Spring 2020

A Broadly Distributed, Novel Class of Diiron Enzymes Involved in Para-Aminobenzoate Synthesis

Joseph M. Zambelas
University of South Carolina - Columbia

A BROADLY DISTRIBUTED, NOVEL CLASS OF DIIRON ENZYMES INVOLVED IN *PARA*-AMINOBENZOATE SYNTHESIS

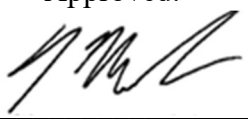
By

Joseph M. Zambelas

Submitted in Partial Fulfillment
of the Requirements for
Graduation with Honors from the
South Carolina Honors College

May, 2020

Approved:



Thomas Makris, PhD
Director of Thesis

Hippokratis Kiaris, PhD
Second Reader

Steve Lynn, Dean
For South Carolina Honors College

Abstract

The overall structure and ligand arrangement of diiron enzymes influence their function and catalytic potential, allowing for them to catalyze an expansive breadth of high energy transformations. The diiron enzyme CADD from *Chlamydia trachomatis* is involved in a novel biosynthetic pathway for p-aminobenzoic acid (pABA), a precursor for tetrahydrofolate. The unique ligand structure and reactivity of CADD and homolog NE1434 provide an intriguing means to contrast structure function relationships with other known diiron enzymes that activate dioxygen. In this work, studies of CADD and NE1434 orthologs and comparison to other diiron enzymes reveal several sequence motifs that are likely critical for its distinct function.

Table of Contents

Introduction

Diiron proteins are a set of powerful enzymes responsible for a wide array of catalytic reactions such as regiospecific oxidation of aromatic derivatives [1], methane hydroxylation [2], and alkane hydroxylation [3]. Diiron enzymes react with O_2 to form highly oxidizing reactive intermediates, which enable them to perform catalytically challenging reactions. The study of these proteins allows us to better discover the reactivity of the diiron core and its ability to initiate catalysis on unreactive substrates. Diiron proteins contain two coupled iron atoms bound to the protein by amino acid ligands. CADD is a diiron protein with a novel ligand structure [4], however, few studies have been published that characterize CADD or its mechanism of action. Thus far, little is known about the functional role of the diiron cluster of CADD.

Chlamydia protein associating with death domains (CADD) is an O_2-dependent diiron protein named for its ability to induce apoptosis within infected mammalian cells. CADD has also been identified as having a key role in tetrahydrofolate biosynthesis in *Chlamydia trachomatis* [5]. This diiron protein carries out a unique synthetic strategy to make *para*-aminobenzoate (pABA), a necessary precursor to folate (Vitamin B_9) and tetrahydrofolate synthesis [6]. Considering humans and most other bacteria lack this biosynthetic machinery, CADD may serve as a highly selective drug target. Considering the catalytic potential of diiron clusters, we hypothesize that the diiron site is involved in a reaction to generate pABA. As well, CADD resembles only a few other thoroughly classified enzymes, mainly decarboxylase UndA [7]. Given the novel ligand structure of CADD, the unprecedented reaction of pABA biosynthesis, and since this ligand motif has synthesized two vastly different products, it is likely that the mechanism of action for this protein is unique and will add further to literature on diiron binding motifs and tunability. CADD is an oxygen dependent enzyme and forms a long lived intermediate designated CADD-O_2 [7]. Oxygen

activation by diiron cluster typically manifests cyclically. Diiron catalysis is assumed to follow a known diiron intermediate pathway. Upon introduction of oxygen, the diiron core is oxidized from the diferrous state to the diferric peroxo state termed “intermediate P” [8]. From here, a number of oxidation states may persist, but it is currently hypothesized that CADD utilizes this peroxo-intermediate to induce catalysis. After unknown substrate oxidation, the diiron cluster would be left in a diiron (III) state without oxygen forming an “ox intermediate.” It then relaxes to the diferrous state [8].

Homologs of CADD have been identified in many bacteria including *Nitrosomonas* and *Lactobacillus* species [9]. Recently, a paper was published describing a protein in *Nitrosomonas europaea* with a similar role in pABA biosynthesis to CADD [5,9]. This protein, designated NE1434, has yet to be isolated and characterized *in vitro.* An analog for CADD may have more manageable properties since, currently, CADD has low iron loading. As well, a CADD analog would elucidate further characteristics of the CADD family and add to the literature understanding of diiron tunability.

The goal of this work was to isolate the protein encoded by gene *ne1434* and characterize it relative to CADD. This entails primary isolation and growth using vector plasmid mediated expression. As well, NE1434 characteristics and properties will be estimated using protein sequence analysis as well as structural prediction and comparison. Conserved residues and structures in CADD and NE1434 orthologs illuminate key motifs that may be responsible in the unprecedented reaction of pABA biosynthesis.

