Supplementary Information for

Extended conformations of bifurcating electron transfer flavoprotein constitute up to half the population, possibly mediating conformational change.

Sharique A. Khan, Alan Hicks, Wellington C. Leite, James Byrnes, Biswajit Gorai, Maria-Andrea Mroginski, Hugh O'Neill*, Anne-Frances Miller*

Contents

Methodological details:

Detailed explanation of the genetic algorithm strategy.

Tables

Comparison with open and closed structures, for each of the conformations identified by optimization of ensembles *vs* SANS data.

Figures

Figure S1. Monodispersity of samples.

Figure S2. Testing the genetic algorithm parameters for ensemble optimization based on conformations generated by Bilbo-MD.

Figure S3. Radius of gyration vs. simulation time in MD trajectory using enhanced sampling

Figure S4. Overlay of the crystal structures of ETF in the 'closed' (green) and 'open' conformations.

Figure S5. Comparison of SANS data of ETF reduced with non-physiological donor dithionite *vs* physiological donor NADH.

Figure S6. Overlay of the extended and compact conformers that best explain the different SANS data sets discussed for *Afe*ETF.

Figure S7. Determination of contrast match point of dBCD.

Figure S8. Models of the ETF_2•BCD_2 complex contributing to the ensemble that best explains the SANS data.

Detailed explanation of the genetic algorithm strategy

Genetics algorithms have a long history in meta-heuristic optimization algorithms [1, 2] as well as their application to small-angle scattering (SAS) in general [3-5]. To validate the application of GAs to SANS data, we developed a similar algorithm to that developed in the minimum ensemble search of Bilbo-MD.[3] In this case, the goal of the optimization is to find the minimum set of *k* conformations, $\{I(q)\}_k$, that best fits the experimental scattering profiles. First, the pool of conformers is randomly grouped into the ensembles, $\{I(q)\}_k$. The gene is the ensemble of conformations and the χ_v^2 is the phenotype, given by the following equations:

$$I_{Mod}(q) = \left(c\sum_{j=1}^{k} w_j I_j(q\ \ ) + b\right); \chi_v^2 = \frac{1}{N-(k+2)}\sum_{i=n}^{N}\frac{\left(I_{Obs}(q_i)-I_{Mod}(q_i)\right)^2}{\sigma_{Obs}^2(q_i)}, \quad \text{[Eq.1]}$$

where I_{Mod} is the model data with the respective weight of each scattering profile, w_j, in the ensemble, *c* is a global scaling parameter, *b* is a background offset, I_{Obs} and σ_{Obs} are the observed experimental scattering profiles and errors. After each evaluation step in the generation, the fittest ensemble is updated. Then, parents are selected via their fitness as two distinct sets from the total population. Cross-over occurs with a probability, $p_c = 0.5$, and is performed by switching elements of the sets with opposing indices. For example, in a 3-ensemble fit, index 1 of parent 1 would switch with index 3 of parent 2 if the split occurs between profile 1 and 2 in the ensemble. After cross-over, the children are selected for mutation with probability, $p_m = 0.1$, by selecting a random scattering profile in the set and exchanging it out with another profile from the pool. If the child generated by the cross-over or mutation is unique, defined as not sharing identical scattering profiles, and valid, having a reasonable fit i.e., all of the weights are greater than a cutoff (1e-6), then the children become parents for the next generation. As stated earlier, the process will continue for a number of generations (Supporting Figure S2), but can also be re-iterated by randomly selecting different initial ensembles in case the algorithm gets trapped in a local optimum.

The initial ensemble of conformers can be generated any number of ways, but in this study two separate MD simulations were performed. The first was performed using the Bilbo-MD MD workflow on the Sibyls computer resources [3]. The MD simulations started from the 4KPU PDB structure with the His_6 tag added with AlphaFold2. The base domain of ETF, EtfA1-205 and EtfB1-11,36-219, was selected as the fixed domain for Bilbo-MD sampling. The two flexible linkers between the base and head domains in EtfA and EtfB were left flexible, along with the loop, EtfB12-35. All other domains were held rigid. These domains were selected because they had high B-factors in 4KPU. A radius of gyration (R_g) range from 20 to 35 Å was selected for the R_g restraints and the MD simulation generated 1200 conformations, split into six 3 Å R_g bins, with 200 conformations per bin.

Another set of conformations was generated with the full FAD complement bound in the OX state, using metadynamics.

The SAXS and SANS profiles were calculated from the MD ensembles using Pepsi-SAXS/SANS [6] with 500 q-vectors spanning from 0.0 to 0.5 $Å^{-1}$ out to the 50th multipole order. The SANS profile was calculated in 100% D_2O. Deuterated PDBs for the Pepsi-SANS calculation were generated using the SCOMAP-XD [7] deuteration algorithm at 100% solvent D_2O. Individual scattering profiles were fit to the scattering profile using the *scipy curve_fit*

module in *python* with the Levenburg-Marquadt algorithm. For the genetic algorithm, the SAS profiles were fit to the model using the *lmfit* python module with restraints such that the sum of the weights were equal to one, i.e., $\sum_{j=1}^{k} w_j = 1$. The algorithm was tested running using 4 numbers of generations (25, 50, 75, and 100) and with up to 5 iterations. We found that it is possible for the algorithm to find an optimal solution equally likely regardless of the number of generations (Figure S2), so 75 generations were used in this study for efficiency and sampling.

