

# SCIENTIFIC REPORTS



OPEN

## Open and Lys–His Hexacoordinated Closed Structures of a Globin with Swapped Proximal and Distal Sites

Received: 24 November 2014

Accepted: 22 May 2015

Published: 22 June 2015

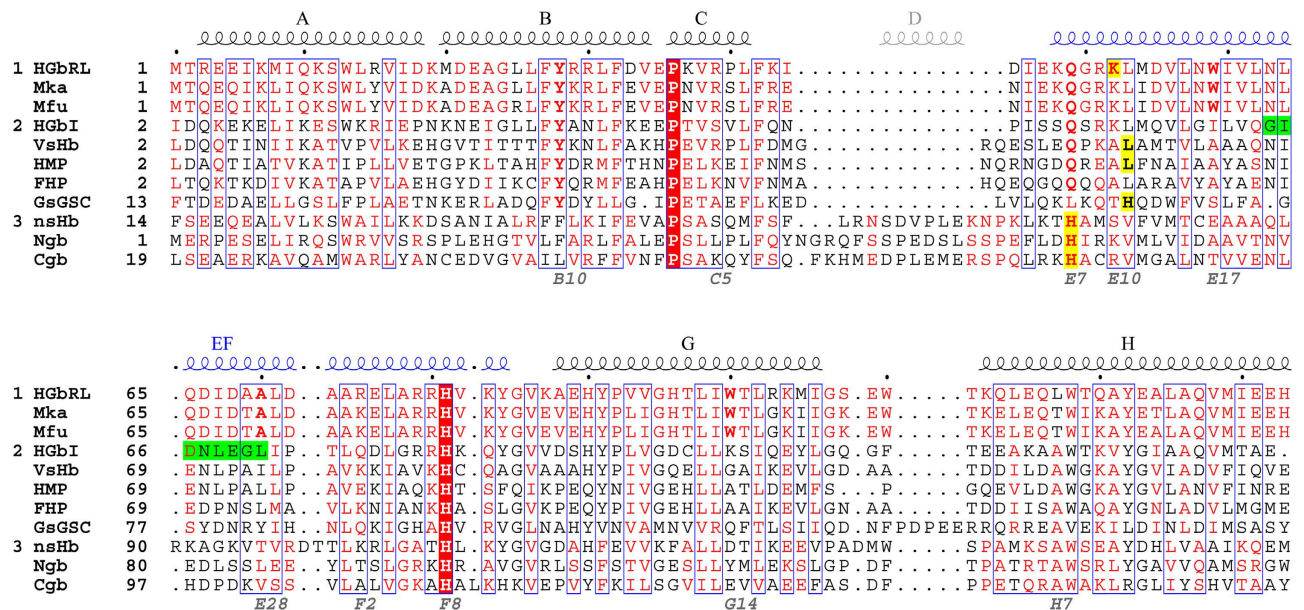
Aik-Hong Teh<sup>1</sup>, Jennifer A. Saito<sup>2</sup>, Nazalan Najimudin<sup>3</sup> & Maqsudul Alam<sup>2,4</sup>

Globins are haem-binding proteins with a conserved fold made up of  $\alpha$ -helices and can possess diverse properties. A putative globin-coupled sensor from *Methylobacterium thermophilum*, HGbRL, contains an N-terminal globin domain whose open and closed structures reveal an untypical dimeric architecture. Helices E and F fuse into an elongated helix, resulting in a novel site-swapped globin fold made up of helices A–E, hence the distal site, from one subunit and helices F–H, the proximal site, from another. The open structure possesses a large cavity binding an imidazole molecule, while the closed structure forms a unique Lys–His hexacoordinated species, with the first turn of helix E unravelling to allow Lys52(E10) to bind to the haem. Ligand binding induces reorganization of loop CE, which is stabilized in the closed form, and helix E, triggering a large conformational movement in the open form. These provide a mechanical insight into how a signal may be relayed between the globin domain and the C-terminal domain of HGbRL, a Roadblock/LC7 domain. Comparison with HGbI, a closely related globin, further underlines the high degree of structural versatility that the globin fold is capable of, enabling it to perform a diversity of functions.

The bacterium *Methylobacterium thermophilum*, isolated from the Hell's Gate (Tikitere) geothermal area in New Zealand, is an aerobic methanotrophic bacterium growing optimally at 60 °C and pH 2.0<sup>1</sup>. Its genome encodes five globins — one double-domain HGbRL and four single-domain HGbI–IV<sup>2</sup> — and thus far only the single-domain HGbI<sup>3</sup> and HGbIV<sup>4</sup> have been studied. HGbI possesses a very high affinity for O<sub>2</sub> and an unusually low autoxidation rate<sup>3</sup>. Its structure consists of a typical but compact globin fold, with the highly conserved Tyr29(B10) and Gln50(E7) in the distal site flexible enough to bind a molecule of either O<sub>2</sub> or acetate<sup>3</sup>. HGbIV, on the other hand, is a truncated haemoglobin with two unique extensions at both termini — a Pre-A loop at the N terminus and a helix I at the C terminus<sup>4</sup>. Its distal site, intriguingly, is comparatively large and polar, containing a specifically conserved His70(B9)–His71(B10) motif that in a crystal structure was well positioned to bind a phosphate ion<sup>4</sup>.

The globin fold, having one of the most versatile architecture, can bind a variety of different ligands in the distal site. Besides small ligands such as O<sub>2</sub>, CO and NO, ligands as large as econazole or even a phospholipid have been reported in the structures of yeast flavohaemoglobin<sup>5</sup> (Yhb) and *Ralstonia eutropha* flavohaemoglobin (FHP)<sup>6–8</sup>. In some globins, their His(E7) bind directly to the haem in the absence of an exogenous ligand, forming a bis-His hexacoordinated species first observed in the structure of a non-symbiotic plant haemoglobin<sup>9</sup>. Neuroglobin, meanwhile, employs a novel haem-sliding mechanism for

<sup>1</sup>Centre for Chemical Biology, Universiti Sains Malaysia, 10 Persiaran Bukit Jambul, 11900 Bayan Lepas, Penang, Malaysia. <sup>2</sup>Department of Microbiology, University of Hawaii, 2538 McCarthy Mall, Honolulu, HI 96822, United States. <sup>3</sup>School of Biological Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia. <sup>4</sup>Advanced Studies in Genomics, Proteomics and Bioinformatics, University of Hawaii, 2565 McCarthy Mall, Honolulu, HI 96822, United States. Correspondence and requests for materials should be addressed to A.H.T. (email: aikhong@usm.my)