Methods

Expression Trials of NE1434

Gene *ne*1434 was codon optimized by company Genewiz and inserted into a pQE60 vector (Figure 1) using NcoI and BamHI restriction sites which adds a hexahistidine n-terminal tag. A transformation was performed to insert this plasmid into BL-21 (DE3) competent *Escherichia coli* cells. These cells were then grown in petri dishes containing ampicillin expressed in 5-mL cultures of Luria Bertani (LB) broth with and without 200 µM IPTG at a gradient of temperatures: 18˚C, 25 ˚C, and 37 ˚C. All LB solution contained 200 µL/mL of ampicillin. The cells were transferred into 15-mL falcon tubes and spun down. The supernatant was discarded, and the cells were resuspended in 15 mL of Sigma Aldrich Bugbuster Protein Extraction Reagent. These samples were vortexed, incubated for 5 minutes, and spun down at 4000 rpm for ten minutes. Then, 5 µL of each supernatant was mixed with a 5x Laemmli buffer solution and analyzed using SDS-PAGE.

Large scale expression

A 6-L batch of NE1434 protein was expressed in BL-21 (DE3) cells, grown in LB within six 2-L Erlenmeyer flasks with 200 µL/mL ampicillin and induced using IPTG. The cells were then spun down via centrifugation at 6000 rpm for 10 minutes and the pellet was resuspended in lysis buffer. All buffers after resuspension contained 25 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) and 100 mM NaCl with a pH of 7.5. Lysis buffer had an additional 10 mM imidazole. These cells were placed on ice and underwent seven rounds of sonication, each round lasting two minutes sonicating and five minutes on ice. These cells were spun down at 16000 rpm. The supernatant was then loaded onto a conditioned Nickel-Nitrilotriacetic Acid (Ni-NTA) column and washed using 10 column volumes (CV) of a 20 mM imidazole buffer. The protein was eluted

using 400 mM imidazole via fraction collector. A sample was taken for each step of the purification process to track protein location. Insoluble cell material post-lysis was digested in 8 M urea and then treated with, along with all other samples, Laemmli buffer and then analyzed via SDS-PAGE.

Sequence Alignment and Cladogram Generation

The NE1434 protein sequence was used to conduct a Basic Local Alignment Search Tool (BLAST) search within the NCBI database [11] for all proteins with sequence similar enough to the protein to have an E-value below 10^{-3}. During this search, redundant reference sequences were removed. A set of 31 different proteins were randomly selected from this set and then used to construct a sequence alignment using Multiple Sequence Comparison by Log- Expectation (MUSCLE) algorithm in Mega software [12,13,14]. From this alignment, a Maximum Likelihood cladogram was constructed using the Jones-Taylor-Thornton Model. The rates among sites were uniform and the ML heuristic method used was nearest-neighbor-interchange.

Structural Alignment

All structural alignments were performed using PyMOL and the a prior align function within it [15]. The align first creates a sequence alignment and from this builds a structural superposition followed by successive refinements to reject structural outliers. NE1434 protein structure was predicted using Expasy Swiss [16,17].

Results & Discussion

Expression Trials of NE1434 and Large-Scale Attempt

After initial transformation of each the pQE60 vector plasmid, small satellite colonies of a similar coloration appeared in the petri dishes after overnight growth. It is currently theorized that these small colonies may have excreted their plasmids and so ceased to grow with ampicillin present. The process was repeated using freshly transformed proteins to confirm there was no contamination with similar results.

Temperature trials revealed that expression decreased with temperature and the strongest band around 28 kDa (the expected size of NE1434) appeared most prominently under 18˚C conditions (Figure 2). The presence of IPTG had little to no effect on the expression of the protein. This is likely due to the "leaky" nature of the pQE60 vector. There is no *lacI* gene present within the vector as seen in Figure 1. This gene encodes for the LacI protein which is responsible for repressing the expression of the inserted gene.

After the 6-L expression, SDS-PAGE analysis revealed a 28 kDa band remained within the insoluble cell lysis (Figure 3). This may indicate that NE1434 is present within the lysate. Given the structure of the protein, it seems unlikely that it is insoluble, rather it may be simply expressing poorly, possibly due to issues with the vector or DNA sequence.

Sequence Alignment & Cladogram

A BLAST search of the NE1434 primary sequence resulted in over 1300 proteins with an E-value below 10^{-3}. A random sample of 31 proteins was taken with E-values ranging from 10^{-60} to 10^{-3}. Identity values ranged from 18-47%. A sequence alignment and subsequent cladogram (Figure 4) were used to predict the evolutionary relationship between the sample proteins. Diiron

proteins CmlI from *Streptomyces venezuelae* and UndA from *Pseudomonas protegens* were also added to the sequence alignment since they are later used for structural comparisons. CmlI is a well classified protein that has been known to catalyze a reaction with pABA [18]. It has 21% identity and an E-value of 0.94 when compared to NE1434. Protein UndA is similar in character to NE1434 in structure but catalyzes the decarboxylation of medium chain fatty acids and has 27% identity and an E-value of 2.8.

Cladogram analysis suggests that proteins related to CADD may be indicators of pathogenic tendencies within bacteria. CADD has the ability to induce the death of infected cells and so those proteins that are similar to it in structure may retain this property. Most of the bacteria found from the random sample are in some manner harmful to other organisms. For example, *Sternotrophomonas maltophilia* is a direct human pathogen [19], the *Rickettsia* endosymbiont is known to live in ticks [20], and *Rickettsia slovaca* causes tick-borne lymphadenopathy (TIBOLA) in humans [21]. More distantly related bacteria also demonstrate pathogenic tendencies. *Klebsiella pneumoniae* is an opportunistic pathogen [22]. Obligate intra-cellular *Deltaproteobacteria* have been isolated [23]. However, not all of these bacteria are pathogenic. *Nitrsomonas europaea,* from which NE1434 is derived, has no reported pathogenic tendencies and is a nitrogen fixing bacteria [24]. However, given the large variation in bacteria, and the vague description of sample proteins, such as *Deltaproteobacteria* protein, statistics-based analysis needs to be applied if a connection between protein structure and pathogenicity is to be made.

