outer environment and faster O_2 displacement. Both conformations are in keeping with the RR measurements on the fully oxygenated ferrous *Av*GReg. All in all, the crystal structure of the *Av*GReg-Gb ferric form reveals that the B-C-E region can undergo to structural disorder, probably favoured by the Gly-rich sequence at the BE region boundaries (not present in other GCSs), and this protein flexibility is compatible with the presence of an *open* conformation. From RR and kinetics follows that there is a second more closed (and probably more rigid) conformation that is not populated in the crystals. This closed conformation involves Tyr(44)B10.

Furthermore, the presence of the transmitter domain in the full-length *Av*GReg relaxes the heme proximal coordination when compared to the isolated *Av*GReg-Gb sensor domain. Indeed, the aquo-met crystal structure is in keeping with the RR measurements on the ferrous unligated protein,

indicating a staggered conformation of the proximal His(127)F8 side chain and, therefore, a higher strain on the Fe-His bond for the isolated globin domain relative to the full-length protein. Such relaxation at the F-helix proximal His(127)F8 may be also part of the signal transduction mechanism between domains upon O_2 binding.

In contrast to *Av*GReg, no dramatic change in the sensor domain of *Bpe*GReg can be ascribed to the absence/presence of the transmitter domain. Indeed, the RR spectra measured on *Bpe*GReg-Gb and *Bpe*GReg (and their *Bpe*GReg-Gb* and *Bpe*GReg* Cys-mutants) are very similar. This suggest that the signal transduction mechanism for the *Bpe*GReg globin domain activation involves structural modifications in protein regions far from the heme pocket. The measured k'_{O2} and k_{O2} values suggest that *Bpe*GReg is capable of selective and more efficiently binding O_2 when a gas mixture is present, as is also the case for myoglobin (1). The low K_d (O_2) values indicate that, under physiological conditions, the sensor will permanently bind O_2. A biphasic behavior for O_2 dissociation, as seen for *Av*GReg, is not recorded for *Bpe*GReg, in line with the RR observation of only one (*closed*) conformation for the stabilization of the distal O_2 ligand. However, flash photolysis experiments indicate that the CO binding kinetics differs between the isolated *Bpe*GReg-Gb (one exponential) and the full-length *Bpe*GReg (two exponential). This result proves that the overall protein dynamicity changes when the transmitter domain is present, probably due to the presence of structural interactions between sensor and transmitter domains, likely at the oligomerization interface, necessary for the signal transduction process as also seen by Rivera et al (50).

Finally, the crystal structure of the aquo-met form of *Bpe*GReg-Gb* shows that the water ligand is not stabilized by any distal residue, the Tyr(43)B10 hydroxyl group is pointing away from the heme distal side, and the propionates are not interacting with the protein. These data provide a complementary picture of the recently published O_2-bound structure (PDB code: 6M9A) (51) where, instead, the strong O_2-binding to Tyr(43)B10 is coupled with heme distortion, interaction between Lys(64)E6 and the heme propionate, and interaction between Ser(68)E10 and Tyr(43)B10 through a distal water molecule, not present in the *Bpe*GReg-Gb* aquo-met form. Thus our data support the

proposed mechanism that differences in ligation state at the distal side is potentially propagated through interactions with residues on the E helix, allowing for signaling to the DGC output domain. Furthermore, a second crystal form of ferric *Bpe*GReg-Gb* reveals the presence of two possible dimerization modes, both based on the G-H helical bundle. The first dimerization mode is that typically found in GCSs, the second differs by a rotation of 60° between monomers and is stabilized/forced by the presence of an extra heme molecule, probably due to a somewhat higher heme availability in solution at the more alkaline crystallization condition used (see Material and Methods, condition 2), covalently linking two His residues (His155) at the end of the H helices (the region which is linked to the middle domain of the full-length protein). Although this dimerization mode could be classified as a crystallization artifact, it clearly demonstrates that the G-H surface of a monomer is intrinsically characterize by high plasticity, being able to face, fit and bind its counterpart in more than one way. The different dimerization mode is also associated with a different distal site organization, with Tyr(43)B10 swinging in and out the distal site. Thus, a transmission mechanism through a partial rotation at the G-H interface is conceivable, whereby a different relative position of the H-helices upon ligand binding would directly affect the following middle domain and/or affect the global oligomerization state of the full-lengh protein (dimer with low diguanylate cyclase activity *vs* tetramer with high activity) (51).

Our *in vitro* data on DGC activity in *Av*GReg demonstrate that O_2-binding to *A. vinelandii* GCS triggers c-di-GMP biosynthesis, as described for *B. pertussis* GCS (20,33,51,70). In bacteria, C-di-GMP is an essential second messenger involved, *in vivo*, in the regulation (translational and post-translation) of motility, exo-polysaccharide production and biofilm formation (26,34,49,52). Thus O_2 binding to GCS as *A. vinelandii* and *B. pertussis*, increases c-di-GMP production and activates biofilm formation.

Innovation

Biofilms are considered an important medical and industrial issue and, consequently, GCS with DGC activity are becoming a class of enzymes of increasing scientific interest. This paper extends the knowledge on the GCS proteins. It provides a comprehensive biochemical and structural characterization of *A. vinelandii* and *B. pertussis* O_2 sensors involved in the control of biofilm formation and biofilm maintenance through production of the c-di-GMP second messenger.

Materials and Methods

Cloning, expression, mutagenesis and protein purification

Full-length *Av*GReg (472 residues) and its N-terminal globin domain *Av*GReg-Gb (residues 1-170) were amplified by PCR, and cloned into the pBAD-a vector as previously described (14,47,64). In addition, for the *Av*GRegYB10F variant, Tyr(44)B10 was mutated into Phe using the QuickChange™ site-directed mutagenesis method (Stratagene) (15). All were expressed as N-terminal His-tagged proteins in *E. coli* TOP10 cells (Invitrogen) (14,47) and purified from the soluble fraction using affinity chromatography. This procedure is very different from the one used to obtain †*Av*GReg in (64). In that case, the procedure involved the purification of the expressed, denatured proteins from inclusion bodies, with a subsequent *in vitro* folding in the presence of hemin. This approach deviates significantly from the here presented method based on *in-vivo* folding of the proteins in *E. coli* followed by affinity purification.

Full-length *Bpe*GReg (475 residues) and its N-terminal globin domain (*Bpe*GReg-Gb) (residues 1-169) were amplified by PCR and cloned into the pET23-a vector using NdeI and XhoI restriction sites. The C-terminal His-tagged fusion proteins were expressed in *E. coli* BL21(DE3)pLysS (Invitrogen) as done for the *Av*GReg proteins (14,47). To prevent aggregation during the crystallization process, four out of six cysteine residues present in the globin domain (Cys16, Cys45, Cys114, Cys154) were mutated into Ser using the QuickChange™ site-directed mutagenesis method (Stratagene) (15). The selection for the Cys residue to be mutated was done based on the *in silico* modelling of *Bpe*GReg-Gb obtained by using the I-TASSER server (72). The mutated proteins are further termed and *Bpe*GReg* (full length molecule) and *Bpe*GReg-Gb* (globin domain). The *Av*GReg proteins were kept in 50mM Tris-HCl pH 7.5, 50mM NaCl, 10mM $MgCl_2$ and 5% glycerol, whereas the *Bpe*GReg proteins were kept in 20mM Tris-HCl pH 8, 50mM NaCl and 40% glycerol. Both the globin domains were kept in 50 mM Tris-HCl pH 8.5, 10 mM NaCl and 0.5 mM EDTA.

UV/Visible absorption spectroscopy

The optical absorption measurements were obtained by using a Varian Cary 5E UV-Vis-NIR spectrometer. The spectra were recorded from 350 nm to 750 nm at room temperature. The ferrous O_2-ligated forms were measured, without additional treatment (as-purified), soon after protein purification, while the unligated form was obtained by adding 1% v/v sodium dithionite saturated solution to the sample, after N_2 bubbling.

Resonance Raman spectroscopy

The resonance Raman (RR) spectra were measured on a Dilor XY-800 Raman spectrometer in low-dispersion mode, associated to a liquid N_2-cooled CCD detector (slit width = 200 μm), with a Kr^+ laser operating at 413.1 nm as the excitation source (Spectra-Physics BeamLok 2060). The applied laser power was 10 mW (*Bpe*GReg proteins) and 15 mW (*Av*GReg variants). The samples had a final concentration of 40 μM, and were magnetically stirred at 500 rpm to avoid local heating and photochemical decomposition. Six spectra were recorded with 120-180 s integration time, depending on the protein sample. To remove spikes due to cosmic rays, the highest and lowest data points for each frequency value were removed and averaged over the remaining values with an in-house written program.

Ligand binding kinetics and data analysis

Laser flash photolysis experiments were performed to measure the CO and O_2 binding kinetics at 20 °C using a laser photolysis system (Edinburgh Instruments LP920, UK) equipped with a frequency-doubled, Q-switched Nd:YAG laser (Quanta-Ray, Spectra Physics). A laser flash of 532 nm was used.

The carbon monoxide (CO)–ferrous Hb complexes were prepared in sealed 4 × 10 mm quartz cuvettes with 1 ml of 100 mM potassium phosphate buffer (pH 7.0) containing 1 mM EDTA. The buffer was equilibrated with mixtures of CO and N_2 in different ratios to obtain CO concentrations

of 50–800 μM by using a gas mixer (HighTech System; Bronkhorst, The Netherlands). Saturated sodium dithionite solution (10 μl) was added and the protein was injected to a final concentration of ~4 μM. Formation of the CO–ferrous Hb complex was verified by UV/visible absorption spectroscopy. Recombination of the photo-dissociated CO ligand was monitored at 418 nm (*Bpe*GReg) or 421 nm (*Av*GReg).

O_2/CO competition studies were performed with the ferrous oxygenated (as-purified) form of the GCS. Sealed cuvettes containing 100 mM phosphate buffer pH 7.0 were flushed with O_2/CO mixtures (O_2 concentrations between 250 and 812.5 μM, and CO concentrations between 350 and 800 μM). The protein was added to the experimental cuvettes immediately prior to the photolysis experiment. The CO/O_2 displacement was followed at 418 nm and 436 nm for *Bpe*GReg and at 421 nm and 436 nm for *Av*GReg.

The data obtained were analyzed with MATLAB® (The MathWorks, Inc., Natick, Massachusetts, USA) and fitted with Origin (OriginLab, Northampton, Massachusetts, USA). The time courses for CO rebinding followed a mono exponential decay and were fitted to the minimum reaction mechanism (scheme 1) with equations as described elsewhere (42,64).

$$GCS + CO \underset{k_{CO}}{\overset{k'_{CO}}{\rightleftarrows}} GCS - CO \qquad \text{(Scheme 1)}$$

The O_2/CO competition time courses fitted with a two-exponential process (Scheme 2, (42)), valid also for the globin domain alone:

$$GCS - O_2 + CO \xrightarrow{k'_{CO},k_{O2}} GCS - CO + O_2 \qquad \text{(Scheme 2a)}$$

$$GCS + O_2 \xrightarrow{k'_{O2}} GCS - O_2 \qquad \text{(Scheme 2b)}$$

The values of k'_{O2}, k_{O2}, and k'_{CO} were obtained according to the equations:

$$k_f = k'_{O2}[O_2] + k'_{CO}[CO] + k_{O2} \qquad \text{(Equation 1)}$$

$$k_s = \frac{k_{O2}}{1 + \frac{k'_{O2}[O_2]}{k'_{CO}[CO]}} \quad \text{(Equation 2)}$$

where k_f is the fast observed reaction constant k_{obs}, and k_s is the slow k_{obs} at a given ligand concentration.

The slower O_2 dissociation rate constant was determined with a ligand replacement method using stopped flow, as already described (41). Experiments were performed at 20 °C in triplicates on an Applied Photophysics SX20 stopped-flow spectrometer apparatus equipped with the Pro-Data™ software suite. After measurements, the raw data were further analyzed with Origin (OriginLab, Northampton, Massachusetts, USA). The oxygenated protein solution was rapidly mixed with an O_2/CO mixture (the final protein concentration after mixing was 2.5 µM) in 100 µM potassium phosphate buffer pH 7.5, 1 mM EDTA.