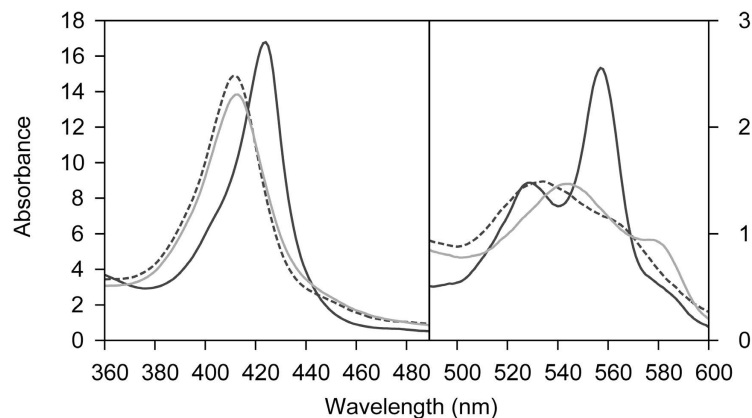


**Figure 1. Structural alignments of globins.** HGbRL is unique to the genus *Methylococcoides* (Group 1), presently found only in two other bacteria, *M. kamchatkense* (Mka) and *M. fumariolicum* (Mfu) (aligned by protein sequences). Its globin domain is closely related to bacterial globins (Group 2) with conserved Tyr(B10) and Gln(E7), as well as lacking helix D that is present in eukaryotic globins (Group 3). Thus far N-HGbRL is the only globin containing fused helices E and F (EF) among all the known structures, with the corresponding loop EF of HGbI (green) straightened into part of the elongated helix. Haem hexacoordination in eukaryotic globins invariably involves His(E7) (yellow), as seen in nonsymbiotic plant haemoglobin (nsHb), neuroglobin (Ngb) and cytoglobin (Cgb). In bacterial globins, on the other hand, hexacoordination involves residues at either position E10 or E11, such as Lys52(E10) in N-HGbRL, His66(E11) in GsGSC, and Leu57(E11) in both VsHb and HMP for haem shielding. Following the conventional numbering for the distal Gln/His(E7) and proximal His(F8), the fused helices E and F are numbered E6–E31 followed by F1–F11.

its bis-His hexacoordination<sup>10</sup>. In the unligated structure of *Escherichia coli* flavohaemoglobin (HMP), interestingly, the sixth coordination site is physically occupied by Leu57(E11)<sup>11</sup>.

Globins have also been found fusing to the N termini of a variety of proteins, forming multi-domain proteins which are widely distributed in prokaryotes and eukaryotes alike. In flavohaemoglobins, which function in cellular responses to nitrosative stress, the globin domain is fused to a ferredoxin reductase-like module consisting of a FAD- and an NADH-binding domain<sup>12</sup>. In HemAT, a globin-coupled sensor (GCS) involved in *Bacillus subtilis* aerotaxis, the C-terminal domain is a methyl-accepting chemotaxis protein<sup>13</sup>. RsbR from *B. subtilis*, on the other hand, comprises a C-terminal STAS domain and forms a gigantic stressosome with RsbS and RsbT, which plays a role in the bacterium's general stress response<sup>14</sup>. The N-terminal globin domain of RsbR, however, not only lacks helices C and D but also has the strictly conserved haem-binding His(F8) in the proximal site replaced by Ala74 — it no longer binds any haem<sup>15</sup>. While *B. subtilis* RsbR represents an extreme example of a globin that has evolved to lose its haem, the RsbR homologues from *Sapropira grandis* still retain their haem and are able to bind O<sub>2</sub><sup>16</sup>.

HGbRL from *M. infernorum* is a putative globin-coupled sensor of 274 residues, consisting of an N-terminal globin domain and a C-terminal Roadblock/LC7 domain. It is unique to the genus *Methylococcoides*, and presently is found only in two other closely related bacteria, *M. kamchatkense*<sup>17</sup> and *M. fumariolicum*<sup>18</sup>. The globin domain, N-HGbRL, is 38% identical to HGbI, with both belonging to the same group of bacterial globins that lack helix D and conserve both Tyr(B10) and Gln(E7) (Fig. 1). The structures of several bacterial globins from this group, such as HMP<sup>11</sup>, FHP<sup>6–8</sup> and *Vitreoscilla stercoraria* haemoglobin (VsHb)<sup>19</sup>, have displayed a high degree of versatility in their distal sites involving mainly loop CE and helix E. HMP, for instance, moves its helix E inwards in the closed form to occupy the distal site<sup>11</sup>, while FHP moves its outwards in the open form, expanding the distal site to bind a ligand as large as a phospholipid<sup>7</sup>. Here we further demonstrate the architectural plasticity of the globin fold by reporting the structure of an unprecedented site-swapped dimeric globin, N-HGbRL, which has been solved in both the open and closed forms with several novel features.



**Figure 2. Absorption spectra.** Ferric *N*-HGbRL after purification (dashed line) is reduced to the deoxy form (black) in an excess of sodium dithionite, which gradually turns into the oxy form after 20 minutes (grey).