We ran the GA for 2-, 3-, and 4- conformer ensembles over 75 generations and 5 iterations. These settings were chosen as it was equally likely for the genetic algorithm to find an optimal solution at 25 generations as it was 100 generations, so by running 75 steps and 5 repeat simulations with different initial conditions we efficiently sampled the combinatorial landscape for 2-, 3-, and 4-conformer ensembles. However, we did not see any improvement in the χ_ν^2 with the larger ensemble sizes, so 2-conformer ensembles are reported throughout.

Tables

Table S1: Comparison with open and closed structures[a], for each of the conformations identified by optimization of ensembles *vs* SANS data.

Sample	SAS technique	Source(s) of Conformers	Extended conformation[b]			Compact conformation[b]		
			R_g (Å)	RMSD[c] *vs* 6FAH (open, Å)	RMSD[c] *vs* 4KPU (closed, Å)	R_g (Å)	RMSD *vs* 6FAH (open, Å)	RMSD *vs* 4KPU (closed, Å)
Oxidized ETF	SANS	Bilbo-MD	32.3	18.2	17.4	25.1	6.7	4.9
		metadynamics	30.8	14.0	15.9	24.5	7.0	7.2
		combined	32.6	19.0	18.2	24.9	5.9	8.2
Reduced ETF (NADH)	SANS	Bilbo-MD	34.6	16.7	14.1	25.5	8.4	0.9
		metadynamics	30.5	13.4	15.6	25.2	8.3	1.3
		combined	34.8	16.7	14.3	25.5	8.4	0.9
Reduced ETF (Dithionite)	SANS	Bilbo-MD	34.8	16.7	14.3	25.5	8.4	0.9
		metadynamics	30.5	13.4	15.6	25.2	8.3	1.3
		combined	34.8	16.7	14.3	25.5	8.4	0.9
ETF-dBCD complex								
ETF_2-$dBCD_2$ complex	SANS	combined	29.3	15.4	15.9	25.6	8.4	2.8
Free/singly bound ETF						24.5	7.5	7.6

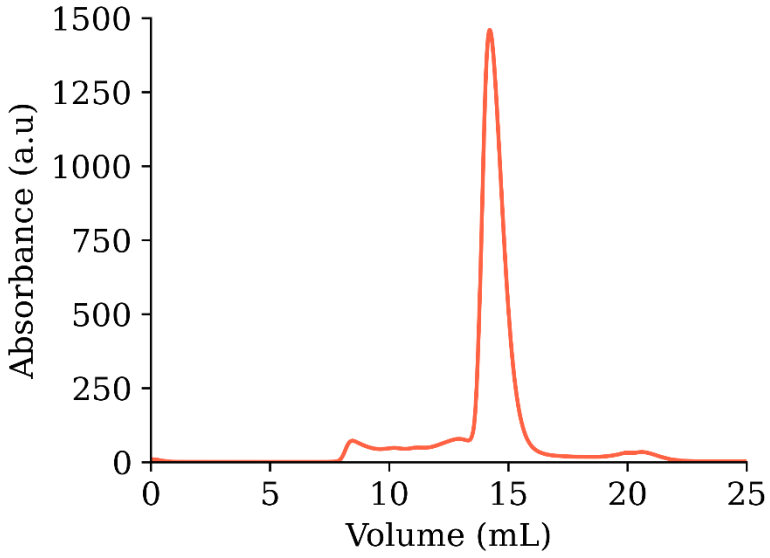
[a] The open conformation is exemplified by 6FAH and the closed conformation by 4KPU.

[b] The extended and compact conformations obtained from each modeling method are listed.

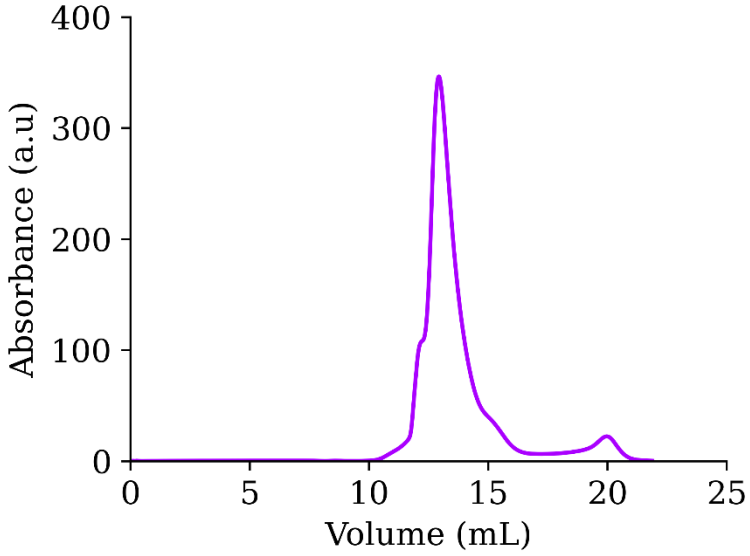
[c] RMSDs were calculated between the backbone and CB atoms after alignment to the backbone and CB atoms of the reference structure, 4KPU or 6FAH, using *mdtraj*.

