Supporting Information

The circularly permuted globin domain of Androglobin

Brandon J. Reeder[1*], Giuseppe Deganutti[1,3], John Ukeri[1], Silvia Atanasio[1], Dimitri A. Svistunenko[1], Christopher Ronchetti[1], Juan Carlos Mobarec[1], Marten H. Vos[2], Michael T. Wilson[1], and Christopher A Reynolds[1,3*]

Materials

Horse heart Mb, sodium oleate, 4,4'-dithiodipyridine, tris(2-carboxyethyl)phosphine (TCEP) and all buffer materials were purchased from Sigma-Aldrich (Merck), Poole, UK. Isopropyl β-D-thiogalactopyranoside (IPTG) and 5-aminolaevulinic acid were from Molekula, Gillingham, UK. Proli-Nonoate was purchased from Cayman Chemicals via Cambridge Bioscience UK. Carbon monoxide was purchased from BOC, UK.

The androglobin alignment

The alignment was carried out using a novel multi-template profile alignment based on multiple pairwise alignments.[1] The alignments for Adgb, alpha-Hb, beta-Hb, Cygb, Mb, and Ngb were determined by (i) collecting sequences from the NCBI (https://www.ncbi.nlm.nih.gov/) non-redundant database using BLASTp,[2] (ii) sifting the BLASTp output for keywords, e.g. Cygb, so that there was no cross-contamination by other globins, (iii) eliminating 'predicted' (i.e. mRNA-based) sequences so only high quality sequences were included, (iv) alignment using Clustal Omega[3] and (v) manual editing using Jalview.[4] Sequences more than 90 % similar to other sequences were removed. The structural alignment of alpha-Hb, beta-Hb, Cygb, Mb and Ngb (PDB codes: 1BZ1, 1DXT, 1UT0, 3RGK, and 4MPM respectively) was determined using the align module of Modeller;[5] since these proteins align to give a well-defined sequence alignment, the alignment of Adgb to one of these 5 was essentially equivalent to alignment to the others and this forms a key aspect of the multi-template alignment. Profiles for the main helices A-B, E-H were generated for each protein with a central region of 13 residues and a flank of 25 residues on either side. For each helix, the profile alignment between Adgb and each of alpha-Hb, beta-Hb, Cygb, Mb, and Ngb was carried out by doing all possible pairwise alignments and counting the votes for the top alignment, as described elsewhere[1, 6] and illustrated in Figure S1. The averaged votes were scaled between 0 and 1 and reported in Figure 1 (left hand side) for each helix and for each protein alpha-Hb, beta-Hb, etc. A consensus alignment was determined by multiplying the 5 sets of votes together, as shown in Figure 1 (right hand side). Control alignments were also carried out for all possible pairs of alpha-hemoglobin, beta-hemoglobin, cytoglobin, myoglobin, and neuroglobin; these were in agreement with the structural alignment (results not shown). The alignment is shown in Figure 2, along with the 2012 alignment 1 (for the key sections where it differs from the one reported here).

Modelling the androglobin structure

Two initial models of Adgb-GD were generated using Modeller.[7] One was built on the assumption that Adgb was not cyclically permuted but rather had the same topology as the other Hb molecules (i.e. assuming residues for helix A-H are contiguous in the sequences). The second was built taking into account the unique circularly-permuted Adgb topology, but omitting the IQ domain, residues 891-933. The alignment was taken from Figure 2; the Protein Data Bank (PDB) codes for the structural templates were 1BZ1, 1DXt, 1UT0, 3RGK, and 4MPM. Both structures were built with and without the Cys787 and Cys978 disulfide bond. In each case, 2000 Modeller structures were generated and the model with the lowest DOPE score was selected for MD studies. Initial 0.5 μs MD simulations on these models suggested that the structure was stable (results not shown) and on this basis the expression studies were initiated.