Diguanylate cyclase assay

A HPLC-based DGC assay was adapted from (54) and performed in triplicate to quantify c-di-GMP synthesis by *Av*GReg. Shortly, 10 µM of enzyme was incubated at 37 °C with 20 to 65 µM GTP in 50 mM TRIS-HCl pH 7.5, 50 mM NaCl and 10 mM $MgCl_2$. Aliquots were taken at different times and the reaction was rapidly stopped with one-fourth volume of 0.5 M EDTA pH 8.0, which ensures repeatability in the measurement of more time points in a few seconds time frame, as required for initial rate determination of enzymatic processes. For deproteination, the stopped reaction mixture was heated for 5 min at 95 °C and centrifuged for 10 minutes at 11,300 x g. 20 µl of the supernatant was injected onto an Adsorbosphere Nucleotide-Nucleoside RP C18 HPLC column (Alltech, 250x4.6 mm) equipped with the same type guard column (Alltech/Grade, 7.5x4.6 mm). As mobile phase, 0.15 M NaH_2PO_4 and 40% acetonitrile was used. A relatively rapid linear gradient from 0% to 35% acetonitrile, for 10 minutes at 1 ml min^{-1}, was applied to separate GTP, c-di-GMP and pGpG, with

which we could efficiently process the high number of samples produced. Known amounts of all three molecules were used for standardization.

Given the high sensitivity of *Av*GReg to product inhibition, low substrate concentrations were used. When the substrate concentration [S] << K_m, then [S] is almost negligible and the Michaelis-Menten's equation becomes:

$$v_i = \frac{V_{max}[S]}{K_m}$$

Given that $V_{max} = k_{cat}[E]_0$, in our experimental conditions the enzyme specificity constant (k_{cat}/K_m) can be calculated from Michaelis-Menten plots.

Crystallization, structure determination, and refinement

Crystallization of *Av*GReg-Gb and *Bpe*GReg-Gb* was achieved using the sitting drop vapour diffusion method. *Av*GReg-Gb crystals were obtained by equilibrating at 4 °C the protein solution (30 mg ml^{-1}) against 33% PEG 6000 and 10 mM Na-citrate. The crystals diffracted up to 2.83 Å resolution, using synchrotron radiation (Elettra - Sincrotrone Trieste S.C.p.A, Trieste, Italy; XRD1 beamline). They belong to the monoclinic $P2_1$ space group and two *Av*GReg-Gb molecules per asymmetric unit (estimated solvent content 33%) (Table 4). *Bpe*GReg-Gb* crystals grew at 4 °C by using two different ionic and pH conditions: 1.6 M magnesium sulphate and 0.1 M MES pH 6.5 (condition 1: *Bpe*GReg-Gb*) or 0.8 - 1 M ammonium sulphate and 0.1 M TRIS-HCl pH 6.0 - 9.0 (condition 2: *Bpe*GReg-Gb*-2). The crystals grown under condition 1 and 2 diffracted up to 3.20 Å and 1.55 Å resolution, respectively, using synchrotron radiation (at SOLEIL and ESRF facilities, France). Both crystal forms belong to the *C*2 space group, but with different cell parameters, and contain two protein molecules in the asymmetric unit (estimated solvent content 48.25 % for *Bpe*GReg-Gb*, and 34.69 % for *Bpe*GReg-Gb*-2) (Table 4). MOSFLM (31), SCALA (18), and the CCP4 suite programs (71) were used for reducing, scaling and analysing all the collected data.

The *Bpe*GReg-Gb*-2 structure was solved with a combination of single wavelength anomalous dispersion (based on the heme-Fe atom anomalous signal) and molecular replacement. Few regions of the structure were built based on the SAD data. This partial model was used to determine the remaining phases with Phaser molecular replacement program (35). The molecular model was checked manually with COOT (17) and refined with REFMAC5 (37,38) to the maximum resolution (Table 4). The *Bpe*GReg-Gb*-2 refined structure was used as starting model (as a monomer) to solve the *Bpe*GReg-Gb* and the *Av*GReg-Gb dimeric structures with Phaser (35). The resulting structures were remodeled by using COOT (17) and refined with REFMAC5 (37,38) (Table 4). Their stereo chemical quality was assessed with MolProbity (9) and the quaternary assemblies were identified with PISA (30). Finally, the atomic coordinates and the structure factors have been deposited in the Protein Data Bank (www.rcsb.org) with entry codes 4UII (*Av*GReg-Gb), 6I2Z (*Bpe*GReg-Gb*), and 4UIQ (*Bpe*GReg-Gb*-2).

Acknowledgements

This work was supported by the Fund of Scientific Research – Flanders (FWO) grant no. G.0247.09N and supported by the University of Antwerp through the GOA biofilm project 25624 (GOA BOF UA 2011-2014) and the GOA-BOF project 28312. F.G. is a PhD fellow of the Fund of Scientific Research – Flanders (FWO). A.D.S. was funded by a PhD grant of the Innovation by Science and Technology (121339) (IWT, Belgium). The authors thank Miroslaw Tarnawski and Thomas Barends for making the *Ec*DosC model (full-length) available. The authors have no conflict of interest to declare.

Abbreviations:

*Af*GcHK = globin-coupled histidine kinase from *Anaeromyxobacter sp.* Fw109-5

*Av*GReg = globin-coupled regulator from *Azotobacter vinelandii*

*Av*GReg-Gb = *Av*GReg globin domain

*Av*GRegYB10F = *Av*GReg with Tyr(44) mutated to Phe

*Av*GReg178 = recombinant and reconstructed (heme + apoprotein) globin domain

*Bj*FixL = *Bradyrhizobium japonicum* nitrogen fixation gene expression regulator

*Bs*HemAT = *Bacillus subtilis* heme-based aerotaxis transducer

*Bpe*GReg = globin-coupled sensor from *Bordetella pertussis*

*Bpe*GReg* = *Bpe*GReg with cysteines (Cys16, 45, 114, 154) mutated to serines

*Bpe*GReg-Gb = *Bpe*GReg globin domain

*Bpe*GReg-Gb* = *Bpe*GRegGb with cysteines (Cys16, 45, 114, 154) mutated to serines

c-di-GMP, cyclic-di-(3',5')-GMP

DGC = diguanylate cyclase

*Ec*DosC = *E. coli* globin-coupled sensor with DGC activity

*Ec*DosC-heme = isolated globin domain of the *E. coli* globin-coupled sensor

GCS = globin-coupled sensor

*Gs*GCS^{162} = sensor domain from the *Geobacter sulfurreducens* globin-coupled sensor

PDE = phosphodiesterase

*Rm*FixL* = soluble truncated domain of the *Rhizobium meliloti* nitrogen fixation gene expression regulator

Rmsd = root mean squared deviation

RR, Resonance Raman spectroscopy

*Rr*CooA = *Rhodospirillum rubrum* carbon monoxide oxidation activator; sGC, soluble guanylyl cyclase

*Sw*Mb = sperm whale myoglobin

REFERENCES

1. Antonini, E. and Brunori M. Hemoglobin and myoglobin in their reactions with ligands. 1971: North-Holland Pub. Co.
2. Aono S, Kato T, Matsuki M, Nakajima H, Ohta T, Uchida T, and Kitagawa T. Resonance Raman and ligand binding studies of the oxygen-sensing signal transducer protein HemAT from *Bacillus subtilis*. *J Biol Chem* 277: 13528-13538, 2002.
3. Bart MJ, van Gent M, van der Heide HG, Boekhorst J, Hermans P, Parkhill J, and Mooi FR. Comparative genomics of prevaccination and modern *Bordetella pertussis* strains. *BMC Genomics* 11: 627, 2010.
4. Bart MJ, Zeddeman A, van der Heide HG, Heuvelman K, van Gent M, and Mooi FR. Complete genome sequences of *Bordetella pertussis* isolates B1917 and B1920, representing two predominant global lineages. *Genome Announc* 2: pii: e01301-14, 2014.
5. Brunori M and Schuster TM. Kinetic studies of ligand binding to hemoglobin and its isolated subunits by the temperature jump relaxation method. *J Biol Chem* 244: 4046-4053, 1969.
6. Burns JL, Deer DD, and Weinert EE. Oligomeric state affects oxygen dissociation and diguanylate cyclase activity of globin coupled sensors. *Mol Biosyst* 10: 2823-2826, 2014.
7. Burns JL, Rivera S, Deer DD, Joynt SC, Dvorak D, and Weinert EE. Oxygen and bis(3',5')-cyclic dimeric guanosine monophosphate binding control oligomerization state equilibria of diguanylate cyclase-containing globin coupled sensors. *Biochemistry* 55: 6642-6651, 2016.
8. Centeno JA. Evidence of dithionite contribution to the low-frequency resonance Raman spectrum of reduced and mixed-valence cytochrome c oxidase. *Arch Biochem Biophys*. 292: 624-628, 1992.
9. Chen VB, Arendall WB, III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, and Richardson DC. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66: 12-21, 2010.

10. Conover MS, Sloan GP, Love CF, Sukumar N, and Deora R. The Bps polysaccharide of *Bordetella pertussis* promotes colonization and biofilm formation in the nose by functioning as an adhesin. *Mol Microbiol* 77: 1439-1455, 2010.
11. Das TK, Couture M, Ouellet Y, Guertin M, and Rousseau DL. Simultaneous observation of the O---O and Fe---O_2 stretching modes in oxyhemoglobins. *Proc Natl Acad Sci* USA 98: 479-484, 2001.
12. Das TK, Friedman JM, Kloek AP, Goldberg DE, and Rousseau DL. Origin of the anomalous Fe-CO stretching mode in the CO complex of *Ascaris* hemoglobin. *Biochemistry* 39: 837-842, 2000.
13. Deinum G, Stone JR, Babcock GT, Marletta MA. Binding of nitric oxide and carbon monoxide to soluble guanylate cyclase as observed with Resonance raman spectroscopy. *Biochemistry* 35: 1540-1547, 1996.
14. Dewilde S, Mees K, Kiger L, Lechauve C, Marden MC, Pesce A, Bolognesi M, and Moens L. Expression, purification, and crystallization of neuro- and cytoglobin. *Methods Enzymol* 436: 341-357, 2008.
15. Dewilde S, Kiger L, Burmester T, Hankeln T, Baudin-Creuza V, Aerts T, Marden MC, Caubergs R, and Moens L. Biochemical characterization and ligand binding properties of neuroglobin, a novel member of the globin family. *J Biol Chem* 276: 38949-38955, 2001.
16. Dufour D, Leung V, and Lévesque CM. Bacterial biofilm: structure, function and antimicrobial resistence. *Endod Top* 22: 2-16, 2012.
17. Emsley P, Lohkamp B, Scott WG, and Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66: 486-501, 2010.
18. Evans P. Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62: 72-82, 2006.
19. Freitas TA, Hou S, and Alam M. The diversity of globin-coupled sensors. *FEBS Lett* 552: 99-104, 2003.

20. Germani F, Moens L, and Dewilde S. Haem-based sensors: a still growing old superfamily. *Adv Microb Physiol* 63: 1-47, 2013.

21. Gibson QH, Olson JS, McKinnie RE, and Rohlfs RJ. A kinetic description of ligand binding to sperm whale myoglobin. *J Biol Chem* 261: 10228-10239, 1986.

22. Gilles-Gonzalez MA, Gonzalez G, Perutz MF, Kiger L, Marden MC, and Poyart C. Heme-based sensors, exemplified by the kinase FixL, are a new class of heme protein with distinctive ligand binding and autoxidation. *Biochemistry* 33: 8067-8073, 1994.

23. Guttenplan SB and Kearns DB. Regulation of flagellar motility during biofilm formation. *FEMS Microbiol Rev* 37: 849-871, 2013.