Further analysis of the cladogram uncovered a likely protein homolog from *Verrucomicrobia* bacterium protein (referred to from here on as Verr). This protein has 47% identity with NE1434 and 48% identity with CADD. The sequence identity between Verr and CADD is higher than the identity between NE1434 and CADD. The predicted structure of Verr

contains an identical set of ligands that bind to the diiron center, also known as the primary coordination sphere. As well, an amino acid residue that is interacting with the primary coordination sphere, a residue of the secondary coordination sphere, is retained. So, Verr may be a decent candidate for homolog expression trials and further study. An added homolog will both increase our understanding of small variations that may affect rate and chemospecificity of the catalyzed reaction. As well, protein Verr may have qualities such as increased stability that would be beneficial for future experiments.

In order to find functionally significant residues, a sequence alignment of the 9 most closely related proteins from the sample group was constructed. This revealed amino acids that are conserved in between the alpha helices and about a cavity in the center of the protein (Figure 5a). This cavity is estimated to be a likely site for substrate binding [4], and so these residues may be involved in binding of the substrate. Many of these are aromatic, primarily tyrosine, tryptophan, and phenylalanine residues. This may be due to the increased protein stability from aromatic-aromatic interactions [25]. As well, all of the ligands from the primary coordination sphere were conserved along with one from the secondary coordination sphere. This suggests that the secondary coordination sphere may not be particularly significant for pABA synthesis. As well, a number of phenylalanine and tyrosine residues near the activation site were retained.

If these residues are important exclusively for binding pABA substrate, they will not likely be conserved in proteins that utilize radically different substrates. These residues were compared to the rest of the sample population. Internal aromatic residues were moderately well conserved, along with all of the ligands.

To further identify the importance of Y, W, and F in NE1434 and similar proteins, a structural comparison was performed between CADD and NE1434 to determine how well all of

these residues were conserved about the protein. The NE1434 structural model was found to have remarkably similar residue overlap between the alpha helix bundles and close to the pocket, but little overlap was seen outside of it (Figure 6).

To further identify critical residues for CADD and NE1434 function, more sequence alignment analysis was performed with proteins that bind to product. Diiron proteins CmlI from *Streptomyces venezuelae* and AurF monooxygenase from *Streptomyces thioluteus* both bind pABA [26], the likely product of NE1434, and so may have binding motifs that would illuminate important residues within NE1434 structure. Sequence alignment between these three proteins revealed 19 conserved amino acids out of 250. These conserved residues demonstrate a seemingly random distribution about the protein (Figure 5b), and these similarities may be statistical in nature. Only three amino acids about the central cavity were retained. This suggests that binding substrate in NE1434 differs from CmlI and AurF.

However, the similarities between these proteins require further study. Given the vast differences between these types of strucutres, a sequence alignment may not reveal important similarities given the differences between length and strucutre. For instance, the ligand motif for AurF and CmlI differ by a single nucleotide from NE1434, however, in the sequence alignement, only two of the ligands are conserved.

To note, a proline residue is present within the H6 helix that may be, in part, responsible for a well conserved curve throughout most of the sample proteins. This residue is well conserved within most of the protiens tested.

Structural Comparisons

The Swiss model predicted structure of NE1434 as compared with CADD revealed significant similarities between the proteins (Figure 7). The only notable differences lie in small loop variations and an alpha helix extension between the H2 and H3a helixes of NE1434. This, along with a sequence alignment, suggest that tertiary structure will have little impact on variations in protein function for CADD and NE1434 and may catalyze similar reactions. However, Expasy Swiss software, in part, predicted the structure of NE1434 from CADD. So the structures are likely less similar than predicted.

Diiron protein UndA (pdb file: 4WWZ) catalyzes the decarboxylation of moderate chain fatty acids to olefins. The protein has a similar seven-helix motif and are both approximately 250 residues in length. The four alpha helix bundle about the diiron center differ in placement by fractions of angstroms. The H7 alpha helix, outside of the central four alpha helices, has increased length in UndA, which decreases inertia and limits the motility of H7. This likely affects interactions between the four-helix motif and substrate. The H5 to H6 and the H1a to H1b loop contain significantly more helix character than in CADD which may affect motility of the helices and increase rigidity. As well, the ligand motif is off by one additional ligand in UndA. Slight variations in ligand motif can correlate to drastic variation in reaction trajectory. However, close to the central cavity, UndA and NE1434 both appear to have a high hydrophobic residue content. Though there appear to be fewer polar residues in UndA to accommodate the fatty acid substrate.