A region of the IQ calmodulin-binding motif (IQ refers to the first 2 amino acids of the motif) was predicted to be helical by both PsiPred[8] and Jpred.[9] Consequently, this region was docked to the above permuted Adgb-GD model using cluspro,[10] giving two main poses, one on either side of the heme; from these structures two alternative templates were generated for inclusion with the other structural templates, (i) Ala904-Ala928 docked to helices G and H (complex 1) and (ii) Ala904-Ala928 docked to helices E and F (complex 2). Residues Glu905 – Ala928 were restrained to be helical within Modeller. In each case, 2000 structures were generated and the two models with the lowest DOPE score was selected for further MD studies (the DOPE score indicated that complex 2 was more stable than complex 1).

Molecular dynamics studies of complex 1 and complex 2

The AlphaFold 2 [11] model of the human androglobin, which does not present the heme bound to Adgb-GD, was retrieved at https://alphafold.ebi.ac.uk/entry/Q8N7X0. Complex 1, complex 2, and the Adgb-GD of AlphaFold 2 model (residues Glu785 to Lys985) were prepared for molecular dynamics (MD) simulations with the CHARMM36 force field.[12] An in-house procedure that combines HTMD[13] and Tcl (Tool Command Language) scripts allowed the addition of hydrogen atoms by means of the

pdb2pqr[14] and propka[15] software (considering a simulated pH of 7.0); the protonation of titratable side chains was checked by visual inspection. TIP3P water molecules[16] were added to the simulation box by means of the VMD Solvate plugin 1.5 (Solvate Plugin, Version 1.5. at http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/) with a padding of 15 Å. Overall charge neutrality was finally reached by adding Na^+/Cl^- counter ions (final ionic strength of 0.150 M), using the VMD Autoionize plugin 1.3 (Autoionize Plugin, Version 1.3 at http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/). The proximal histidine (His824) was patched to the heme iron atom according to the CHARMM36 PHEM patch.

The MD engine ACEMD[17] was employed for both the equilibration and productive simulations. Equilibration was achieved in two steps: after 500 cycles of conjugate-gradient minimization, the systems were simulated for 4 ns employing an integration time step of 2 fs in the isothermal-isobaric conditions (NPT), using the Berendsen barostat [18] (target pressure 1 atm) and the Langevin thermostat [19] (target temperature 300 K) with a low damping of 1 ps^{-1}. Positional restraints of 1 kcal mol^{-1} $Å^{-2}$ on protein and heme atoms were gradually reduced to zero.

MD trajectories were computed with an integration time step of 4 fs in the canonical ensemble (NVT) at 300 K, using a thermostat damping of 0.1 ps^{-1} and the M-SHAKE algorithm [20] to constrain the bond lengths involving hydrogen atoms. The cut off distance for electrostatic interactions was set at 9 Å, with a switching function applied beyond 7.5 Å. Long range Coulomb interactions were handled using the particle mesh Ewald summation method (PME)[21] by setting the mesh spacing to 1.0 Å. Complex 1 and 2 were simulated in triplicate for 500 ns each replica, while a single 2 µs replica was performed on the AlphaFold 2 model.

Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) were computed using VMD[22]. Distances between cysteine residues were computed using MDTraj.[23]