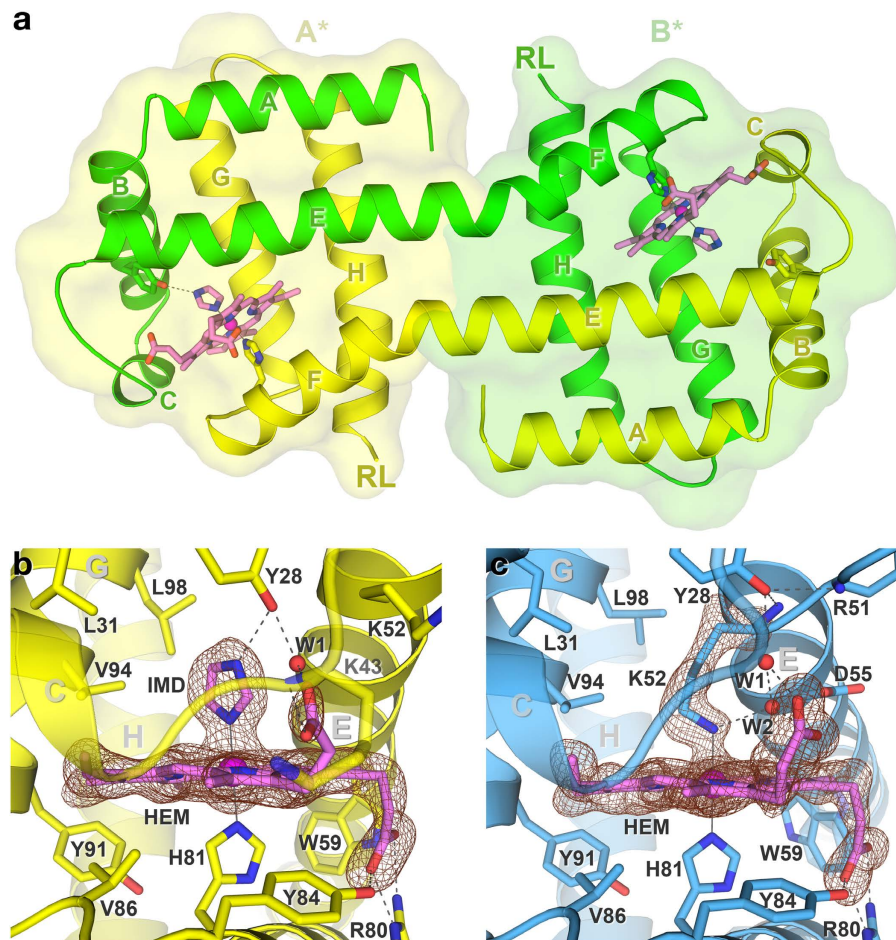
## Results and Discussion

**Hexacoordination of *N*-HGbRL.** The purified dimeric *N*-HGbRL was in the ferric state, whose spectrum revealed a predominantly hexacoordinated form with a Soret band at 412 nm and a visible peak at 534 nm (Fig. 2). In the presence of an excess of sodium dithionite that also removed dissolved oxygen, *N*-HGbRL was reduced but in the hexacoordinated form too, with the Soret band shifting to 424 nm and two distinct peaks emerging at 529 and 557 nm (Fig. 2). Both HgbI<sup>3</sup> and HgbIV<sup>4</sup> have been reported to form a hexacoordinated species as well. The internal sixth ligand, however, did not seem to form a strong interaction with the haem at least in the ferrous state, as under the experimental conditions the deoxy *N*-HGbRL slowly turned oxygenated within 20 minutes, with the Soret band and the two peaks shifting to 413, 544 and 576 nm (Fig. 2).

**Site-swapped *N*-HGbRL with fused helices E and F.** *N*-HGbRL consists of seven  $\alpha$ -helices, helices A–H, lacks helix D and forms a dimer. It displays a typical bacterial globin fold except for a very untypical feature — its helices E and F unexpectedly fuse into an elongated helix (Fig. 3a). Loop EF is straightened into part of the  $\alpha$ -helical conformation, and the elongated helices from both subunits form an antiparallel pair. Instead of being sandwiched between helices E and F from a single globin subunit, a haem group in *N*-HGbRL is sandwiched by two subunits. A single *N*-HGbRL globin fold, denoted subunit A\*, is therefore made up of helices A–E from one subunit and helices F–H from the other. In other words, both the proximal and distal sites of a single globin fold are constituted mainly by two separate subunits, representing the first ever site-swapping example in the globin family. The two C-terminal Roadblock/LC7 domains, meanwhile, will be positioned at the opposite ends of the dimer's helices H, but may still lie in the same side to interact with one another.

The structure of *N*-HGbRL was first solved in the open form in space group *C*2, with two subunits in the asymmetric unit. In the distal site, an imidazole molecule from the Ni<sup>2+</sup>-affinity chromatography binds to the haem and the highly conserved Tyr28(B10), which also coordinates a water molecule, W1 (Fig. 3b). Several imidazole-bound globin structures have been reported too, including those of V<sub>s</sub>Hb<sup>20</sup> and sperm whale myoglobin<sup>21</sup>, and even larger ligands with an imidazole moiety bound in a similar pattern to the haem, such as the econazole molecules bound to the structures of FHP<sup>8</sup> and Yhb<sup>5</sup>, have also been observed. The distal site of *N*-HGbRL is lined by all hydrophobic residues (Phe27, Leu31, Phe42, Leu53, Val56, Val94 and Leu98) except for Tyr28(B10) and Lys52(E10). The A-ring propionate of the haem also hydrogen-bonds to the side chains of Arg80 and Tyr84, and the D-ring propionate to Lys43 N and the Tyr28-bound water W1 (Fig. 3b).

Further optimization subsequently yielded crystals of the closed form, also in the same space group *C*2 but with only a subunit in the asymmetric unit. Intriguingly, the first turn of helix E in the distal site has unravelled, allowing Lys52(E10) to close in to bind to the haem at the sixth coordination site and to a water molecule, W2 (Fig. 3c). Tyr28(B10), meanwhile, has been driven away to hydrogen-bond to both the N of Arg51(E9) and Lys52(E10), as well as to W1. Proposed as a mechanism to regulate ligand affinity, haem hexacoordination has been widely observed in several globin structures including nonsymbiotic plant haemoglobin<sup>9</sup>, fruit fly haemoglobin<sup>22</sup>, *Caenorhabditis elegans* neural globin GLB-6<sup>23</sup>, cytoglobin<sup>24</sup> and neuroglobin<sup>25,26</sup> that normally involves a distal His(E7) residue, or in truncated haemoglobin<sup>27</sup> an His(E10) residue. Haem coordination by lysine, however, has thus far been reported only in the structures of cytochrome *c* nitrite reductase (NrfA)<sup>28</sup> and tetrathionate reductase<sup>29</sup>, but both in the proximal site. Presently the only structures of haem hexacoordination by a proximal histidine and a distal lysine are found in a complex of NrfA and NrfH, a membrane-bound cytochrome *c* quinol