24. Hirota S, Ogura T, Appelman E.H., Shinzawaitoh K, Yoshikawa S, and Kitagawa T. Observation of a new oxygen-isotope-sensitive Raman band for oxyhemoproteins and its implication in heme pocket structures. *J Am Chem Soc* 116: 10564-10570, 1994.

25. Hu SZ, Smith K, and Spiro TG. Assignment of protoheme resonance raman spectrum by heme labeling in myoglobin. *J Am Chem Soc* 118: 12638-12646, 1996.

26. Jenal U, Reinders A and Lori C. Cyclic di-GMP: second messenger extraordinaire. *Nat Rev Microbiol* 15:271-284, 2017.

27. Kerr EA, Yu NT, Bartnicki DE, and Mizukami H. Resonance Raman studies of CO and O_2 binding to elephant myoglobin (distal His(E7)----Gln). *J Biol Chem* 260: 8360-8365, 1985.

28. Kitanishi K, Kobayashi K, Kawamura Y, Ishigami I, Ogura T, Nakajima K, Igarashi J, Tanaka A, and Shimizu T. Important roles of Tyr43 at the putative heme distal side in the oxygen recognition and stability of the Fe(II)-O_2 complex of YddV, a globin-coupled heme-based oxygen sensor diguanylate cyclase. *Biochemistry* 49: 10381-10393, 2010.

29. Kitanishi K, Kobayashi K, Uchida T, Ishimori K, Igarashi J, and Shimizu T. Identification and functional and spectral characterization of a globin-coupled histidine kinase from *Anaeromyxobacte*r sp. Fw109-5. *J Biol Chem* 286: 35522-35534, 2011.

30. Krissinel E, and Henrick K. Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372: 774-797, 2007.

31. Leslie AG. The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* 62: 48-57, 2006.

32. Martinez LC and Vadyvaloo V. Mechanisms of post-transcriptional gene regulation in bacterial biofilms. *Front Cell Infect Microbiol* 4: 38, 2014.

33. Martinkova M, Kitanishi K, and 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* 288: 27702-27711, 2013.

34. Maunders E and Welch M. Matrix exopolysaccharides; the sticky side of biofilm formation. *FEMS Microbiol Lett* 363:fnx120, 2017.

35. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ. Phaser crystallographic software. *J Appl Crystallogr* 40: 658-674, 2007.

36. Mishra M, Parise G, Jackson KD, Wozniak DJ, and Deora R. The BvgAS signal transduction system regulates biofilm development in *Bordetella*. *J Bacteriol* 187: 1474-1484, 2005.

37. Murshudov GN, Vagin AA, and Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240-255, 1997.

38. Murshudov GN, Vagin AA, Lebedev A, Wilson KS, and Dodson EJ. Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* 55: 247-255, 1999.

39. Nagai K, Kitagawa T, and Morimoto H. Quaternary structures and low frequency molecular vibrations of haems of deoxy and oxyhaemoglobin studied by resonance raman scattering. *J Mol Biol* 136: 271-289, 1980.

40. Nardini M, Pesce A, Thijs L, Saito JA, Dewilde S, Alam M, Ascenzi P, Coletta M, Ciaccio C, Moens L, and Bolognesi M. Archaeal protoglobin structure indicates new ligand diffusion paths and modulation of haem-reactivity. *EMBO Rep* 9: 157-163, 2008.

41. Olson JS. Stopped-flow, rapid mixing measurements of ligand binding to hemoglobin and red cells. *Methods Enzymol* 76: 631-651, 1981.

42. Olson JS, Foley EW, Maillet DH, Paster EV, Measurement of Rate Constants for Reactions of O_2, CO and NO with Hemoglobin. In: Nagel RL (eds) Hemoglobin Disorders, Methods in Molecular Biology, Humana Press Inc., Totowa NJ, 82, 65- 91, 2003.

43. Othman S, Richaud P, Vermeglio A, and Desbois A. Evidence for a proximal histidine interaction in the structure of cytochromes c in solution: a resonance Raman study. *Biochemistry* 35: 9224-9234, 1996.

44. Pesce A, Thijs L, Nardini M, Desmet F, Sisinni L, Gourlay L, Bolli A, Coletta M, Van DS, Wan X, Alam M, Ascenzi P, Moens L, Bolognesi M, and Dewilde S. HisE11 and HisF8 provide bis-histidyl heme hexa-coordination in the globin domain of *Geobacter sulfurreducens* globin-coupled sensor. *J Mol Biol* 386: 246-260, 2009.

45. Pesce A, Tilleman L, Dewilde S, Ascenzi P, Coletta M, Ciaccio C, Bruno S, Moens L, Bolognesi M, and Nardini M. Structural heterogeneity and ligand gating in ferric *Methanosarcina acetivorans* protoglobin mutants. *IUBMB. Life* 63: 287-294, 2011.

46. Pesce A, Tilleman L, Donne J, Aste E, Ascenzi P, Ciaccio C, Coletta M, Moens L, Viappiani C, Dewilde S, Bolognesi M, and Nardini M. Structure and haem-distal site plasticity in *Methanosarcina acetivorans* protoglobin. *PLoS ONE* 8: e66144, 2013.

47. Pesce A, Nardini M, Dewilde S, Ascenzi P, Burmester T, Hankeln T, Moens L, and Bolognesi M. Human neuroglobin: crystals and preliminary X-ray diffraction analysis. *Acta Crystallogr D Biol Crystallogr* 58: 1848-1850, 2002.

48. Poole RK and Hill S. Respiratory protection of nitrogenase activity in *Azotobacter vinelandii*--roles of the terminal oxidases. *Biosci Rep* 17: 303-317, 1997.

49. Purcell EB and Tamayo R. Cyclic diguanylate signaling in Gram-positive bacteria. *FEMS Microbiol Rev* 40: 753-773, 2016.

50. Rivera S, Burns JL, Vansuch GE, Chica B, and Weinert EE. Globin domain interactions control heme pocket conformation and oligomerization of globin coupled sensors. *J Inorg Biochem* 164: 70-76, 2016.

51. Rivera S, Young PG, Hoffer ED, Vansuch GE, Metzler CL, Dunham CM, and Weinert EE. Structural insights into oxygen-dependent signal transduction within globin coupled sensors. Inorg. Chem. 57: 14386-14395, 2018.

52. Römling U and Galperin MY. Discovery of the second messenger Cyclic di-GMP. *Methods Mol Biol* 1657:1-8, 2017.

53. Ryan RP. Cyclic di-GMP signalling and the regulation of bacterial virulence. *Microbiology* 159: 1286-1297, 2013.

54. Ryjenkov DA, Tarutina M, Moskvin OV, and Gomelsky M. Cyclic diguanylate is a ubiquitous signaling molecule in bacteria: insights into biochemistry of the GGDEF protein domain. *J Bacteriol* 187: 1792-1798, 2005.

55. Setubal JC, Dos SP, Goldman BS, Ertesvag H, Espin G, Rubio LM, Valla S, Almeida NF, Balasubramanian D, Cromes L, Curatti L, Du Z, Godsy E, Goodner B, Hellner-Burris K, Hernandez JA, Houmiel K, Imperial J, Kennedy C, Larson TJ, Latreille P, Ligon LS, Lu J, Maerk M, Miller NM, Norton S, O'Carroll IP, Paulsen I, Raulfs EC, Roemer R, Rosser J, Segura D, Slater S, Stricklin SL, Studholme DJ, Sun J, Viana CJ, Wallin E, Wang B, Wheeler C, Zhu H, Dean DR, Dixon R, and Wood D. Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. *J Bacteriol* 191: 4534-4545, 2009.

56. Shimizu T, Huang D, Yan F, Stranava M, Bartosova M, Fojtikova V, and Martinkova M. Gaseous O_2, NO, and CO in signal transduction: structure and function relationships of heme-based gas sensors and heme-redox sensors. *Chem Rev* 115: 6491-6533, 2015.

57. Sondermann H, Shikuma NJ, and Yildiz FH. You've come a long way: c-di-GMP signaling. *Curr Opin Microbiol* 15: 140-146, 2012.

58. Song S, Boffi A, Chiancone E, and Rousseau DL. Protein-heme interactions in hemoglobin from the mollusc *Scapharca inaequivalvis*: evidence from resonance Raman scattering. *Biochemistry* 32: 6330-6336, 1993.

59. Spiro TG and Li XY. Resonance Raman spectroscopy of metalloporphyrins. *Biological applications of Raman spectroscopy* 3: 1-37, 1988.

60. Takahashi S, Ishikawa, Takeuchi N, Ikeda-Saito M, Yoshida Y, and Rousseau DL. Oxygen-bound heme-heme oxygenase complex: evidence for a highly bent structure of the coordinated oxygen. *J Am Chem Soc* 117: 6002-6006, 1995.

61. Takahashi S, Wang J, Rousseau DL, Ishikawa K, Yoshida T, Takeuchi N, and Ikeda-Saito M. Heme-heme oxygenase complex: structure and properties of the catalytic site from resonance Raman scattering. *Biochemistry* 33: 5531-5538, 1994.

62. Tamura K, Tanaka Y, Oue S, Tsukamoto K, Nomura M, Tsuchiya T, Adachi S, Takahashi S, Iizuka T, and Shiro Y. Nature of endogenous ligand binding to heme iron in oxygen sensor FixL. *J Am Chem Soc* 118: 9434-9435, 1996.

63. Tarnawski M, Barends TR, and Schlichting I. Structural analysis of an oxygen-regulated diguanylate cyclase. *Acta Crystallogr D Biol Crystallogr* 71: 2158-2177, 2015.

64. Thijs L, Vinck E, Bolli A, Trandafir F, Wan X, Hoogewijs D, Coletta M, Fago A, Weber RE, Van Doorslaer S, Ascenzi P, Alam M, Moens L, and Dewilde S. Characterization of a globin-coupled oxygen sensor with a gene-regulating function. *J Biol Chem* 282: 37325-37340, 2007.

65. Tomita T, Ogura T, Tsuyama S, Imai Y, and Kitagawa T. Effects of GTP on bound nitric oxide of soluble guanylate cyclase probed by resonance Raman spectroscopy. *Biochemistry* 36: 10155-10160, 1997.

66. Tsai AL, Berka V, Martin E and Olson JS. A "Sliding Scale Rule" for Selectivity among NO, CO and O_2 by Heme Protein Sensors. *Biochemistry* 51:172-186, 2012.

67. Uchida T, Ishikawa H, Takahashi S, Ishimori K, Morishima I, Ohkubo K, Nakajima H, and Aono S. Heme environmental structure of CooA is modulated by the target DNA binding. Evidence

from resonance Raman spectroscopy and CO rebinding kinetics. *J Biol Chem* 273: 19988-19992, 1998.

68. Uchida T and Kitagawa T. Mechanism for transduction of the ligand-binding signal in heme-based gas sensory proteins revealed by resonance Raman spectroscopy. *Acc Chem Res* 38: 662-670, 2005.
69. Vinogradov SN and Moens L. Diversity of globin function: enzymatic, transport, storage, and sensing. *J Biol Chem* 283: 8773-8777, 2008.
70. Wan X, Tuckerman JR, Saito JA, Freitas TA, Newhouse JS, Denery JR, Galperin MY, Gonzalez G, Gilles-Gonzalez MA, and Alam M. Globins synthesize the second messenger bis-(3'-5')-cyclic diguanosine monophosphate in bacteria. *J Mol Biol* 388: 262-270, 2009.
71. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, and Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67: 235-242, 2011.
72. Yang, J., et al., The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 12: 7-8, 2015.
73. Yeh SH. Pertussis: persistent pathogen, imperfect vaccines. *Expert Rev Vaccines* 2: 113-127, 2003.
74. Yeh SR, Couture M, Ouellet Y, Guertin M, and Rousseau DL. A cooperative oxygen binding hemoglobin from *Mycobacterium tuberculosis*. Stabilization of heme ligands by a distal tyrosine residue. *J Biol Chem* 275: 1679-1684, 2000.
75. Zhang W and Phillips GN, Jr. Structure of the oxygen sensor in *Bacillus subtilis*: signal transduction of chemotaxis by control of symmetry. *Structure* 11: 1097-1110, 2003.