Androglobin globin domain expression and purification

The gene for the Adgb-GD was synthesized by Epoch Life Sciences Inc, Missouri City, USA and incorporated into expression vector pET28a (Merck). The vector was transformed into BL21 DE3 cells (Sigma-Aldrich, UK) by heat shock method (42 ºC, 90 s). Protein was expressed in shaking flasks containing 1.4 L Luria-Bertani media with kanamycin sulfate (50 µg ml-1), shaken at 180 rpm, 37 ºC. When optical density was ~1.0 at 600 nm, protein expression was initiated by addition of 500 µM IPTG. Ferric citrate (50 µM) and 5-aminolaevulinic acid (250 µM) were added to the media to augment heme synthesis. Carbon monoxide gas was bubbled through the solution for 30 s and the flasks were sealed using a rubber bung and incubated for a further 18 h, shaken at 80 rpm, 32 ºC. Cells were isolated by centrifugation (8000 g, 20 min), and cells were resuspended in water and frozen at -20 ºC. Defrosted cells were lysed using an Avestin Emulsiflex C3 homogenizer at 15000-20000 psi, 2 passes. Following centrifugation (30000 g, 20 min, 4 ºC) supernatant was separated and buffer added to a final concentration of 20 mM sodium phosphate, 500 mM sodium chloride, 0.25 % SDS and 20 mM imidazole, pH 7.4. The cell pellet contains significant amounts of Adgb, so the pellet was resuspended in 20 mM sodium phosphate, 500 mM sodium chloride, 1 % SDS and 20 mM imidazole, pH 7.4. The suspension was centrifuged (30000 g, 20 min, 4 ºC) and the supernatant diluted to 0.25 % SDS with 20 mM sodium phosphate, 500 mM sodium chloride and 20 mM imidazole, pH 7.4. The two supernatant solutions were pooled.

The His-tagged protein was purified using a GE Healthcare immobilized metal (nickel) affinity column (5 ml), washed with phosphate buffer, 500 mM sodium chloride, 0.25 % SDS with 20 mM imidazole and eluted with phosphate buffer, 500 mM sodium chloride with 500 mM imidazole. Imidazole was removed by dialysis (1 mM sodium tetraborate, pH 9.5, 3 changes) and the His-tag was cleaved through incubation with bovine thrombin (Sigma–Aldrich; 10 units/mg of protein) at room temperature (22 ºC) overnight with gentle mixing in 20 mM Tris pH 8.4, 150 mM NaCl, 2.5 mM $CaCl_2$. Tag-free protein was purified using the nickel-affinity column (20 mM sodium phosphate, 500 mM sodium chloride, 0.25 % SDS and 20 mM imidazole, pH 7.4), dialysed and concentrated using a Whatman 3 kDa spin filter.

During purification some loss of heme was experienced and required heme reconstitution. Approximately 2x excess hemin (from a 10 mM hemin solution dissolved in 10 mM sodium hydroxide) was added based on a calculated 280 nm extinction coefficient of 32.5 mM^{-1} cm^{-1} for the apo protein and left overnight at 4 °C followed by gel filtration to remove excess heme (Sephadex G25, 5x1 cm column, 10 mM sodium phosphate pH 7.4).

Determination of androglobin globin domain absorption coefficient

Unknown concentrations of ferric Adgb-GD, along with a known concentration of Mb (calculated using the ferrous protein, molar absorption coefficient (ε_{435nm} = 121 mM^{-1} cm^{-1})[24] had their optical properties recorded (ferric, deoxyferrous, ferrous-NO, -CO) before they were subjected to reverse-phase HPLC using an Agilent 1290 HPLC fitted with a diode array spectrophotometer. The column used was a Zorbax Stablebond 300C3 250 mm × 4.6 mm fitted with a 12 mm × 4.6 mm guard column, with a water/TFA – acetonitrile/TFA gradient as previously described[25]. The concentration of heme in the unknown Adgb-GD sample was calculated from the integrated area under the peak (~15 min elution time) and compared to the integral of the heme peak of the known concentration of Mb. This was repeated on three independent samples. The ε_{412nm} of ferric Adgb-GD pH 7.4 was calculated to be 106 mM^{-1} cm^{-1} and was used for concentration determination in all experiments.