Figures

Elution profile of ETF alone



Elution profile of complex of ETF_2•BCD_2



SEC complex

Figure S1. Monodispersity of samples. Size exclusion chromatography (SEC) profiles of recombinant *Afe*ETF (above) and *Afe*ETF_2•$dBCD_2$ complex (below). After Ni-NTA affinity chromatography, proteins were separated from higher aggregates or monomers by SEC over a Cytiva Superdex 200 Increase 10/300 GL column, in their respective buffers. Only the protein in the main peak was used for SANS or SAXS (see Experimental section). This was fractions from 14 to 15 mL for ETF and fractions from 12.5 to 14 mL for the *Afe*ETF_2•$dBCD_2$ complex.

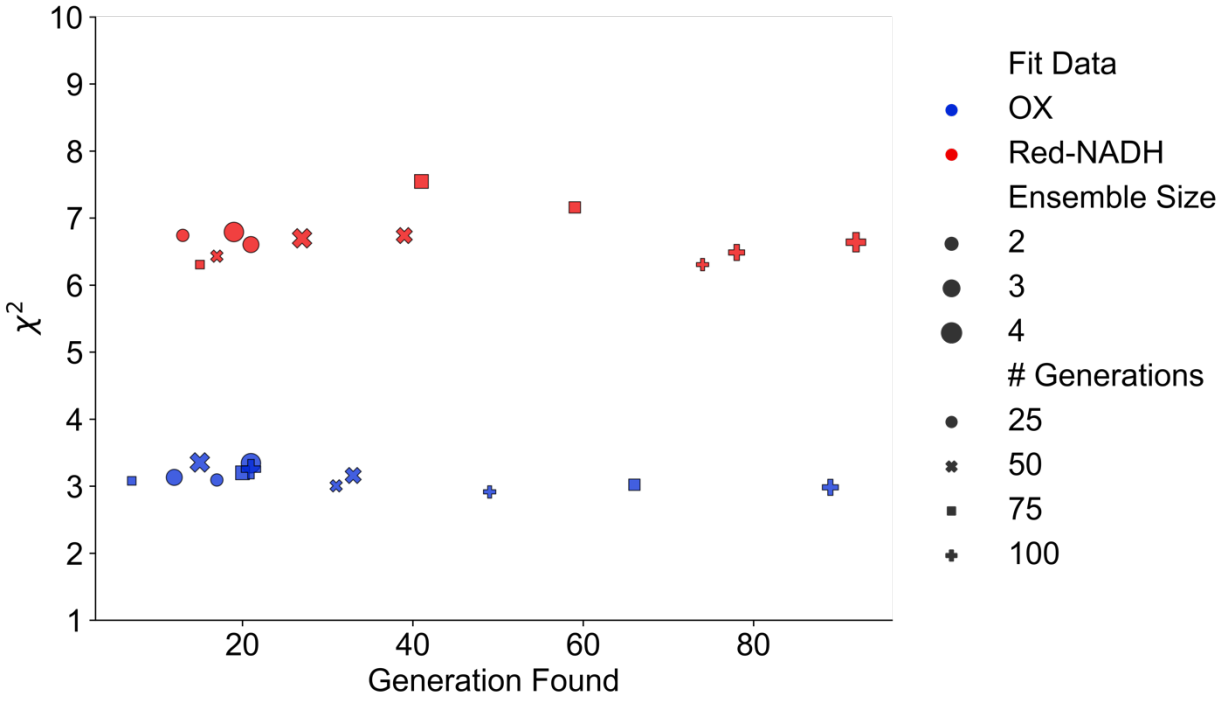


Figure S2. Testing the genetic algorithm parameters for ensemble optimization based on conformations generated by Bilbo-MD. For both the OX (blue) and the RED state (red, reduced using NADH) data, 25 generations appears to suffice for identification of an optimal ensemble, since allowance of more generations did not yield systematically lower χ^2 values, although the search did tend to continue longer. Cutoffs compared allowed up to 25 (circles), 50 ('X'), 75 (squares) or 100 generations ('+'). Comparison of the χ^2 values obtained from different ensemble sizes (sizes of the markers) tended to increase as the ensemble size increased for fits to data on both OX and RED (by NADH) *Afe*ETF. Therefore, the manuscript discusses only ensembles comprised of 2 structures.