The *Deltaproteobacteria* protein (Referred to as Delta) differs more significantly from NE1434. Delta does not contain a diiron center but rather is more akin to a pyrroloquinoline quinone-synthase (PqqC). PqqCs catalyze the formation of pyrroloquinoline quinone (PQQ), a three-ring compound with high carboxylate content. In Delta, the residues positionally equivalent to ligands have been replaced by glycine and alanine. This plus differences in position have

removed any possibility of iron-loading. In between the alpha helix bundles, the Y, W, and F residues, however, remain conserved and have a similar distribution and positioning to NE1434. PqqC type proteins utilize eight-electron oxidation reactions, creating a need to release oxidative stress.

In order to better determine how NE1434 and CADD proteins bind to substrate, a PqqC compound with product in the active site was found from *Klebsiella pneumoniae pneumoniae* (pdb file: 3HLX). Molecule PQQ is significantly different from pABA. However, given similarities in protein and product structure, seeing how PQQ binds to PqqC may illuminate important features of NE1434 binding. Though in NE1434, residues about the active site are less polar than in the PqqC protein.

N-oxygenase CmlI is capable of catalyzing a reaction with pABA as a substrate, so similarities between CmlI and NE1434 may reveal significant motifs and residues for substrate binding. Structural comparison between the two proteins revealed a shared four-helical scaffolding about the diiron center (Figure 8). As well, the ligand structure is similar in positioning and orientation (Figure 9). However, the internal cavity of CmlI is smaller than that of NE1434 given the more tightly packed bundle structure, so pABA may be less motile in CmlI than in NE1434, changing overall reaction catalysis.

Future Directions

In order to avoid issues that were found in the expression trials of NE1434, the gene *ne*1434 will be cloned into vector pET21-b (Figure 10) using a BamHI to HindIII restriction site scheme. This would give the added benefit of utilizing a T7 promoter, a promoter used for CADD

expression, and one that is more compatible for growth in BL-21 (DE3) cells. The primers that will be used are as follows:

5’-CGACGATGGGACGTGTTCGAATCTAGA-3’ and 5’-GAGAAAGGATCCATGGCCACCAACACC-3’

These proteins will then be expressed in BL-21 (DE3) competent cells and grown under a number of different conditions.

Currently, the substrate of CADD is unknown. However, given present knowledge of CADD products, docking calculations with pABA may be useful in determining likely position and orientation of the compound about the diiron cluster. Such a procedure would be completed using program AutoDock [27]. When using this program, a file for protein NE1434 will be loaded into a program along with a file containing a model of pABA. A location will be specified within the protein to determine its area of best fit. This will illuminate important residues for protein binding and may assist in locating substrate. As well, a comparison between the internal cavities of CADD and NE1434 to CmlI, AurF, and UndA. This may serve to illustrate further similarities between these proteins including cavity size relative to substrate and cavity residue composition. These may illuminate not only substrate character, but also possible structural recombination in the proteins. CmlI has an exceedingly small cavity with respect to its substrate. So, it may be that CmlI and CADD will shift structure in a similar manner.

One limiting factor found with the experimental design of the sequence alignment procedures was that the proteins used varied significantly, so much so that many of the sequences tested differed in function. Some of the proteins that were predicted to be evolutionarily similar to NE1434 were more akin to PqqC types than CADD. This likely skewed the conserved residue results and so the sequence alignment bears repeating with functionally similar proteins.

The function and importance of the aromatic residue chains conserved in these proteins is ripe for further elucidation. Chains of tryptophan and tyrosine residues have been found to transfer electrons into and out of proteins [28]. Given the importance of such a redox chain, it is probable that these conserved residues act as a conduit to transport electrons. Measuring the distances between aromatic residues would reveal more about the potential for these residues to carry electrons and relieve oxidative stress. Mutagenesis at these locations may also reveal the important impact that these residues have on the stability of these proteins.

The four-helix motif found in all of the proteins tested is prime for further study. Additional structural alignment and sampling from a broader array of proteins may reveal important specifics regarding this motif and its effects on reaction chemospecificity. Further research regarding internal cavity size and residue composition will add to our understanding of these proteins.

Figures & Tables

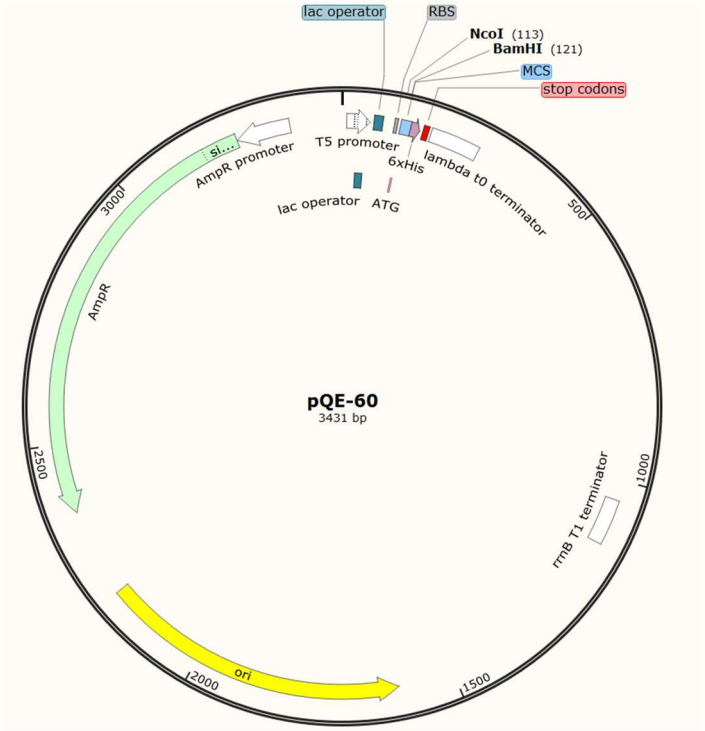


Figure 1: pQE60 vector with the NcoI and BamHI sites labeled. [10]

