|  |  |  |  |
| --- | --- | --- | --- |
| Date | Agenda | Notes | Action Items |
| 12/05/2023 | **Initial Meeting**   1. Revised details for the project (given the split into two) 2. Tools to learn and where to find (Gremlin mentioned before is this the BakerLab software for contact maps? AlphaFold as well – Docker in BBK system available? Or own PC?) 3. Expectations for scheduling meetings as a part-time student (frequency?)    * Book a time for this meeting now?    * How often does he want to meet, shall we schedule something recurring now pr leave this ad-hoc? 4. Question from reading: Is there no AmiC bound with peptidoglycan structure? Is that possible to model? | * Sarmini (other student) – also part time * Meeting frequency: Every 2 weeks (on a Friday?) * To be more up to date about what is known   + Proteosome scale work on gram negatives (E coli worked quite well)   + Focus on more current papers * First month: catch up with the literature, state of knowledge on the different proteins (inform on what we’ve found) * Activation process * David Baker’s lab (genome scale ppi predictions), not every possible pair (local pairs only) – looking into possible gene interactions - [10.1126/science.aaw6718](https://doi.org/10.1126/science.aaw6718) – competitor to AlphaFold (more recent paper, maybe not in E.coli) * Deborah Marks: EVFold: EVComplex (coevolution to predict complexes) * Set up of AF in the department – GoogleCollab version also available.   AmiC structure: Review papers, amidases, what are structures? (search by family/function, sequence)  Some don’t have copies (called AmiB in some bacteria, CwlV in others)  BLAST: restrict search in PDB  Pfam/Interpro: search for structures here too (search by sequence)  AmiB: which are AmiB and which are homologs of AmiC?  Identify conserved features, how do we classify what is similar to AmiC and involved in the pathway?  Amidase families (1: bacteriophage, 2: incl. AmiD, recycling?, 3: may or may not be related to family2) | Action 1: Curate current knowledge (Literature search into the different proteins, focus in on AmiC and AmiB)  Action 2: Curate a list of proteins which interact with these in gram negative bacteria (Identify proteosome papers on gram negative bacteria and possible interactions with AmiC and AmiB)  Action 3: Read the David Baker and Deborah Marks papers on genome scale predictions/interactions and protein complex prediction, make notes. |
| 26/05/2023 | 1. Feedback on orthologs of AmiC found   * Found 19 Amidase\_3 family structures * 146 structures more broadly with reported NAMLAA activity   + Eukaryotic as well! * Used blastp search, amidase\_3 family search, and AMIN domain search for AmiC (type Iva pilus machinery hits – same in Pfam for AMIN domain?) * Pfam and Interpro for the Amidase\_3 family   + 6290 species (Pfam)   + 74,000 sequences (37 reviewed by UniProt)   + 61,000 alphafold predictions   + 1013 domain architectures incl. Amidase\_3 + AMIN, Amidase\_3 + LysM   + 23 pathways (peptidoglycan recycling, also biosynthesis pathways) * Structures: 10 gram positive, 4 gram negative, some bacteriophage. * Related domains which came up multiple times in wider searching (see notes) * Did a very quick look in Chimera at the 19 Amidase\_3 structures; N-terminal domains quite different with AMIN unique to the E.coli structure, also looked like gram negatives had this auto-inhibiting helix in the active site which gram positives maybe didn’t have?   1b: Questions from structure searches   * Predicted structures as well? (AlphaFold) * Worth exploring AMIN architecture as well as Amidase 3?   2. Feedback on recent literature   * AmiC and NlpD interactions at septal site as part of divisome * AmiB and AmiA also (with EnvC) * Note on the order of involvement after PG synthesis, elongation, and cross-wall linkages w/Fts proteins * AmiC might be exported as signal peptide, involvement in antimicrobial effects? * Possible coevolution with bepA in E.coli (assembly enhancing protease) – 2019 Baker lab paper * Candidate interactions in DolP and ActS? (possibly at the active site, after NlpD, or at AMIN domain?) * Still working through literature (several other papers around AmiC in E.coli found)   3. Possible next steps   * Looking into sequence difference between isolates of deposited AmiC amidases in NCBI (evolutionary divergence?) * PPI with other pathways (FtsN, how AmiC might interact with DolP and ActS, antimicrobial pathways?) * Differences between Amidase\_2 and Amidase\_3 protein family * Further research into AMIN and Amidase\_3 family pathways – get a grip into what these proteins are interacting with across their predicted pathways, and if the literature backs this up (proteome papers – bepA with AmIC?) * Look into other PG binding domains (eg PG\_binding\_domain and SPOR domain) and the overlap in architecture with Amidase\_3/Amidase\_2? | * [PF01471](https://www.rcsb.org/search?q=rcsb_polymer_entity_annotation.annotation_id:PF01471%20AND%20rcsb_polymer_entity_annotation.type:Pfam&rt=polymer_entity) – PG\_binding\_1 domain (found in some of the Amidase\_2 family) – look into? * PF05036 – SPOR domain (binding to PG, involved with sporulation and division proteins like FtsN?) * PF01832 – Glycosaminidase (hydrolysing peptidoglycan) * PF00877 – NipC domain in TB (related to NIpD??) * PF12123 – CBC\_PlyG (cell wall binding domain in bacteria and viruses) * PF05257 – CHAP domain (related to amidase function, cell wall metabolism in bacteria, amidase domain of E. coli glutathionylspermidine synthetase?)   