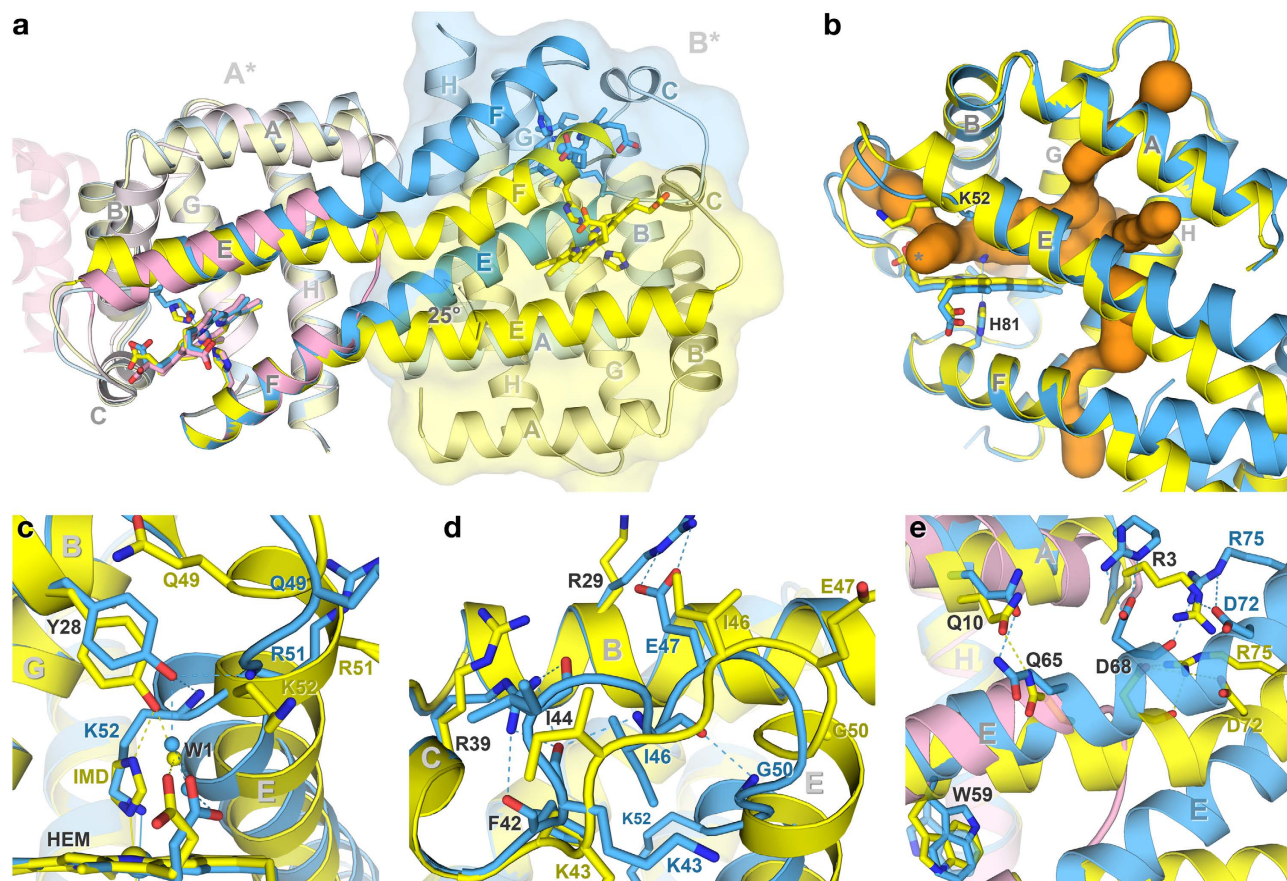


**Figure 3. Structures of *N*-HGbRL in the open and closed forms.** (a) Site-swapped dimer of *N*-HGbRL. Both helices E and F unexpectedly fuse into an elongated helix, resulting in a globin fold, denoted A\*, being made up of helices A–E from one subunit (green) and F–H from the other (yellow). While the two C-terminal Roadblock/LC7 (RL) domains are positioned at the opposite ends, they may still lie on the same side to interact with each other. (b) Open form of *N*-HGbRL. An imidazole (IMD) molecule binds to the haem (HEM) and the conserved Tyr28(B10). (c) Closed form of *N*-HGbRL. Helix E moves inwards and unravels its first helical turn for Lys52(E10) to bind to the haem, forming a unique Lys–His hexacoordinated structure. The  $F_o - F_c$  omit maps for both structures are contoured at 3  $\sigma$ .

dehydrogenase<sup>30</sup>, as well as in the M100K mutant of cytochrome *c*-550 from *Paracoccus versutus*, whose haem-coordinating Met100 was replaced by lysine<sup>31</sup>. The closed form of *N*-HGbRL is hence the first globin structure with Lys–His hexacoordination.

**Distal site reorganization and a large conformational shift upon ligand binding.** Both subunits A\* of the open and closed structures superpose well at helices A–E (rmsd 1.22 Å) and considerably better at helices F–H (rmsd 0.46 Å), with the distal site undergoing some conformational changes. A ~25° bending of helix E results in a large shift of subunits B\* between the two forms (Fig. 4a), creating a huge pocket of about 250 Å<sup>3</sup> in the distal site of the open form that connects to several tunnels (Fig. 4b). In the closed form, meanwhile, Lys48(E6)–Arg51(E9) are untangled and no longer form the first turn of helix E, which tilts towards the haem and binds to it (Fig. 4c). Loop CE is also pulled towards helices B and C, stabilized by new hydrogen bonds between Arg39(C5) from helix C and Phe42–Ile44, between Gly50(E8) from the unravelled turn of helix E and Ile46, and between Arg29(B11) from helix B and Glu47 (Fig. 4d). The side chain of Trp59(E17) has two alternate conformations in the closed form — one with an occupancy of 0.6 facing the solvent and another of 0.4 flipping 180° inwards — but in the open form it flips inwards, appearing to act as a switch between the two forms (Fig. 4e). In the open form, interaction between subunits B\* from two adjacent asymmetric units, Arg39(C5) from one asymmetric unit and Glu47 from the other, causes loop CE to shift relative to that of subunit A\* (Supplementary Fig. S1).