Figure legends

Figure 1. Resonance Raman spectra of *Av*GReg.

RR spectra of *Av*GReg (a,b) and *Av*GReg-Gb (c,d) in the low-frequency region (**A**) and high-frequency region (**B**) and RR spectra of *Bpe*GReg (a,b) and *Bpe*GReg-Gb (c,d) in the low-frequency region (**C**) and high-frequency region (**D**). Traces (a,c) show the RR spectra of the proteins as purified (oxygenated Fe(II) state), traces (b,d) show the RR spectra of the dithionite-reduced proteins (ferrous state). Wavenumbers marked with asterisk indicate reminiscent contributions of dithionite as described in (8).

Figure 2. Ligand binding kinetics curves for *Av*GReg.

(**A**) Experimental curves of the CO-rebinding to *Av*GReg (black: 200 μM CO, red: 400 μM CO, green: 600 μM CO, blue: 800 μM CO). The dependence of the pseudo-first order constants of CO binding (k_{obsCO}) to *Av*GReg on the relative ligand concentration is reported in the inset. (**B**) Stopped-flow O_2/CO replacement experimental curves of *Av*GReg (black: 50% CO - 50% O_2, red: 60% CO - 40% O_2, green: 70% CO - 30% O_2, blue: 80% CO - 20% O_2). The *fast* and the *slow* replacement phases are detectable for all the gas ratios used. (**C**) Dependence of the apparent k_f constants of O_2 binding to *Av*GReg on the relative ligand concentration. (**D**) Dependence of the apparent k_s constants of CO binding to *Av*GReg on the relative ligand concentration. Experimental values are reported with open circles, while the theoretical ligand-binding mechanism is represented with a straight line.

Figure 3. Ligand binding kinetics curves for *Bpe*GReg.

(**A**) Experimental CO-rebinding decays to *Bpe*GReg (black: 200 μM CO, magenta: 400 μM CO, green: 600 μM CO, blue: 800 μM CO) with the corresponding fitting curves (in red color). The dependence of the second order constants of CO binding (k_{obsCO}) to *Bpe*GReg on the relative ligand concentration is reported in the inset. Experimentally obtained values are indicated with closed (fast

rebinding) and open (slow rebinding) circles, the theoretical fitting lines are reported in red. (**B**) Dependence of the apparent k_f constants of O_2 binding to *Bpe*GReg on the relative ligand concentration. (**C**) Dependence of the apparent k_s constants of CO binding to *Bpe*GReg on the relative ligands concentration ratio. Experimental values are reported with open circles, while the theoretical ligand-binding mechanism is represented with red lines.

Figure 4. Enzymatic characteristics of the DGC activity of *Av*GReg.

(**A**) Reaction rate. (**B**) Summary of 3 independent measurements.

Figure 5. Tertiary structures.

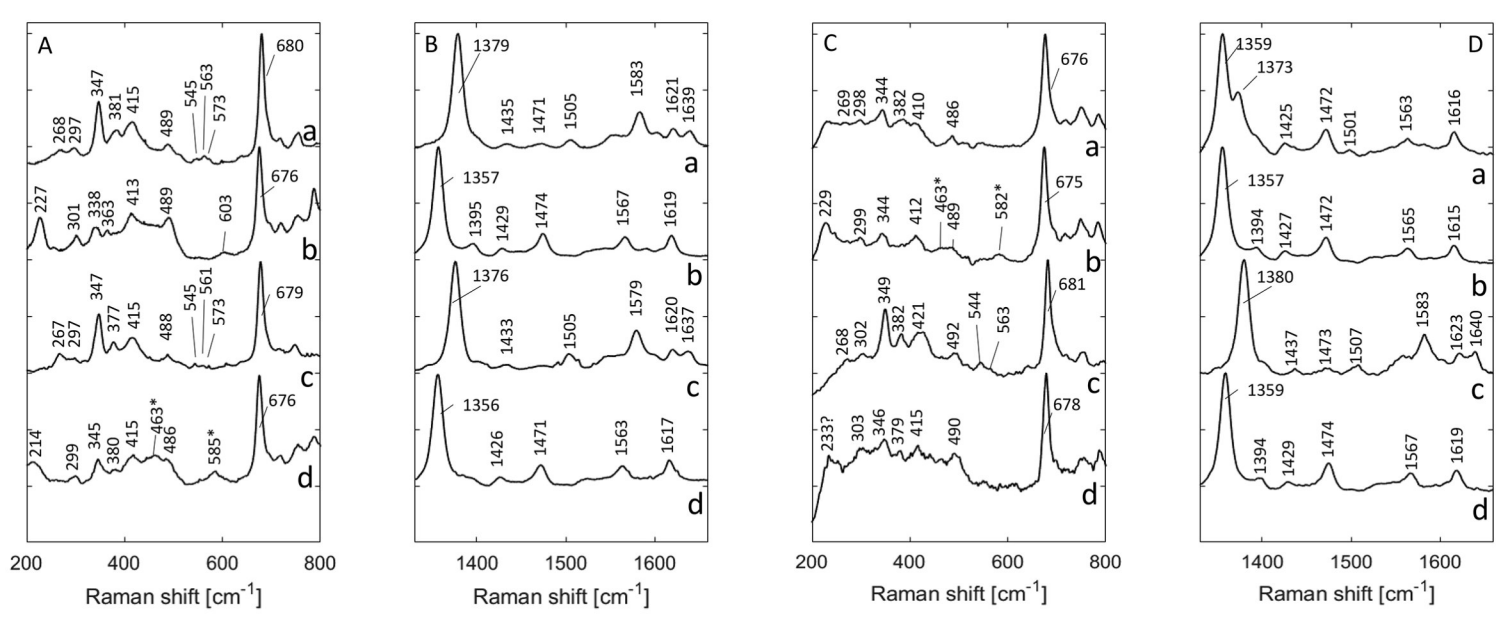
Ribbon representation of the tertiary structure of (**A**) *Av*GReg-Gb (green) and (**B**) *Bpe*GReg-Gb* (cyan). Helices are labelled accordingly with the globin fold nomenclature. The extra-turn of the F helix of *Av*GReg-Gb is shown in magenta. Proximal His residues and Fe^{3+}-ligated water molecules are indicated. (**C**) Cα-trace overlay of the *Av*GReg-Gb A (green) and *Bpe*GReg-Gb* (cyan). Regions displaying the main structural differences are highlighted.

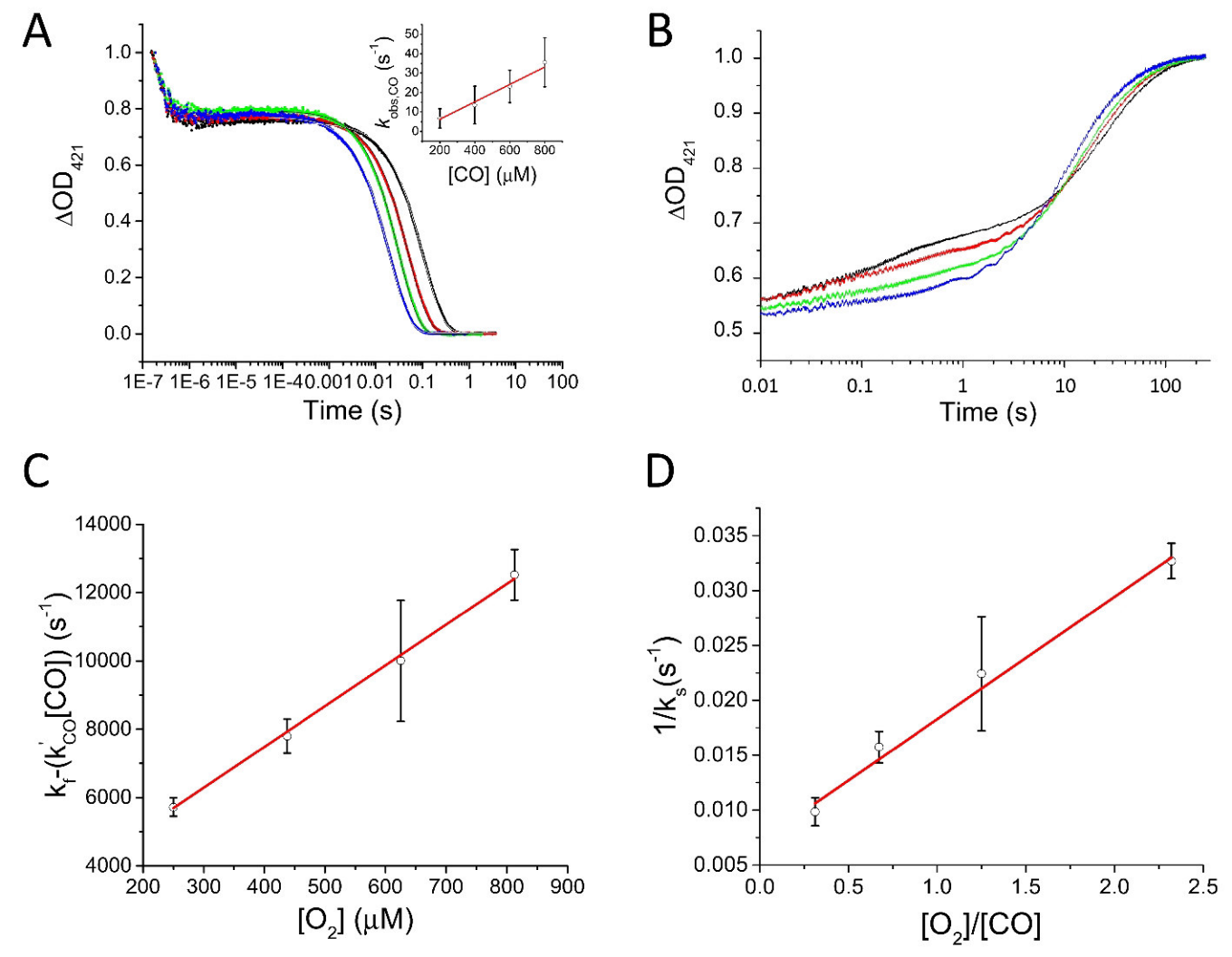
Figure 6. Quaternary structures.