Determination of androglobin globin domain disulfide

The extent of free cysteine sulfhydryls was determined by the dithiodipyridine assay as described previously.[26] Protein (10 µM) was titrated with 4,4'-dithiodipyridine and the absorbance monitored at 324 nm. This was repeated following protein disulfide reduction by TCEP. The titration curve was fitted to a fractional saturation curve using the least-squares method by Microsoft Excel solver program as previously described.[27]

Electron Paramagnetic Resonance Spectroscopy

Adgb aliquots (80 µM, 250 µl) were placed in Wilmad SQ EPR tubes. Nitric Oxide was added immediately prior to freezing by addition of Proli-Nonoate (1 mM, giving 2 mM NO) in 10 mM sodium hydroxide. Tubes were flash-frozen in dry-ice-cooled methanol and frozen samples were transferred to liquid nitrogen (77 K). All EPR spectra were measured using a Bruker EMX EPR spectrometer (X-band) equipped with a spherical high-quality Bruker ER 4122 SP9703 resonator and an Oxford Instruments liquid helium system. The modulation frequency was 100 kHz. The EPR spectra of the blank samples (frozen water) were subtracted from the EPR spectra of the protein samples using WinEPR v. 2.22 (Bruker Analytik, GmbH) to eliminate the baseline caused by the walls of the resonator, quartz insert or quartz EPR tube.

Stopped flow spectroscopy

Stopped-flow absorbance spectroscopy was used to measure the kinetics of change in heme iron co-ordination using an Applied Photophysics SX20 stopped-flow spectrometer fitted with a diode array spectrometer with time resolution of 1.25 ms. For the measurement of NO binding, buffer (100 mM sodium phosphate buffer pH 7.4) was degassed using a glass tonometer connected to a supply of argon gas and a vacuum pump via a custom build glass tap system to allow repeated degassing cycles. The degassed buffer was transferred anaerobically to a 10 ml glass syringe. Proli-NONOate was prepared in 25 mM sodium hydroxide to a concentration of 40 mM and then purged with argon gas. Proli-NONOate was transferred using a glass Hamilton syringe to the degassed buffer (12.5–200 µM), under these conditions the Proli-NONOate degrades to generate two molecules of NO. This was rapidly mixed in a 1:1 ratio with 10 µM Adgb (5 µM final concentration) in 100 mM sodium phosphate buffer (pH 7.4) in degassed buffer with minimal dithionite added to keep the protein reduced and oxygen free. For the nitrite reductase activity Adgb (10 µM, 5 µM final) was rapidly mixed in a 1:1 ratio with sodium nitrite (2-40 mM, 1-20 mM final). Both Adgb and nitrite solutions were oxygen free by use of excess sodium dithionite.

Femtosecond laser flash photolysis

Ultrafast broad-band transient absorption experiments at a 500 Hz repetition rate, were performed as previously described,[28] using pump pulses centered at 570 nm and white-light continuum probe pulses. The sample was continuously rastered by a Lissajous scanner. Global analysis of the data in terms of multi-exponential functions was performed using Glotaran.[29]

[1] A. Lock, R. Forfar, C. Weston, L. Bowsher, G. J. Upton, C. A. Reynolds, G. Ladds, A. M. Dixon, *Biochim Biophys Acta* **2014**, *1838*, 3036-3051.
[2] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, *J Mol Biol* **1990**, *215*, 403-410.
[3] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson, D. G. Higgins, *Mol Syst Biol* **2011**, *7*, 539.
[4] M. Clamp, J. Cuff, S. M. Searle, G. J. Barton, *Bioinformatics* **2004**, *20*, 426-427.
[5] N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M. Y. Shen, U. Pieper, A. Sali, *Curr Protoc Protein Sci* **2007**, *Chapter 2*, Unit 2 9.

[6] aB. Taddese, G. J. Upton, G. R. Bailey, S. R. Jordan, N. Y. Abdulla, P. J. Reeves, C. A. Reynolds, *Plant Physiol* **2014**, *164*, 287-307; bH. A. Watkins, M. Chakravarthy, R. S. Abhayawardana, J. J. Gingell, M. Garelja, M. Pardamwar, J. M. McElhinney, A. Lathbridge, A. Constantine, P. W. Harris, T. Y. Yuen, M. A. Brimble, J. Barwell, D. R. Poyner, M. J. Woolley, A. C. Conner, A. A. Pioszak, C. A. Reynolds, D. L. Hay, *J Biol Chem* **2016**, *291*, 11657-11675.