These changes provide a mechanical insight into how *N*-HGbRL may sense a signal. When a signalling ligand binds the haem in the closed form, it dislodges Lys52(E10) and pushes helix E outwards,



**Figure 4. Conformational changes between the open and closed forms.** (a) Superposition of open (yellow) and closed (blue) forms of subunits A\* of *N*-HGbrL with HGBI (pink) at helices F–H. A bending of  $\sim 25^\circ$  of the fused helices E and F leads to a large shift in subunits B\* of the two *N*-HGbrL forms. Dimerization of HGBI involves helices B of the first and the second (transparent) subunits. (b) Tunnels and distal pocket of *N*-HGbrL. Besides connecting to several tunnels (orange), the pocket in the distal site of the open form is also directly exposed to the solvent through an opening (\*). (c) Conformational changes in distal site. Helix E moves outwards in the open form to allow Tyr28(B10) to bind to imidazole, whereas it moves towards the haem in the closed form and drives Tyr28(B10) away from the distal site. (d) Conformational changes in loop CE. While folding compactly into the distal site in the closed form, loop CE as well as helix E in the open form extend outwards to open up the distal site. (e) Conformational changes along the fused helices E and F between the two forms.

inducing part of loop CE to fold into the extra turn of helix E of the open form (Fig. 4c). This may probably expose Trp59(E17) wholly to the solvent and prompt it to flip inwards, further facilitating helices E and F to move away from helix A (Fig. 4a). The side chain of Asp68(E26) breaks its bonding with Arg3 and in turn binds to Arg75(F2), promoting the bending of the fused helices E and F (Fig. 4e). The presence of cooperativity between the two subunits can also be deduced from these structures — the conformational changes induced by ligand binding in subunit A\* drives helix F in subunit B\* towards helix E, forcing the latter's fused helices to bend and hence to trigger the reorganization that opens up the closed distal site of subunit B\* (Fig. 4a). Once the ligand dissociates, loop CE may tend to assume the closed conformation as it is stabilized by the interactions, absent in the open form, with Arg29(B11), Arg39(C5) and Gly50(E8) that pulls Lys52(E10) towards the haem (Fig. 4d).

**A new globin subgroup.** The huge pocket of *N*-HGbrL's open structure may be able to bind a comparatively large ligand. FHP likewise has a very huge pocket and forms an open and closed state, characterized by rearrangement of loop CE, an extra turn of helix E and a reorientation of the C-terminal NADH domain in the open state<sup>6,8</sup>. However, FHP's reportedly 'closed' state is not totally closed — its two closed structures still bind some large azole derivatives — and it is also unclear if FHP's helix E can move further inwards to bind to the haem when an exogenous ligand is actually absent. On the other hand, HMP has been solved only in a completely closed structure, with its haem totally shielded by Leu57(E11)<sup>11</sup>. The single-domain VSHb also shields its haem with Leu57(E11) in the closed form, but only slight movements of its helix E and a sideways rotation of Leu57(E11) are already sufficient to

open up its distal site to bind a molecule of azide, thiocyanate or imidazole<sup>19,20</sup>. Nevertheless, as HMP's Leu57(E11) is closer to the haem than Vshb's Leu57(E11) is, in HMP it may require a larger outward movement of helix E for ligand binding.

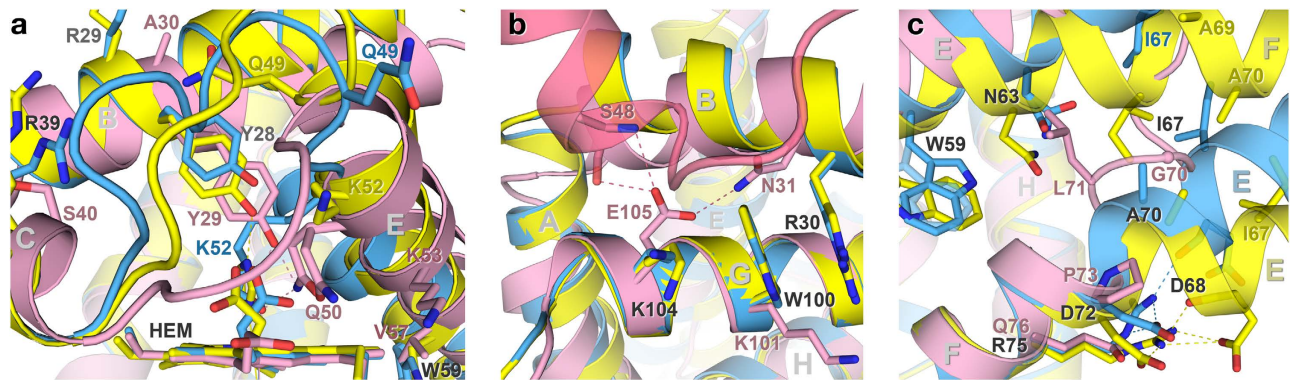
*N*-HGbRL and these multi-domain globins probably constitute a subgroup of novel bacterial globins that share distinct features such as an open and a closed form, a huge distal pocket, haem hexacoordination or shielding, and, upon ligand binding, reorganization of loop CE and helix E which may possibly lead to a large movement of a neighbouring domain. A signature of this subgroup may be haem hexacoordination or shielding with a residue at site E10 or E11, instead of E7, such as Lys52(E10) in *N*-HGbRL and Leu57(E11) in HMP. Haem hexacoordination normally involves the invariant His(E7), and only recently has an E11 residue been reported, i.e. His66(E11) in *Geobacter sulfurreducens* globin-coupled sensor (*GsGCS*)<sup>32</sup>. These globins notably do not possess helix D, whereas globins with haem-coordinating His(E7) usually do. The absence of helix D may not only increase the flexibility in this region, but a shorter sequence may also be more energetically favoured for a concerted conformational change involving loop CE and helix E. Still, at present as only *N*-HGbRL has been solved in both forms, more structures of globins in either form are needed for further studies.

**Architectural differences between *N*-HGbRL and HGbI.** Despite an identity of 38%, the unsuspected large structural differences between *N*-HGbRL and HGbI<sup>3</sup> have eclipsed their sequence closeness. HGbI contains several buried residues [Gly79(F5), Ser102(G15), Phe111, Ala115(H4), Ala118(H7)] which are replaced by larger residues [Ala78(F5), Thr101(G15), Trp110, Leu114(H4), Leu117(H7)], or solvent-exposed residues [Ile60(E17), Lys101(G14)] by hydrophobic residues [Trp59(E17), Trp100(G14)], in *N*-HGbRL (Fig. 1). This could possibly render an HGbI-like fold unstable for *N*-HGbRL. Nevertheless, helices F–H from both globins superpose considerably well with an rmsd of 0.76 Å (Fig. 4a), and their proximal sites are totally conserved. When comparing their highly conserved Gln(E7)–Leu(E15) regions, however, helix E of *N*-HGbRL's open form is moved forwards by about a helical turn relative to that of HGbI, leading to a structural 'frameshift' of three residues. The O<sub>2</sub>-binding Gln50(E7) of HGbI, for instance, is structurally aligned not to *N*-HGbRL's corresponding Gln49(E7) but to Lys52(E10), and HGbI's Val57(E14) is aligned to *N*-HGbRL's Trp59(E17), which may have a role as a pivot for the shift (Fig. 5a). The shift has lengthened *N*-HGbRL's loop CE and enlarged its distal site. An interlocking pair of Arg30(B12)–Trp100(G14) has also been observed to shift *N*-HGbRL's helix B by about half a helical turn (Fig. 5b). The side chains of Trp100(G14) and Thr101(G15) further push helix B outwards, resulting in a similar shift in helix A of a helical turn which is coupled to the shift in helix E.