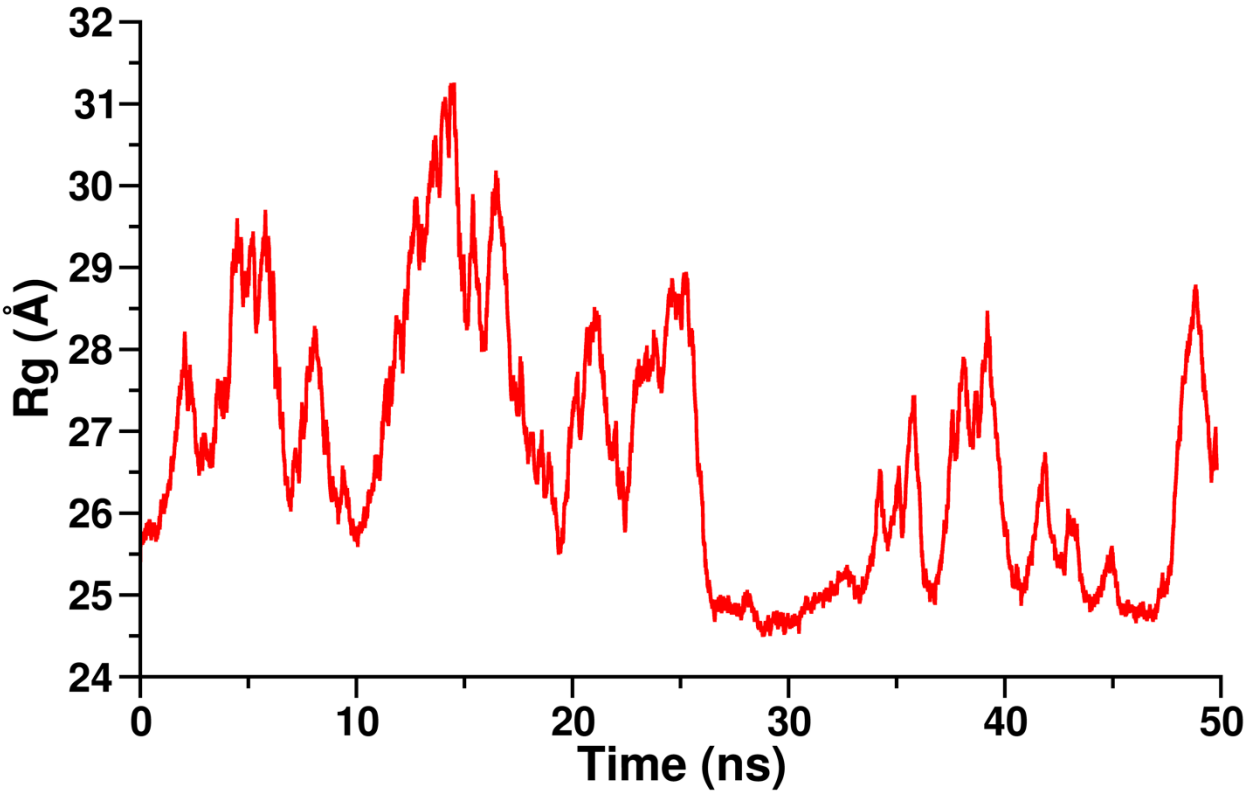


Figure S3. Radius of gyration vs. simulation time in MD trajectory using enhanced sampling. For details please see the Experimental section.

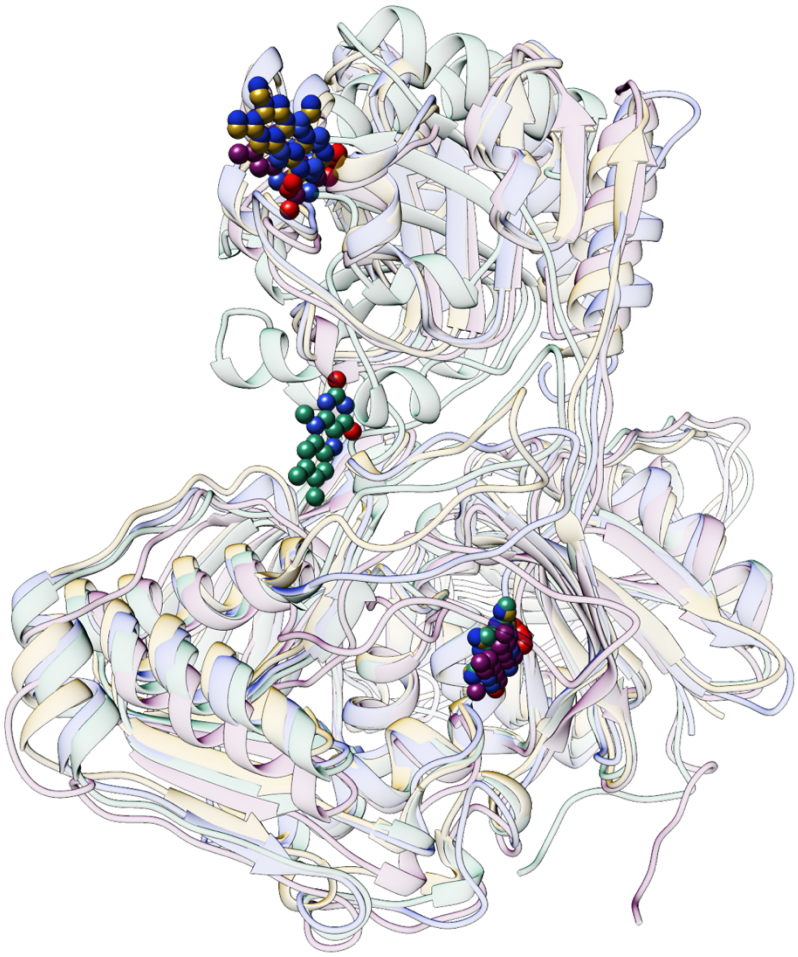


Figure S4. **Overlay of the structures of ETF in the 'closed' (green) and 'open' (all others) conformations**. Ribbons are 70% transparent but nevertheless convey the shape and dimensions of the different ETFs. The flavin head groups are in ball-and-stick to facilitate appreciation of the different distances of electron transfer between the Bf-FAD (lower down in figure, all 4 are superimposed) and ET-FAD (up and to the left, the 3 open conformations place their ET flavin further from the Bf-flavin than does the closed conformation). The green ribbon depicts the structure of A*fe*ETF (4KPU), the 6FAH structure of closed *Awo*ETF$_{CAR}$ from the complex with Caffeoyl CoA DH is in blue, the 5OL2 structure of closed *Cdi*ETF from the complex with BCD is in gold, and the 7QH2 cryo-EM structure of *Awo*ETF$_{LDH}$ from the complex with LDH is in burgundy. An 'intermediate' conformation has also been described for a bifurcating ETF that engages a quinone reductase,[8] instead of a CoA dehydrogenase as for the structures shown here, including *Afe*ETF.

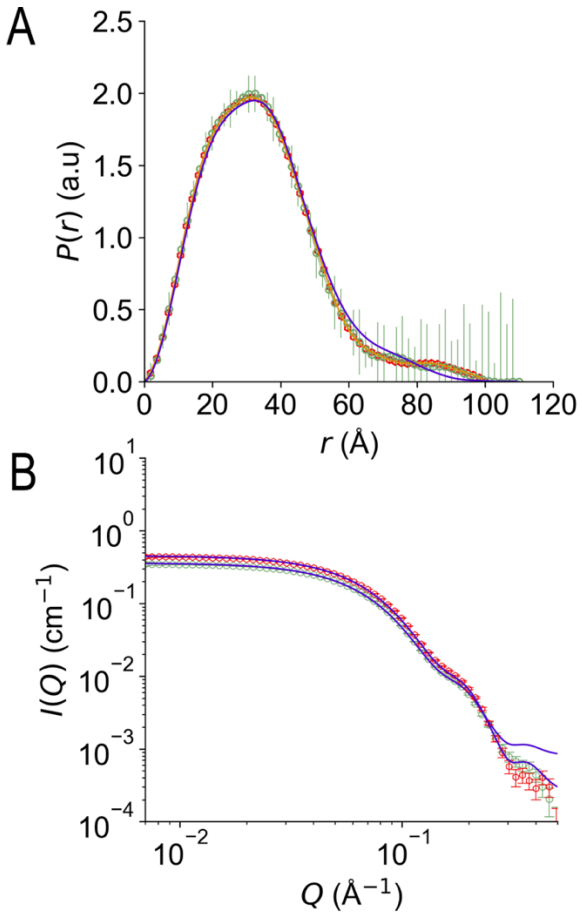
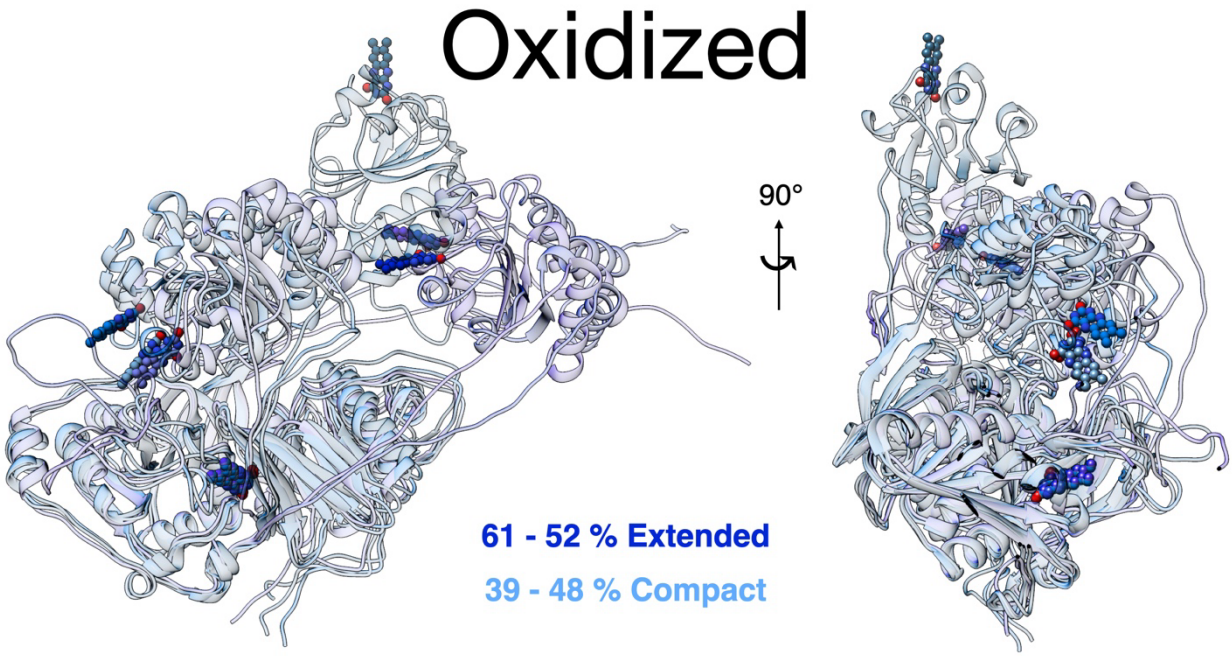
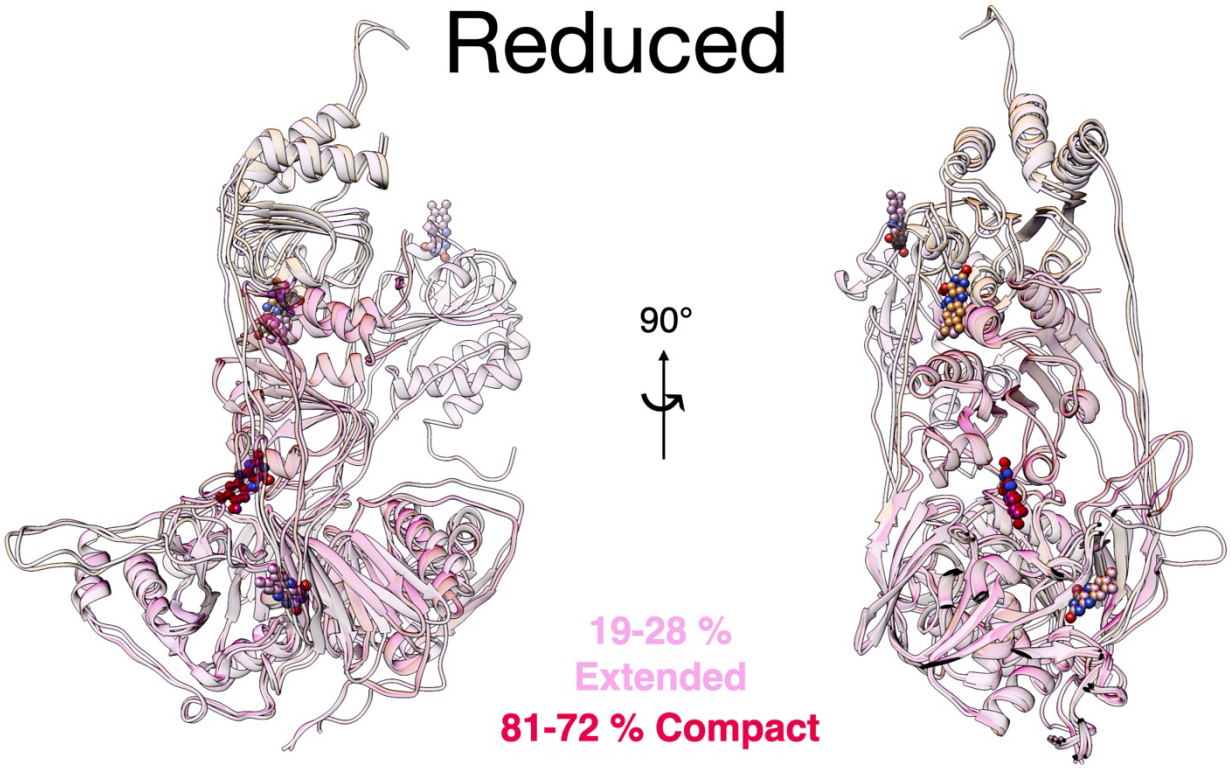


Figure S5. Comparison of SANS data of ETF reduced with the non-physiological donor dithionite *vs* the physiological donor NADH. Panel A: Overlay of pairwise distance distribution functions (*P(r)* profiles) of *Afe*ETF treated with dithionite (green), and *Afe*ETF treated with NADH (red circles), wherein the prevalence of scattering sites being separated by a particular distance is plotted *vs* the distance separating the two scattering sites, *r*. The solid blue line provides the prediction from the best-fitting two-conformation ensemble based on the combination of conformations generated by Bilbo-MD and metadynamics. Panel B: Overlay of scattering profiles of dithionite-reduced *Afe*ETF (green circles) and NADH-reduced ETF (red circles). Each set of data is accompanied by the theoretical scattering calculated based on the ensemble that best fit it (blue lines) based on conformations generated by both Bilbo-MD and metadynamics (combined model). Application of the GA to NADH- and dithionite-reduced *Afe*ETF scattering data produced almost the same conformations. Error bars are shown as solid vertical lines. For *P(r)* profiles (panel A) they are standard deviations based on multiple fits to the data using a series of Monte Carlo simulations[9] while for the SANS profiles (panel B), they are derived from counting statistics errors ($N^{1/2}/N$), where N is the number of detector counts.

Figure S6. Overlay of the extended and compact conformers that best explain the different SANS data sets discussed, for *Afe*ETF. In each case, the ribbon in the darker shade is the most populated conformer, and the populations are also shown (the ranges of values stem from the differences between the results obtained with different starting pools of conformers). Table 2 provides the individual values as well as R_g and D_{max}.



A. Oxidized state conformations chosen by the GA from the different pools of MD conformers. Regardless of the pool of conformers used, an extended conformation dominates the population (52 - 61%) and so is displayed using darker ribbons. The minority compact conformation is present in lighter ribbons. Ribbons are 70 % transparent to facilitate comparison of structures and flavin positions. Conformers obtained based on the Bilbo-MD generated conformer pool are in bright blues, those obtained based on the metadynamics generated pool are in teal blues and those based on the combination of both pools are in periwinkle blues. Corresponding flavins are likewise in darker and lighter shades, and depicted using heavy balls-and-sticks to facilitate assessment of their separation from one-another. The flavin that is lowest in the figure is a superposition of the Bf-FADs of all 6 conformations.



B. Reduced state conformations chosen by the GA. As in 'A', the dark colour is used for the dominant conformer, in this case the compact one. Indistinguishable conformations were obtained for *Afe*ETF reduced by NADH and dithionite. Ribbons are 70 % transparent to facilitate comparison of structures and flavin positions. Red/orange are used for results identified from a pool of conformers generated by Bilbo-MD, fuchsia/orchid pink are used for results emerging from a pool of conformers generated by metadynamics. Use of the two pools together resulted in identification of the same compact conformation as found based on Bilbo-MD generated pool but a different extended conformation, shown in pale pink. Corresponding flavins are likewise in darker and lighter shades, and depicted using heavy balls-and-sticks to facilitate assessment of their separation from one-another. The flavin that is lowest in the figure is a superposition of the Bf-FADs of all 5 conformations.

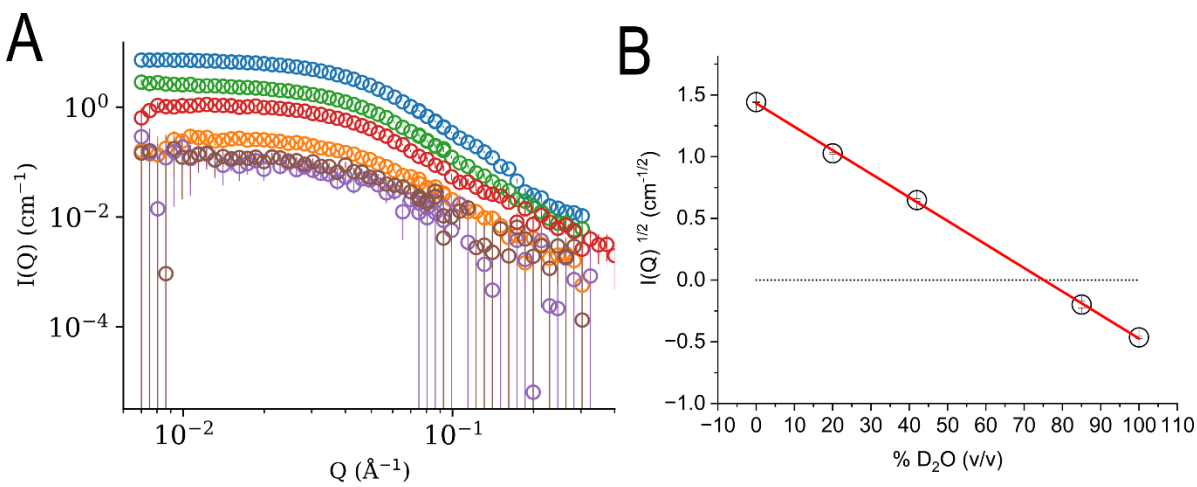


Figure S7. Determination of contrast match point of dBCD. A: SANS profiles of 70 % deuterated BCD (dBCD) in buffers containing D_2O at 0 % level (blue), 20 % (green), 42 % (red), 80 % (purple), 85 % (brown) and 100 % D_2O (orange). Error bars are shown as solid vertical lines and are derived from counting statistics errors ($N^{1/2}/N$), where N is the number of detector counts. For the contrast match point determination plot, errors are propagated from the Guinier analysis. B: Plot of the neutron scattering length density *vs* percent of D_2O. The intercept at 74.5 % D_2O provides the match point for the dBCD.