[7] B. Webb, A. Sali, *Curr Protoc Protein Sci* **2016**, *86*, 2 9 1-2 9 37.

[8] D. W. Buchan, F. Minneci, T. C. Nugent, K. Bryson, D. T. Jones, *Nucleic Acids Res* **2013**, *41*, W349-357.

[9] C. Cole, J. D. Barber, G. J. Barton, *Nucleic Acids Res* **2008**, *36*, W197-201.

[10] H. Mirzaei, S. Zarbafian, E. Villar, S. Mottarella, D. Beglov, S. Vajda, I. Paschalidis, P. Vakili, D. Kozakov, *J Chem Theory Comput* **2015**, *11*, 1063-1076.

[11] aJ. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Zidek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583-589; bK. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Zidek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K. Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, *Nature* **2021**, *596*, 590-596.

[12] J. Huang, A. D. MacKerell, Jr., *J Comput Chem* **2013**, *34*, 2135-2145.

[13] S. Doerr, M. J. Harvey, F. Noe, G. De Fabritiis, *J Chem Theory Comput* **2016**, *12*, 1845-1852.

[14] T. J. Dolinsky, J. E. Nielsen, J. A. McCammon, N. A. Baker, *Nucleic Acids Res* **2004**, *32*, W665-667.

[15] M. H. Olsson, C. R. Sondergaard, M. Rostkowski, J. H. Jensen, *J Chem Theory Comput* **2011**, *7*, 525-537.

[16] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J Chem Phys* **1983**, *79*.

[17] M. J. Harvey, G. Giupponi, G. D. Fabritiis, *J Chem Theory Comput* **2009**, *5*, 1632-1639.

[18] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *J Chem Phys* **1984**, *81*, 3684-3690.

[19] R. J. Loncharich, B. R. Brooks, R. W. Pastor, *Biopolymers* **1992**, *32*, 523-535.

[20] V. Krautler, W. F. Van Gunsteren, P. H. Hunenberger, *J Comput Chem* **2001**, *22*, 501-508.

[21] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *Journal of Chemical Physics* **1995**, *103*, 8577-8593.

[22] W. Humphrey, A. Dalke, K. Schulten, *J Mol Graph Model* **1996**, *14*, 33-38.

[23] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernandez, C. R. Schwantes, L. P. Wang, T. J. Lane, V. S. Pande, *Biophys J* **2015**, *109*, 1528-1532.

[24] E. Antonini, M. Brunori, in *Frontiers in Biology, Vol. 21* (Eds.: A. Neuberger, E. L. Tatum), North-Holland publishing company, Amsterdam-London, **1971**, pp. 13-52.

[25] aP. Beckerson, M. T. Wilson, D. A. Svistunenko, B. J. Reeder, *Biochem J* **2015**, *465*, 127-137; bB. J. Reeder, D. A. Svistunenko, M. A. Sharpe, M. T. Wilson, *Biochemistry* **2002**, *41*, 367-375.

[26] C. K. Riener, G. Kada, H. J. Gruber, *Anal Bioanal Chem* **2002**, *373*, 266-276.

[27] B. J. Reeder, D. A. Svistunenko, M. T. Wilson, *Biochem J* **2011**, *434*, 483-492.

[28] L. Lobato, L. Bouzhir-Sima, T. Yamashita, M. T. Wilson, M. H. Vos, U. Liebl, *J Biol Chem* **2014**, *289*, 26514-26524.

[29] J. J. Snellenburg, S. P. Laptenok, R. Seger, K. M. Mullen, I. H. M. van Stokkum, *J Stat Softw* **2012**, *49*, 1-22.

[30] D. Hoogewijs, B. Ebner, F. Germani, F. G. Hoffmann, A. Fabrizius, L. Moens, T. Burmester, S. Dewilde, J. F. Storz, S. N. Vinogradov, T. Hankeln, *Mol Biol Evol* **2012**, *29*, 1105-1114.