Other strategic substitutions in *N*-HGbRL, particularly around Asn63(E21)–Ala70(E28) corresponding to HGbI's loop EF, are also employed for the formation and accommodation of the fused helices E and F. The conversion of HGbI's Ser51(E8) to Gly50(E8) has not only added flexibility for helix E's unwinding, but also avoided clashes with helix B. Substituting HGbI's Gly64 and Gly70 with Asn63(E21) and Ala69(E27), in turn, increases the propensity for helix formation in loop EF, while replacing HGbI's conserved Pro73 with Asp72(E30) also contributes to helix forming by providing a free backbone N (Fig. 5c). Arg75(F2), a lengthened substitution for HGbI's Gln76, notably helps straighten loop EF by binding to Asp68(E26) and Asp72(E30) (Fig. 5c). Other residues such as Ile67(E25), Ala70(E28), Ala73(E31) and Ala74(F1) that replace HGbI's larger Leu68, Leu71, Thr74 and Leu75, meanwhile, are better accommodated in their surroundings. Clashes between the antiparallel pair of helices EF in the dimer are avoided, in particular, by the Ile67(E25)–Ile67(E25) pair in the closed form and the Ala70(E28)–Ala70(E28) pair in the open form (Fig. 5c). Meanwhile, the lengthened loop CE as well as the bulky Arg29(B11) and Arg39(C5), which replace HGbI's Ala30(B11) and Ser40(C5) and help stabilize loop CE in the closed form, have abolished the mostly hydrophobic dimer interface of HGbI, which is made up of helices B from each subunit (Supplementary Fig. S2).

HGbI also has a smaller distal pocket which is connected to the solvent through several tunnels<sup>3</sup>, but the distal pocket of the open form of *N*-HGbRL is readily accessible from the solvent through an opening (Fig. 4b). In HGbI's pocket, Tyr29(B10) is positioned closer to the haem and able to bind an O<sub>2</sub> molecule, which also binds Gln50(E7). Although spectrally ferrous *N*-HGbRL also binds O<sub>2</sub> (Fig. 2), it is unclear if its corresponding Tyr28(B10) could directly coordinate O<sub>2</sub>, since this would definitely require a larger outward movement of helix E to make room for Tyr28(B10) to move in. On the other hand, the hexacoordinated spectra of HGbI<sup>3</sup> are a very perplexing observation, as no histidine is present in the vicinity of the distal site and the only lysine nearby, Lys53(E10), is positioned outside the pocket (Fig. 5a). It would require considerable conformational changes of HGbI's helices B and E for Tyr29(B10) to make way for Lys53(E10) to move in to bind the haem, and this certainly would also disrupt the dimerization of HGbI. A structure of the hexacoordinated HGbI will no doubt prove interesting to reveal perhaps another new mechanism for haem hexacoordination.

**Evolution of HGbRL from HGbI and HRL.** As observed in the reorientation of FHP's NADH domain<sup>8</sup>, the large conformational movement of *N*-HGbRL between its open and closed forms, by the same token, is also expected to trigger changes in HGbRL's C-terminal Roadblock/LC7 domain. Roadblock/LC7 proteins form one of the light chains of dynein in eukaryotes, a large multisubunit complex that transports cellular cargo along microtubules. In bacteria, they constitute an ancient protein superfamily which



**Figure 5. Structural comparison with HGBI.** (a) Distal site comparison. Despite their highly conserved proximal sites, the distal site of HGBI (pink) differs considerably from those of the open (yellow) and closed (blue) forms of *N*-HGBRL. A forward shift of a helical turn in *N*-HGBRL's helix E results in the lengthening of loop CE. (b) Arg30–Trp100 interlocking pair. The interlocking pair causes *N*-HGBRL's helix B to move outwards and backwards by half a helical turn, subsequently dragging helix A backwards by about a helical turn which is coupled to the shift in helix E. The similar region in HGBI interacts with the second subunit (transparent). (c) Straightening of loop EF. Strategic residue replacements in *N*-HGBRL straighten the region corresponding to HGBI's loop EF, and reduce clashes between the antiparallel pair of the fused helices EF in the dimer.

may function in NTPase regulation<sup>33</sup>. The Roadblock/LC7 protein from *Myxococcus xanthus*, MglB, is the cognate GTPase-activating protein (GAP) of the GTPase MglA, and they are involved in regulating the dynamic switching of cell polarity in bacterial motility<sup>34,35</sup>. The *M. infernorum* genome also encodes a pair of genes similar to the MglB–MglA operon, but whose functions are still unknown — a single-domain Roadblock/LC7 protein (GenBank: ACD84470) 34% identical to the C-terminal domain of HGBRL, and an MglA-like GTPase (GenBank: ACD84469) encoded 183 bp upstream.

HGBRL is unique to the genus *Methylacidiphilum*, and could have evolved from the duplication and subsequent fusion of the genes of HGBI and the Roadblock/LC7 protein. With both domains having their own carbon copies in the genome — hence able to relieve themselves of the same functions — HGBRL was therefore free to have developed a new role of its own. Its globin domain, upon binding a yet to be identified ligand, is likely to mechanistically relay the signal to the Roadblock/LC7 domain through the former's conformational changes. The downstream pathway might involve cell motility as mediated by the Roadblock/LC7 domain, but *M. infernorum* has been reported as a non-motile rod<sup>1</sup>. Alternatively, since HGBRL is also encoded in the same operon as an S-adenosyl methionine (SAM)-dependent methyltransferase (GenBank: ACD83143), signalling from HGBRL may perhaps lead to the methylation of a target molecule by the methyltransferase. Further characterization of this protein, including identification of the native ligand for the globin domain, will definitely help relate the unusual structural features of HGBRL to its function.