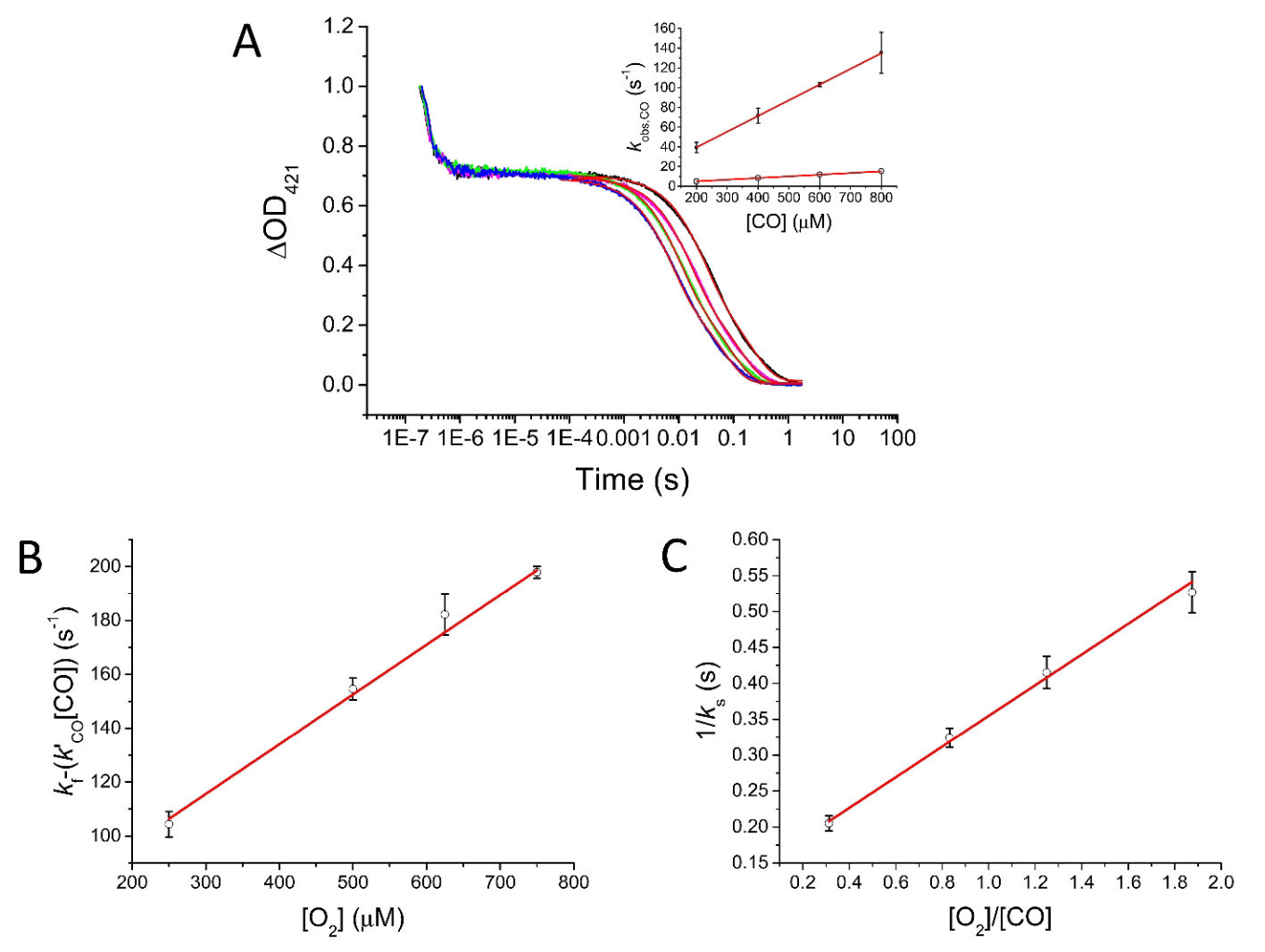
Top view of (**A**) the *Av*GReg-Gb dimer, and (**B**) the *Bpe*GReg-Gb* dimer, with the two subunits interacting mainly through the G- and H-helices (helical bundle), shown as ribbon models. The heme groups are shown in red. (**C**) Dimeric arrangement of *Bpe*GReg-Gb*-2. The dimeric interface involves the G- and H-helices as in the *Bpe*GReg-Gb* dimer, but the subunits are oriented differently. (**D**) Relative orientation of the B subunits of *Bpe*GReg-Gb* (blue) and *Bpe*GReg-Gb*-2 (pink), after superimposing the two A chains (omitted in the picture for clarity). The rotation of 60° is highlighted, taking the position of the G helices as reference. (**E**, **F**) The heme group at the top of the *Bpe*GReg-Gb*-2 dimer is modelled in two conformations (green and yellow) with 0.5 occupancy each, with the Fe-atom hexacoordinated to the His155 side chain of the two subunits of the dimer.

Figure 7. Proximal and distal sites.

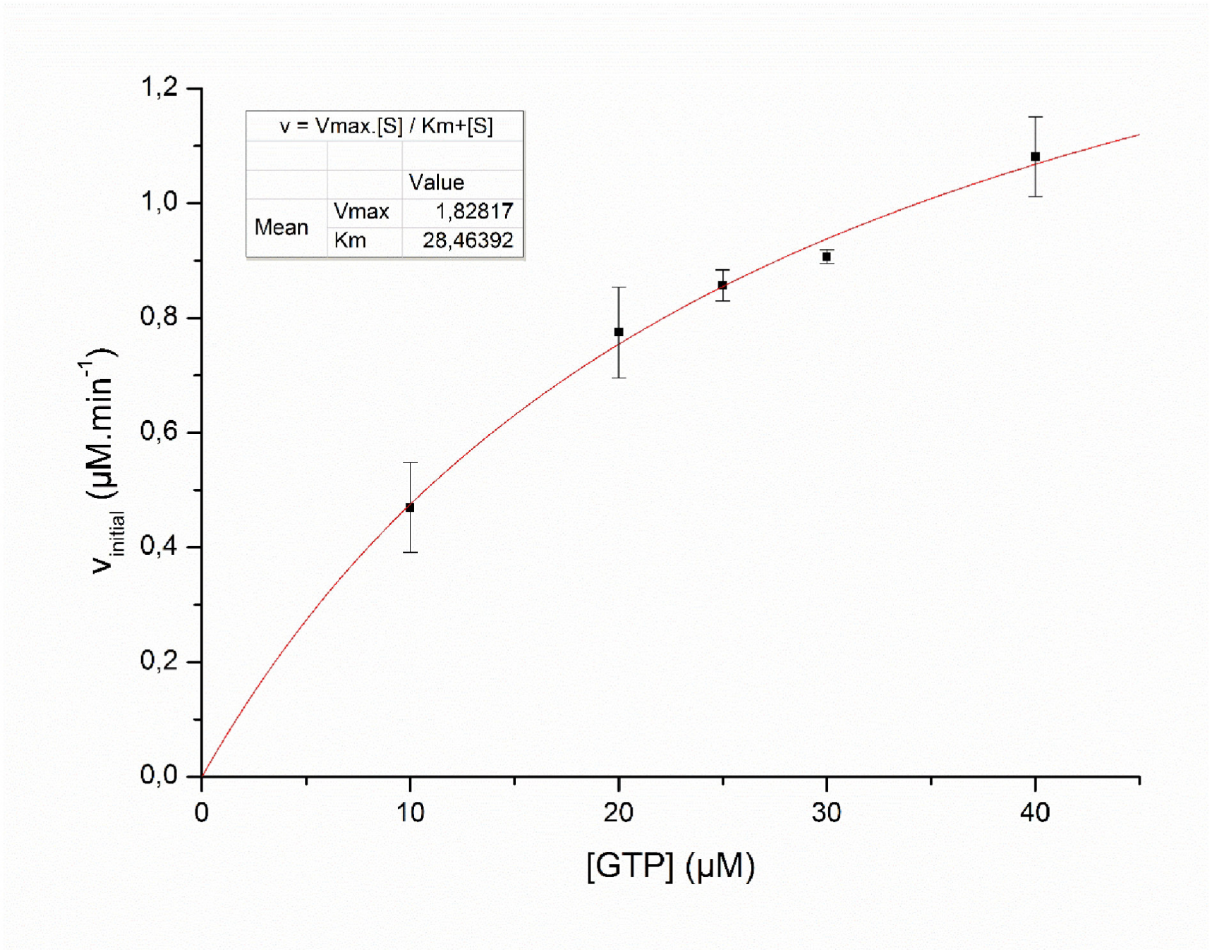
The figure shows the main residues building the heme proximal and distal sites for (**A**) *Av*GReg-Gb (A chain), (**B**) *Bpe*GReg-Gb* (A chain), (**C,D)** *Bpe*GReg-Gb*-2 (A and B chains). The *Av*GReg-Gb and *Bpe*GReg-Gb* distal sites host a water molecule (gray) coordinated to the heme-Fe^{3+}. In *Bpe*GReg-Gb*-2 a coordination with Tyr43 is present in the A chain, while an imidazole molecule is bound in the B chain.


A
680
347
268
297
381
415
489
545
563
573
a
676
227
301
338
363
413
489
603
b
679
347
267
297
377
415
488
545
561
573
c
676
214
299
345
380
415
463*
486
585*
d
200
400
600
800
Raman shift [cm-1]
B
1379
1435
1471
1505
1583
1621
1639
a
1357
1395
1429
1474
1567
1619
b
1376
1433
1505
1579
1620
1637
c
1356
1426
1471
1563
1617
d
1400
1500
1600
Raman shift [cm-1]
C
676
269
298
344
382
410
486
a
675
229
299
344
412
463*
489
582*
b
681
349
268
302
382
421
492
544
563
c
678
233?
303
346
379
415
490
d
200
400
600
800
Raman shift [cm-1]
D
1359
1373
1425
1472
1501
1563
1616
1357
1394
1427
1472
1565
1615
a
1380
1437
1473
1507
1583
1623
1640
b
1359
1394
1429
1474
1567
1619
c
d
1400
1500
1600
Raman shift [cm-1]


A
ΔOD421
Time (s)
1E-7 1E-6 1E-5 1E-4 0.001 0.01 0.1 1 10 100
kobs,CO (s-1)
[CO] (μM)
B
ΔOD421
Time (s)
C
kf-(k'CO[CO]) (s-1)
[O2] (μM)
D
1/ks (s-1)
[O2]/[CO]

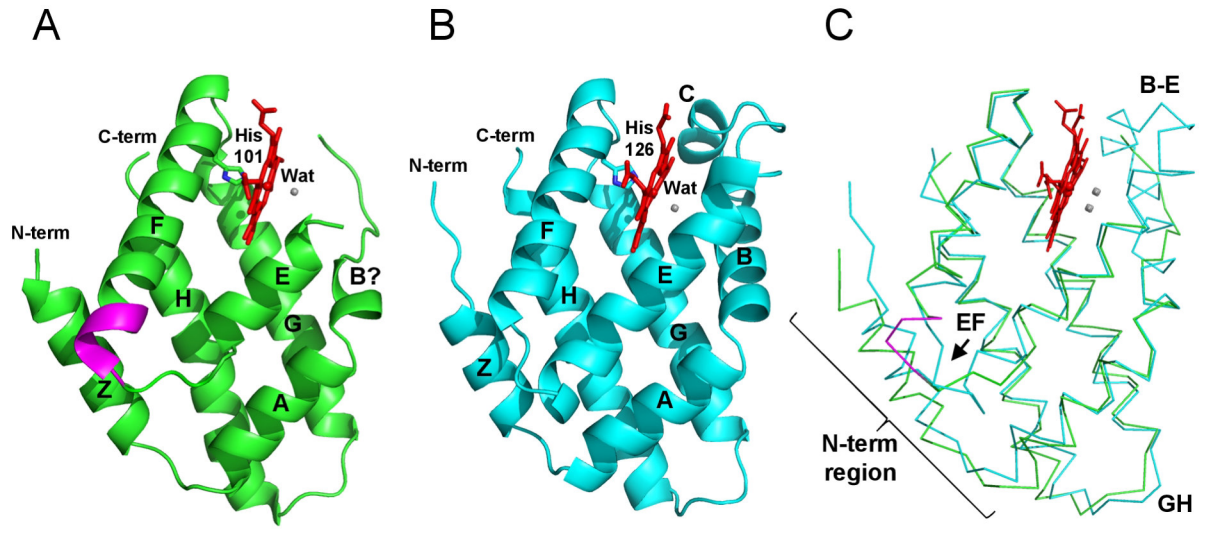

A
ΔOD421
Time (s)
1E-7 1E-6 1E-5 1E-4 0.001 0.01 0.1 1 10 100
kobs,CO (s-1)
[CO] (μM)
B
kf-(k'CO[CO]) (s-1)
[O2] (μM)
C
1/ks (s)
[O2]/[CO]

