# Troxidoreductase, a protein of unknown function found in *Archaeoglobus fulgidus* shows EC 1 and EC 3 characteristics

Tate Rosenhagen and Dr. William Conrad
Lake Forest College
Lake Forest, Illinois 60045

**Abstract**
Troxidoreductase, a protein of unknown function found in *Archaeglobus fulgidus*, is commonly found in and believed to be responsible for the souring of oil fields via sulfur production. With a hypothesized EC 3 function we tested structure and function by using structural comparison (Pymol/Promol) and ligand binding analysis (Autodock) to examine similarities with proteins of known function and BLASTP and Pfam to search for families and genetic sequence similarities in the DNA. Results showed active site similarities between Troxidoreductase and some EC 3 proteins, but closes structural and ligand binding similarities with EC 1 proteins, leading to the hypothesis that Troxidoreductase is an Oxidoreductase that also has function of a Hydrolase, which could explain the souring of oil fields through the production of sulfur gas when water is present.

**Introduction**
Oil production is currently a vital industry, as many of the machines used every day around the world are powered by oil-based fuel. Thus, the efficient usage, but also efficient production, of oil cannot be understated, as an oil field turning sour can cost the producers and the customer money, as well endangering workers through the production of sulfur gas. Oil fields are an extreme environment in which few organisms can be found. Those surviving there are often considered extremophiles as they are also found exclusively in extreme conditions, such as on the ocean floor around hydrothermal vents. One such organism able to survive in these conditions is *Archaeoglobus fulgidus*, a sulfur-loving extremophile. This archaebacteria has been found in the depths of the biosphere, but also in hot deep oil reserves, and is hypothesized to cause the missing first step of hydrocarbon oxidization in oil fields, which then in turn leads to the souring of said oil field or reservoir (Khelifi et al. 2014). *Archaeglobus fulgidus* has been found to create a biofilm in response to stressors in the environment which may help it attach to metal in order to stimulate growth in a metal deficient environment, thus aiding in growth as it is a metal-reducing organism (Lapaglia and Hartzell 1997). As a hyper thermophilic sulfur reducing organism, there is reason to believe that *A. fulgidus* may be the first step of hydrocarbon oxidization in oil fields, which in turn leads to the souring of oil fields and reservoirs (Khelifi et al. 2014). The souring of oil fields leads to increased oil prices to the average consumer, as a high sulfide concentration can lead to rejection by the oil refiners, lowering the supply of oil in the market. It also produces hydrosulfide gas, an extremely toxic gas that lowers air quality and must be monitored during the oil mining process. Furthermore, it can corrode the infrastructure of an oil mining machine and plug the reservoir through the precipitation of iron sulfide, all increasing the cost of the production of oil (Gieg, Jack, & Foght 2011). In an effort to examine this possible reaction, we have chosen to examine hypothetical binding site Troxidoreductase on hypothetical protein AF_1432. AF_1432 is part of the HD superfamily of metal dependent phosphorylases and a hypothetical protein believed to be found in *Archaeglobus fulgidus* (Aravind and Koonin 1998). HD domains have been found to play a role in antiviral defense in the human protein SAMHD1 (Beloglazova et al. 2013). The SAMHD1 protein was found to limit replication of the HIV-1 genome, but not HIV-2, and the purified HD domain from humans and mice contained dGTP-stimulated tryptophohydralayse activity (Beloglazova et al. 2013). Troxidoreductase's enzyme consortium number is 3.1.3.89 where each number separated by periods signifies something about the protein. The 3. Signifies the enzyme is a hydrolase, and the 1 signifies it is an esterase, so it acts on ester bonds. 3.1.3 signifies phosphoric-monoester hydrolases, and the 3.1.3.89 specifies that Troxidoreductase shows specificity towards deoxyribonucleoside 5'-monophosphates ("ENZYME - 3.1.3.89 5'-Deoxynucleotidase.") From this information, there is reason to believe the *Archaeglobus fulgidus* may cause the vital missing step of the souring of oil fields. Due to the organism of origin, as well as the EC number, we hypothesize that Troxidoreductase will have characteristics of a metal dependent phosphorylase due to its origin in *A. fulgidus* but perform some sort of role of a hydrolase due to its EC 3 classification.