## Methods

**Cloning, protein expression and purification.** The gene encoding *N*-HGBRL (residues 1–133; GenBank: ACD83144) was amplified by PCR using primers containing the NdeI and BamHI restriction sites, with a His-tag sequence inserted before the stop codon. The PCR product was cloned into the pET-3a expression vector (Novagen), and the resulting vector was transformed into Rosetta 2(DE3)pLysS (Novagen). The cells were grown in LB medium containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol. Expression was induced with 50 µM isopropyl-β-D-thiogalactopyranoside, supplemented with 100 mg/l FeSO<sub>4</sub>·7H<sub>2</sub>O and 17 mg/l δ-aminolevulinic acid. After overnight incubation at 37 °C, the cells were harvested by centrifugation and resuspended in 50 mM Tris-HCl, 200 mM NaCl, pH 8.0, sonicated and centrifuged to remove cell debris. The supernatant was then applied to a Ni<sup>2+</sup>-charged HiTrap IMAC HP column (GE Healthcare), and *N*-HGBRL was eluted with a gradient of up to 1.0 M imidazole. Fractions containing the protein were pooled and further purified by size-exclusion chromatography on a Superdex 200 column (GE Healthcare).

**Absorption spectra.** Extensive buffer exchange was carried out to remove traces of imidazole. The absorption spectrum for the met form of *N*-HGBRL was measured in 50 mM Tris-HCl, pH 8.0 at room temperature. The deoxy form was obtained by adding freshly prepared sodium dithionite to a final concentration of 100 mg/ml, and the absorption spectra were measured every ten minutes for an hour.

	Open	Closed
<b>Data collection</b>		
Space group	C2	C2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	98.6, 44.3, 81.9	67.2, 54.4, 44.7
$\beta$ (°)	119.7	106.2
Resolution (Å)	1.94 (1.99–1.94)	1.65 (1.69–1.65)
Completeness (%)	94.3 (99.4)	89.6 (87.3)
Redundancy	2.7 (2.6)	2.3 (2.1)
Mean <i>I</i> / $\sigma$ ( <i>I</i> )	13.9 (3.0)	22.0 (11.0)
<i>R</i> <sub>merge</sub>	0.044 (0.307)	0.025 (0.061)
<b>Refinement</b>		
No. of reflections	21,703	16,812
<i>R</i> <sub>work</sub>	0.183	0.151
<i>R</i> <sub>free</sub>	0.234	0.169
R.m.s.d bonds (Å)	0.007	0.009
R.m.s.d angles (°)	1.027	1.201
Mean <i>B</i> -factors (Å <sup>2</sup> )		
Protein	25.8	18.7
Ligands	20.4	16.6
Water	32.1	30.6
Ramachandran plot		
Favoured (%)	98.6	98.6
Allowed (%)	1.4	1.4
Values for the outer shell are given in parentheses.		

**Table 1.** Data collection and refinement statistics.

**Crystallization and data collection.** *N*-HGbRL was concentrated to 10–20 mg/ml prior to crystallization. Using the sitting-drop vapour diffusion technique, the imidazole-bound crystals were formed at 20 °C in 3.5 M 1,6-hexanediol, 0.1 M sodium citrate pH 5.6, and cryoprotected with 12% glycerol. Crystals of the hexacoordinated closed form, meanwhile, were formed at 4 °C on a big salt crystal in 50% 2-methyl-2,4-pentanediol, 0.1 M Tris-HCl pH 8.0, 0.2 M (NH<sub>4</sub>)<sub>2</sub>HPO<sub>4</sub>. Diffraction data for both types of crystals, flash-cooled at 100 K, were collected on a Rigaku MicroMax-007 HF X-ray generator ( $\lambda = 1.5418$  Å) equipped with an R-Axis IV<sup>++</sup> area detector, and processed with XDS<sup>36</sup> in the same space group C2 but with different cell parameters.

**Structure determination.** The HGbI structure<sup>3</sup> (PDB: 3S1I) was used as an initial model to solve the closed form by molecular replacement with Molrep<sup>37</sup>, which was then used to solve the imidazole-bound form. These structures were subsequently built and refined using Phenix<sup>38</sup> and Coot<sup>39</sup>, with TLS refinement introduced into the closed form. Data collection and refinement statistics are summarized in Table 1. Structural alignments were performed with the Dali server<sup>40</sup>, manually edited and presented with ESPript<sup>41</sup>. Tunnels and the distal pocket were calculated with CAVER<sup>42</sup>, and figures were generated using PyMOL (<http://www.pymol.org>).

## References

- Dunfield, P. F. *et al.* Methane oxidation by an extremely acidophilic bacterium of the phylum Verrucomicrobia. *Nature* **450**, 879–882 (2007).
- Hou, S. *et al.* Complete genome sequence of the extremely acidophilic methanotroph isolate V4, *Methylococcoides burtonii*, a representative of the bacterial phylum Verrucomicrobia. *Biol. Direct* **3**, 26 (2008).
- Teh, A. H. *et al.* Hell's Gate globin I: an acid and thermostable bacterial hemoglobin resembling mammalian neuroglobin. *FEBS Lett.* **585**, 3250–3258 (2011).
- Jamil, F. *et al.* Crystal structure of truncated haemoglobin from an extremely thermophilic and acidophilic bacterium. *J. Biochem.* **156**, 97–106 (2014).
- El Hammi, E. *et al.* Active site analysis of yeast flavohemoglobin based on its structure with a small ligand or econazole. *FEBS J* **279**, 4565–4575 (2012).
- Ermler, U., Siddiqui, R. A., Cramm, R. & Friedrich, B. Crystal structure of the flavohemoglobin from *Alcaligenes eutrophus* at 1.75 Å resolution. *EMBO J.* **14**, 6067–6077 (1995).