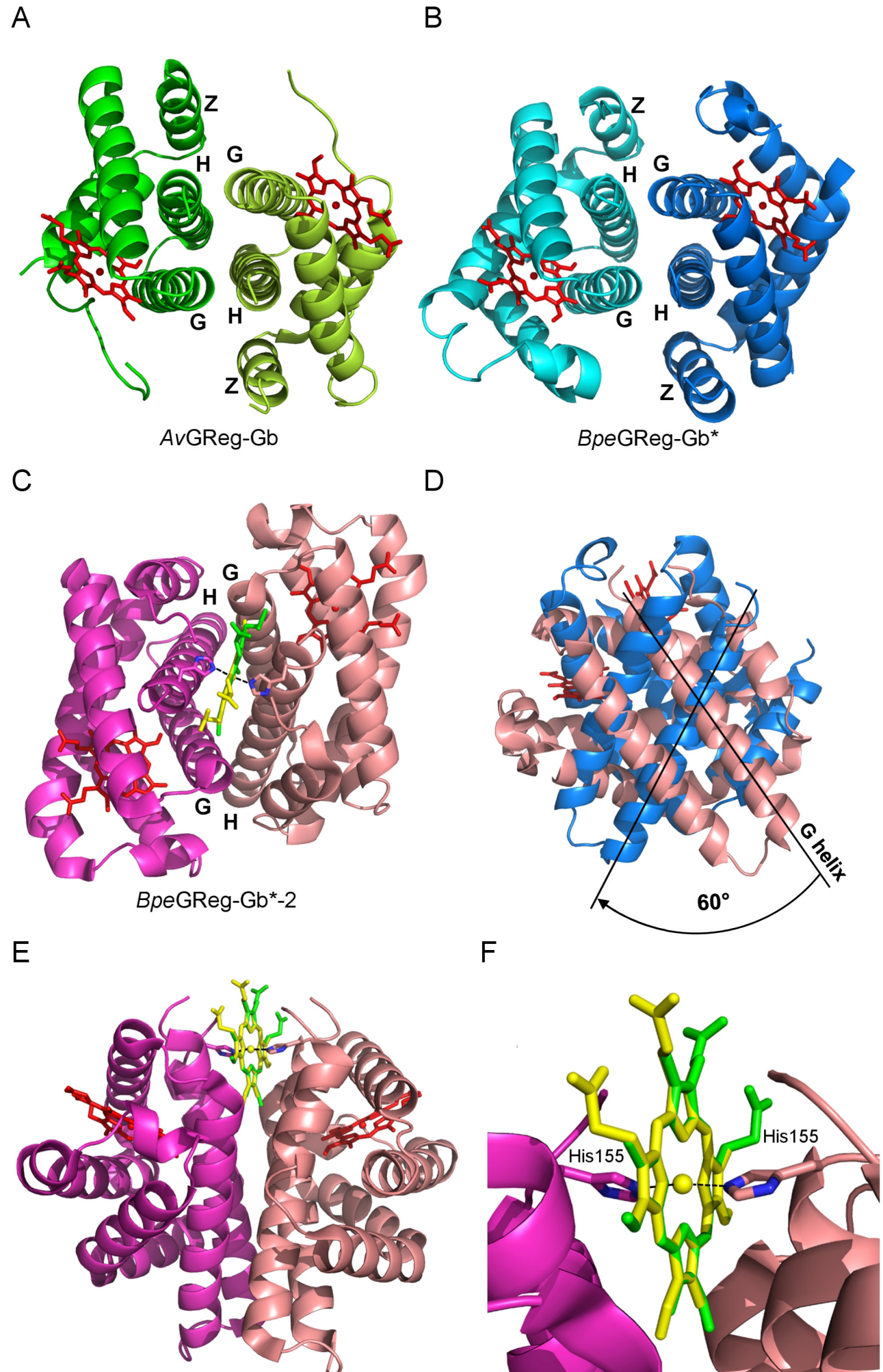
A



B

	Measurement 1	**Measurement 2**	**Measurement 3**	**Average**	**STDEV**
V_{max}	1.86 µM.min^{-1}	1.90 µM.min^{-1}	1.70 µM.min^{-1}	1.82 µM.min^{-1}	0.08 µM.min^{-1}
K_m	23.6 µM	30.9 µM	26.5 µM	27.0 µM	3.0 µM
K_{cat} / K_m	0.787 µM^{-1}.min^{-1} 1.31 x10^4 M^{-1}.s^{-1}	0.614 µM^{-1}.min^{-1} 1.02 x10^4 M^{-1}.s^{-1}	0.640 µM^{-1}.min^{-1} 1.06 x10^4 M^{-1}.s^{-1}	0.680 µM^{-1}.min^{-1} 1.13 x10^4 M^{-1}.s^{-1}	0.076 µM^{-1}.min^{-1} 0.13 x10^4 M^{-1}.s^{-1}


A
C-term
His
101
Wat
N-term
F
E
B?
H
G
Z
A
B
C
C-term
His
126
N-term
Wat
F
B
E
H
G
Z
A
C
B-E
EF
N-term
region
GH


A
Z
H
G
G
H
Z
AvGReg-Gb
B
Z
H
G
G
H
Z
BpeGReg-Gb*
C
G
H
G
H
BpeGReg-Gb*-2
D
G helix
60°
E
F
His155
His155

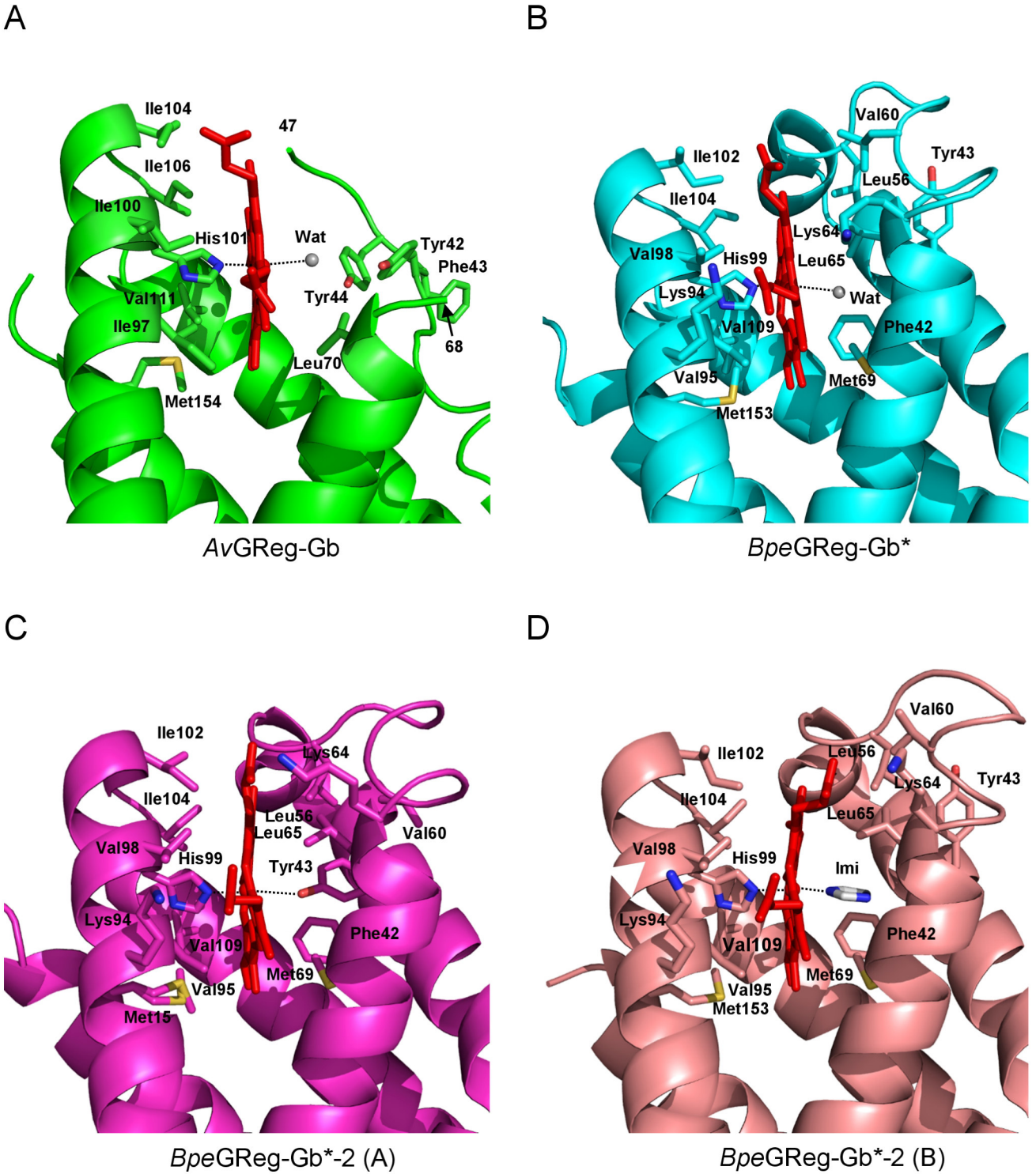

A
Ile104
47
Ile106
Ile100
His101
Wat
Tyr42
Phe43
Val111
Tyr44
Ile97
68
Leu70
Met154
AvGReg-Gb
B
Val60
Ile102
Tyr43
Leu56
Ile104
Lys64
Val98
His99
Leu65
Lys94
Wat
Val109
Phe42
Val95
Met69
Met153
BpeGReg-Gb*
C
Ile102
Lys64
Ile104
Leu56
Leu65
Val60
Val98
His99
Tyr43
Lys94
Val109
Phe42
Met69
Val95
Met15
BpeGReg-Gb*-2 (A)
D
Val60
Ile102
Leu56
Lys64
Tyr43
Ile104
Leu65
Val98
His99
Imi
Lys94
Val109
Phe42
Met69
Val95
Met153
BpeGReg-Gb*-2 (B)

Table 1. UV/VIS maxima of the different forms of *Av*GReg and *Bpe*GReg

***Av*GReg** (wavelength in nm)						
As purified (Fe^{2+}-O_2)	413			541		580
Reduced (Fe^{2+})			434		560	
CO-ligated		421		542	568	

***Bpe*GReg** (wavelength in nm)						
As purified (Fe^{2+}-O_2)	415			547		582
Reduced (Fe^{2+})			432		561	
CO-ligated		422		544	570	

Table 2. Resonance Raman vibrational modes values of different hemoproteins

Hemoprotein	Vibrational modes (cm^{-1})					Ref
	ν_4	ν_3	ν_2	ν_{Fe-His}	ν_{Fe-O2}	
O_2-*Av*GReg	1379	1505	1583	-	563/573	This study
*Av*GReg(FeII)	1357	1474	1567	227	-	
O_2-*Av*GReg-Gb	1376	1505	1579	-	561/573	
*Av*GReg-Gb(FeII)	1356	1471	1563	214(b)	-	
O_2-*Bpe*GReg	1373	1501	n.d.	-	n.d.	This study
*Bpe*GReg (FeII)	1357	1472	1565	229	-	
O_2-*Bpe*GReg-Gb	1380	1507	1583	-	563	
*Bpe*GReg-Gb (FeII)	1359	1474	1567	233	-	
O_2-*Bpe*GReg*	1373	1502	1577	-	556	This study
*Bpe*GReg* (FeII)	1356	1472	1564	229	-	
O_2-*Bpe*GReg-Gb*	1375	1502	1578	-	560	
*Bpe*GReg-Gb* (FeII)	1356	1472	1564	230(b)	-	
O_2-*Sw*Mb	1377	1507	1584	-	569	(60,61)
*Sw*Mb(FeII)	1356	1473	1564	220	-	
†*Av*GReg(FeIII)	1375	1506	1588	-	-	(64)
†*Av*GReg(FeII)	1361	1470	1592	n.d.	-	
†O_2-*Av*GReg178	1372	1502	1575	-	n.d.	
†*Av*GReg178(FeII)	1354	1470	1562	220	-	
O_2-*Bs*HemAT	1372	1501	1578	-	560	(2)
*Bs*HemAT(FeII)	1352	1469	1558	225	-	
O_2-*Ec*DosC-heme	1377	1499	1581	-	565	(28)
*Ec*DosC-heme(FeII)	1353	1470	1564	227	-	
O_2-*Af*GcHK	1375	1501	1579	-	557	(29)
*Af*GcHK(FeII)	1354	1470	1558	223	-	
O_2-*Rm*FixL*	1376	1502	1577	-	n.d.	(62)
*Rm*FixL*(FeII)	1355	1470	1558	212	-	
*Rr*CooA(FeII)	1359	1491	1579	n.d.	-	(67,68)
sGC(FeII)	1358	1471	1562	204	-	(13,65)

Error is 2 cm^{-1}, b=broad peak, n.d.=not detected

*Av*GReg and *Av*GReg-Gb: *in vivo* folded *A. vinelandii* full-length and truncated globin-coupled regulator; *Bpe*GReg and *Bpe*GReg*: globin-coupled sensor from *Bordetella pertussis* wild-type and with mutated cysteines (Cys16, 45, 114, 154); *Bpe*GReg-Gb and *Bpe*GReg-Gb*: isolated globin domain of *Bpe*GReg and *Bpe*GReg*: *Sw*Mb: *Sperm whale* myoglobin; †*Av*GReg and †*Av*GReg178: *in vitro* refolded *A. vinelandii* full-length and truncated globin-coupled regulator; *Bs*HemAT: heme-based aerotaxis transducer from *Bacillus subtilis*; *Ec*DosC-heme: isolated heme-binding domain from *Ec*DosC, a globin-coupled sensor with DGC activity from *E. coli*; *Af*GcHK: globin-coupled histidine kinase from *Anaeromyxobacter sp.* Fw109-5; *Rm*FixL*: soluble truncated domain of the *Rhizobium meliloti* nitrogen fixation gene expression regulator; *Rr*CooA: *Rhodospirillum rubrum* carbon monoxide oxidation activator; sGC: soluble guanylyl cyclase.