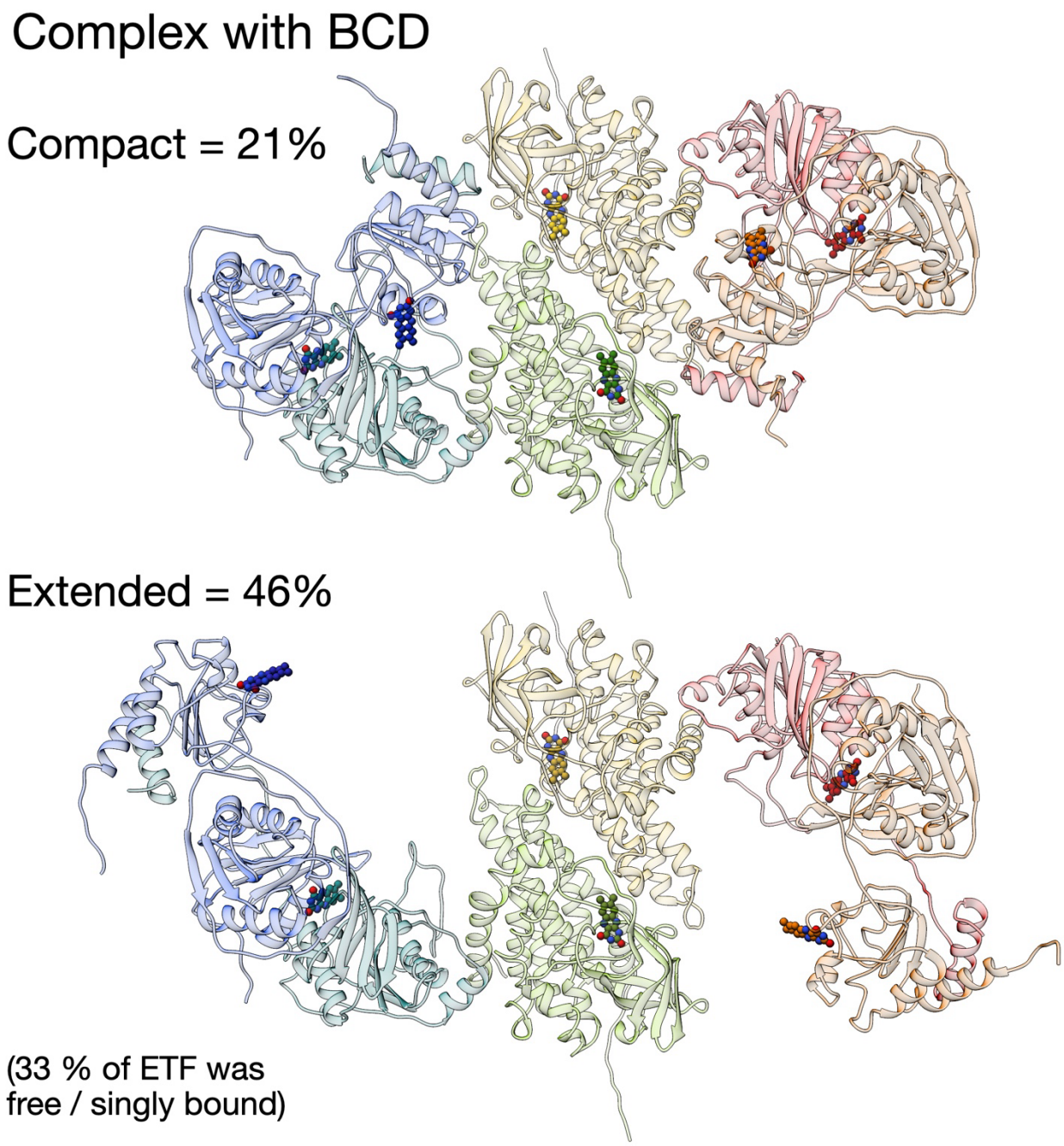


Figure S8. Models of the *Afe*ETF$_2$·BCD$_2$ complex contributing to the ensemble that best explains the SANS data. Above: compact conformation accounting for 21 % of the scattering. Below: extended conformation accounting for 46 % of the scattering. The best-fitting ensemble also included a compact conformation of *Afe*ETF that could be free *Afe*ETF or could be a single *Afe*ETF bound alone to a BCD$_2$ (i.e., a BCD dimer that only has a single ETF bound), as 33 % of the population. BCD$_2$ is in line green and yellow ribbons, one ETF is in blue and teal, the other is in red and orange. Ribbons are 60 % transparent to facilitate location of the flavin head groups (heavy ball-and-stick depictions) whose near/distant positions relative to one another in ETF is a characteristic of the closed/open conformations.

References Cited in the Supplementary Information

1. J. H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor MI, U.S.A., 1975.
2. K. F. Man, K. S. Tang and S. Kwong, Genetic algorithms: concepts and applications [in engineering design], *IEEE Trans. on Industrial Electronics*, 1996, **43**, 519-534.
3. M. Pelikan, G. L. Hura and M. Hammel, Structure and flexibility within proteins as identified through small angle X-ray scattering., *Gen. Physiol. Biophys.* , 2009, **28**, 174–189.
4. G. Tria, H. D. T. Mertens, M. Kachala and D. I. Svergun, Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering, *IUCrJ* 2015, **2**, 207-217.
5. P. Bernadó, E. Mylonas, M. V. Petroukhow, M. Blackledge and D. I. Svergun, Structural characterization of flexible proteins using small-angle X-ray scattering., *J. Am. Chem. Soc.*, 2007, **129**, 5656-5664.
6. S. Grudinin, M. Garkavenko and A. Kazennov, Pepsi-SAXS: an adaptive method for rapid and accurate computation of small-angle X-ray scattering profiles, *Acta Crystallogr D Struct Biol*, 2017, **73**, 449-464.
7. A. Hicks, P. Abraham, W. Leite, Q. Zhang, K. L. Weiss, H. O'Neill, L. Petridis and J. C. Smith, SCOMAP-XD: atomistic deuterium contrast matching for small-angle neutron scattering in biology, *Acta Cryst.*, 2023, **D79**, 420-434.
8. X. Feng, G. J. Schut, G. L. Lipscomb, H. Li and M. W. W. Adams, Cryoelectron microscopy structure and mechanism of the membrane-associated electron-bifurcating flavoprotein Fix/EtfABCX, *Proc Natl Acad Sci U S A*, 2021, **118**.
9. D. I. Svergun and J. S. Pedersen, Propagating Errors in Small Angle Scattering Data Treatment., *J. Appl. Cryst.*, 1994, **27**, 241-248.