7. Ollesch, G., Kaunzinger, A., Juchelka, D., Schubert-Zsilavec, M. & Ermler, U. Phospholipid bound to the flavohemoprotein from *Alcaligenes eutrophus*. *Eur. J. Biochem.* **262**, 396–405 (1999).
8. El Hammi, E. *et al.* Structure of *Ralstonia eutropha* flavohemoglobin in complex with three antibiotic azole compounds. *Biochemistry* **50**, 1255–1264 (2011).
9. Hargrove, M. S. *et al.* Crystal structure of a nonsymbiotic plant hemoglobin. *Structure* **8**, 1005–1014 (2000).
10. Vallone, B., Nienhaus, K., Matthes, A., Brunori, M. & Nienhaus, G. U. The structure of carbonmonoxy neuroglobin reveals a heme-sliding mechanism for control of ligand affinity. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 17351–17356 (2004).
11. Ilari, A., Bonamore, A., Farina, A., Johnson, K. A. & Boffi, A. The X-ray structure of ferric *Escherichia coli* flavohemoglobin reveals an unexpected geometry of the distal heme pocket. *J. Biol. Chem.* **277**, 23725–23732 (2002).
12. Poole, R. K. & Hughes, M. N. New functions for the ancient globin family: bacterial responses to nitric oxide and nitrosative stress. *Mol. Microbiol.* **36**, 775–783 (2000).
13. Hou, S. *et al.* Myoglobin-like aerotaxis transducers in Archaea and Bacteria. *Nature* **403**, 540–544 (2000).
14. Marles-Wright, J. *et al.* Molecular architecture of the “stressosome,” a signal integration and transduction hub. *Science* **322**, 92–96 (2008).
15. Murray, J. W., Delumeau, O. & Lewis, R. J. Structure of a nonheme globin in environmental stress signaling. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17320–17325 (2005).
16. Saw, J. *et al.* Complete genome sequencing and analysis of *Saprospira grandis* str. Lewin, a predatory marine bacterium. *Stand. Genomic Sci.* **6**, 84–93 (2012).
17. Islam, T., Jensen, S., Reigstad, L. J., Larsen, O. & Birkeland, N. K. Methane oxidation at 55°C and pH 2 by a thermoacidophilic bacterium belonging to the *Verrucomicrobia* phylum. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 300–304 (2008).
18. Khadem, A. F. *et al.* Draft genome sequence of the volcano-inhabiting thermoacidophilic methanotroph *Methylacidiphilum fumarolicum* strain SolV. *J. Bacteriol.* **194**, 3729–3730 (2012).
19. Tarricone, C., Galizzi, A., Coda, A., Ascenzi, P. & Bolognesi, M. Unusual structure of the oxygen-binding site in the dimeric bacterial hemoglobin from *Vitreoscilla* sp. *Structure* **5**, 497–507 (1997).
20. Bolognesi, M. *et al.* Anticooperative ligand binding properties of recombinant ferric *Vitreoscilla* homodimeric hemoglobin: a thermodynamic, kinetic and X-ray crystallographic study. *J. Mol. Biol.* **291**, 637–650 (1999).
21. Lionetti, C., Guanziroli, M. G., Frigerio, F., Ascenzi, P. & Bolognesi, M. X-ray crystal structure of the ferric sperm whale myoglobin: imidazole complex at 2.0 Å resolution. *J. Mol. Biol.* **217**, 409–412 (1991).
22. de Sanctis, D. *et al.* Bishistidyl heme hexacoordination, a key structural property in *Drosophila melanogaster* hemoglobin. *J. Biol. Chem.* **280**, 27222–27229 (2005).
23. Yoon, J. *et al.* Structure and properties of a bis-histidyl ligated globin from *Caenorhabditis elegans*. *Biochemistry* **49**, 5662–5670 (2010).
24. de Sanctis, D. *et al.* Crystal structure of cytoglobin: the fourth globin type discovered in man displays heme hexa-coordination. *J. Mol. Biol.* **336**, 917–927 (2004).
25. Pesce, A. *et al.* Human brain neuroglobin structure reveals a distinct mode of controlling oxygen affinity. *Structure* **11**, 1087–1095 (2003).
26. Vallone, B., Nienhaus, K., Brunori, M. & Nienhaus, G. U. The structure of murine neuroglobin: Novel pathways for ligand migration and binding. *Proteins* **56**, 85–92 (2004).
27. Hoy, J. A., Kundu, S., Trent, J. T., 3rd, Ramaswamy, S. & Hargrove, M. S. The crystal structure of *Synechocystis* hemoglobin with a covalent heme linkage. *J. Biol. Chem.* **279**, 16535–16542 (2004).
28. Einsle, O. *et al.* Structure of cytochrome *c* nitrite reductase. *Nature* **400**, 476–480 (1999).
29. Mowat, C. G. *et al.* Octaheme tetrathionate reductase is a respiratory enzyme with novel heme ligation. *Nat. Struct. Mol. Biol.* **11**, 1023–1024 (2004).
30. Rodrigues, M. L., Oliveira, T. F., Pereira, I. A. & Archer, M. X-ray structure of the membrane-bound cytochrome *c* quinol dehydrogenase NrfH reveals novel haem coordination. *EMBO J.* **25**, 5951–5960 (2006).
31. Worrall, J. A., van Roon, A. M., Ubbink, M. & Canters, G. W. The effect of replacing the axial methionine ligand with a lysine residue in cytochrome *c*-550 from *Paracoccus versutus* assessed by X-ray crystallography and unfolding. *FEBS J.* **272**, 2441–2455 (2005).
32. Pesce, A. *et al.* HisE11 and HisF8 provide bis-histidyl heme hexa-coordination in the globin domain of *Geobacter sulfurreducens* globin-coupled sensor. *J. Mol. Biol.* **386**, 246–260 (2009).
33. Koonin, E. V. & Aravind, L. Dynein light chains of the Roadblock/LC7 group belong to an ancient protein superfamily implicated in NTPase regulation. *Curr. Biol.* **10**, R774–776 (2000).
34. Leonardy, S. *et al.* Regulation of dynamic polarity switching in bacteria by a Ras-like G-protein and its cognate GAP. *EMBO J.* **29**, 2276–2289 (2010).
35. Zhang, Y., Franco, M., Ducret, A. & Mignot, T. A bacterial Ras-like small GTP-binding protein and its cognate GAP establish a dynamic spatial polarity axis to control directed motility. *PLoS Biol.* **8**, e1000430 (2010).
36. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
37. Vagin, A. & Teplyakov, A. Molecular replacement with MOLREP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25 (2010).
38. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
39. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
40. Holm, L. & Rosenstrom, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–549 (2010).
41. Gouet, P., Courcelle, E., Stuart, D. I. & Metz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308 (1999).
42. Chovancova, E. *et al.* CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **8**, e1002708 (2012).

## Acknowledgments

This work was supported by the Post Genome Studies grant (1002/PCCB/910208) from Universiti Sains Malaysia.

## Author Contributions

A.H.T., N.N. and M.A. designed the research. A.H.T. and J.A.S. performed the experiments. A.H.T. analysed the data and wrote the paper.

### Additional Information

**Accession codes:** Atomic coordinates have been deposited with the Protein Data Bank under accession codes 3WFW for the hexacoordinated closed form, and 3WFX for the imidazole-bound open form.

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Teh, A.-H. *et al.* Open and Lys–His Hexacoordinated Closed Structures of a Globin with Swapped Proximal and Distal Sites. *Sci. Rep.* **5**, 11407; doi: 10.1038/srep11407 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>