**Methods:**

**Pymol/Promol Search**
In this lab we used a VPN to access Lake Forest College's programs remotely via a remote desktop. This was necessary in order to use Pymol and Promol, as both of these programs have strict licenses and can only be run on Lake Forest's network. Once Pymol and Promol were opened, motif finder in Promol was selected and our two proteins, 2O14 and the unknown, were searched separately. The search can be narrowed by selecting different libraries to search your protein in, as well as limiting the search by EC number if your EC number is known. 2O14 has an EC number of 3, so 2O14's search was limited in terms of EC 3 for time's sake. Searches typically take 5-10 minutes. Pymol then shows a search result box, as well as an image of your searched protein, which can be compared against other proteins that have come up in your search. We then checked the show alignment box and calculate RMSD box and compared our searched protein with every protein result that had a Levenshtein score of 0. From there 3 RMSD values were calculated in the Pymol molecular graphics system box, the first being the RMSD of all atoms, the second being just alpha carbons, and the third being alpha and beta carbons. The RMSD of alpha and beta carbons will be of particular note to us this lab. Finally, after choosing the three matches with the lowest RMSD value (alpha and beta carbon) were selected to compare residue alignments. This was performed by clicking on the structures that lined up together, and the amino acid residues that matched would show up in the pymol graphic box. Alternatively, by clicking the small s in the pymol viewer window, and then clicking the structures that overlap again. The AA sequences will be displayed at the top of the viewer, with the overlap of the structure being highlighted in red.

**BLASTP Search**
First, the amino acid sequences for our unknown protein and 2O14 from rcsb.org. The sequences were run through the BLASTP search and the amino acids were broken into 3 amino acid 'words' compared to other protein sequences of known functions. Other proteins were sorted by number and percentage of alignments, and thus the proteins most similar to our proteins could be seen and their functions could be examined in hopes of understanding our proteins functions. A SMARTBLAST was also performed to see the organisms in which our proteins are found. Finally, a Pfam search was performed on our proteins amino acid sequences to determine the superfamilies and domains found within our amino acid sequence. The Pfam search was largely supplemental to the BLASTP search in our findings.

**Dali Search**
In order to compare our proteins on Dali, we searched proteins by name (2o14 and Troxidoreductase) on Dali, which found similar structures without taking into account their side chains in order to save space. The search took a few hours, and we then examined one protein from each molecule to amount to a total of three similar proteins and examined their amino acid structures and their 3D cartoon structure comparisons. We then used the active sites found in the previous lab, found their location in the amino acid sequence, and created a five-letter amino acid code by choosing the five amino acids before our active site. This was necessary as Dali and RCSB amino acid number does not line up exactly. We then searched this amino acid sequence in order to find our active sites location in the Dali search and checked to see if the similar proteins had the same active sites. This may give insight into the function of our protein.

**Autodock search**
To begin with, PyRx was loaded in order to run an autodock test on our proteins of choice. We then went to RCSB, downloaded the PDB file of our protein, and uploaded it into PyRx. This molecule was then converted to a macromolecule. After that, we went back to RCSB to search for ligands of interest in the same EC class. These ligands ideal SDF data files were downloaded and many had to be edited from a txt to SDF file. Once 4-6 ligands were downloaded and translated to SDF files, they were inserted into PyRX via the import chemical table file – sdf function. After they were all imported, they were right clicked to minimize them all, and right clicked again to convert all to autodock ligand. We then used Vina Wizard to run the docking function, whereas we clicked all the ligands and our one protein of interest (macromolecule). The processing time took a few minutes, and we then sorted by affinity to find our best binding site match, showing both the ligand and the specific conformation of the ligand. We downloaded the list

by saving a CSV file which can be used in Excel later, and then saved the whole workspace as a tar.gz file. Finally, we downloaded our best match, macromolecule and ligand, again as a PDB and our best match ligand as a PDB file. We then uploaded these files into Pymol for a better visualization of the binding site of our ligand and macromolecule. By selecting the show surface of the macromolecule and highlighting the binding sites found in the previous Pymol lab, we can change them a different color. Any errors and issues were resolved with assistance from Dr. Conrad.

**Protein Purification**

First, we bound our protein of interest with the binding solution, which bound specifically to our POI. The binding buffer was 50 mM Na2, HPO4, 300mM NaCl, and pH 8.0, supplemented with 10mM imidazole. This binding buffer was chosen to reduce non-specific interactions between our protein of interest and other proteins, thus ensuring our binding solution only binds to our POI. Then we used a wash solution to elute any proteins not bound to our binding solutions, thus leaving us with hopefully pure protein bound in our column. Finally, we used an elution solution to elute our bound POI, which contained an excess of imidazole to force our protein out of the binding column. We expected to see our protein in the fractions after the lysate, and in the elution fractions with the last elution fraction having no protein, meaning we got it all out. We collected every fraction to ensure we did not have our protein elute too early or have other proteins present in our wash solution.

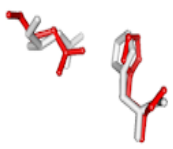
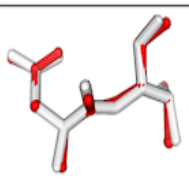
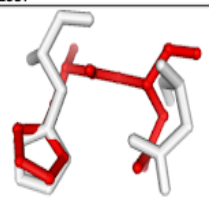
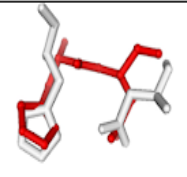
**Protein Kinetics (2o14 and YfkN)**

3GVE (YfkN) elution 1 was dialyzed into 50 mM sodium phosphate buffer (pH 7.6). This was done by using the Pierce 3 kDa MWCO centrifugal filter concentrator and spin-concentrated the samples 3x using 20 ml of buffer per 1 ml of elution. The protein abundance was measured using absorbance at 280 nm with a path length of 1 cm. The beer's law equation was used to calculate the concentration of the sample in molarity, and then converted to mg / ml. Using a 96 well plate, the following reagents were added with varying volume: 160 ul of 50 mM sodium phosphate buffer (pH 7.4) and X uL of 10 mM para nitrophenol acetate in acetonitrile. The baseline absorbance at 405 nm wavelength for each well was measured. 20 ul of enzyme was added to each well, and then proceeded to read A405 across the entire 96 well plate. The plate was read every 41 seconds in order to observe the change in absorbance over time in each well. This was repeated for 2o14 as well.

**Figures:**

**Pymol and Promol shows EC 1 and EC 3 structural characteristics**

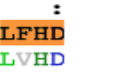
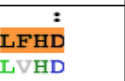
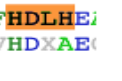
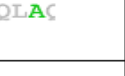
*Table 1. Table containing three best alignments from a Pymol/Promol search with RMSD values for alpha and beta carbons. Troxidoreductase and 2O14 are shown in red, protein being compared shown in grey.*

| Alignment Ranking | 2O14 (red) w/____ (grey) | 2O14 RMSD (alpha and beta carbons) | 1YOY (red) w/____ (grey) | 1YOY RMSD (alpha and beta carbons) |
|---|---|---|---|---|
| 1 |  1BWR | 0.712 |  1S3I | 0.65 |
| 2 |  1A30 | 0.616 |  1C9U | 0.93 |
| 3 |  2ACY | 1.046 |  1BG6 | 1.13 |

**Dali search finds best match as another hypothetical protein in the same archaebacteria**

Troxidoreductase

*Table 2. Table of Dali search results and comparisons of Troxidoreductase (green) with known proteins (other colors). 3 best matches, each from a different organism, was chosen for comparison.*

| Chain/Description | 4l1j-a HD DOMAIN-CONTAINING PROTEIN 2 | 1xx7-a OXETANOCIN-LIKE PROTEIN | 2par-b 5'-DEOXYNUCLEOTIDASE YFBR |
|---|---|---|---|
| Image Capture (Cartoon) |  |  |  |
| AS: His 76 (All residues off by two on fasta) LFHD |  Yes |  Yes |  Yes |
| AS: Asp 77 LFHD |  Yes |  Yes |  Yes |
| AS: Asp 126 DAD |  No Match |  Yes |  Yes |
| AS: Glu 80 HDLH |  Yes |  Yes |  No |
| AS: Asp 124 DAD |  No |  No |  No |
| # of active site match | 3/5 | 4/5 | 3/5 |
| Good Fit? | No | Yes | No |
| Z- Score | 14.9 | 14.4 | 13.9 |
| RMSD | 2.2 | 2.3 | 2.6 |
| LALI | 138 | 132 | 136 |
| NRES | 188 | 172 | 178 |
| %ID | 23 | 33 | 21 |

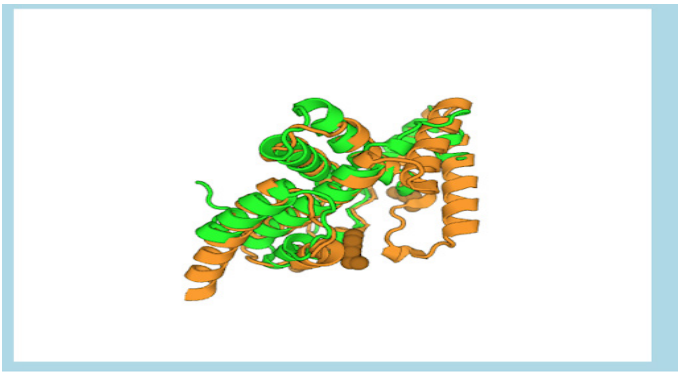


*Figure 1. Dali structural comparison of Troxidoreductase (green) and 4L1J (orange). Structure is in cartoon format.*

**BLASTP search results point to HD domain presence**

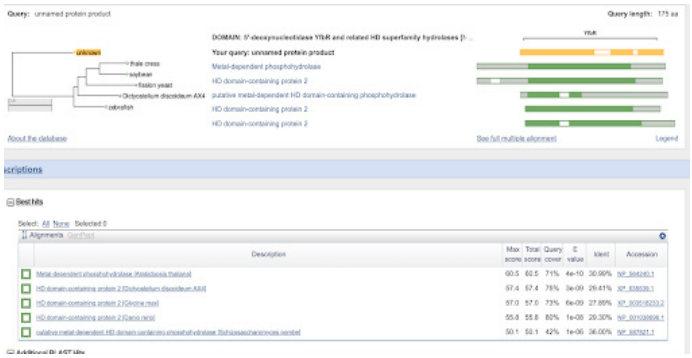


*Figure 2. Phylogenic tree and family description from BLASTP search results of Troxidoreductase.*

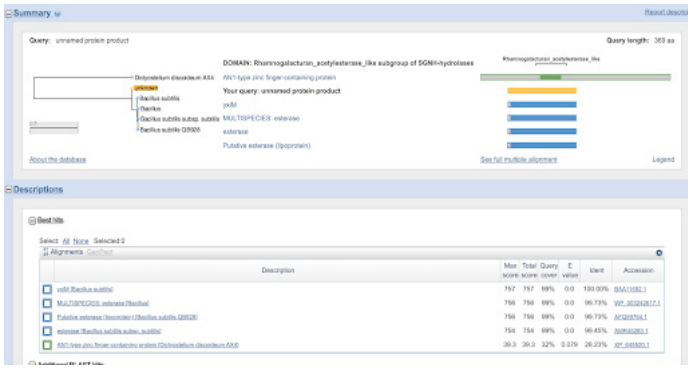


*Figure 3. Phylogenetic tree for YXIM based on known protein sequences in organisms of origin. Here we see a large grouping of Bacillus species, suggesting a highly conserved function within Bacillus.*

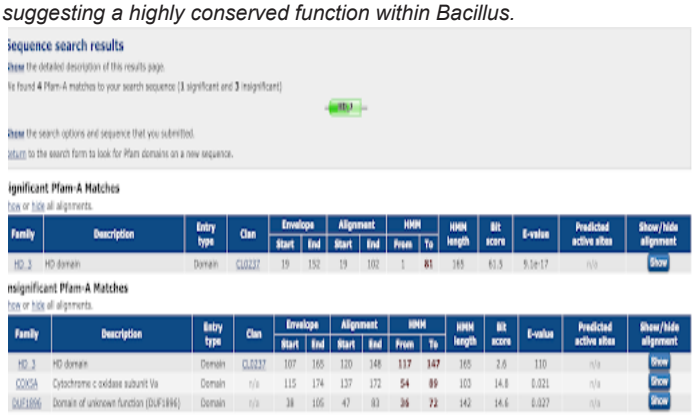


*Figure 4. Family descriptions of Troxidoreductase from BLASTP search when compared with domains and families of known function.*

**Autodock results show lack of binding at hypothesized EC 3 binding sites, leading to EC 1 function hypothesis**

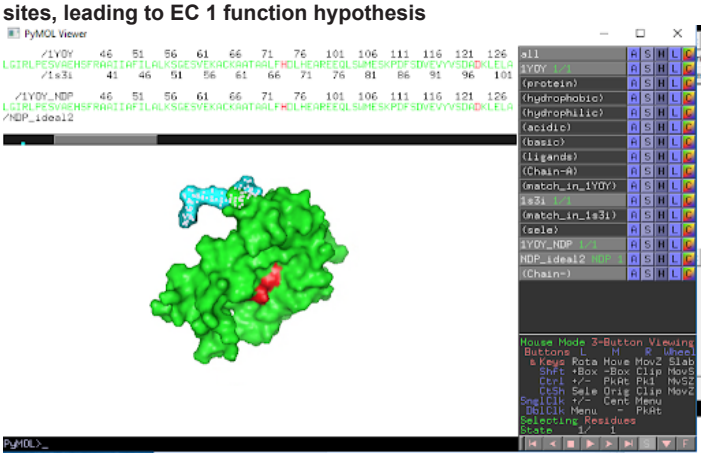


*Figure 5. Pymol visualization of the Troxidoreductase and a ligand (NBD). AF protein Troxidoreductase is seen in green, previous binding sites seen in red, and NDB ligand shown in blue. The ligand does not bind to the hypothesized active site, suggesting another possible binding site and function.*

**Protein kinetics shows lack of enzymatic activity in unknown protein and unusable data with 2o14.**

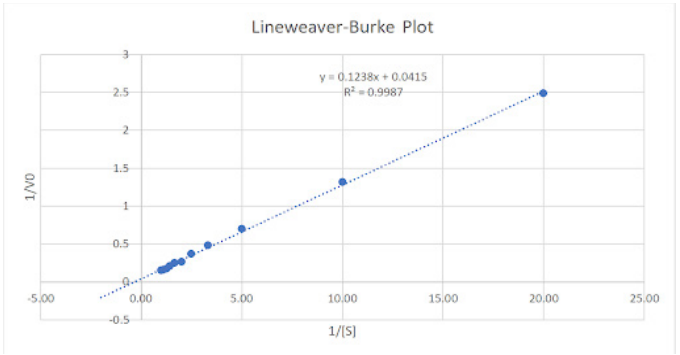



*Figure 6. Line-Weaver Burke plot shows unusable data for 2014 elution 1. Line of best fit has a y-intercept of +0.0415, showing that we cannot use this data to draw conclusions of our protein's functionality.*

**Results:**

**Pymol/Promol search shows structural alignment with both EC 1 and EC 3 proteins.**

After the initial research on RCSB, we examined Troxidoreductase in a Pymol/Promol search to compare structural similarities with proteins of known function. We did this as a first step to confirm that our hypothesized function of a hydrolase, due to its EC 3 classification, was correct. However, contradictory to our initial EC 3 classification, the three best matches were all proteins with an EC 1 classification. Initially, we believed we may have run this test incorrectly, but after running it with our positive control YXIM, we confirmed we were running our test correctly and had fairly good RMSD values as see in Table 1. After confirming the test was run correctly, we did more research on the best fit for AF_1432's Promol search and found that 1S3I, while classified as an EC1, had characteristics of an EC 3 as well, as it has functions of an oxidoreductase and a hydrolase. With this context, it does make sense that Troxidoreductase could show characteristics of an EC1 and EC3 type enzyme.

**BLASTP search shows presence of conserved HD domain and few similarities with proteins of known function in sample organisms.**

In an effort to add further details to explain the function of our protein, a BLASTPP search was performed on our protein of interest and YXIM, as a positive control. We hypothesized that our unknown protein would have domains of a hydrolase family and possibly an oxidoreductase due to the EC 1 classification in the Pymol/Promol lab. In Figure 2 we found that the phylogenetic tree did not produce good results, as can be seen with the highest identity score being a protein in Zebra fish with 30% match, followed by matches in Soybean and Thale Cress among other organisms. While a wide variety of organisms who point to a highly conserved mechanisms, the identity percentage of 30% at its highest instead points to a highly specific mechanism, as no other organisms in this databank had this mechanism. To ensure this was not simply an error run in the test, we ran the test with YXIM as well, and found a 100% match with 2014 in *Bacillus subtillis*, showing the test was run correctly (Figure 3). However, the most important result of this test was in Figure 4, where the HD domain was found to be present in Troxidoreductase. The HD domain is a known phosphohydrolase, which again helps us in hypothesizing that it is an EC 3 classification.

**Dali search results are inconclusive as the only protein of good match was itself.**

While the previous tests provide context as to the function of Troxidoreductase, finding the listed domains and structural similarities, we also decided to use a Dali search to look for other matches found through Dali's global alignment and ignorance of side chains. We hypothesized that the best match would again be a phosphohydrolase and, unfortunately, the best fit for our unknown protein was another hypothetical of protein of unknown function, 1xx7, but the hypothesized function was a metal dependent phosphorylase, which is what we believe our unknown protein to be (Table 2). The

next match with a protein of known function, which was a poor match to 4l1j, does provide some more context as it is a hydrolase found in homo sapiens, but it only had 3 out of 5 active site matches and a poor percent ID. Thus, we can continue with our hypothesis that our unknown protein is a metal dependent phosphorylase that may also some features of an oxidoreductase.

**Autodock lab shows best binding ligands are ligands of 1xx7, an EC 1 protein that was most similar to Troxidoreductase in the Pymol/Promol lab.**

Proceeding with our experiments, we performed an autodock test to examine ligand binding and compare then to the ligand binding sites hypothesized in the Pymol/Promol lab. In Figure 5 we see our protein Troxidoreductase in green, the proposed EC 3 binding site in red, and the ligand binding in blue. NBD, the ligand of best match, was found to be the best fit in our protein, and coincidentally is not an EC 3 ligand, but an EC 1 ligand from the protein of closest resemblance to Troxidoreductase, 1xx7, an EC 1 protein. This result lends itself to believing that Troxidoreductase has some features and an active site similar to a hydrolase, but the main active site, or the one who's ligands matched better in this search, was an oxidoreductase active site. Similar to its EC 1 counterpart, 1xx7, Troxidoreductase may help or perform two different reactions, but we can confidently say that a ligand of an EC 1 protein binds best to Troxidoreductase, thus suggesting EC 1 function.

**Overall experiment is inconclusive, shows characteristics of EC 1 and 3, leading to hypothesis that Troxidoreductase is a water dependent oxidoreductase.**

While we cannot make concrete conclusions from our data, we have been able to make some progress in understanding the function(s) of Troxidoreductase in these experiments. While not fitting simply into one EC class, Troxidoreductase appears to have active sites of an EC 3 protein as well as ligand binding affinity for EC 1 ligands (DMB).

**Protein kinetics lab shows a reactive control enzyme but unusable data.**

While we were unable, due to covid restrictions, to complete the purification and kinetics of our given protein, we were, as a class, able to complete the purifications and kinetic tests with 2o14, our class control. For our 2o14 sample, we saw enzymatic increase over time and an increase of enzymatic activity with an increase in concentration of the protein. However, as seen in Figure 6, the line-weaver burke plot had a positive Y intercept, meaning that the data cannot be used. In regard to our data, we would expect a very similar line weaver plot for our data to be successful, except for a negative Y intercept. From there we would be able to interpret their Km and Vmax.

**Discussion:**

While we cannot make concrete conclusions of our protein's functions based on our experiments, we can make a hypothesis based on the sum of our experiments. As previously stated, Troxidoreductase showed characteristics of EC 1 and EC 3 functions, with an HD domain similar to a hydrolase but structural similarities with an oxidoreductase. Upon further research, the EC 1 protein of best fit, 1S3I, which has a PDB classification of a hydrolase and oxidoreductase performs a catalytic reaction ending with the creation of carbon dioxide ((Bank, n.d.)). With this information and the data from our experiments, several assumptions can be made about the mechanism and role of Troxidoreductase. From our binding site comparison and our Dali search, we believe that the binding site seen in Figure 5 is used for a phosphohydrolase mechanism as the binding site falls within the HD domain. However, the ligand binding to the protein in Figure 5 is not binding to the active site highlighted, as this ligand of best fit is from an EC1 protein. This seemingly nullified our hypothetical phosphohydrolase mechanism, which was based on the EC3 hydrolase classification, and pointed towards a different mechanism all together. With Troxidoreductases similarities to 1S3I, and the ligand 1xx7 binding on a different region to our proposed active site, we also believe that our protein performs an oxidation reduction reaction along with its hydrolase function. This combination of reactions would make sense as our protein is believed to cause the souring of oil fields by producing sulfur gas, most likely by an oxidation reduction reaction as *Archaeglobus* bacteria are known to be sulfate reducing bacteria, and these oilfields tend to be soured with the addition of water, suggesting the hydrolase reaction may play some part in this reaction. However, we cannot make these statements with much confidence due to the fact that many of our lab results were not in agreement with each other and we were unable to take this specific protein, Troxidoreductase, to the completion of the lab, and thus are unable to propose a plausible mechanism at this time.

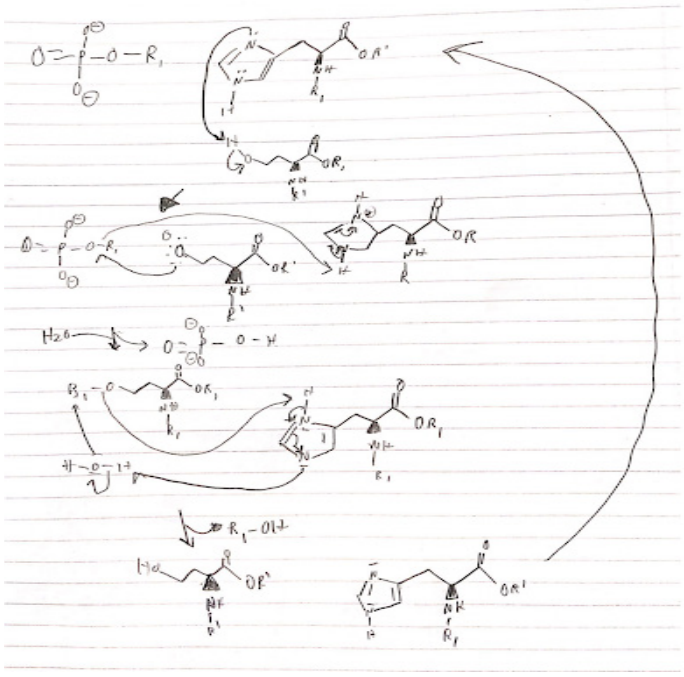


*Figure 7. Hypothesized EC 3 mechanism for Troxidoreductase involving the Histidine 76 and Aspartic Acid 124 residues. This mechanism was not found in our experimental trials; thus, we cannot say that this mechanism is present. The mechanism found seems to be of EC 1 nature and more tests are required to find the correct mechanistic steps. Authors Note: Mechanism drawn by hand as Chemdraw failed to work.*

If we were able to continue this lab, or study Troxidoreductase in the future, one possible experiment would be to test varying concentrations of our proteins in oil, both with and without water, and monitor each solution over time for production of sulfur, similar to our temperatures (83 degrees Celcius) to mimic the conditions within an oil field, as they are usually at high temperature and pressure (Beeder et al. 1994). If no wells had any production of sulfur, Troxidoreductase would not be the missing link in the oil souring chain of bacteria. If sulfur was produced in both the oil and oil + water wells, Troxidoreductase's oxidation-reduction reaction would not be water dependent, and if sulfur was produced in only the oil + water wells, then we would know the hydrolase reaction is necessary. It is our hypothesis that only the wells containing oil and water will produce sulfur, as in previous research oil fields were found to only sour with the addition of water, and that 1S3I requires the hydrolase reaction to produce carbon dioxide as previously stated. The next continuation of our future experiment would be to determine if inhibition of the HD domain found in the EC 3 active site would inhibit sulfur production from Troxidoreductase's reaction. This would have great significance on the oil industry, as the inhibition of sulfur production in an oil field would protect oil workers and lower the price of oil due to a decrease in unusable oil, or oil that needs to be refined even further, making it more costly. To do this, we would use the same experimental set up as in the previous experiment, but only with oil and water, and with the addition of varying amounts of six select SAMHD1 inhibiting compounds from the research paper *Identification of Inhibitors of the dNTP Triphosphohydrolase SAMHD1 Using a Novel and Direct High-Throughput Assay* (Mauney et al. 2018). The wells would contain increasing concentrations of each compound and monitored for production of sulfur gases. We expect if a compound successfully inhibits the HD domain in the active site, no sulfur would be produced in those wells. If a compound or compounds were successful in inhibiting the production of sulfur, we could then move on to studying the compounds effects on crude oil and seeing if the compound is a viable product for the oil industry to treat or prevent souring of oil fields. If none of the wells show inhibition of sulfur production, either the compounds did not successfully block the HD domain of Troxidoreductase, or the HD domain is not needed for the production of sulfur. In either case, more studies on Troxidoreductase are needed to determine what needs to be inhibited to limit the production of sulfur and if this organism is in fact the missing link in the souring of oil fields.