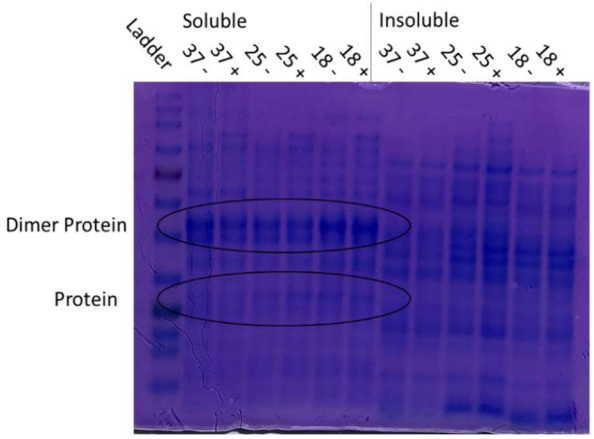


Figure 2: Temperature trials for NE1434 expression.

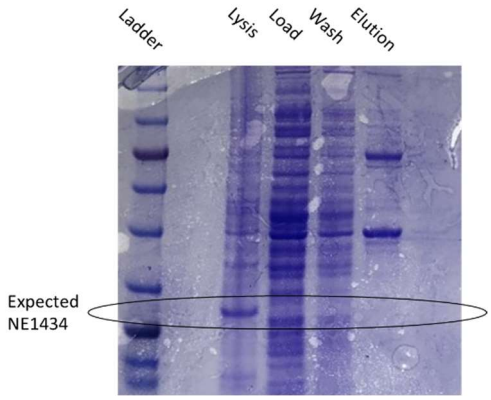


Figure 3: SDS-PAGE analysis of purification of NE1434. Note that a strong band at the expected size of NE1434 is present only for the lysis sample. This suggests that the NE1434 protein, if present, is kept, for some reason, insoluble.

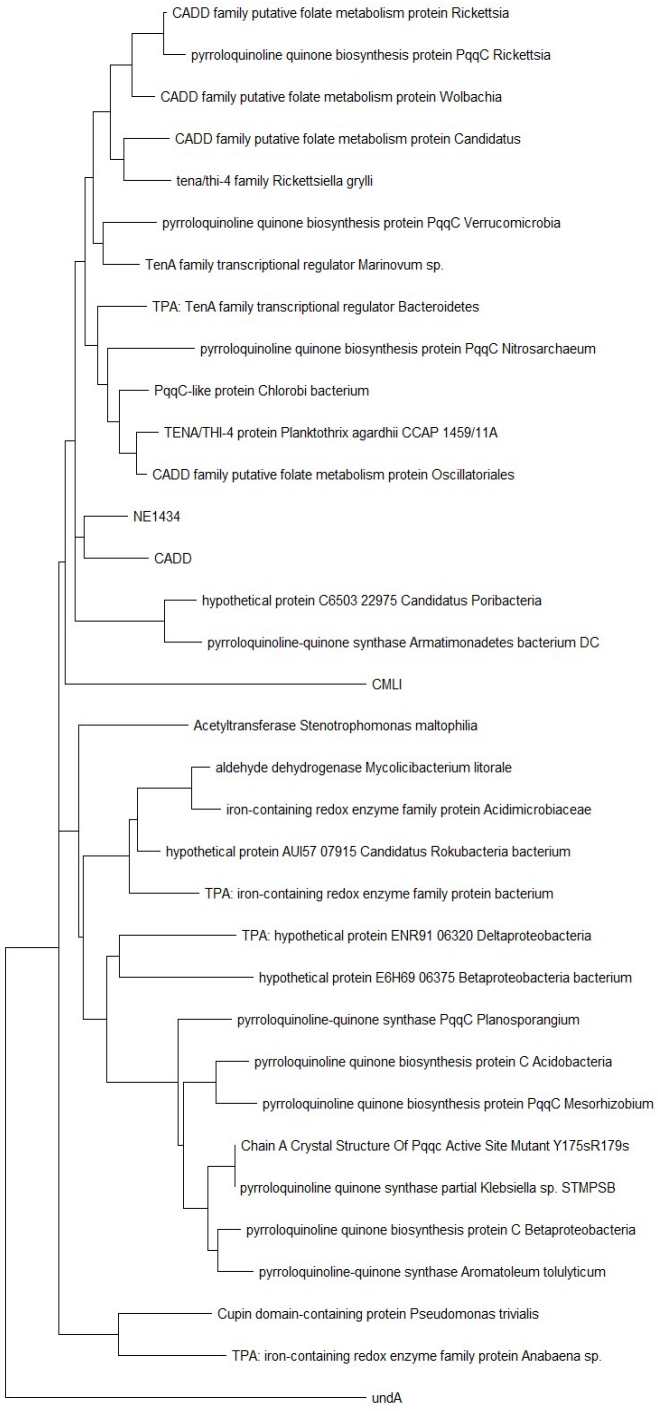
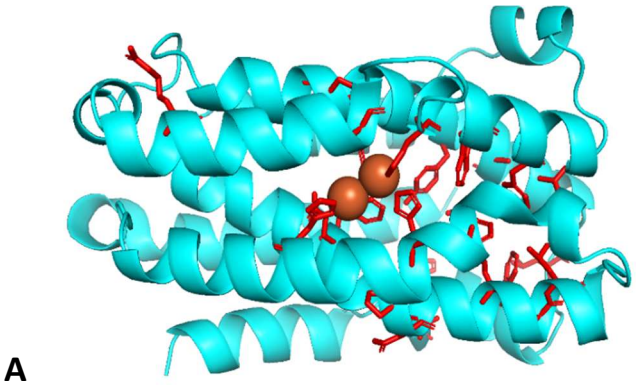
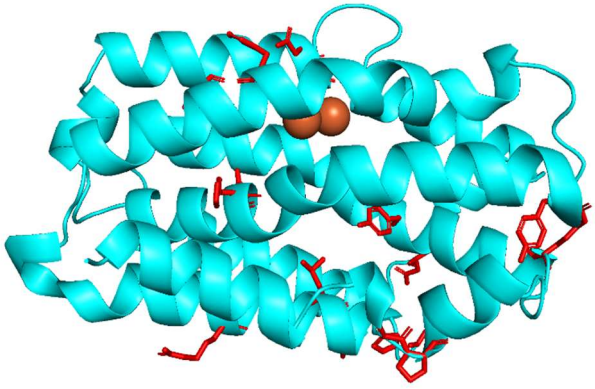


Figure 4: Cladogram of a variety of different proteins all with E values below E-value below 10^{-3} compared to NE1434.



A



B.

Figure 5: a) The red residues here are conserved between the nine closely related proteins and are displayed on the structure of protein NE1434. Note that most of the residues between the closely related proteins are ligands about the orange diiron center and between the alpha helices on the right of the image. b) These residues are conserved between NE1434, CmlI, and AurF and are scattered about the protein in a seemingly statistically random way.

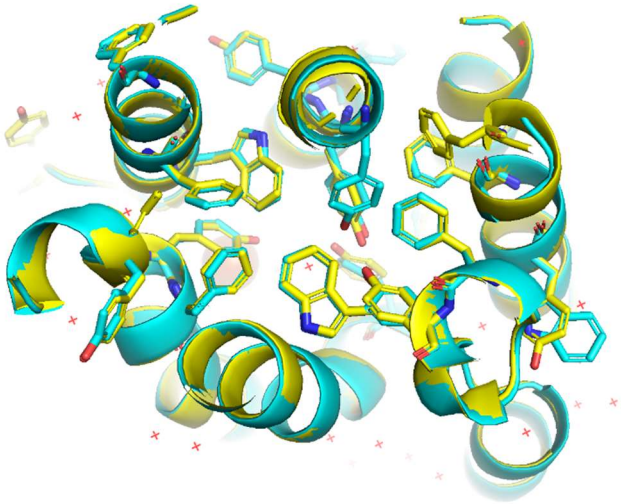


Figure 6: Structural comparison of Y, W, and F residues outside of the internal cavity in NE1434 (light blue) and CADD (yellow).

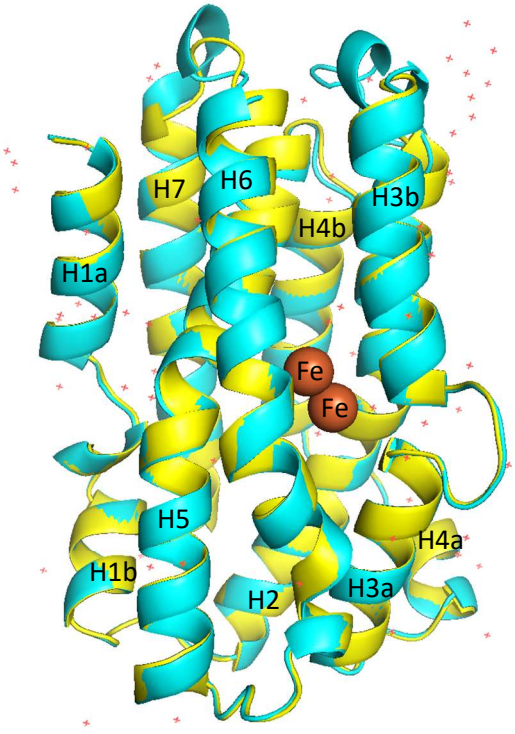


Figure 7: Structural alignment of CADD in yellow and NE1434 in blue. Helices H1-H7 are labeled along with the two irons within the protein.

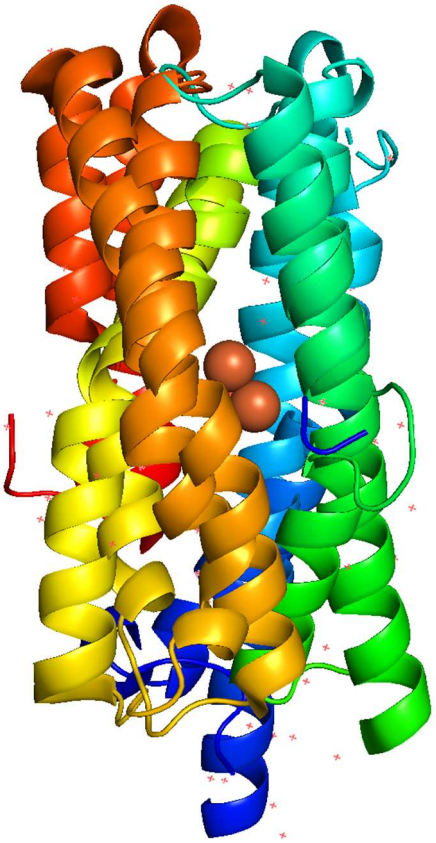


Figure 8: Alignment of NE1434 and CmlI four-helix bundles, the structures are colored for sake of differentiation between the helices.

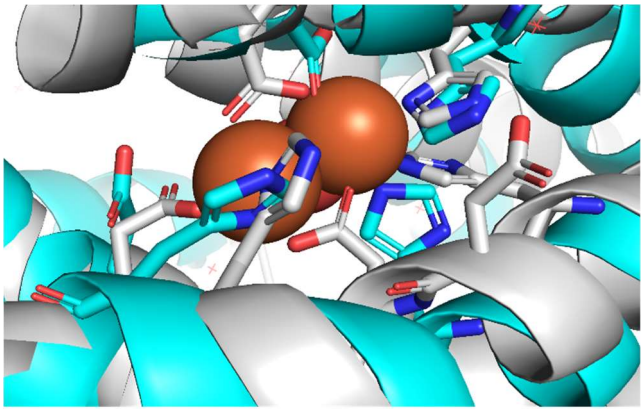


Figure 9: The ligand structure of NE1434 (blue) and CmlI (white).

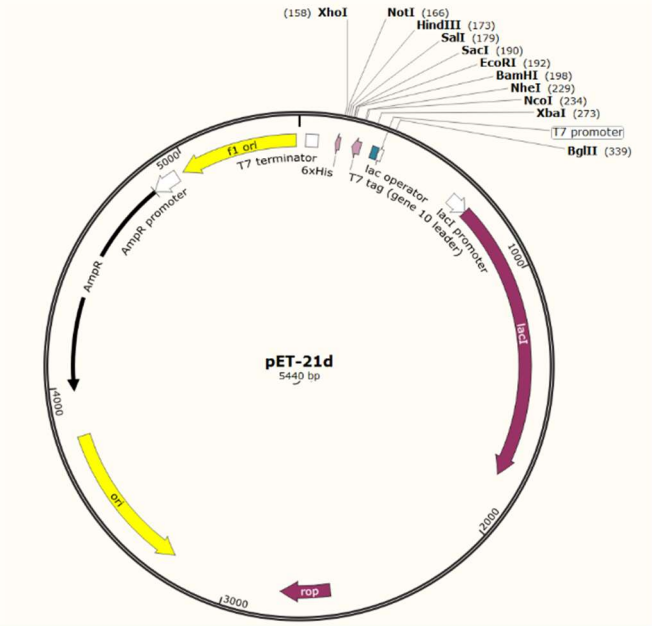


Figure 10: pET21 vector with BamHI and HindIII restriction sites labeled. [10]

Sources Cited

1. Whited, G. M., & Gibson, D. T. (1991). Toluene-4-monooxygenase, a three-component enzyme system that catalyzes the oxidation of toluene to p-cresol in Pseudomonas mendocina KR1. *Journal of Bacteriology*, *173*(9), 3010–3016. doi: 10.1128/jb.173.9.3010-3016.1991
2. Lipscomb, J. D. (1994). Biochemistry of the Soluble Methane Monooxygenase. *Annual Review of Microbiology*, *48*, 371–399. Retrieved from https://www.annualreviews.org/doi/pdf/10.1146/annurev.mi.48.100194.002103
3. Austin, R. N., Chang, H.-K., Zylstra, G. J., & Groves, J. T. (2000). The Non-Heme Diiron Alkane Monooxygenase of Pseudomonasoleovorans(AlkB) Hydroxylates via a Substrate Radical Intermediate. *Journal of the American Chemical Society*, *122*(47), 11747–11748. doi: 10.1021/ja001500v
4. Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Robinson, H., Bossy-Wetzel, E., Reed, J. C., & Liddington, R. J. (2004). Structure of the Chlamydia Protein CADD Reveals a Redox Enzyme That Modulates Host Cell Apoptosis. *The Journal of Biological Chemistry*, 29320–29324. doi: 10.1074/jbc.M401268200
5. Adams, N. E., Thiaville, J. J., Proestos, J., Juarez-Vazquez, A. L., McCoy, A. J., Barona-Gomez, F., … Maurelli, A. T. (2014). Promiscuous and Adaptable Enzymes Fill "Holes" in the Tetrahydrofolate Pathway in Chlamydia Species. *American Society of Microbiology*, *5*(4). doi: 10.1128/mBio.01378-14
6. Green, J. M., & Matthews, R. G. (2007). Folate Biosynthesis, Reduction, and Polyglutamylation and the Interconversion of Folate Derivatives. EcoSal. doi: 10.1128/ecosal.3.6.3.6

7. Manley, O. M.; Fan, R.; Lightfoot, V. C.; Guo, Y.; Makris, T. M. CADD, a dinuclear iron enzyme that forms an atypical peroxo-diferric species. *Journal of the American Chemical Society, in preparation*.
8. Kurtz, D. M. (1997). Structural similarity and functional diversity in diiron-oxo proteins. Journal of Biological Inorganic Chemistry, 2(2), 159–167. doi: 10.1007/s007750050120
9. Satoh, Y., Kuratsu, M., Kobayashi, D., & Dairi, T. (2014). New gene responsible for para-aminobenzoate biosynthesis. *Journal of Bioscience and Bioengineering*, *117*(2), 178–183. doi: 10.1016/j.jbiosc.2013.07.013
10. SnapGene software (from Insightful Science; available at snapgene.com)
11. National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2017 Apr 06]. Available from: https://www.ncbi.nlm.nih.gov/
12. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *BMC Bioinformatics*, *%=5*. doi: 10.1186/1471-2105-5-113
13. Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Nucleic Acids Research*, *32*(5). doi: 10.1186/1471-2105-5-113
14. Sudhir Kumar, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. Molecular Biology and Evolution 35:1547-1549
15. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
16. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T. SWISS-

MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296-W303 (2018).

17. Bienert, S., Waterhouse, A., de Beer, T.A.P., Tauriello, G., Studer, G., Bordoli, L., Schwede, T. The SWISS-MODEL Repository - new features and functionality. Nucleic Acids Res. 45, D313-D319 (2017).
18. Komor, A. J., Jasniewski, A. J., Que, L., & Lipscomb, J. D. (2018). Diiron monooxygenases in natural product biosynthesis. *Natural Product Reports*, *35*(7), 646–659. doi: 10.1039/c7np00061h
19. Brooke, J. S. (2012). Stenotrophomonas maltophilia: an Emerging Global Opportunistic Pathogen. *Clinical Microbiology Reviews*, *25*(1), 2–41. doi: 10.1128/cmr.00019-11
20. Perlman, S. J., Hunter, M. S., & Zchori-Fein, E. (2006). The emerging diversity of Rickettsia. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1598), 2097–2106. doi: 10.1098/rspb.2006.3541
21. Parola, P., Rovery, C., Rolain, J. M., Brouqui, P., Davoust, B., & Raoult, D. (2009). Rickettsia slovaca and R. raoultii in Tick-borne Rickettsioses. *Emerging Infectious Diseases*, *15*(7), 1105–1108. doi: 10.3201/eid1507.081449
22. Patel, G., Huprikar, S., Factor, S. H., Jenkins, S. G., & Calfee, D. P. (2008). Outcomes of Carbapenem-Resistant Klebsiella pneumoniae Infection and the Impact of Antimicrobial and Adjunctive Therapies. *Infection Control & Hospital Epidemiology*, *29*(12), 1099–1106. doi: 10.1086/592412
23. Schmitz-Esser, S., Haferkamp, I., Knab, S., Penz, T., Ast, M., Kohl, C., … Horn, M. (2008). Lawsonia intracellularis Contains a Gene Encoding a Functional Rickettsia-Like

ATP/ADP Translocase for Host Exploitation. *Journal of Bacteriology*, *190*(17), 5746–5752. doi: 10.1128/jb.00391-08

24. Chain, P., Lamerdin, J., Larimer, F., Regala, W., Lao, V., Land, M., … Arp, D. (2003). Complete Genome Sequence of the Ammonia-Oxidizing Bacterium and Obligate Chemolithoautotroph Nitrosomonas europaea. *Journal of Bacteriology*, *185*(9), 2759–2773. doi: 10.1128/jb.185.9.2759-2773.2003
25. Anjana, R., Vaishnavi, M. K., Sherlin, D., Kumar, S. P., Naveen, K., Kanth, P. S., & Sekar, K. (2012). Aromatic-aromatic interactions in structures of proteins and protein-DNA complexes: a study based on orientation and distance. *Bioinformation*, *8*(24), 1220–1224. doi: 10.6026/97320630081220
26. Korboukh, V. K., Li, N., Barr, E. W., Bollinger, J. M., & Krebs, C. (2009). A Long-Lived, Substrate-Hydroxylating Peroxodiiron(III/III) Intermediate in the Amine Oxygenase, AurF, from Streptomyces thioluteus. *Journal of the American Chemical Society*, *131*(38), 13608–13609. doi: 10.1021/ja9064969
27. Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. (2009) Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *J. Computational Chemistry* 2009, **16**: 2785-91.
28. Gray, H. B., & Winkler, J. R. (2015). Hole hopping through tyrosine/tryptophan chains protects proteins from oxidative damage. *Proceedings of the National Academy of Sciences*, *112*(35), 10920–10925. doi: 10.1073/pnas.1512704112