Table S1. The mean percentage identity for the pairwise alignments for each helix-helix profile alignment between androglobin and the other hemoglobins.

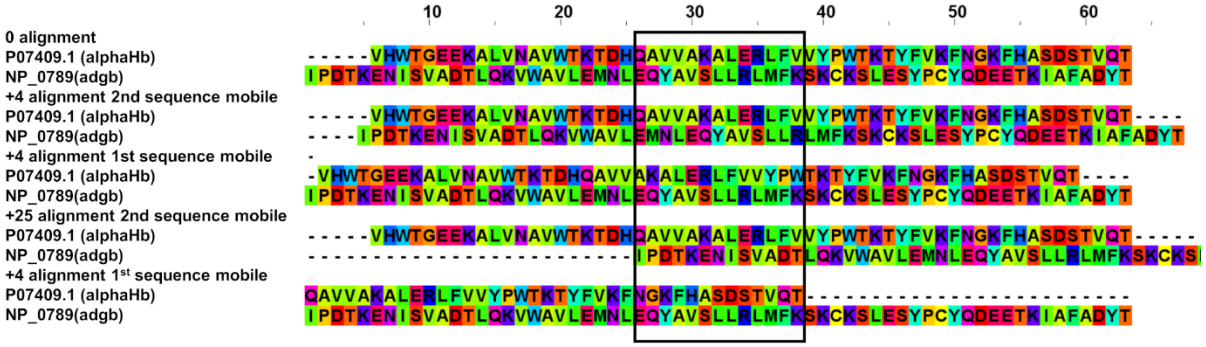
	A helix	B helix	E helix	F helix	G helix	H helix
α-Hb (alpha)	15.3	16.2	15.1	27.6	13.7	10.3
β-Hb (beta)	15.2	11.1	12.1	21.5	14.2	9.7
Cygb (cyto)	16.6	23.7	19.8	21.5	13.9	11.2
Mb (myo)	21.1	18.7	22.2	25.8	8.9	11.3
Ngb (Neuro)	15.1	15.5	9.1	25.0	13.6	7.8

Table S2. Frequency analysis of amino acid CD1 assignment. Human androglobin sequence was aligned with 771 androglobin sequences. Left: amino acid frequency as previously published based on residue 769. Right: revised amino acid frequency based on residue 975.

CD1 Assignment: Residue 769		**CD1 Assignment: Residue 975**	
Amino Acid	Frequency (%)	Amino Acid	Frequency (%)
Phe	75.38	Tyr	81.84
Cys	22.02	Phe	16.78
Other	2.60	Other	1.38

Table S3. Wavelength maxima and extinction coefficients of Adgb-GD. Buffer was 100 mM sodium phosphate buffer pH 7.4.

Oxidation and ligation state	**Soret peak, nm ($mM^{-1}cm^{-1}$)**	**Visible beta peak, nm ($mM^{-1}cm^{-1}$)**	**Visible alpha peak, nm ($mM^{-1}cm^{-1}$)**
Ferric	412 (106)	534 (12.8)	566 (10.1)
Deoxyferrous	426 (108)	531 (12.2)	560 (17.8)
Ferrous-CO	422 (180)	540 (16.0)	560 (15.0)
Ferrous-NO	395 (95)	540 (14.4)	559 (14.1)



