SUPPLEMENTAL ONLINE MATERIALS

SUPPLEMENTAL TABLES

Table S1: Figures of merit of models generated based on group consensus sequences.[a]

Group	# sequences	Template	Clashes[b] (percentile)	Geometry[b] (percentile)	Ramachandran, favored[b]
Fix (2D2)	53	4kpu	88th %	85th %	93%
Bf (2A & 2B, non-basal[c])	118	4kpu	88th %	92 %	96%
Non-Bf (group 1)	449	1efv	99th %	98th%	96%
Group 2 (non-basal[c])	310	4kpu	86th%	90th %	96%
Group 2 (non-basal[c])•NAD	310	4L2I	80th%	86th%	95%

[a] FAD, AMP and NAD^+ moieties were grafted into the models from the related template files (listed below).

[b] Clashes, amino acid geometries, and Ramachandran allowed-ness were evaluated via the MolProbity Server (http://molprobity.biochem.duke.edu) (1).

[c]Non-basal sequences represent more than 95% of all of the group sequences and exclude isolated sequences that cluster close to the branching point in each group. Omission of such sequences is performed to increase the signal-to-noise of the analyses that are directed within the group. We employed this precaution for those analyses whose objective was to learn what residues were strongly correlated within the group and to construct a consensus sequence for the group. This practice is based on the understanding that basal sequences are those that are least similar to the consensus sequence and carry the highest statistical risk of possessing residues that could randomly be those characteristic of another group (they also tend to have poor bootstrap support).

Consensus Amino Acid Sequences used to generate models

Group 2AB, Includes known Bifurcating Etfs

>EtfB-G2AB

MKIVVCIKQVPDTTEVKIDPVTGTLIRDGVPSIMNPDDKNALEEALRLKEEYGGK
VTVITMGPPQAKAALREALAMGADEAILLSDRAFAGADTLATSYTLAAAIKKLG
DYDLIICGRQAIDGDTAQVGPQIAEHLGIPQVTYVEKIEVVGDDSLTVKRALEDG
YEVIEVKTPCLLTVIKELNEPRYPSVKGIFEAYDKDKEIKVWSADDIEVDESKLGL
KGSPTQVKKSFTPEAKGAGEILEGLSPEEAADKLVEKLKEKHII

>EtfA-G2AB
MSIDKSDYKGVWVFAEQREGKIQPVSLELLGKGRELADKLGVEVTAVLLGSNV
KDLAKELIAYGADKVYVVDDPELKDYTTEPYTKAICDLINEYKPEIVLVGATTIG
RDLAPRVAARLRTGLTADCTSLDIDEETKLLLMTRPAFGGNIMATIICPNHRPQM
ATVRPGVMKKLERDESRKGEIIKVEVDLTESDIRTKVLEIVKKAKETVDIEEADII
VSGGRGVGSPENFELLEELADLLGGEVAASRAAVDAGWIDADHQVGQTGKTVR
PKLYIACGISGAIQHLAGMQDSDYIIAINKDPDAPIFKVADYGIVGDLYKVVPELIE
KIKANSLK

Group 2D2, Diazotrophs
>EtfB-G2D
MHIVVCIKQVPDSAQIRVHPVTNTIMRQGVPTIINPYDLFALEEALRLRDRFGGEV
TVLTMGPPMAEDALRKALSYGADRAVLLTDRAFAGSDTLATSYALAAAIRKIGE
EFPVDIVFTGKQTIDGDTAQVGPGIAKRLGLQQLTYVSKIVSIDLAAREITVERRA
EGGVQVLKTKLPCLITMLEGTNEIRRGSMDDALRAARAEIVKWSAADAGIEDVS
KCGLKGSPTVVKKVFAPTPRAEKAEMIETADKTPRDLAEALIAKIFTRQPKLEAE
LAFRAAA

>EtfA-G2D
MSTANKEPAAPAKGRAGMKKELPEHFKAYKHVWVFIELERGQVHPVSWELLGE
GRKLADKLGVELAGVVLGPPGEALEAAAAEAFAYGADLAYLVEDPVLADYRNE
PYTKALTDLVNTYKPEILLLGATTLGRDLAGSVATTLLTGLTADCTELDIDADRS
LAATRPTFGGSLLCTIYTLNYRPQMATVRPRVMAMPERDESRTGRIIEHKLGMVE
EDIVTKVLDFIPDRQSNKANLAYADVVVAGGLGLGNAENFQLVKDLARVLGAE
VGCSRPLVQKGWVPADRQIGQTGKTIRPKLYIAAGISGAIQHRVGVEGADLIVAI
NTDPNAPIFDFAHYGIVGDAIRLLPALTEAFRRRLSPHSRDRLAS

Group 1 non-basal (non-Bf)
>EtfB-G1nb
MKVLVPVKRVVDYNVKVRVKADGSGVDLANVKMSMNPFDEIAVEEAVRLKEA
GVATEVVAVSIGPAQAQETLRTALAMGADRAILVETDEELEPLAVAKLLKAVVD
KEQPQLVILGKQAIDDDSNQTGQMLAALLGWPQATFASKVEVADGKATVTREV
DGGLETLSLKLPAVVTTDLRLNEPRYASLPNIMKAKKKPLDTVTPADLGVDVAP
RLKTLKVEEPAKRSAGVKVADVAELVEKLKNEAKVI

>EtfA-G1nb
MTILVIAEHDNASLKAATLNTVTAAAKIGGDVHVLVAGSGAGAAAEAAAKIAG
VSKVLLADAAAYAHGLAENVAALVVSLAGDYSHILAPATATGKNVLPRVAALL
DVAQISDITAVVSADTFERPIYAGNAIATVQSSDAKKVITVRTTAFDAAAAEGGS
AAVEAVAAAADAGLSSFVGRELAKSDRPELTSAKIIVSGGRGLGSGENFTKVLEP
LADKLGAAVGASRAAVDAGYVPNDWQVGQTGKIVAPQLYIAVGISGAIQHLAG
MKDSKVIVAINKDEEAPIFQVADYGLVADLFEAVPELEKAL

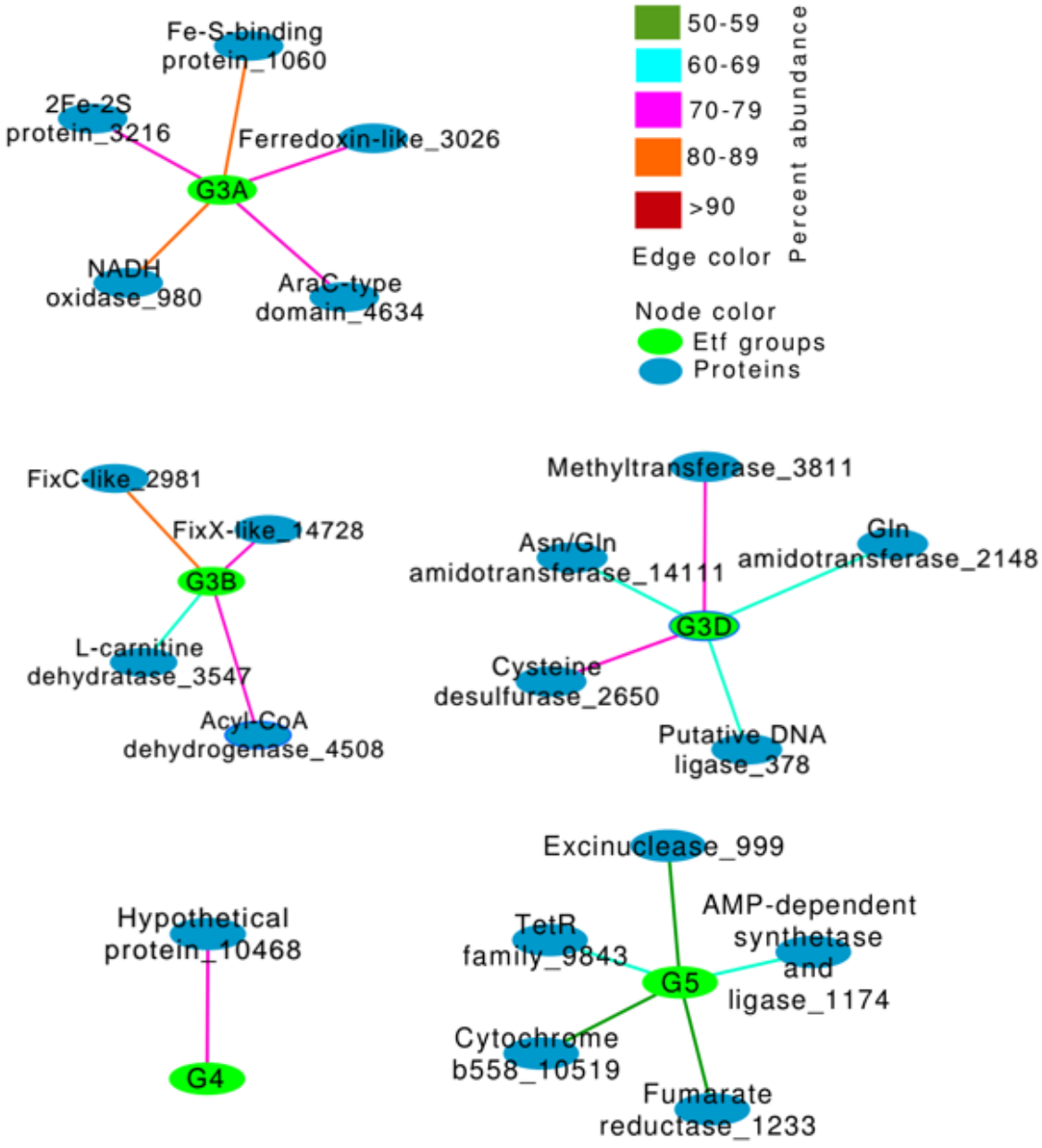
Group 2 non-basal
>EtfB-G2nb

MHIVVCIKQVPDTTQVRIDPVTGTLIREGVPSIINPYDLHALEEALRLKDKFGGKV
TVLTMGPPQAEEALREALAMGADEAILLSDRAFAGADTLATSYALAAAIRKIGEE
DVDLIFCGKQAIDGDTAQVGPGIAERLGIPQVTYVEKIEEVDLDKTITVKRRLEGG
YEVVEVKLPCLITVLKELNEPRYPSLPGKLRAARAEIKVWSAADLGDVDPSKIGL
KGSPTKVKKVFTPEARKEGEIIEGGDPEEADDDAAELLVEKLKEKPILEAKCKGC
GKCVKECPEDLAD

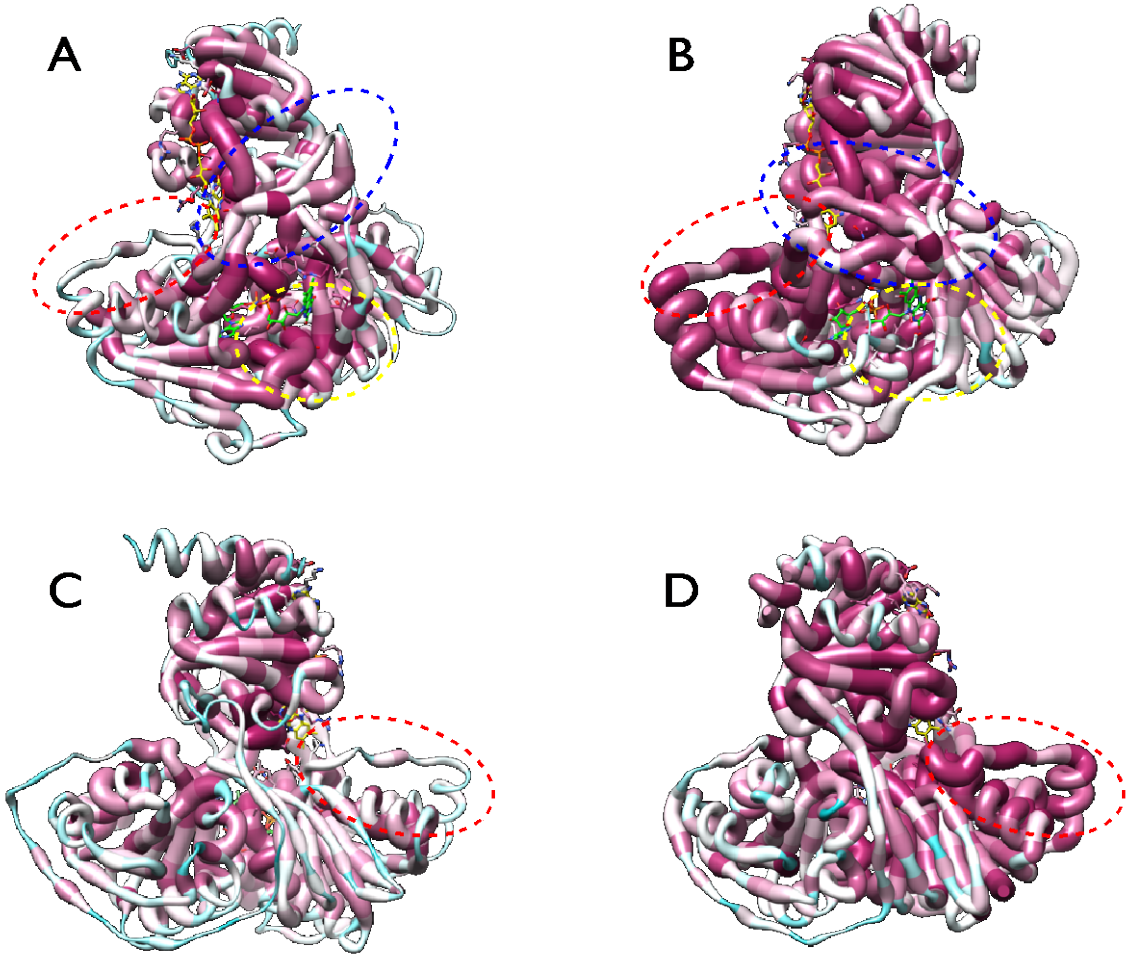
>EtfA-G2nb
YKGVWVFIEQEEGEVHPVSLELLGKGRKLADKLGVELAAVLLGSVVEDLAKELF
AYGADKVYVVDDPVLKDYRTEPYTKALTDLINKYKPEIILLGATTIGRDLAPRVA
TRLKTGLTADCTELDIDPETRLLLQTRPAFGGNIMATIVCPNHRPQMATVRPGVM
KKPERDEGRTGEIIEEEVDLTEEDILTKVLEVIKDREKEKVNLAEADIIVAGGRGL
GSKENFKLLEELADVLGGEVGASRAAVDAGWIPHDRQVGQTGKTVRPKLYIAC
GISGAIQHLVGMQDSDLIIAINKDPNAPIFDVADYGIVGDLFEVVPALTEALKKRL
AAKAK

Table S2. Group designation and taxonomy of organisms that encode for *etfAB* sequences in their genomes identified in available genome sequences (see the excel file in the supplemental online documents)

SUPPLEMENTAL FIGURES



Supplemental Figure 1. Network analysis of proteins encoded by flanking genes (+/- 20 genes) of *etf*. Only proteins (n=20) encoded by ≥50% of the Etf G3, G4 and G5 encoding genomes (i.e., relative frequency of ≥50%) were considered in this analysis. Here, a node represents either a group or a subgroup designation (denoted by green color) or an identified protein (denoted by blue color) within the gene neighborhood, while edge color represents the abundance of the proteins in the group.



Supplemental Figure 2. G2 is marked by greater amino acid conservation in the region of the Bf flavin and NAD binding, whereas G1 is distinguished by greater conservation in the recognition loop. Models of the consensus Etf of G2 (A and C) and G1 (B and D) are shown, based on consensus amino acid sequences determined excluding basal sequences in both cases (see Table S1). Panels A and B present a 'front' view whereas panels C and D present the 'back'. Amino acid identify conservation is color-coded as burgundy for positions at which a single identity is found, transitioning to teal for positions at which the maximum diversity of amino acid identity is found. Amino acid functional type conservation is depicted via worm widths with wide worms indicating positions where amino acids of a single functional type are retained and narrow worms indicating positions where diverse functional types are found, within each group. The FAD shared by both groups of Etf is in yellow sticks with heteroatoms in CPK colours. Green sticks depict the presumed Bf FAD of G2 Etfs, which has also been modeled into the G1 model structure, where it can be accommodated without steric clashes with surrounding amino acids. The red dashed oval identifies the recognition loop (2) which is more conserved in G1 than in G2, suggesting different interaction modes with partner proteins or more diverse partner proteins in group 2. The blue dashed oval identifies the ‘thumb’ loop of EtfB (residues 12-30) that interacts with domain I (EtfA) and in group 1 drapes over the site in which the Bf flavin binds. This loop appears to occupy different locations in G1

vs. G2 Etfs, however this is based on only two different crystal structures for G2 and four from G1 Etfs. The yellow dashed oval identifies the region in which the Bf flavin and NAD bind in the Etf structure 4L2I.pdb of Chowdhury et al. (3). High conservation is seen in the yellow oval in G2 but not G1.

REFERENCES

1. **Chen, V. B., W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson.** 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr **66:**12-21.
2. **Toogood, H. S., D. Leys, and N. S. Scrutton.** 2007. Dynamics driving function: new insights from electron transferring flavoproteins and partner complexes. Febs j **274:**5481-504.
3. **Chowdhury, N. P., A. M. Mowafy, J. K. Demmer, V. Upadhyay, S. Koelzer, E. Jayamani, J. Kahnt, M. Hornung, U. Demmer, U. Ermler, and W. Buckel.** 2014. Studies on the mechanism of electron bifurcation catalyzed by electron transferring flavoprotein (Etf) and butyryl-CoA dehydrogenase (Bcd) of *Acidaminococcus fermentans*. J. Biol. Chem. **289:**5145-57.