Table 3. Kinetics and equilibrium constants for ligand binding to some hemoproteins

Hemoprotein	CO	O_2			
	k'_{CO} (10^6 M^{-1} s^{-1})	k'_{O2} (10^6 M^{-1} s^{-1})	k_{O2} (s^{-1})	$K_d = \frac{k_{O2}}{k'_{O2}}$ (10^{-6} M)	Ref
*Av*GReg	0.48±0.01	12.2±1.5	1732±377 0.53±0.05	141 0.04	This study
*Av*GReg-Gb	0.20±0.01	-	-	-	This study
*Av*GRegYB10F	0.31±0.01	20.2±1.3	18389±556 166±18	910 8.22	This study
*Bpe*GReg	0.60±0.01 0.01±0.01	24.02±1.5	4.0±0.1	0.17	This study
*Bpe*GReg-Gb	0.11±0.01	-	-	-	This study
†*Av*GReg	1	- -	10.6 0.73	- -	(64)
†*Av*GReg178	1	424 5.2	10.6 0.13	0.025 0.025	(64)
*Gs*GCS[162]	6.80	-	23	-	(44)
*Bs*HemAT	0.34	19	1900 87	100 4.6	(75)
*Bpe*GReg	1.03 0.12	7.0	4.5	0.64	(70)
*Ec*DosC	0.22	0.9	13	14	(28)
*Ec*DosC-heme	0.40	1.4	22	16	(28)
*Af*GcHK	0.05	1.3 0.15	0.10	0.08 0.67	(29)
*Bj*FixL	-	0.14	20	140	(22)
*Sw*Mb	0.50	14	11	1.27	(5, 21)

*Av*GReg and *Av*GReg-Gb: *in vivo* folded *A. vinelandii* full-length and truncated globin-coupled regulator; *Av*GRegYB10F: mutant of the two proteins in which the Tyr(44) at position B10 is replaced by Phe; *Bpe*GReg and *Bpe*GReg-Gb: full-length and truncated globin-coupled sensor from *Bordetella pertussis*; †*Av*GReg and †*Av*GReg178: *in vitro* refolded *A. vinelandii* full-length and truncated globin-coupled regulator; *Gs*GCS[162]: sensor domain from the *Geobacter sulfurreducens* globin-coupled sensor; *Bs*HemAT: heme-based aerotaxis transducer from *Bacillus subtilis*; *Ec*DosC: globin-coupled sensor with DGC activity from *E. coli*; *Ec*DosC-heme: *Ec*DosC-isolated heme-binding domain; *Af*GcHK: globin-coupled histidine kinase from *Anaeromyxobacter sp.* Fw109-5; *Bj*FixL: *Bradyrhizobium japonicum* nitrogen fixation gene expression regulator; *Sw*Mb: *Sperm whale* myoglobin.

Table 4. Data collection and refinement statistics

Data collection:	***Av*GReg-Gb**	***Bpe*GReg-Gb***	***Bpe*GReg-Gb*-2**
Space group	$P2_1$	$C2$	$C2$
Cell dimensions:			
a, *b*, *c* (Å)	47.9, 49.5, 62.8	105.0, 86.7, 41.8	70.0, 53.1, 81.2
α, β, γ (°)	90.0, 91.8, 90.0	90.0, 93.3, 90.0	90.0, 93.2, 90.0
Resolution (Å)	30.5 - 2.83 (2.98-2.83)[a]	52.4-3.20 (3.37-3.2)	42.3-1.55 (1.63-1.55)
No. reflections	22,497	17,105	145,853
Unique reflections	6,845	6,186	42,168
R-merge[b] (%)[a]	9.3 (49.8)	24.6 (52.7)	6.8 (49.8)
I/σ(I)[a]	8.6 (2.5)	3.4 (2.0)	9.8 (2.7)
Completeness (%)[a]	95.8 (97.7)	99.2 (99.5)	97.6 (97.5)
Multiplicity[a]	3.3 (3.4)	2.8 (2.8)	3.5 (3.5)
Refinement statistics:			
R-factor[c]/R-free (%)	25.5/29.7	20.9/24.0	15.6/20.4
Protein residues	258 (A: 6-47, 68-157) (B: 7-45, 70-156)	306 (A:3-157) (B: 3-56, 61-157)	307 (A: 5-158) (B: 3-155)
Heme groups	2	2	3
Water molecules	4	30	171
Glycerol molecules	-	-	2
Imidazole molecules	-	-	1
r.m.s.deviation from ideality:			
Bond lengths (Å)	0.011	0.010	0.017
Bond angles (°)	1.97	1.45	1.81
Ramachandran plot (%):			
Most favored	88.1	92.3	98.7
Additionally allowed	11.9	7.7	1.3
Disallowed regions	0.0	0.0	0.0

[a] Highest resolution shell parameters are in parentheses

[b] R-merge = $\Sigma_h\Sigma_i \mid I_{hi} - \langle I_h \rangle \mid / \Sigma_h\Sigma_i I_{hi}$

[c] R-factor = $\Sigma_h ||F_{obs}| - |F_{calc}|| / \Sigma_h |F_{obs}|$, with F_{obs} being the observed and F_{calc} the calculated structure factor amplitudes

Supplementary legends

Figure S1. Resonance Raman spectra of *Bpe*GReg* and *Bpe*GReg-Gb*.

RR spectra of *Bpe*GReg* (a,b) and *Bpe*GReg-Gb* (c,d) in the low-frequency region (**A**) and high-frequency region (**B**). Traces (a,c) show the RR spectra of the proteins as purified (oxygenated Fe(II) state), traces (b,d) show the RR spectra of the dithionite-reduced proteins (ferrous state). Wavenumbers marked with asterisk indicate reminiscent contributions of dithionite as described in (8). These signals hamper the analysis of the low-frequency region of the RR spectrum of *Bpe*GReg-Gb*.

Figure S2. Detailed Resonance Raman spectrum of O_2-*Av*GReg.

The figure zooms in on the region in which the Fe-O_2 modes are visible.

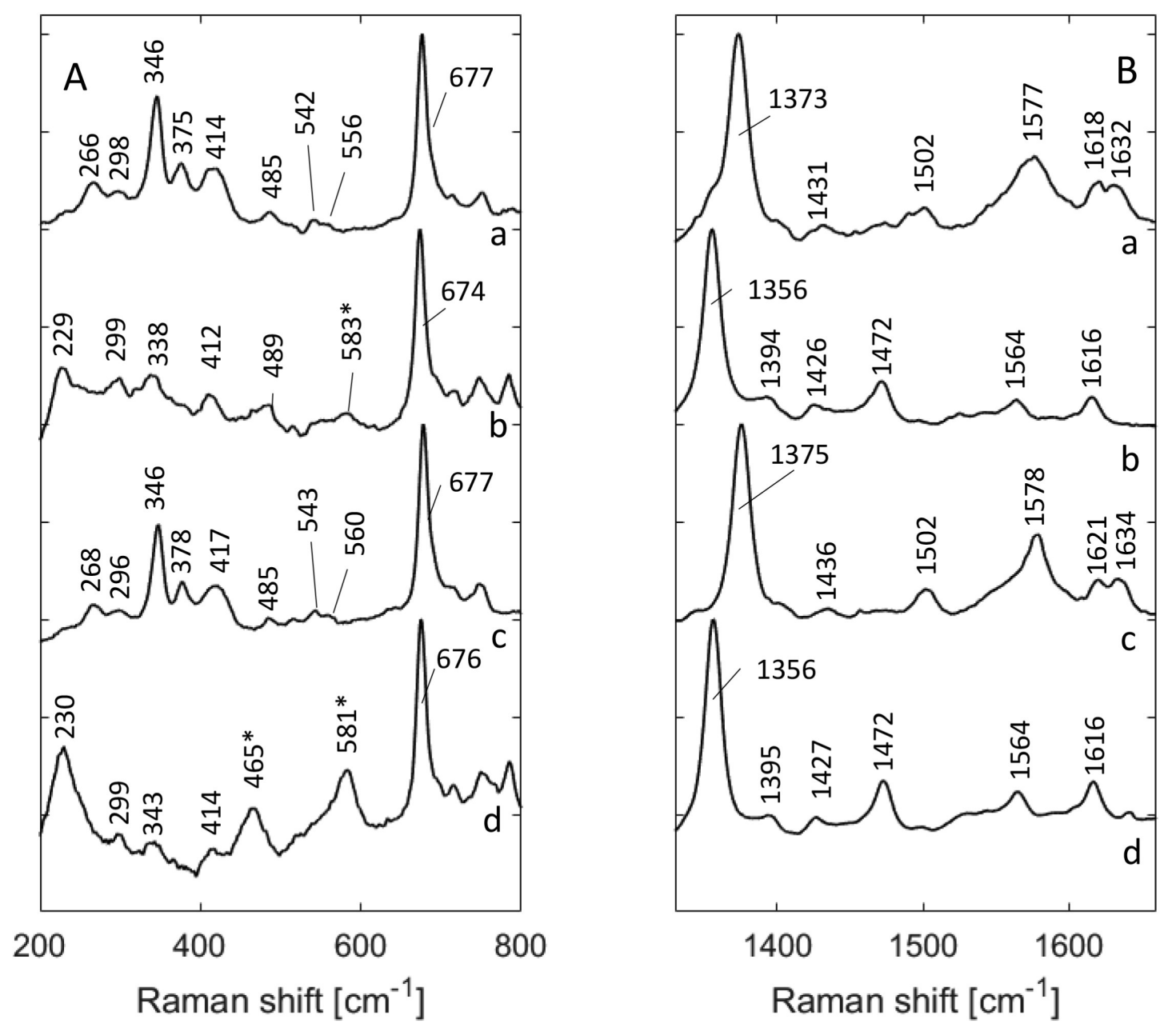
Figure S3. UV/Vis absorption (A) and Resonance Raman (B,C) spectra of ferric *Av*GReg.