240

241 **Figure S1. The basis of the profile alignment for the alpha haemoglobin – androglobin**
242 **Helix 2 alignment, illustrated by a single pair of sequences.** All pairwise alignments of
243 helix 2 were carried out using a helix 2 segment of 13 residues flanked by 25 residues on
244 either side. All possible alignments were scored over the central 13 residue alignment box
245 using the BLOSUM 62 substitution matrix by moving the second alignment 1 position at a time
246 so that every consecutive 13 residue section of the sequence passed through the alignment
247 box, giving 51 alignments in total. The alignment with the highest score is given 1 vote. The
248 process is repeated by keeping the second sequence fixed and moving the first alignment 1
249 position at a time so that every consecutive 13 residue section of the sequence passed
250 through the box, again giving 51 alignments in total. For cases where the zero alignment
251 shown above receives the most votes, moving the first sequence or moving the second
252 sequence is equivalent. However, as shown above for the +4 and +25 alignments, where the
253 second sequences moved 4 or 25 positions to the right gives the most votes, does not
254 correspond to the same calculation for the equivalent +4 or +25 alignment where the first
255 sequence is moved 4 or 25 places to the left because different residues pairs appear in the
256 alignment box. Consequently all alignments are scored by averaging the votes (for moving the
257 second sequence and moving the first sequence) for each position from -25 through to +25
258 over all possible pairs of sequences. For the same reason, alignments were re calculated to
259 ensure that the top scoring alignment occurred at position 0. Two +25 alignment sequences
260 are truncated.

261

262

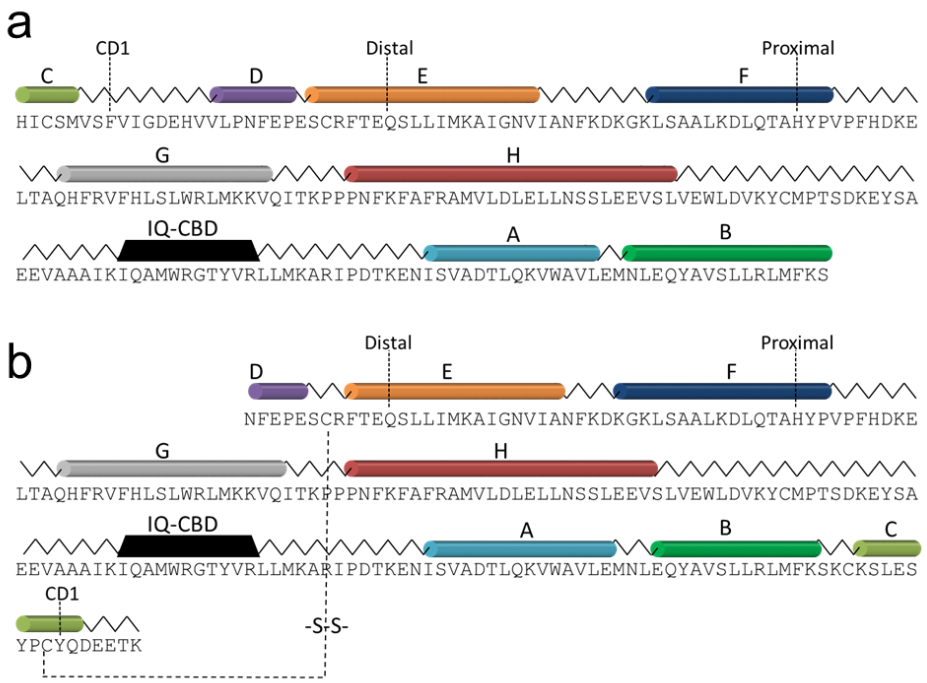


Figure S2. Androglobin heme-binding globin domain structure alignments. (a), Original alignment of alpha helical sequences to the classical 8-helix globin fold and IQ calmodulin binding domain (IQ-CDB) as reported previously (residues 763-968) [30]. Distal glutamine and proximal histidine residues are highlighted in the E and F helical sections, as well as heme-stabilizing CD1 phenylalanine at the N terminal region. **(b),** Alternative sequence alignment with similar/identical A-B and E-H helices but differing C-D alignment (residues 781-985. This revised alignment places the CD1 residue on the C terminal region as a tyrosine. In addition, this predicts the potential for the formation of a disulfide bond (-S-S-) in the CD 'loop' similar in position to that observed in human Ngb. The recombinant expression of the revised sequence allowed stable Adgb-GD protein to be expressed.

