nature research

Corresponding author(s): Thornton

Last updated by author(s): 18/02/2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	NCBI BLAST was used to collect sequences from NCBI databases.
Data analysis	As described in the methods section, MAFFT v7 was used to build sequence alignments. PhyML 3.1 was used to infer phylogeny from globin alignment. PAML 4.1 was used to infer ancestral sequences using maximum likelihood. PyMOL v1.3 was used to visualize and render protein structures. MassHunter was used to perform mass deconvolution on high resolution accuracy mass spec. data. UNIDEC v 1.0 was used to fit molar masses to and estimate molar abundances from native mass spectrometry. SWISS-MODEL (online server: https://swissmodel.expasy.org/) and EMBO PISA v1.48 were used to model protein structures and identify protein-protein contacts. Custom scripts were used to perform statistical analyses on Hydrogen deuterium exchange data, fit dissociation constants to Native MS data and melting curves to circular dichroism data (see methods). DynamX 3.0 (Waters) was used to process HDX-MS data. OriginPro 2016 was used to fit p50s and Hill coefficient parameters to observed oxygen binding data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Reconstructed ancestral sequences have been deposited in Genbank (IDs TBA). Homology model coordinates have been deposited in the Protein Model Database (IDs TBA). Alignment and inferred phylogeny and raw mass spectra have been deposited in Dryad (URL TBA). Scripts for analysis for the HDX permutation analysis and identification of contacts between subunits in modeled structures have been deposited at github (https://github.com/JoeThorntonLab/Hb_evolution).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Not applicable: experiments were performed on purified stocks of recombinantly expressed proteins. Technical replication of assays is described in the manuscript.
Data exclusions	No data were excluded from the analyses.
Replication	HDX experiments were performed in 3 technical replicates per construct. Measurements of oxygen affinity, cooperativity, and allosteric regulation were performed in 3-5 technical replicates per construct. Native mass spectra and size exclusion chromatography were performed across multiple concentrations with one measurement per construct/concentration, as described in the manuscript. Error associated with replication is reported in the figures and figure legends.
Randomization	Not applicable. The experiments were performed on recombinantly expressed and purified proteins, not on individuals sampled from a population and then assigned to groups.
Blinding	Not applicable. The experiments were performed on recombinantly expressed and purified proteins, not on individuals sampled from a population and then assigned to groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging