Electronic Supplementary Material (ESI) for Metallomics.

Supplementary Materials

Defining the domains of Cia2 required for its function in vivo and in vitro

Amanda T. Vo,[a] Nicholas M. Fleischman,[a] Melissa Marquez, Eric J. Camire, Stephanie U. Esonwune, John Grossman, Kelly A. Gay, Jessica A. Cosman, and Deborah L. Perlstein

Department of Chemistry, Boston University, Boston, Massachusetts 02215, United States

SUPPLEMENTAL FIGURES and TABLES

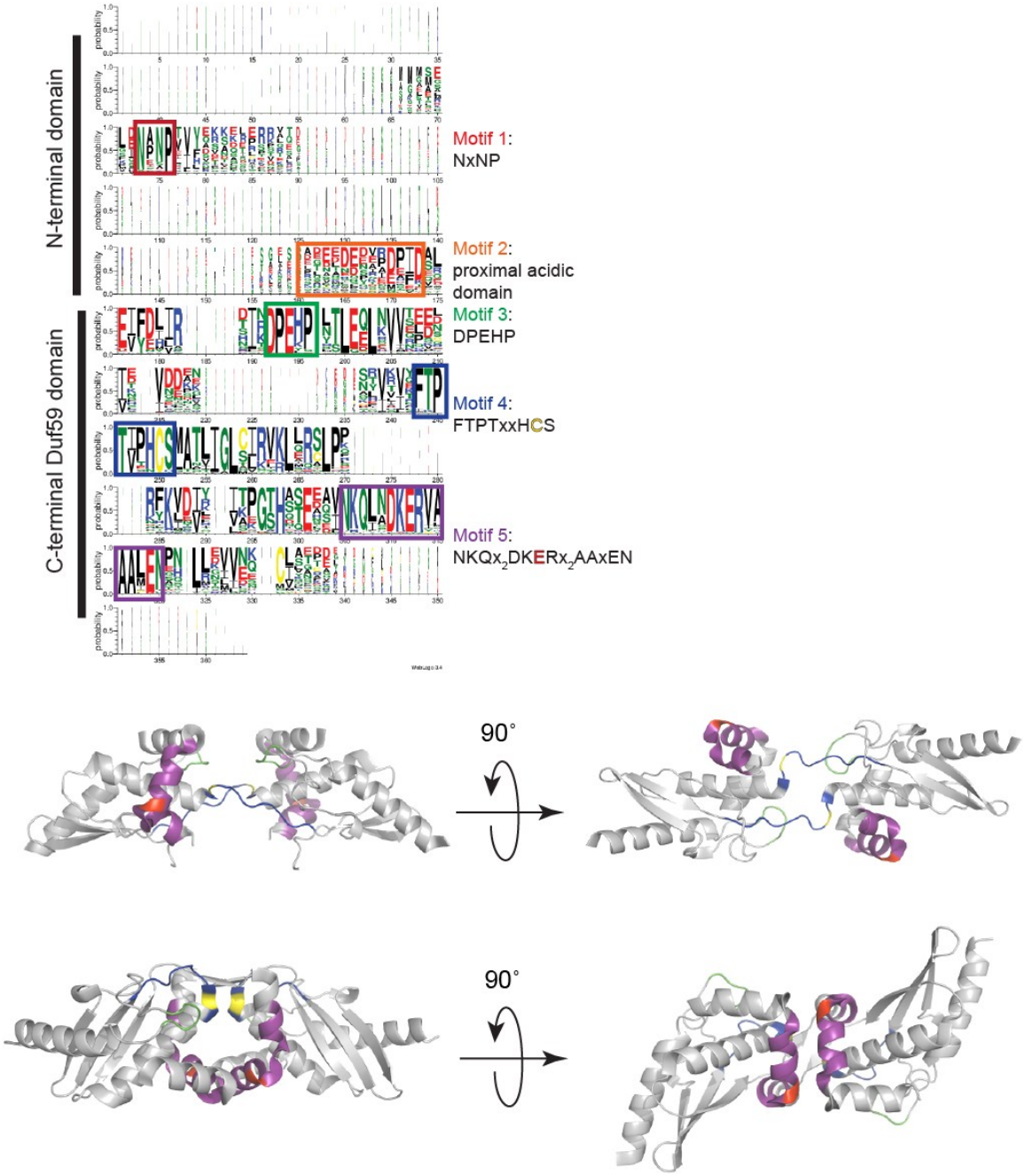
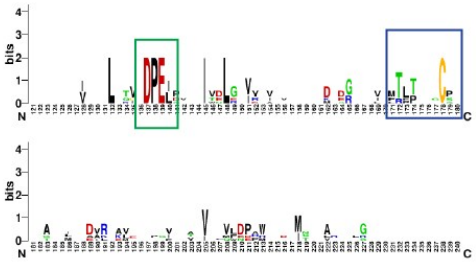


Figure S1. A) Sequence logo representing the alignment of 48 Cia2 homologs.[1] The N-terminal and C-terminal domains are indicated. The five conserved motifs are also indicated: two in the N-terminal domain (motifs 1 and 2 are red and orange, respectfully), two in the DUF59 domain (motifs 3 and 4 are green and blue respectfully with C161 yellow) and one following the DUF59 domain (motif 5 is colored purple with E208 red). B) Conserved regions mapped onto the two dimeric crystal forms of human Cia2a, the major dimer (Top, PDB ID 3UX2) and the major dimer (bottom).[2] Coloring as in part A and Figure 1.

Figure S2. Multiple sequence alignments and sequence logos of DUF59 proteins from bacteria and archaea (Panel A) Cia2a (Panel B), Cia2b (Panel C), Cia2 from organisms encoding one Cia2 ortholog (Panel D). In the alignments, regions predicted to be disordered by the MDFp2 server are red.[3] In the sequence logos, conserved motifs are boxed as in Figure S2 as follows: Motif 1, red; Motif 2, orange; Motif 3, green; Motif 4, blue; Motif 5, purple. Although *A. thaliana* encodes three Cia2 homologs,[4] their sequences cluster with the Cia2 and Cia2b sequences and were therefore all included in the Panel D alignments. The Uniprot numbers for the proteins used in the analysis are: A0CW13; A0DP48; A2DDP4; A7APH3; A7RPY9; A7SPP5; A8J757; A9SK08; A9UT22; B3RT11; B3RVK4; B8C104; C5L218; D2VP27; D3B7Q5; D3BHT1; D7FV08; E1ZKW8; E3KR97; F0Y5Q3; F0Z896; F0ZF45; I1J196; I7M608; M1VDD8; Q01GB5; Q4DHV8; Q6CEL6; R1EZP0; S9U937; V9VIF9; A0A067D6V7; B2WDH4; J3NI82; Q57YF8; Q6PBY9; Q6DHF2; B0EHV7; I1H8W2; A5KA37; G4YYC6; P38829; Q9V968; Q9Y3D0; Q9H5X1; O62252; Q9VTC4.

Figure S2A – bacterial and archaeal DUF59 sequences

```
F8D2V7_HALXS      ------------------------------------------------------------  0
M7AR11_9ACTN      ------------------------------------------------------------  0
PAAD_ECOLI        ------------------------------------------------------------  0
G1WCS9_9BACT      ----------------------------------------------------------MF  2
YITW_BACSU        ------------------------------------------------------------  0
Q5GWP2_XANOR      MRLMECSAVVAWRDGRRRRRCGRLEWPTCARPARPHIQPMYSRSSEPVQFERDCDAVMVP  60

F8D2V7_HALXS      ---------------------------------------------MSETESD----GRAGS  12
M7AR11_9ACTN      ----------------------------------MTALIDRGRDPAGTADPT----HPSDT  23
PAAD_ECOLI        ------------------------------------------------------MQRLATI  7
G1WCS9_9BACT      RGKKERMPKTTRGSWQHGGRQKYGAWLLLFHIICYLCNDNDSKNMTQEEKT---------  53
YITW_BACSU        ------------------------------------------------------------M  1
Q5GWP2_XANOR      QGDSVTLPAGSYGYITQ---ALGGSYTVFVEGNLFRIAGKDGDAIGKEAPPGLELPANAS  117

F8D2V7_HALXS      PAPTRAAVRDRLDRVTDPELDR-SIVELEYIDAIEI-----DGDRV-TVEFTLPTAWCSP  65
M7AR11_9ACTN      AVALTDAVMAALDTVSDPELDQ-PITELRFVRSVLV-----DDEGV-TVHLRLPTSFCSP  76
PAAD_ECOLI        APPQVHEIWALLSQIPDPEIPVLTITDLGMVRNVTQ-----MGEGW-VIGFTPTYSGCP-  60
G1WCS9_9BACT      --KIEERIVDVLKTVYDPEIPV-NIWDLGMIYKIDV-----KDDGNVDLDMTFTAPNCP-  104
YITW_BACSU        EEALKENIMGALEQVVDPELGV-DIVNLGLVYDVDM-----DEDGLTHITMTLTSMGCP-  54
Q5GWP2_XANOR      DEEVEALVWQQLRTCFDPEIPF-NIVDLGLVYEAVVSHREEDNQRRVDVKMTLTAPGCG-  175
                        :   *    ***:    * :*  :             :    : :      *

F8D2V7_HALXS      AFAWMMAVDARDEVESLPAVEEARI--VLDEHMHAEEINRGVNERRSFAESFPD-ADGDV  122
M7AR11_9ACTN      NFAYLMCSDALDALEDIDGIGRVQV--MLDDHHDSDKINAGLAAHAGYKGTFTVEAEQDL  134
PAAD_ECOLI        -ATEHLIGAIREAM-TTNGFTPVQVVLQLDPAWTTDWMTPDARERLREYGISPPAGHSCH  118
G1WCS9_9BACT      -AADFILEDVRTKVDSVEGVAATNVNLVFEPAWDQSMMTEEARVELGFE------------  152
YITW_BACSU        -LAPIIVDEVKKALADLPEVKDTEVHIVWNPPWTRDKMSRYAKIALGIQ------------  102
Q5GWP2_XANOR      -MGEILVDDVRSKVEMIPTIAEADVELVFDPPWGRHMMSEAARLETGML------------  223
                       :       :     . . :    :        :.

F8D2V7_HALXS      EAVRADLDEKARVARQYDAVETLLEAGLEPETIVDLRPADLERYELEGADDGGAGAEAER  182
M7AR11_9ACTN      DELRMTFLRKAHTAAMERAVSPMIAAGTTCV----------------------------DDC  168
PAAD_ECOLI        AHL----PPEVRCPRCASVHTTLISEFGS------------------------------TAC  146
G1WCS9_9BACT      ------------------------------------------------------------  152
YITW_BACSU        ------------------------------------------------------------  102
Q5GWP2_XANOR      ------------------------------------------------------------  223

F8D2V7_HALXS      IAIYVQDRTVAVTVPAAPIDRYL-EKARETGALAAPTEPLFRTPDGEPIDLESFDLVHRR  241
M7AR11_9ACTN      AQLLLRDL------PENPAKVALLRRRVQIGLGVCPNCRVVVDEDGSPLPAEAIPMRLRF  222
PAAD_ECOLI        KALYRCDS------CREPFDYFKCI----------------------------------  165
G1WCS9_9BACT      ------------------------------------------------------------  152
YITW_BACSU        ------------------------------------------------------------  102
Q5GWP2_XANOR      ------------------------------------------------------------  223

F8D2V7_HALXS      GRLAQVNMSSQGGICDGLREARERRFGGEGSATGDAETADD--------------------  282
M7AR11_9ACTN      ARSVRISMEGNGHFCRGLLATRYADDTDCDGTSGPVVTDLRMLPMPTTRTAHTPEPATPG  282
PAAD_ECOLI        ------------------------------------------------------------  165
G1WCS9_9BACT      ------------------------------------------------------------  152
YITW_BACSU        ------------------------------------------------------------  102
Q5GWP2_XANOR      ------------------------------------------------------------  223

F8D2V7_HALXS      ----------------      282
M7AR11_9ACTN      AHTGVPDPHTERSAS       297
PAAD_ECOLI        ----------------      165
G1WCS9_9BACT      ----------------      152
YITW_BACSU        ----------------      102
Q5GWP2_XANOR      ----------------      223
```



Motifs 3 and 4 are diverged in bacterial and archaeal DUF59 proteins but these two motifs form the core of the DUF59 domain signature.

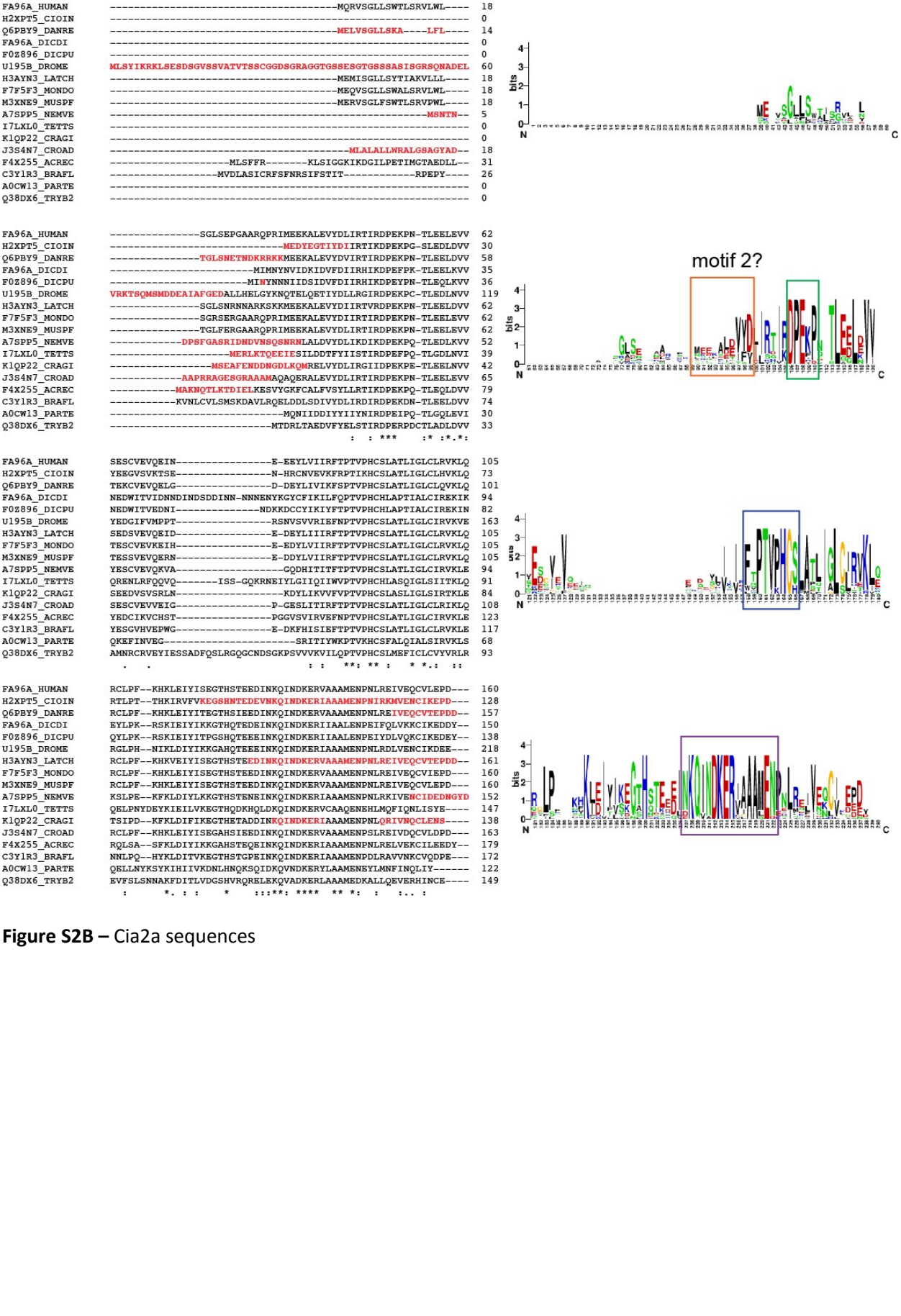


Figure S2B – Cia2a sequences

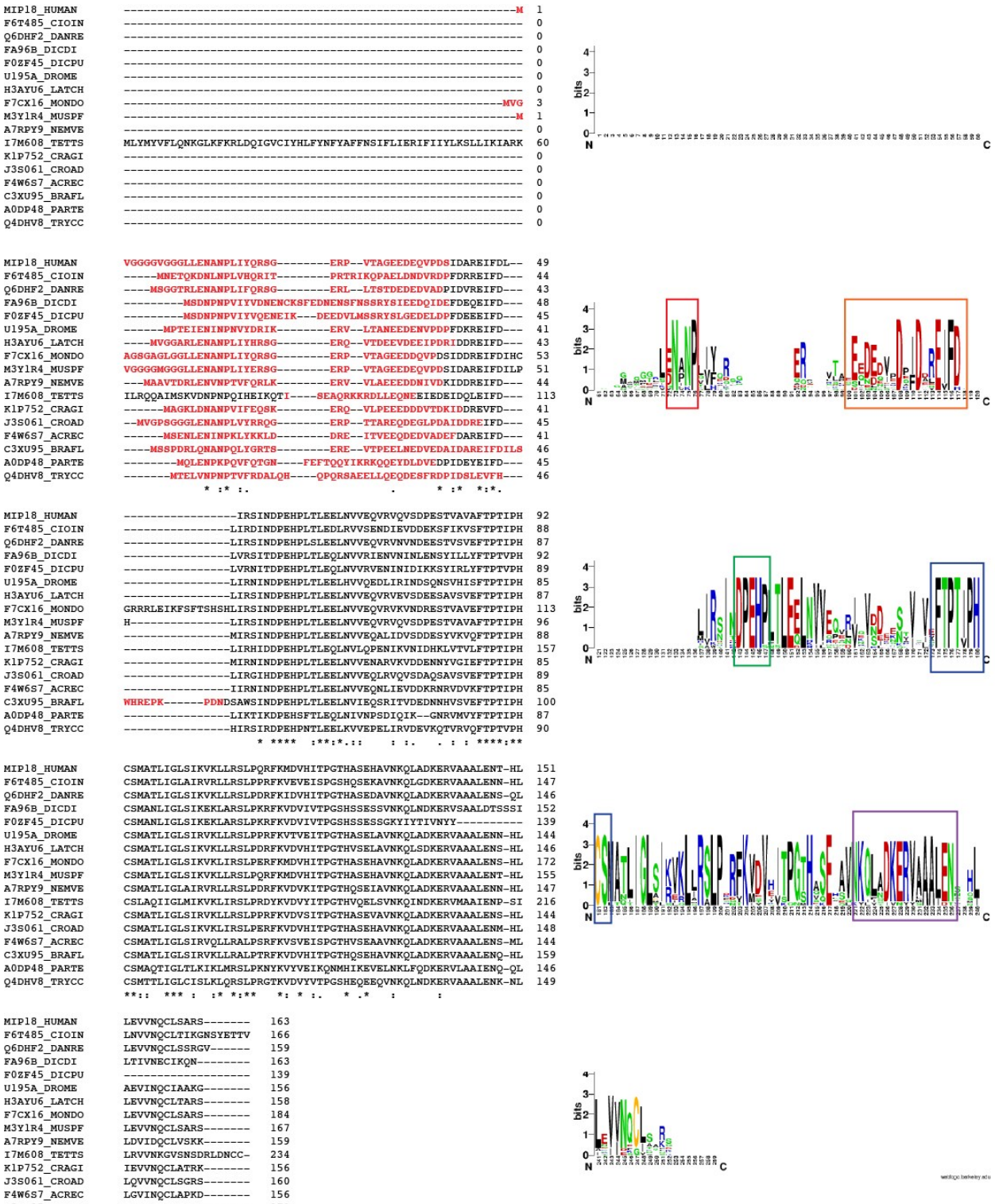


Figure S2C – Cia2b proteins

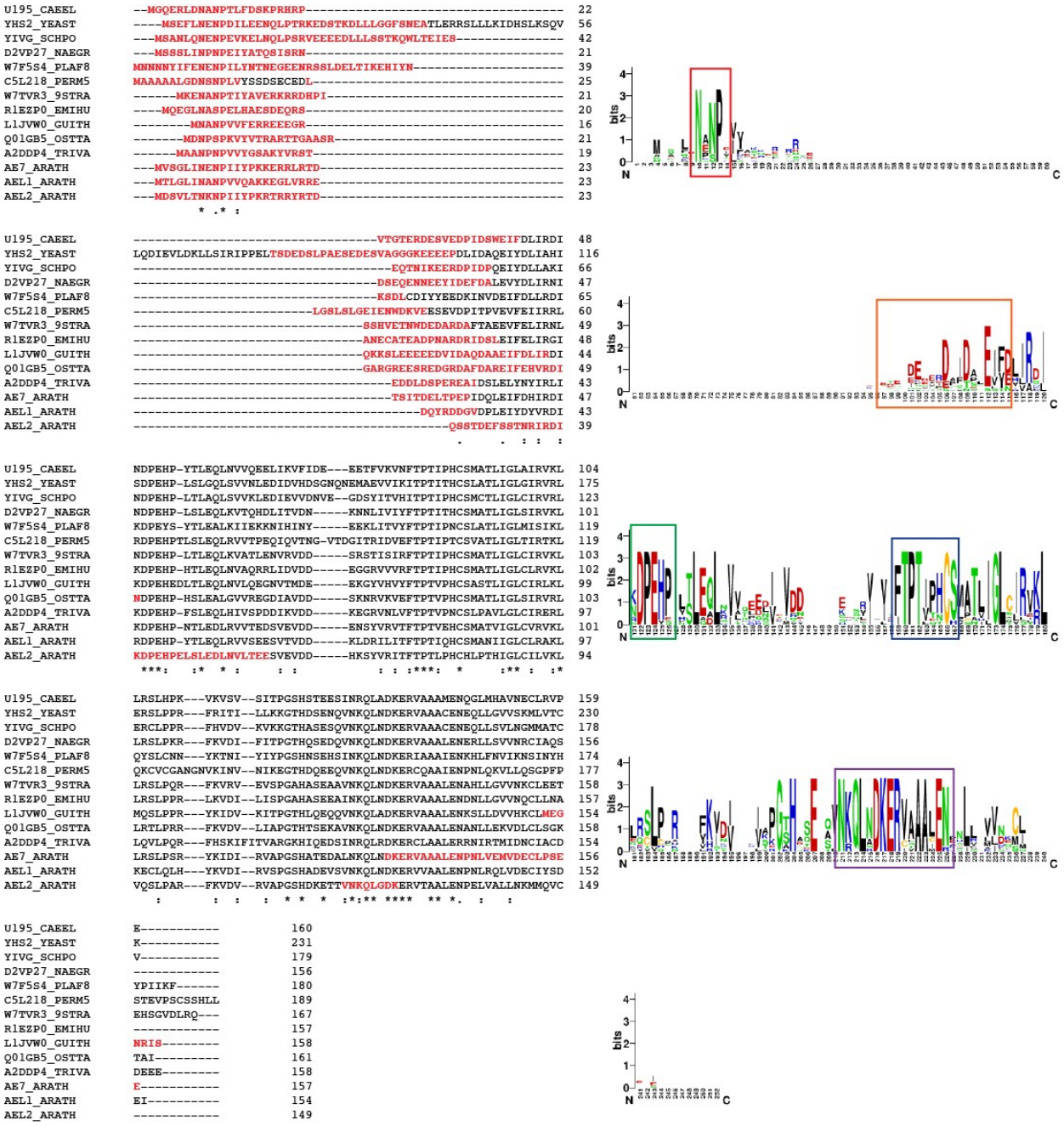


Figure S2D – Cia2 sequences that are not part of a Cia2a/Cia2b paralogous pair**.**

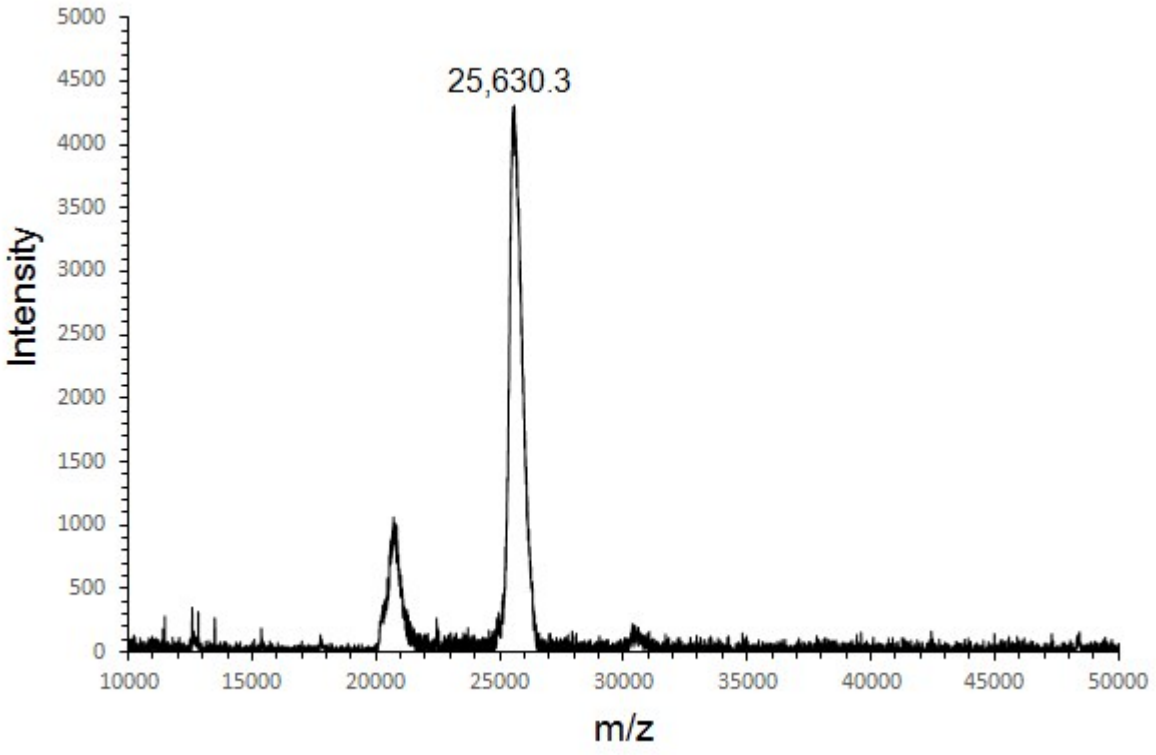


Figure S3. MALDI-TOF MS spectrum of refolded Cia2. The experimentally determined mass is 25,630.3 Da, which is in close agreement with the expected mass of 25,660.23 Da for the [M+H]+ ion.

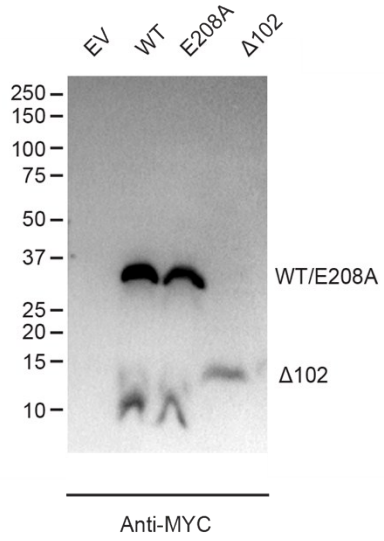


Figure S4. Anti-MYC Western blot of empty vector (EV), wild-type (WT), E208A-Cia2, and Δ102-Cia2. WT and E208A-Cia2 have bands at around 37kDa, consistent with the SDS-PAGE migration for Cia2. Δ102-Cia2 contains a 15kDa band, consistent with its predicted molecular weight. The relative migration of molecular weight markers in kDa are indicated to the left. Both WT and E208-Cia2 have additional lower molecular weight bands that are likely due to proteolysis of Cia2.

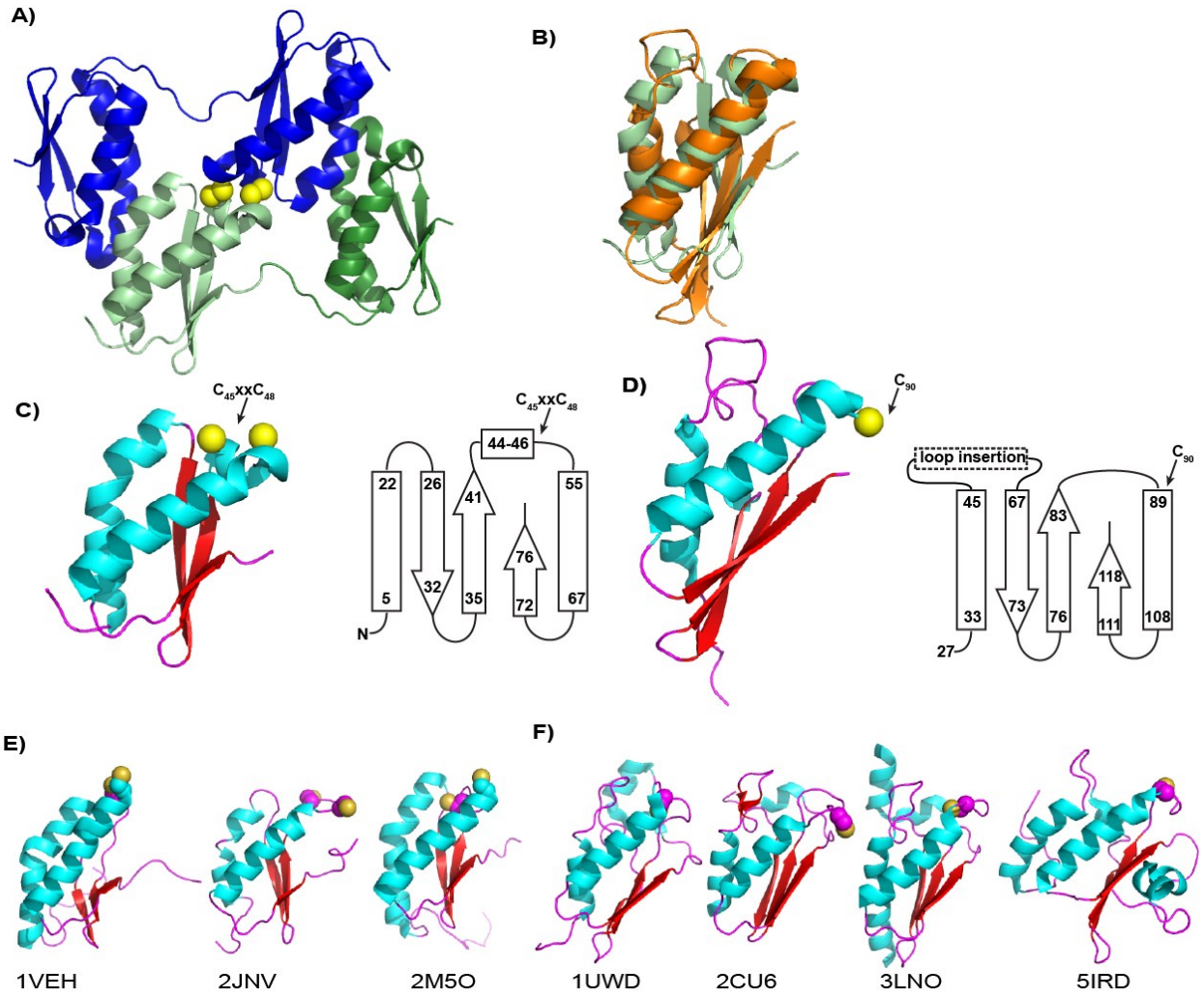


Figure S5. The structural similarity of NFU-domain proteins and DUF59 proteins. A) The structure of arabidopsis CnfU iron sulfur cluster biosynthesis protein[5] (PDB 2Z51). The FeS-binding cysteines, two from each polypeptide, are yellow spheres. The two polypeptide chains are colored blue and green. Each CnfU polypeptide has two NifU domains, colored light and dark green in the green colored polypeptide, where the N-terminal NifU domain has the CxxC motif. B) The FeS binding domain of CnfU (light green in panel A) overlaid with human Cia2a[6] (2M5H, orange). C) The FeS binding domain of CnfU colored by secondary structure and secondary structure map. The location of the CxxC motif is indicated and cysteine sulfurs are yellow spheres. D) Cia2a DUF59 domain (residues 27-119) colored by secondary structure and its secondary structure map. The absolutely conserved cysteine of the DUF59 domain is shown as a yellow sphere. E) All of the known Nfu structures in the protein data bank including the mouse (1VEH), rice (2JNV), and human (5M5O) Nfu proteins. The cysteines of the CxxC motif is shown as spheres. F) All of the known Duf59 structures in the protein data bank including *T. maritima* (1UWD), *T. thermophiles* (2CU6), *B. anthracis* (3LNO), and *M. tuberculosis* (5IRD). The cysteine corresponding to the absolutely conserved cysteine of the DUF59 domain is shown as spheres.

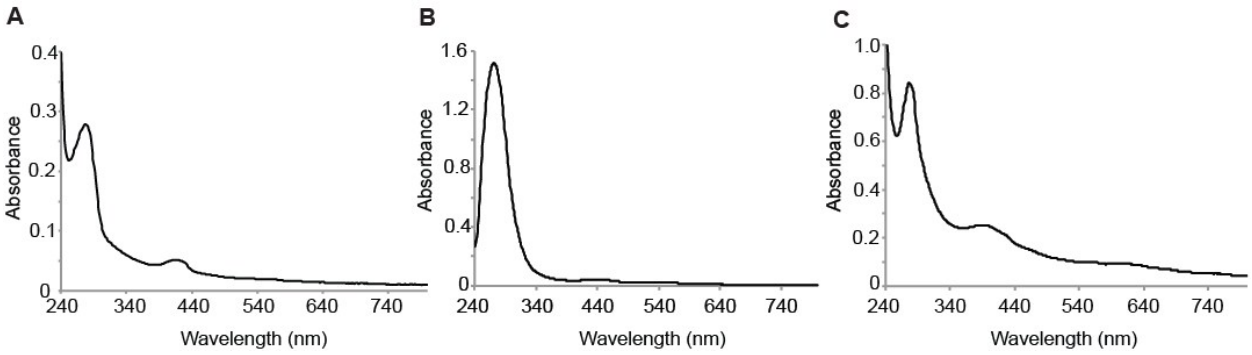


Figure S6. UV-Vis spectra of as isolated SUMO-Cia2 (A), as isolated dtCia2 (B), and chemically reconstituted dtCia2 (C). Features in the low 400 nm region are consistent with $[Fe_2S_2]$ or $[Fe_4S_4]$ binding.

REFERENCES

[1] Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator, *Genome Res. 14*, 1188-1190.

[2] Chen, K. E., Richards, A. A., Ariffin, J. K., Ross, I. L., Sweet, M. J., Kellie, S., Kobe, B., and Martin, J. L. (2012) The mammalian DUF59 protein Fam96a forms two distinct types of domain-swapped dimer, *Acta Crystallogr. D Biol. Crystallogr. 68*, 637-648.

[3] Mizianty, M. J., Stach, W., Chen, K., Kedarisetti, K. D., Disfani, F. M., and Kurgan, L. (2010) Improved sequence-based prediction of disordered regions with multilayer fusion of multiple information sources, *Bioinformatics 26*, i489-496.

[4] Luo, D., Bernard, D. G., Balk, J., Hai, H., and Cui, X. (2012) The DUF59 family gene AE7 acts in the cytosolic iron-sulfur cluster assembly pathway to maintain nuclear genome integrity in Arabidopsis, *Plant Cell 24*, 4135-4148.

[5] Yabe, T., Yamashita, E., Kikuchi, A., Morimoto, K., Nakagawa, A., Tsukihara, T., and Nakai, M. (2008) Structural analysis of Arabidopsis CnfU protein: an iron-sulfur cluster biosynthetic scaffold in chloroplasts, *J Mol Biol 381*, 160-173.

[6] Chen, K. E., Richards, A. A., Ariffin, J. K., Ross, I. L., Sweet, M. J., Kellie, S., Kobe, B., and Martin, J. L. (2012) The mammalian DUF59 protein Fam96a forms two distinct types of domain-swapped dimer, *Acta Crystallogr D Biol Crystallogr 68*, 637-648.