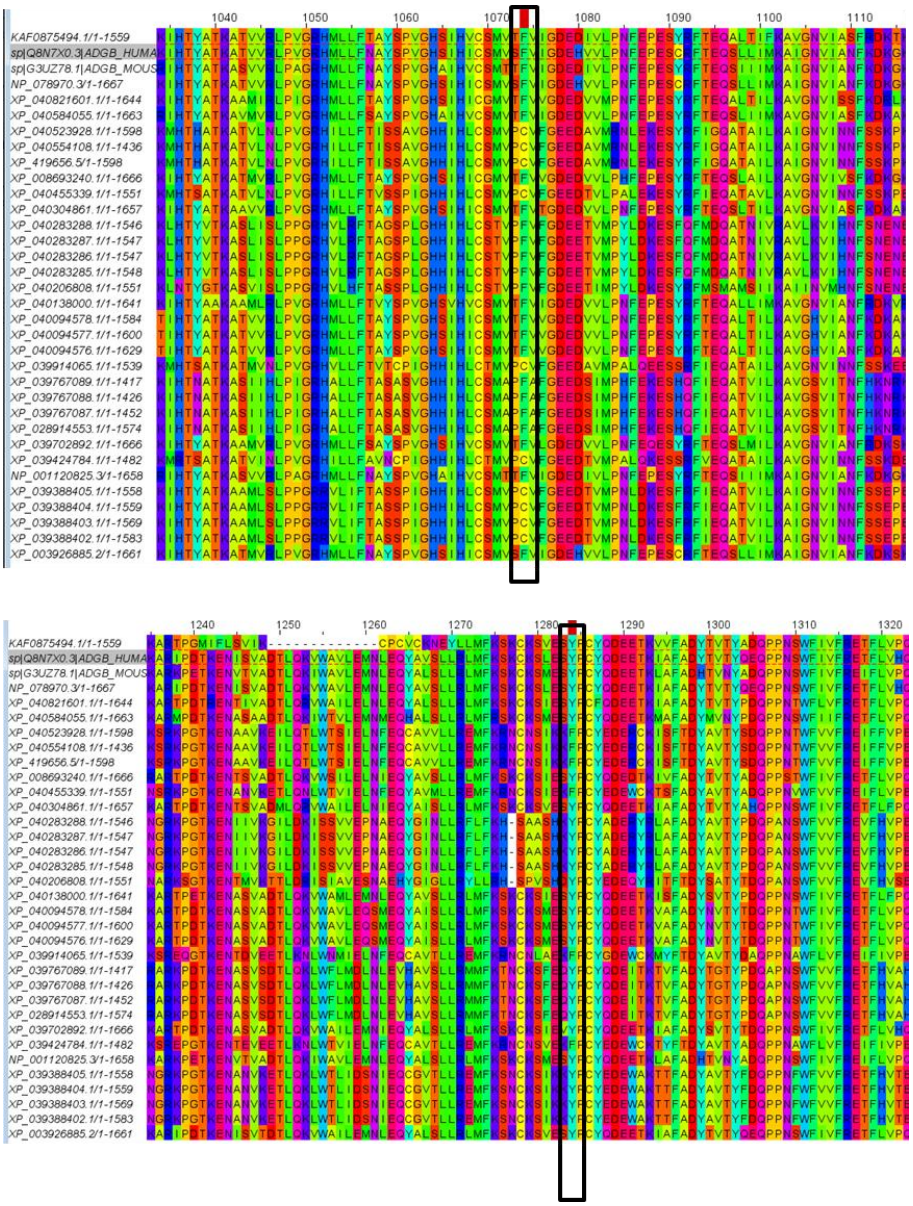


Figure S3. Multiple sequence alignments for androglobin globin domain showing the conservation at the previously proposed CD1 Phe/Cys (Phe770 for human Adgb 2012, Hoogewijs, D *et al*, 2012) or CD1 Tyr/Phe (adgb 2021). Only the first 34 sequences are shown.