Sequence/structure alignment  Quality of structure list?  Mobile element in the active site  MSA: Conservation – superposition – look into more depth with different tools, do gram+ and – separate, look at all pairwise alignments   * Group of structures aligned at once * PDBefold? List of pairs (dl structures and input back) * Cluster analysis with R – k means clustering * See if you can group similar sequences * Should be able to see the loop difference (autoinhibiting) in this)   If differences: Add in alpha fold predictions to gain confidence in obs?  Sequences of these domains (non-redundant) – try to resolve programmatically? (esp. enzymatic domain)  Is it possible to get a good MSA without the orthologs? (ie gene duplication, different functional enzymes – just want cell division function in MSA)   * Query sequences which def. have that function (look up the 19, search with all 19) – how many sequences?   Sequences: NCBI: restrict in non-redundant set – RefSeq focus – trim seq. and search with just the amidase\_3 domain – what is the same/different?  ActS – any structures around this?  Microscopy – AmiC  Tat | Actions:   * Focus on getting quantitative results – esp. for perceived differences between gram negative and gram positive bacteria (using PDBeFold plus another tool?) * If a difference seen, could incorporate AlphaFold predictions? * Look up papers for the 19, ensure their cell division function has been confirmed. * For those with function confirmed, search for similarity in sequences in non-redundant NCBI dataset. (blastp) * Programmatic task: R package for k-means clustering * Programmatic task: Identification of the amidase\_3 domain from sequences (for future cropping, multiple sequence alignment, or searching) * ActS: Any structures of this? Other papers around function? How might the docking/binding of this be modelled (could compare to NlpD as this is more established as a cofactor in AmiC)   + Why might it bind preferentially to AmiB in lower pH? * DolP: The same as above, evidence less confident for this factor however. |
| 09/06/2023 | See presentation but in a nutshell:  NAMLAA functional domain role classifier? |  | See notepad:  Hypothesis: Helix is only in gram negative bacteria for the cell wall division process, as helix self regulates this type of amidase. This type of amidase is not present in gram positive and therefore gram positive amidase sequences will lack this helix.   * Q1: Is this in other gram negative bacteria? * Q2: Is this in gram positive bacteria? * Use RefSeq whole genomes rather than non-redundant sequences in NCBI * Search using cropped domains   Clustering analyses   * K-means clustering of RMSD in the structures identified   Taxonomic map of domain occurrence   * Number of amidase genes in each genome (looking at WGS) * Are there any other conserved features between bacterial sequences looking at amidases? * Get a taxonomy that can be sub-divided, generate taxonomic ID for each bacteria (need to research how to do this) |
| 23/06/2023 | Annotated bibliography  (send ahead of meeting)   * Present findings * Discuss understanding of literature and relevance to project * Note additional studies of relevance.   Results from previous week of research   * Sequence identity of the helix in gram negative and gram positive * Taxonomic map and how to go about this | Annotated Biblio – Good, no updates needed (Milestone 1 completed!)  Next steps:  Re-do alignment with errant sequences removed (as might not be true NAMLAA, since conservation should be higher in enzymatic region, might be affecting alignment as an outgroup).  To find out at what level these sequences might be linked – relate their position in the taxonomy tree.  Assess tax ID for geneus/species for each sequence to allow a tree to be made   * For tax ID: Use Batch Entrez to get tax ID for each species * Taxallnomy REST API using genus/species or tax ID * Is there a link at the class/phylum level? (ie looking at the missing links in the multiple sequence alignment) * Annotate each species based on phylum into gram negative and gram positive species. | Re-do alignment with errant sequences removed (as might not be true NAMLAA, since conservation should be higher in enzymatic region, might be affecting alignment as an outgroup).  To find out at what level these sequences might be linked – relate their position in the taxonomy tree.  Assess tax ID for geneus/species for each sequence to allow a tree to be made   * For tax ID: Use Batch Entrez to get tax ID for each species * Taxallnomy REST API using genus/species or tax ID * Is there a link at the class/phylum level? (ie looking at the missing links in the multiple sequence alignment)   Annotate each species based on phylum into gram negative and gram positive species. |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |