Crystal Structure of 4-Hydroxybutyryl-CoA Synthetase (ADP-forming): A Key Enzyme in the Thaumarchaeal Hydroxypropionate/Hydroxybutyrate cycle

Jerome Johnson[1], Bradley B. Tolar[2,3], Bilge Tosun[1], Yasuo Yoshikuni[5], Christopher A. Francis[2], Soichi Wakatsuki[6,7,*], Hasan DeMirci[1,4,*]

The 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle from ammonia-oxidizing *Thaumarchaeota* is currently considered the most energy-efficient aerobic carbon fixation pathway. The *Nitrosopumilus maritimus* 4-hydroxybutyryl-CoA synthetase (ADP-forming) (Nmar_0206) from this cycle represents one of a number of proteins that exhibit increased efficiency over its crenarchaeal counterparts. Nmar_0206 catalyzes the conversion of 4-hydroxybutyrate (4HB) and Coenzyme-A (CoA) to 4-hydroxybutyryl-CoA through the dephosphorylation of ATP to ADP, as opposed to AMP as seen in *Crenarchaeota*. This enzyme reduces energy requirements on the cell, thus reflecting thaumarchaeal success in adapting to low-nutrient environments.

The first high-resolution structure of Nmar_0206 from *Nitrosopumilus maritimus* SCM1 reveals a highly conserved interdomain stabilizing loop between the CoA-binding and ATP-grasp domains. Phylogenetic analysis suggests the widespread prevalence of this loop and highlights its underrepresentation within the PDB and its structural importance within the ATP-forming Acetyl-CoA superfamily (ACD).

*1. MBGE Department, Koç University, Istanbul 34450 Turkey. 2. Department of Earth System Science, Stanford University, Stanford, CA 94305. 3. Department of Biology and Marine Biology, University of North Carolina Wilmington, Wilmington, NC 28403. 4. Stanford PULSE Institute, SLAC, Menlo Park, CA 94110, USA. 5. The US DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. 6. Department of Structural Biology, Stanford University, Stanford, CA, 94305, USA 7. Biosciences Division, SLAC National Accelerator Laboratory, Menlo Park, CA, 94025, USA *Correspondence and requests for materials should be addressed to S.W. and H.D. (email: soichi.wakatsuki@stanford.edu & hdemirci@ku.edu.tr).*

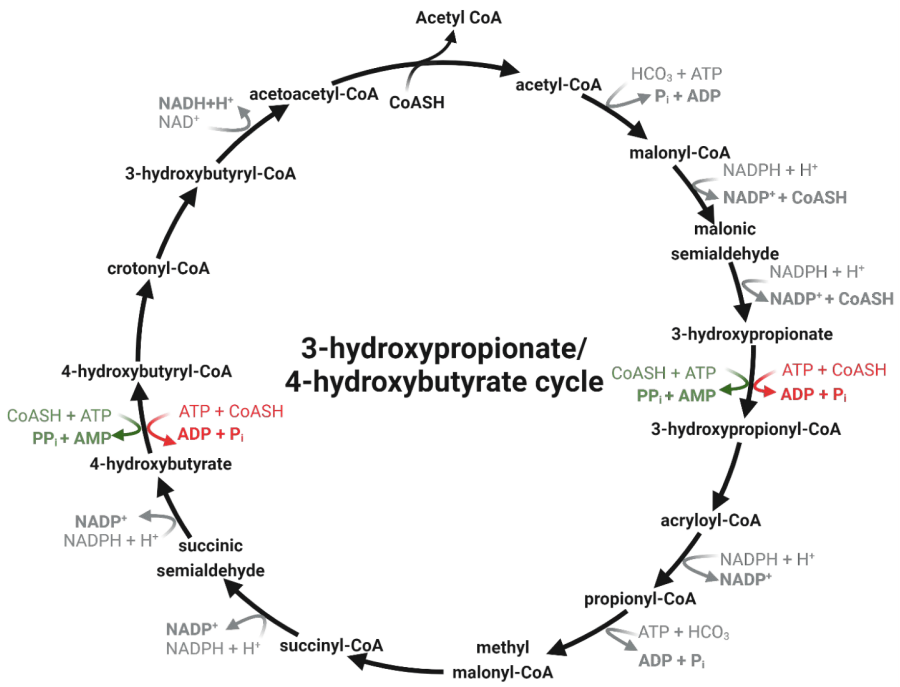

Figure 1 - 3HP/4HB cycle. Green and red arrows indicate differences between the Acetyl-CoA Synthetases (ACDs) of the 3HP/4HB cycles in *Crenarchaeota* and *Thaumarchaeota* respectively. Both of these reactions are catalyzed by 3-hydroxypropionyl-CoA Synthetase and 4-hydroxybutyryl-CoA Synthetase, but *Crenarchaeota* enzymes are AMP-forming while *Thaumarchaeota* enzymes are ADP-forming.

Introduction

Despite growing awareness of global climate change, carbon dioxide emissions are still increasing[1]. In order to address this critical issue, engineered autotrophic carbon-fixation cycles are being developed from a broad array of natural sources as a potential strategy to reduce atmospheric carbon[2,3]. The 3HP/4HB cycle of Thaumarchaeota is the most efficient aerobic carbon fixation cycle and may be responsible for 1% of global carbon fixation[4]. The mechanisms for that increased efficiency were attributed to the unique energy requirements of the enzymes. 4-hydroxybutyryl-CoA (Nmar_0206) and 3-hydroxypropionyl-CoA (Nmar_1309) synthetases from *Nitrosopumilus maritimus* were both shown to be ADP-forming, as opposed to the AMP-forming alternatives present in the crenarchaeal 3HP/4HB cycle[5]. To better understand the

structural elements involved in the catalysis, we have determined and analyzed the crystal structure of Nmar_0206. This may present important scientific and engineering opportunities for improved carbon fixation[5,6].

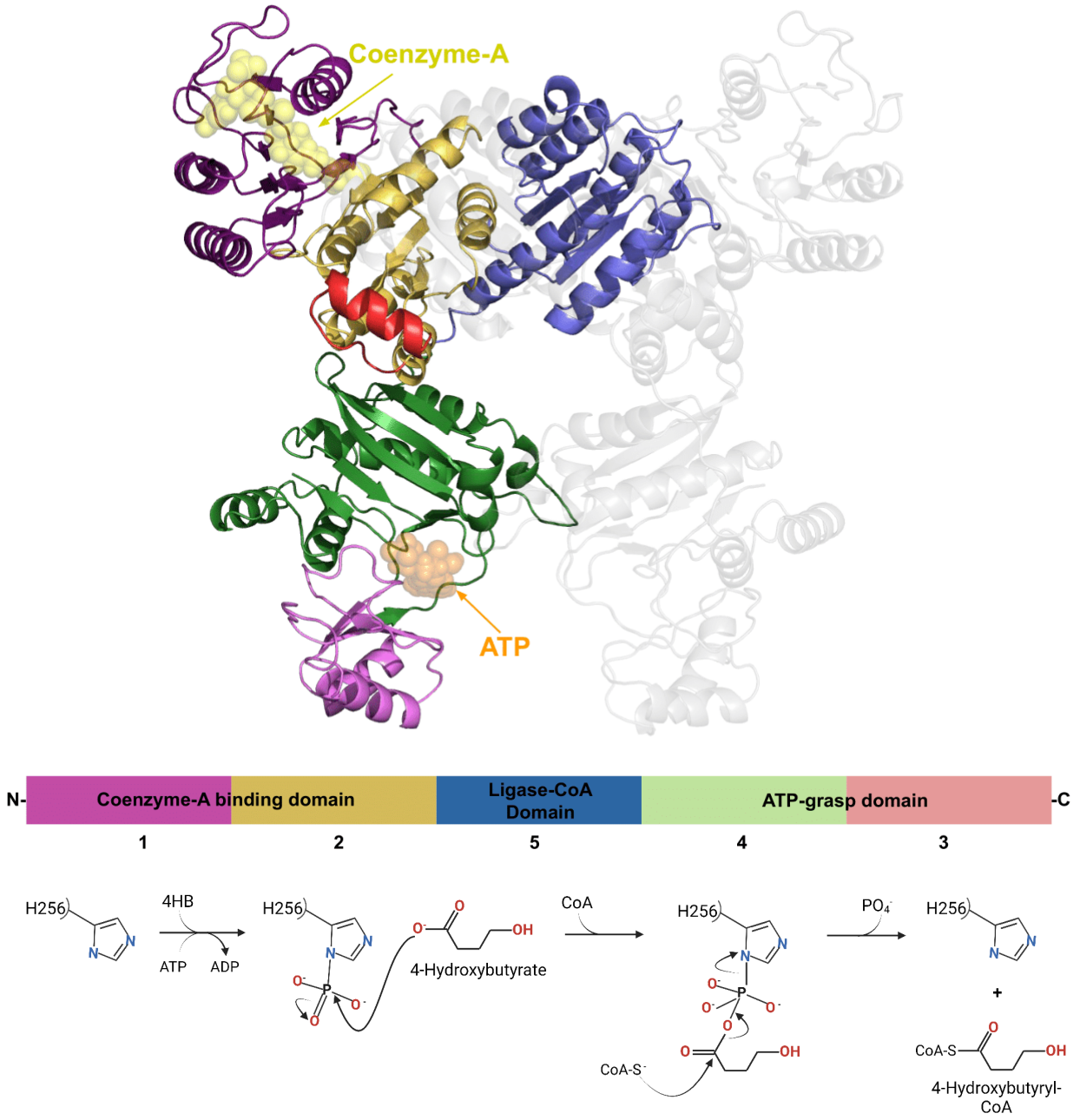

Figure 2 - Crystal structure of Apo-Nmar_0206. The 5 subdomains of 4-Hydroxybutyryl-CoA are differentiated by color. Subdomains 1 and 2 (pink and yellow), support Coenzyme-A binding while 5 (blue), a ligase-CoA domain, supports the ligation of the 4-hydroxybutyryl group with Coenzyme-A. The ATP-grasp domain contains subdomains

4 and 3 (green and salmon) with domain 3 functioning as a 'lid' to close over a bound ATP. The functionally important 'swinging loop' is highlighted red as well. These domains are shown following ecSCS format [1-2-5-4-3], which can be compared with the ecSCS [alpha(1,2)/beta(3,4,5)] where alpha and beta indicate separate chains in that structure. The shuffling of subdomains is a definitive characteristic of the ACD superfamily. The second monomer is transparently overlaid onto the first. Transparent yellow and orange spheres suggest CoA and ATP-binding sites respectively. The lower reaction steps show intermediate phosphohistidine and formation of 4-hydroxybutyryl-phosphohistidine (near sulfinyl tail of 4HB at the interface between domains 1, 2 and 5) before final product formation is shown as part of a simplified reaction mechanism.

Nmar_0206 is a member of the NDP-forming Acetyl-CoA synthetase (ACDs) family which catalyzes the formation of CoA thioesters through the dephosphorylation of ATP to ADP[7,8]. Its structure is a homodimer composed of 5 shuffled subdomains common among ACDs: ADP-grasp and lid domains, 2 CoA-binding domains, and a CoA-ligase domain. The structure confirms elements expected from a member of the ACD superfamily, including domains, loops and active site residues reported as necessary to catalyze the acetylation of CoA. Furthermore, a previously described 'swinging loop' is present, as well as associated power helices, which plays a role in loop stabilization[9,10]. Although the Nmar_0206 structure lacks any of the standard interacting substrates, a bound sulfate is found nearby H256, possibly mimicking the binding of the product phosphate after final 4HB-CoA formation. Additionally, a less common connector loop is found between CoA and ATP-grasping domains, suggesting a mechanism for increased stability. This paper describes for the first time the structure of Nmar_0206, the 4-hydroxybutyryl-CoA Synthetase from the 3HP/4HB cycle of *Thaumarchaeota*; highlighting the unique interdomain linker, interfaces and binding site which differentiate it functionally from other ACDs.

Materials & Methods

Cloning

The Nmar_0206 gene was obtained from Genscript BioTech with a N-terminal histidine tag. It was then cloned into the pET28a vector using *Nde*I and *Bam*HI endonuclease restriction sites, and transformed into *E. coli*, strain BL21(Rosetta-2). Transformed cells were selected for by growth in agar plates containing kanamycin (50 µg/ml) and chloramphenicol (50 µg/ml) at 37℃.

Protein Expression and Purification

Expression of Nmar_0206 was performed using the BL21(Rosetta-2) *E. coli* transformed cells. Cultures were grown overnight in LB media and then diluted 1:100 into 2 L cultures. The cell cultures were grown to an optical density of 0.8 at 600 nm and then expressed overnight at 18°C following induction with 0.7 mM IPTG. Cell paste was then obtained following centrifugation at 4000 rpm, resuspended in a lysis buffer (pH 7.0, 1 M KCl) and sonicated. Soluble protein was maintained at 4°C and purified using Ni-NTA affinity resin (GE Healthcare). A column wash was performed using pH 7.0 HisA containing 300 mM NaCl, 20 mM Tris and eluted with pH 7.5 HisB containing 500 mM imidazole, 300 mM NaCl, 50 mM Tris. Following purification, thrombin and beta-mercaptoethanol were added to Nmar_0206 to remove the N-terminal histidine tag and reduce any disulfide bonds that stabilize oligomers. Reverse Ni-NTA was then performed to remove the thrombin-cleaved hexa-histidine tag. Nmar_0206 purity was confirmed through SDS-PAGE. Following purification, 5 mL of 50% glycerol was added to 30 mL of the enzyme solution for long-term storage.

Crystallization

Crystallization screens were performed using 72-well Terasaki microbatch plates. 0.83 µL of purified Nmar_0206 were pipetted into the bottom of the sitting drop and mixed with an additional 0.83 µL of crystallization screen condition, followed by a 16.6 µL application of Paraffin oil to protect samples from oxidation. Plates were stored in styrofoam containers at room temperature and crystals of Nmar_0206 were obtained 1-2 weeks after initial crystallization screenings. Protein crystals were obtained in a solution of 2.5 µL of 100 mM Na-acetate/acetic acid pH 4.5, 200 mM Lithium sulfate, 50% (v/v) PEG 400, 2.5 µL of the enzyme solution, and 0.5 µL of 100% PEG 400 covered with ~50 µL of Paraffin oil.

Data Collection and Processing

Protein crystals of Nmar_0206 were prepared for crystallography by flash freezing in liquid nitrogen. The data was collected at Stanford Synchrotron Radiation Lightsource Beam-Line 12-2 at SLAC, Menlo Park, California, USA. The detector distance for the enzyme structure was set at 400 mm with an exposure time of 0.2 s per frame, with the X-ray energy set to 12.65 keV. The diffraction data were collected to 2.7 Å resolution at 100 K. The crystal belongs to the space group $P2_12_12_1$ with unit cell dimensions a = 70.53 Å b = 75.92 Å c =357.28 Å $\alpha = 90$ $\beta = 90$ $\gamma = 90$. The diffraction data were processed with the XDS package[11] for indexing and scaled by using XSCALE.

Structure Determination and Refinement

A previously solved structure of ATP dependent dimethylsulfoniopropionate (PDB ID: 7CM9) from *Psychrobacter* sp. D2[12] was used as a search model for automated molecular replacement using *PHASER* within the *PHENIX* software suite[13]. This was followed by simulated-annealing,

individual coordinate and TLS parameters refinement. COOT[14] was then used to confirm residue and water positions, and addition of residues missing in the electron density map. Further refinement was performed to 2.7 Å resolution within COOT while maintaining positions with strong difference densities and Ramachandran statistics for the structure were optimized to 97/2.9/0.1% (most favored/additionally allowed/disallowed). The final R-work and R-free were 0.25 and 0.30 respectively.

Results and Discussion

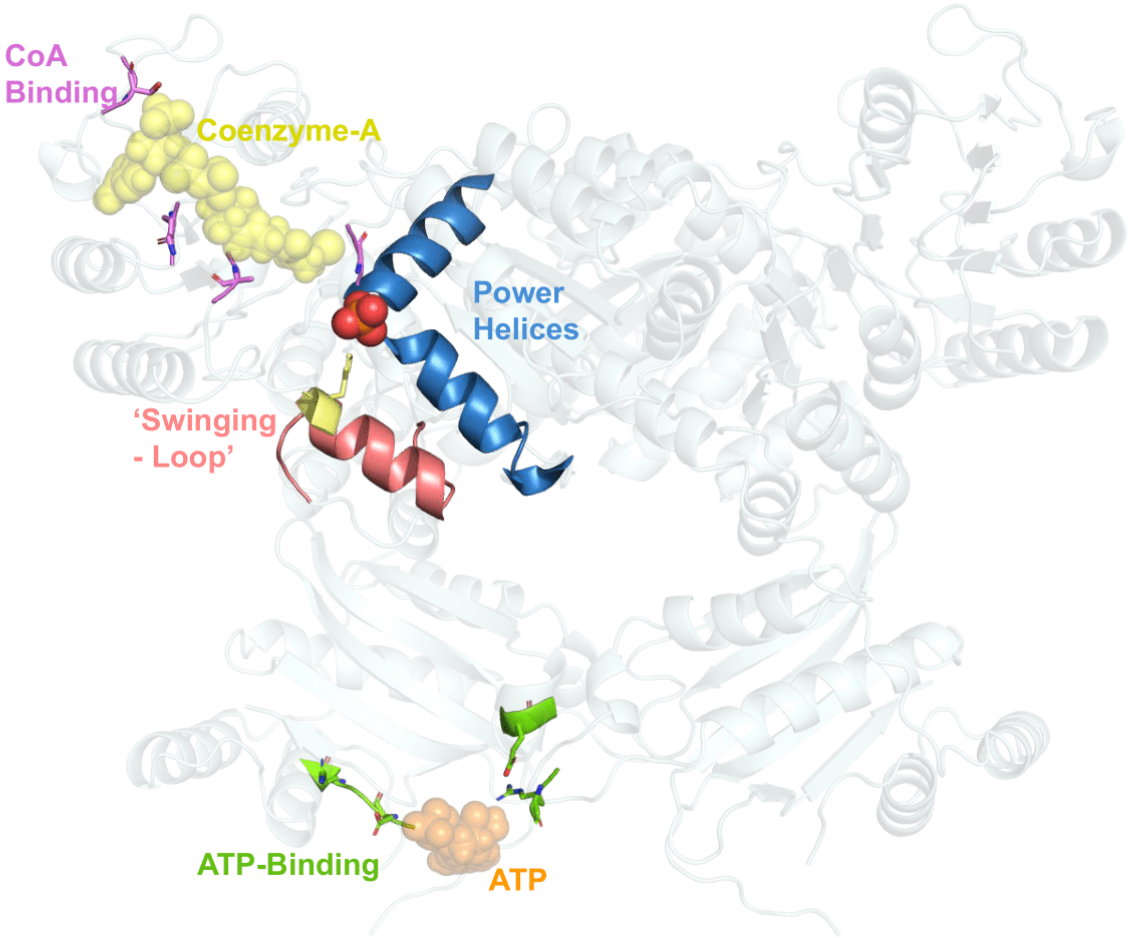

Figure 3 - Nmar_0206 (PDB: 8IUP) including important reactive elements, phosphates and the linker between ATP-grasping and CoA-binding domains. Green residues are proposed to interact with ATP (orange cluster suggests binding site; see pdb:4XYM) as part of the full reaction mechanism [Figure 9]. The salmon colored 'Swinging Loop', which bridges the gap between ATP and the active site, transports a phosphorylated histidine (yellow residue) to the tip of the two power helices. Power helices stabilize a phosphate (sulfate in our structure) bound between the two helices [Figure 4]. Purple residues are proposed to interact with 4-hydroxybutyryl-CoA (orange cluster suggests binding site; see pdb:4XYM) as part of the reaction mechanism [Figure 2].

Overall Crystal Structure of Nmar_0206

The resolved X-ray Crystal structure of Nmar_0206 resulted in a homodimer in the asymmetric unit, with each monomer consisting of 624 out of 698 residues of the full length Nmar_0206 (PDB: 8IUP). Two interfaces can be found between subdomains 4 and 5 of the two monomers [see figure 6]. Like all ACDs, Nmar_0206 consists of shuffled domains conserved within the superfamily. ACD domain structure is usually described in comparison to ecSCS which has a domain structure of [alpha(1,2)/beta(3,4,5)] while Nmar_0206 has the single protein chain [1-2-5-4-3][7]. Specifically, the subdomains 1, 2 and 5 of Nmar_0206 are connected from the N to C terminus. Subdomain 1 consists of a Rossman-like fold, while subdomains 2 and 5 are flavodoxin-like. The subdomains 1 and 2 of one monomer form a CoA-binding pocket with the subdomain 5, a CoA ligase domain, of the other monomer. Subdomain 2 contains an evolutionarily conserved H256 on a flexible swinging loop composed of small amino acids (residues S246-I266), which has been proposed to cross a distance of 31.5 Å to interact with bound ATP in the ATP-grasp domain[9,15]. It is stabilized by two power helices within subdomains 2 and 5 of monomers which present their N termini near H256 generating a positively charged environment into which a reactive phosphate ion is introduced [Figure 4][16,17]. In contrast with most structures within the PDB, domain 4 is attached to the C terminus of subdomain 5 by a 14 residue linker spanning 35 Å between W452-K469 [Figure 5A]. Subdomains 3 and 4 together form an ATP-grasp unit, with subdomain 3 existing within subdomain 4 between residues 503 and 577. Although subdomain 3 is catalytically important in both the binding and transfer of phosphate from ATP to the phosphohistidine swinging loop it was not present within the electron density map suggesting free movement when unbound. Subdomain 4 comprises the ATP-binding site with ATP-interacting residues present [Figure 8; supplementary Figure 3].

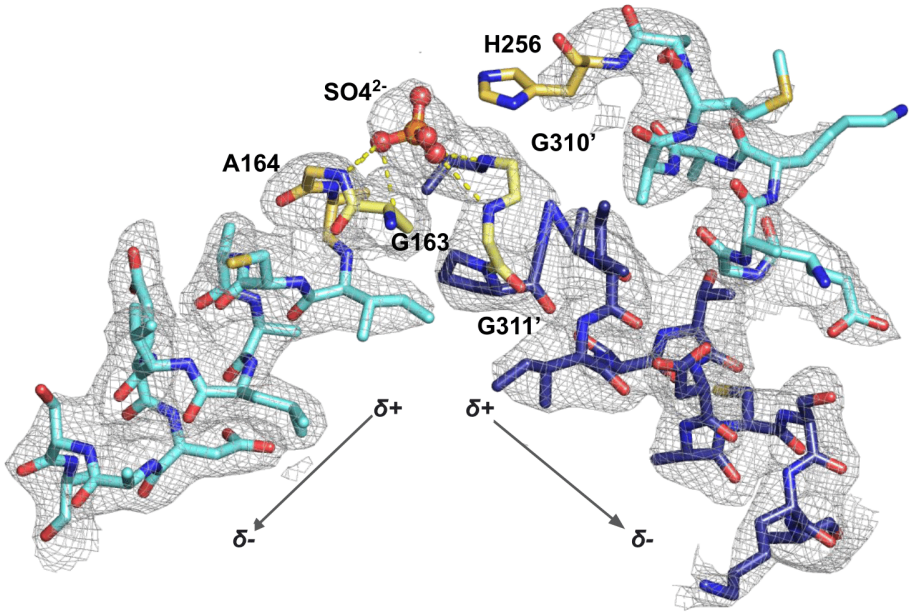

Figure 4 - Power helices and a bound sulfate. Structures can be seen in the 2Fo-Fc electron density map with a contouring level of 1σ and carved to 2 Å. Power helices could stabilize a decoupled phosphate after dissociation from H256 within the reaction mechanism. H256 is not within the electron density map of this structure. Different shades of blue indicate different chains; yellow residues interact with the phosphate.

Bound Sulfate at Active Site of Nmar_0206

Although the Nmar_0206 structure lacks any of the standard interacting substrates, a bound sulfate is found nearby H256 (side chain not found within the electron density) interacting with G163, A164, G310 and G311. At the head of the two power helices, this sulfate could mimic the binding mode of phosphate, one of the products of the reaction. This configuration could represent a final leftover product bound to the structure after releasing 4HB-CoA and ADP. As the sulfate binds deeper inside a cavity containing 4HB and CoA-binding sites, it is unsurprising that 4HB-CoA would be released before a phosphate product. Interestingly, the sulfate molecule is not found interacting with the magnesium ions previously reported to stabilize bound

phosphate[9]. The sulfate was discovered interacting with the amine groups of G163, A164, G310, and G311's backbone.

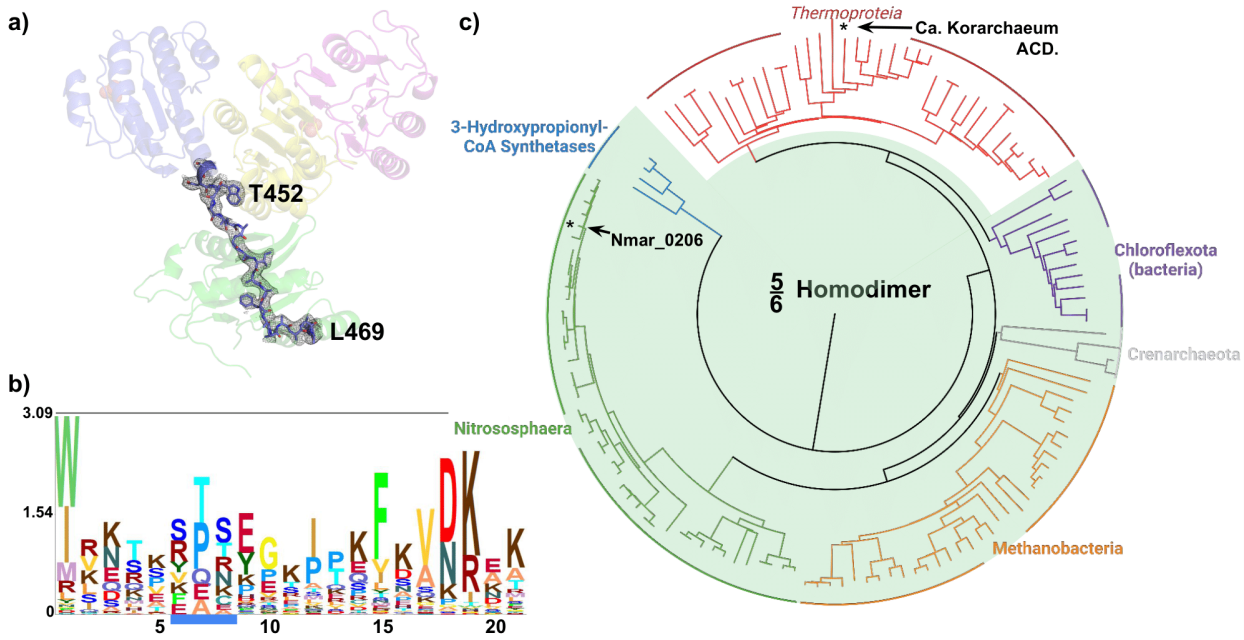

Figure 5 - Interdomain linker and phylogenetic tree. A) 17 residue linker connecting the coenzyme-A binding (blue, yellow and pink representing subdomains 1, 2 and 5) and ATP-grasp domains (green representing subdomain 4) of Nmar_0206 can be seen in the 2Fo-Fc electron density map with a contouring level of 1σ and carved to 2 Å. B) Logos representation of Nmar_0206 interdomain linker. Residues marked blue are not present in our sequence and have less than 10% representation within an MSA of 138 sequences of related structures. C) 138 sequences were used to generate a phylogenetic tree of related ACDs. Asterisks indicate the two structures in the PDB including this Nmar_0206 structure and that of *Ca. Korarchaeum* ACD. Although other structures of ACDs exist in the PDB, they include succinyl- and Citryl-CoA synthetases [Figure 7], all more distantly related. The Nmar_0206 structure is one of only two homodimer ACDs in the PDB.

Connector Loop and Associated Interfaces found between CoA-Binding Domain - ATP-Grasp Domain

The presence of a connector loop between the two substrate binding macrodomains was first shown by Li et al.[12] in PDB 7CM9. Figure 6 shows a stark difference between the CoA-binding and ATP-grasp domain interfaces of the two structures, suggesting that the addition of this connector provides an alternative to the stability that comes from hydrophobic domain interactions. Of 138 sequences used to generate a phylogenetic tree [Figure 5C], 72.5% (or 5% of representative groups) contain the aforementioned linker. Within the PDB, structures exist for DMSP-lyase, succinyl-, citryl-, acetyl-, and now 4-hydroxybutyryl-CoA synthetases. All of these mentioned structures, with the exception of a single DMSP-lyase and this Nmar_0206 structure, are heterotetramers and do not contain the linker.

Figure 5C shows Nmar_0206 within the context of the greater ACD superfamily which plays roles in carbon fixation, acetate metabolism, and ATP generation. As expected, Thaumarchaeota are its closest relatives, the next being methanogens (Methanobacteria) which, along with Crenarchaeota, form a branch separate from other identified ACDs. The 3-hydroxypropionyl-CoA synthetase (Nmar_1309) shows some similarity but is phylogenetically distinct, seen as an outgroup within the tree. These phylogenetic relationships map to evolutionary separations among archaea and bacteria, and show the adaptation of Acetyl-CoA synthetase enzymes to those found within the 3HP/4HB cycle. This can be seen in Figure 5C as the closest relative to the archaeal, heterotetramer Thermoproteia ACDs are those of Chloroflexota, a bacterial phylum. As these Thermoproteia ACDs are homodimers, these might indicate an evolutionary link between homodimer Thaumarchaeota ACDs and the heterotetramer structures. This could suggest two, separate horizontal gene transfer events (HGT); the

heterotetramer Thermoproteia to the homodimer Chloroflexota to the homodimer Thaumarchaeota relatives.

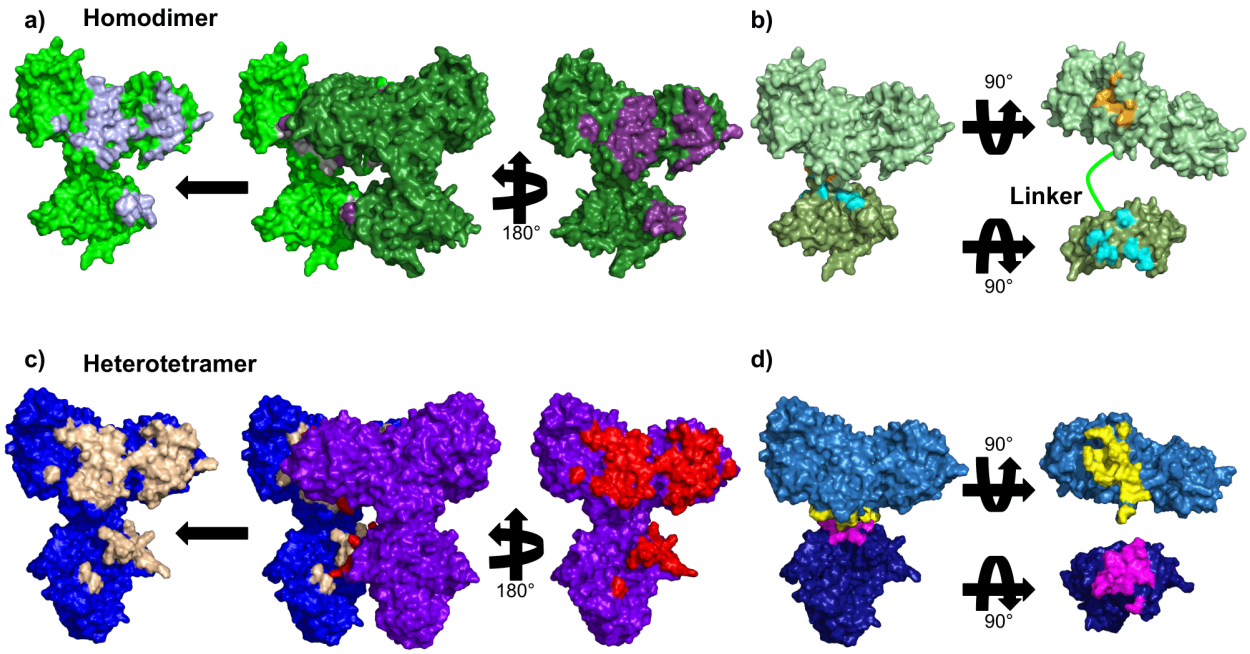

Figure 6 - Interface surfaces of Nmar_0206 and Acetyl-CoA Synthetase (ACS; PDB: 4XYM). Interfaces are displayed between the A) homodimer Nmar_0206 monomers and between the B) Alpha domains and beta domains of ACD. Interfaces between CoA-binding and ATP-grasp domains of B) Nmar_0206 and D) ACD. The homodimer Nmar_0206 shows a significantly reduced interface region when compared to ACD.

The interface between the CoA-binding and ATP-grasp domains is greatly reduced in the homodimer structure than the heterodimer [Figure 6B&D]. As the interface domain plays a role in dimerization, the smaller interface could be expected with covalently fused chains. Assuming that this is the product of the linkage between the two domains, it is worth noting that *Ca. Korachaeum* live at elevated temperatures (78-92°C) whereas *N. maritimus* grows at temperatures of 15-35°C[18]. It is surprising then that a stabilizing linker-loop is present in a mesophilic microorganism and not a hyperthermophile.

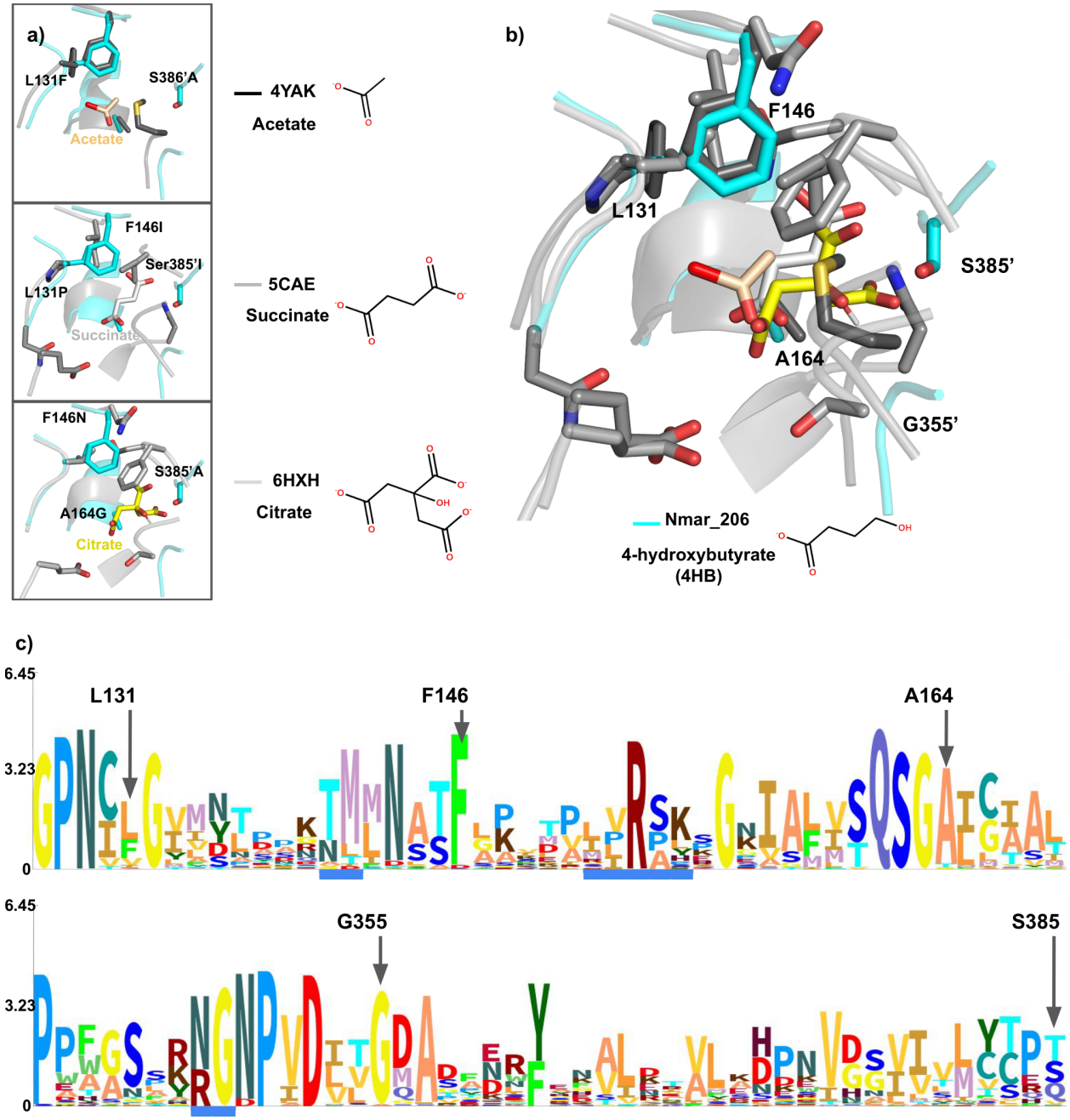

Figure 7 - Comparisons of Nmar_0206 with homologous active sites. Slight differences in this region could affect substrate selectivity. Light blue structure indicates Nmar_0206. A) acetyl-CoA synthetase from *Ca. Korarchaeum* (PDB: 4YAK), succinyl-CoA synthetase from *Sus domesticus* (PDB: 5CAE), and citrate lyase from *Homo sapiens* (PDB: 6HXH) show structural changes surrounding bound substrates. Substrates for Nmar_0206 (4-hydroxybutyrate), 4YAK (acetate), 5CAE (succinate), and 6HXH (citrate) are shown next to associated structures. Alignments were performed using the power helices tips and nearby projecting loops for reference,

namely residues G310-P312, V354-D356, P128-C130, G163-I165, and G187-K189. The Pymol align algorithm was used with 0 outlier rejection cycles and provided RMSD values of 1.12 Å (4YAK), 1.61 Å (6HXH) and 0.77 Å (5CAE) based on their Ca residue positions. Residues within 5 Å of the carboxyl end of each substrate were displayed in stick mode. B) Superposition of all structures highlight changes within the active site which may vary binding affinities to substrates. C) Logos for regions around the active site with highlighted residues found interacting with acetate from acetyl-CoA synthetase.

Residue Shifts within Homologous Active Sites

To accommodate a unique 4HB structure, some residues within the active site have shifted or been replaced. Using residues shown to be interacting with *Ca. Korachaeum's* acetyl-CoA synthetase as reference (PDB: 4YAK), comparisons were made between the Nmar_0206 structure (PDB: 8IUP) and others within the PDB with bound substrates (PDB: 6HXH and 5CAE). Associated residues for L131, F146, A164, G355' and S385' can each be found within 5 Å of the acetyl group within 4YAK. With the exceptions of L131 and F146, these residues each shifted further away from that same location within the Nmar_0206 structure. The alpha carbons of S385', G355', and A164 each shifted 2.5Å, 1.5Å and 1.1Å respectively.

The logos found in Figure 7C highlight residues originally identified as interacting with the acetyl-group from acetyl-CoA within pdb: 4YAK. The sequence alignment used to generate these logos show F146, A164 and G355 as being highly conserved, suggesting important interactions with conserved elements (i.e., carboxyl end of substrate). L131 and S385 are much more variable, possibly indicating residues allowing for substrate promiscuity.

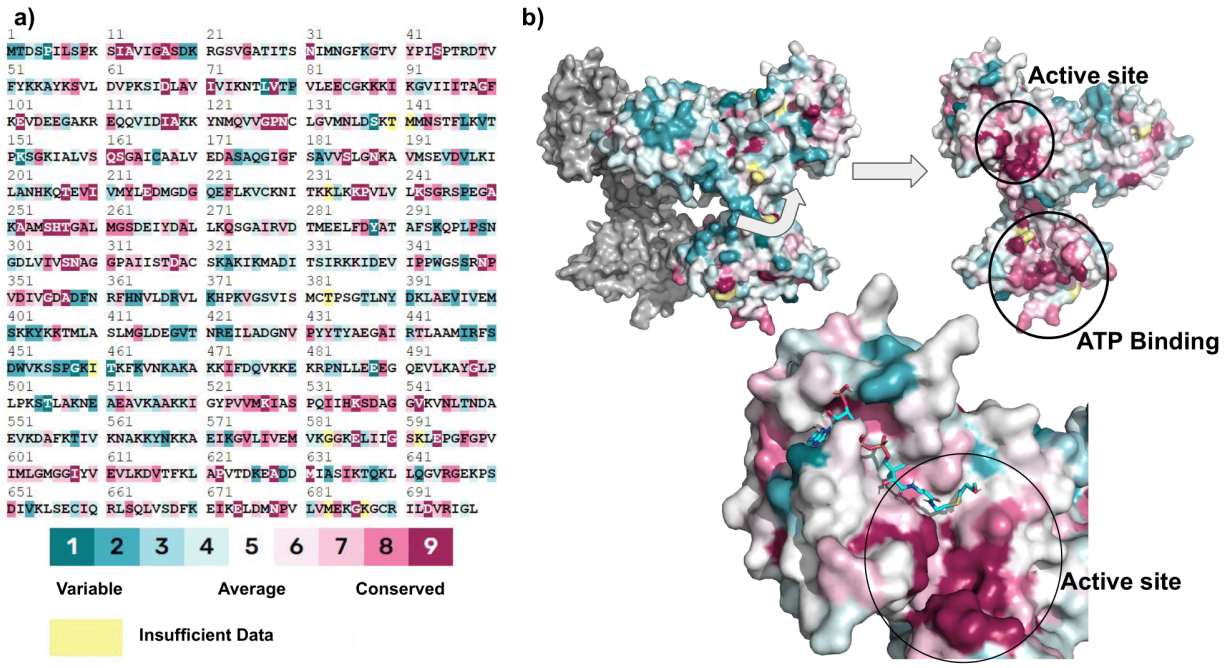

Figure 8 - ConSurf analysis for Nmar_0206. Multiple sequence alignment was built using multiple Blasts for relatives of Nmar_0206 (ranging from 70% to 90% sequence identity) from the NCBI database. Residue conservation is shown on the surface, coloring residues by conservation. 1, dark teal, indicates an extremely low conservation whereas 9, maroon, indicates a high sequence conservation. The information was obtained and is shown using the PyMOL output file from the ConSurf server[26].

Residue Conservation within ACDs

Figure 8 indicates conservation in internal residues between the subunits, particularly at the active site and ATP-binding sites. Although all ACDs utilize CoA as a substrate, the ATP-bind site appears only moderately conserved. Interface residues, discussed in detail following Figure 6, necessitate some conservation to maintain those connections. Although this is seen to some degree within the consurf results near the active site interactions, there is still moderate variability in many residues along the interface.

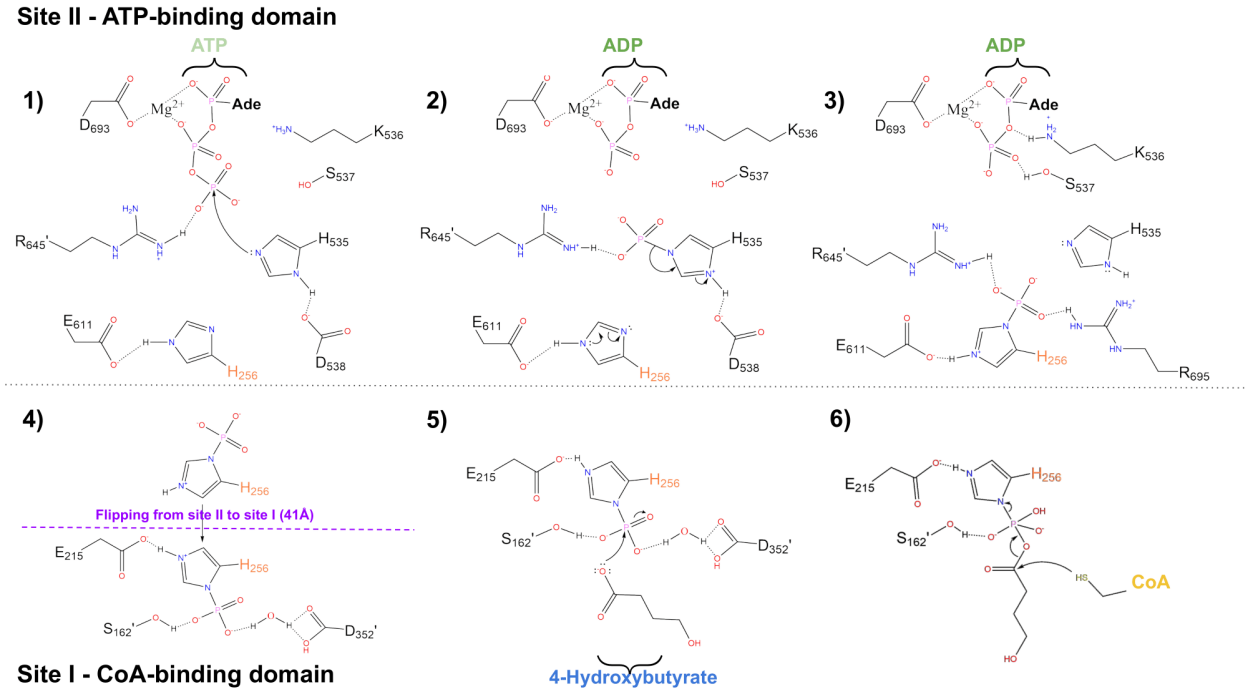

Figure 9 - Proposed reaction mechanism for 4-hydroxypropionyl-CoA formation from ATP, CoA and 4HB. Gamma phosphate of ATP is nucleophilically attacked by H535 before being transferred onto H256 forming phosphohistidine. This Phospho-H256 must then cross a distance of 30 Å into the active site where it interacts with 4HB to form a transient 4HB–phospho-H256. The 4HB group is then attacked by the sulfinyl group of CoA leading to this dissociation of phosphate from H256 and the final formation of 4HB-CoA.

Likely Reaction Mechanism of Nmar_0206

The reaction mechanism for ACDs have been previously established. The first of the reported reaction steps begins when the gamma phosphate of an ATP molecule is removed through a nucleophilic attack by a conserved H535 to form phosphohistidine[9,18]. The phosphate should then be passed to the swinging loop's H256 which forms another phosphohistidine[9,15]. Phospho-H256 should then swing roughly 30 Å back towards the CoA-binding site between subdomains 1 and 2 where it binds to 4HB, forming a transient 4HB-phosphohistidine[20]. The final step of the reaction mechanism involves the sulfur group from CoA performing a nucleophilic attack on the 4HB

group, leading to the release of a free phosphate and 4HB-CoA as the final products[20]. Our structure, albeit missing the ATP-grasp domain 'lid' (domain 3), is bound to phosphate within the active site, suggesting a structural conformation just following the reaction before the release of the free phosphate.

Summary/Conclusion

The adaptation of a highly efficient 3HP/4HB cycle has allowed *N. maritimus* and other ammonia-oxidizing archaea to outcompete their bacterial counterparts in oligotrophic environments[21,22]. The Nmar_0206 structure represents just one component of the greater search for functional elements which differentiate the highly efficient 3HP/4HB cycle of *Thaumarchaeota* from less efficient variants represented in *Crenarchaeota*.

The ACD superfamily is used during carbon fixation, acetate metabolism, and ATP generation in many bacteria and archaea, including organisms present within the phylogenetic tree in Figure 5C. The synthetases specific to the 3HP/4HB cycle of *Thaumarchaeota* include 3-hydroxypropionyl- and 4-hydroxybutyryl-CoA synthetases. The homodimer seen within Nmar_0206 seems to be widespread within many microbial species, suggesting that this is not a unique development within *Thaumarchaeota*. Stability supported by the mentioned linker loop would therefore be relevant within many other structures within the superfamily, suggesting two separate lineages of homodimers and heterotetramers; however, what evolutionary benefit such a fusion/separation would provide is still unclear.

To complete the overall goal of identifying structural components of increased energy efficiency, a similar analysis of an AMP-forming counterpart of 4-hydroxybutyryl-CoA synthetase from *Crenarchaeota* will need to be performed. Previous studies have described AMP-forming substrate binding pocket as being amenable to mutation due to its flexibility and the large size of its binding site[23]. This may suggest that the reason for the evolutionary preference of AMP-forming acetyl-CoA synthetases (62 members in E.C. 6.2.1.) over ACDs (9 members in E.C. 6.2.1.) is for substrate conversion despite their increased efficiency. Future investigations should perform comparative binding site analysis of AMP-forming crenarchaeal counterparts with Nmar_0206 to confirm this.

The enzymes involved in the 3HP/4HB cycle have been selected for energy-efficient attributes, such as the bifunctionality of crotonyl-CoA hydratase/3-hydroxypropionyl-CoA dehydratase[24] and the oxygen-tolerance of 4-hydroxybutyryl-CoA dehydratase[25]. The growing demand for carbon fixation strategies to remediate our rapidly changing atmosphere begs for further tools, such as those which can be produced from the analysis of highly efficient enzymes. The structural characteristics of Nmar_0206 and similar enzymes will help inform scientists and engineers how to adapt these highly efficient enzymes for human needs.

References

1. Friedlingstein, P., Meinshausen, M., Arora, V. K., Jones, C. D., Anav, A., Liddicoat, S. K., & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. Journal of Climate, 27(2), 511–526. https://doi.org/10.1175/jcli-d-12-00579.1

2. Berg, I. A. Ecological Aspects of the Distribution of Different Autotrophic CO 2 Fixation Pathways. Appl. Environ. Microbiol. 77, 1925–1936 (2011).
3. Ducat, Daniel C; Silver, Pamela A (2012). Improving carbon fixation pathways. Current Opinion in Chemical Biology, 16(3-4), 337–344. doi:10.1016/j.cbpa.2012.05.002
4. Ingalls, A. E., Shah, S. R., Hansman, R. L., Aluwihare, L. I., Santos, G. M., Druffel, E. R., & Pearson, A. (2006). Quantifying archaeal community autotrophy in the mesopelagic ocean using natural radiocarbon. Proceedings of the National Academy of Sciences, 103(17), 6442-6447.
5. Könneke, M., Schubert, D. M., Brown, P. C., Hügler, M., Standfest, S., Schwander, T., Schada Von Borzyskowski, L., Erb, T. J., Stahl, D. A., & Berg, I. A. (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO 2 fixation. Proceedings of the National Academy of Sciences, 111(22), 8239–8244. https://doi.org/10.1073/pnas.1402028111
6. Karner, M. B., DeLong, E. F. & Karl, D. M. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. Nature 409, 507–510 (2001).
7. Schäfer, T., Selig, M., & Schönheit, P. (1993). Acetyl-CoA synthetase (ADP forming) in archaea, a novel enzyme involved in acetate formation and ATP synthesis. Archives of Microbiology, 159(1), 72-83.
8. Sánchez LB, Galperin MY, Müller M. Acetyl-CoA synthetase from the amitochondriate eukaryote Giardia lamblia belongs to the newly recognized superfamily of acyl-CoA synthetases (Nucleoside diphosphate-forming) J Biol Chem. 2000;275(8):5794–5803.
9. Weiße, R. H.-J., Faust, A., Schmidt, M., Schönheit, P., & Scheidig, A. J. (2016). Structure of NDP-forming Acetyl-COA synthetase ACD1 reveals a large rearrangement for

phosphoryl transfer. *Proceedings of the National Academy of Sciences*, *113*(5). https://doi.org/10.1073/pnas.1518614113

10. Huang, J., & Fraser, M. E. (2021). Second distinct conformation of the phosphohistidine loop in succinyl-COA Synthetase. Acta Crystallographica Section D Structural Biology, 77(3), 357–368. https://doi.org/10.1107/s2059798321000334
11. Kabsch, W. (2010). xds. Acta Crystallographica Section D: Biological Crystallography, 66(2), 125-132.
12. Li, C.-Y., Wang, X.-J., Chen, X.-L., Sheng, Q., Zhang, S., Wang, P., Quareshy, M., Rihtman, B., Shao, X., Gao, C., Li, F., Li, S., Zhang, W., Zhang, X.-H., Yang, G.-P., Todd, J. D., Chen, Y., & Zhang, Y.-Z. (2021). A novel ATP dependent dimethylsulfoniopropionate lyase in bacteria that releases dimethyl sulfide and acryloyl-COA. ELife, 10. https://doi.org/10.7554/elife.64045
13. Adams PD, Afonine P v., Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung L-W, Kapral GJ, Grosse-Kunstleve RW, et al. 2010. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* 66: 213–221.
14. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D Biological Crystallography 60: 2126–2132.
15. Fraser, M. E., James, M. N., Bridger, W. A., & Wolodko, W. T. (1999). A detailed structural description of Escherichia coli succinyl-CoA synthetase. Journal of molecular biology, 285(4), 1633-1653.

16. Wolodko, W. T., Fraser, M. E., James, M. N., & Bridger, W. A. (1994). The crystal structure of succinyl-CoA synthetase from Escherichia coli at 2.5-A resolution. Journal of Biological Chemistry, 269(14), 10883-10890.
17. Hol, W. G. J., Van Duijnen, P. T., & Berendsen, H. J. C. (1978). The α-helix dipole and the properties of proteins. Nature, 273(5662), 443-446.
18. Qin, W., Heal, K. R., Ramdasi, R., Kobelt, J. N., Martens-Habbena, W., Bertagnolli, A. D., ... & Stahl, D. A. (2017). Nitrosopumilus maritimus gen. nov., sp. nov., Nitrosopumilus cobalaminigenes sp. nov., Nitrosopumilus oxyclinae sp. nov., and Nitrosopumilus ureiphilus sp. nov., four marine ammonia-oxidizing archaea of the phylum Thaumarchaeota. International journal of systematic and evolutionary microbiology, 67(12), 5067-5079.
19. Bräsen, C., Schmidt, M., Grötzinger, J., & Schönheit, P. (2008). Reaction mechanism and structural model of ADP-forming acetyl-CoA synthetase from the hyperthermophilic archaeon Pyrococcus furiosus: evidence for a second active site histidine residue. Journal of Biological Chemistry, 283(22), 15409-15418.
20. Fan, F., Williams, H. J., Boyer, J. G., Graham, T. L., Zhao, H., Lehr, R., Qi, H., Schwartz, B., Raushel, F. M., & Meek, T. D. (2012). On the catalytic mechanism of human ATP citrate lyase. Biochemistry, 51(25), 5198–5211. https://doi.org/10.1021/bi300611s
21. Santoro, A. E.; Buchwald, C.; McIlvin, M. R.; Casciotti, K. L. (2011). Isotopic Signature of N2O Produced by Marine Ammonia-Oxidizing Archaea. Science, 333(6047), 1282–1285. doi:10.1126/science.1208239

22. Wuchter, C., Abbas, B., Coolen, M. J., Herfort, L., van Bleijswijk, J., Timmers, P., ... & Sinninghe Damsté, J. S. (2006). Archaeal nitrification in the ocean. Proceedings of the National Academy of Sciences, 103(33), 12317-12322.
23. Zahn, M., Kurteva-Yaneva, N., Schuster, J., Krug, U., Georgi, T., Müller, R. H., Rohwerder, T., & Sträter, N. (2019). Structures of 2-hydroxyisobutyric acid-COA ligase reveal determinants of substrate specificity and describe a multi-conformational catalytic cycle. *Journal of Molecular Biology*, *431*(15), 2747–2761. https://doi.org/10.1016/j.jmb.2019.05.027
24. Destan, E., Yuksel, B., Tolar, B. B., Ayan, E., Deutsch, S., Yoshikuni, Y., ... & DeMirci, H. (2021). Structural insights into bifunctional thaumarchaeal crotonyl-CoA hydratase and 3-hydroxypropionyl-CoA dehydratase from Nitrosopumilus maritimus. Scientific reports, 11(1), 22849.
25. DeMirci, H., Tolar, B. B., Doukov, T., Petriceks, A., Pal, A., Yoshikuni, Y., ... & Wakatsuki, S. (2020). Structural adaptation of oxygen tolerance in 4-hydroxybutyrl-CoA dehydratase, a key enzyme of archaeal carbon fixation. bioRxiv, 2020-02.
26. Ashkenazy H., Abadi S., Martz E., Chay O., Mayrose I., Pupko T., and Ben-Tal N. 2016 ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucl. Acids Res. 2016; DOI: 10.1093/nar/gkw408; PMID: 27166375

Acknowledgements

This project and the experiments are funded by TÜBİTAK-NSF 2501 bilateral research program (project number 221N3555). H.D. acknowledges support from NSF Science and Technology Center grant NSF-1231306 (Biology with X-ray Lasers, BioXFEL). This publication has been

produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program and the 1001 Scientific and Technological Research Projects Funding Program of the TÜBİTAK (Project Nos. 118C270 and 120Z520). However, the entire responsibility of the publication belongs to the authors of the publication. The financial support received from TÜBİTAK does not mean that the content of the publication is approved in a scientific sense by TÜBİTAK. The work conducted by the U.S. Department of Energy Joint Genome Institute (https://ror.org/04xm1d337), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy operated under Contract No. DE-AC02-05CH11231. SW and CAF acknowledge support from the U.S. Department of Energy (DOE) Office of Science, Biological and Environmental Research; Stanford Precourt Institute; and SLAC Laboratory Directed Research and Development. SSRL is supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences (OBES) under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (NIGMS) (including P41GM103393).https://docs.google.com/document/d/15DKqmvS4hWirc7UkWO6nGVCpnu6zwXQnF-opz30YEp0/edit

Author contributions: *H.D., Y.Y., C.A.F, S.W. conceived the study; H.D., J.J. designed the experiments; H.D., J.J., B.B.T. performed the experiments with input from S.W., J.J. and B.B.T. performed phylogenetic analyses. All the authors took part in the interpretation of results and preparation of the manuscript.*

Conflicts of Interest: The authors declare that there are no conflicts of interest.