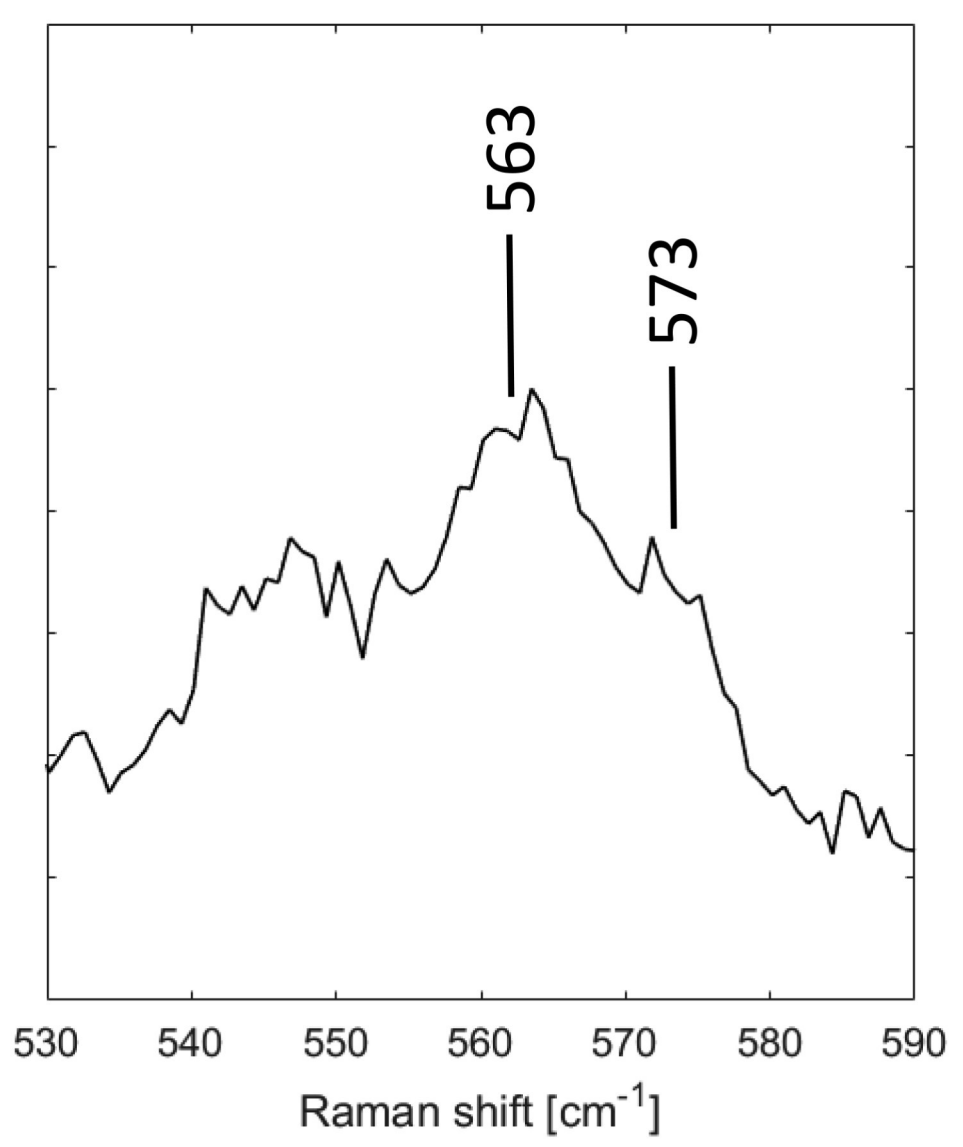
*Av*GReg was oxidized by addition of potassium ferricyanide. (**A**) The Soret and Q-bands confirm the oxidation to a ferric low-spin form. However, the presence of a peak at 646 nm indicates that also partial degradation has taken place (formation of hemochrome derivatives, like biliverdin). (**B,C**) The RR spectra are similar to those observed previously for the *in-vivo* refolded †*Av*GReg (64). RR peaks confirm the assignment from (prominent peaks at ν_3=1505 cm^{-1} confirms the low-spin (hexa-coordination of the heme)). Additionally, the observation of small peaks in the area of 545-565 cm^{-1} indicate that some of the proteins are still in the oxygenated ferrous form. Interestingly, the 573 cm^{-1} peak has disappeared indicating that this open configuration of the heme group oxidizes the easiest.

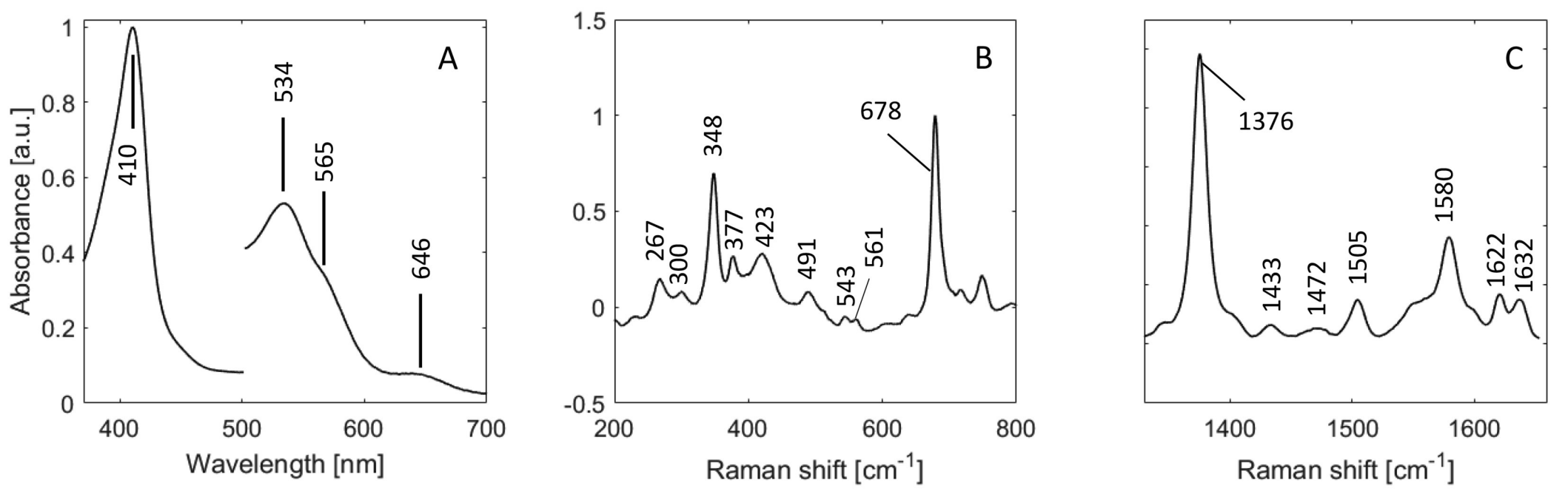
Figure S4. Ligand binding kinetics curves for *Av*GRegYB10F.

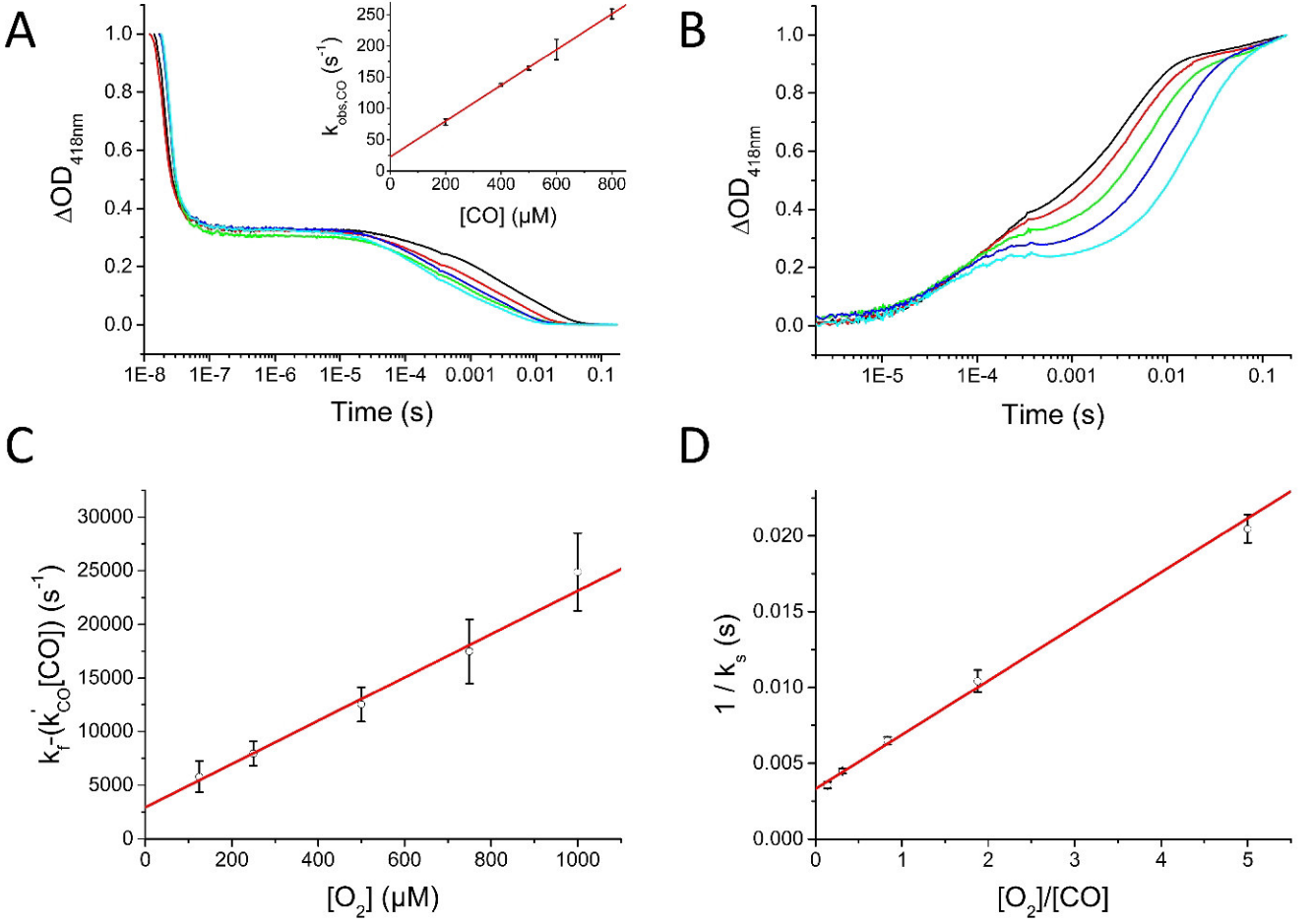
(**A**) Experimental curves of the CO-rebinding to *Av*GRegYB10F (black: 200 µM CO, red: 400 µM CO, green: 500 µM CO, blue: 600 µM CO, cyan: 800 µM CO). The dependence of the pseudo-first

order constants of CO binding (k_{obsCO}) to *Av*GReg YB10F on the relative ligand concentration is reported in the inset. (**B**) Stopped-flow O_2/CO replacement experimental curves of *Av*GRegYB10F (black: 90% CO - 10% O_2, red: 80% CO - 20% O_2, green: 60% CO - 40% O_2, blue: 40% CO - 60% O_2, cyan: 20% CO - 80% O_2). The *fast* and the *slow* replacement phases are detectable for all the gas ratios used. (**C**) Dependence of the apparent k_f constants of O_2 binding to *Av*GRegYB10F on the relative ligand concentration. (**D**) Dependence of the apparent k_s constants of CO binding to *Av*GRegYB10F on the relative ligand concentration. Experimental values are reported with open circles, while the theoretical ligand-binding mechanism is represented with a straight line.


A
346
677
266
298
375
414
485
542
556
a
674
229
299
338
412
489
583*
b
677
346
268
296
378
417
485
543
560
c
676
230
299
343
414
465*
581*
d
200
400
600
800
Raman shift [cm-1]
B
1373
1577
1618
1632
1431
1502
a
1356
1394
1426
1472
1564
1616
b
1375
1578
1621
1634
1436
1502
c
1356
1395
1427
1472
1564
1616
d
1400
1500
1600
Raman shift [cm-1]


563
573
530
540
550
560
570
580
590
Raman shift [cm^{-1}]


A
Absorbance [a.u.]
Wavelength [nm]
410
534
565
646
B
Raman shift [cm^{-1}]
267
300
348
377
423
491
543
561
678
C
Raman shift [cm^{-1}]
1376
1433
1472
1505
1580
1622
1632


A
ΔOD418nm
Time (s)
1E-8
1E-7
1E-6
1E-5
1E-4
0.001
0.01
0.1
kobs,CO (s-1)
[CO] (μM)
B
ΔOD418nm
Time (s)
C
kf-(k'CO[CO]) (s-1)
[O2] (μM)
D
1 / ks (s)
[O2]/[CO]