Structural Work: *Photorhabdus* Virulence Cassettes

PhD Update Talk

October 2016

Joe H

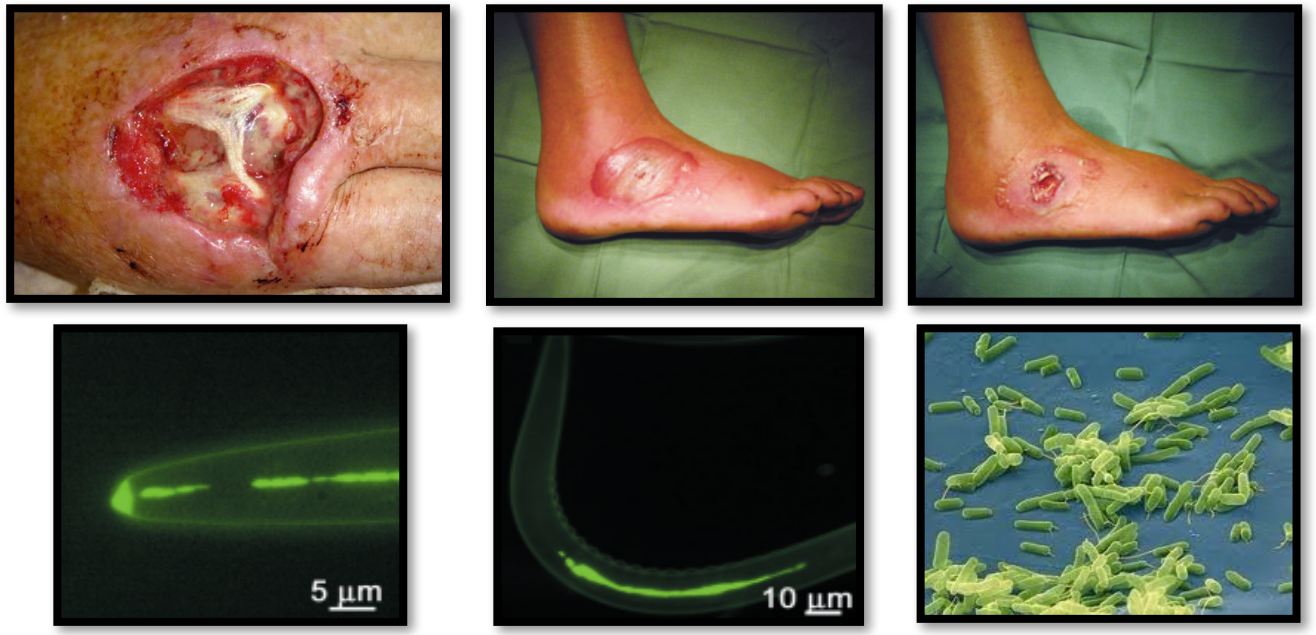
Content

- Quick Refresher
  - Genetic Structure
  - Physical Structure
- Structural Simulations & Analysis
  - Status
  - RMSD Calculation -> Realistic?
    - Correlation with ID?
  - Accurate Measurement
  - Basic protein characteristic calculations
    - Correlate? Clustering?
  - Next Steps
- Tail Fibre Structural Work
  - Trimerisation
  - CD Stability
  - DichroWeb
  - Crystallography
  - Next Steps
- If time…
  - FutuRx & Biotech Stuff
  - Revisit phylogenetics

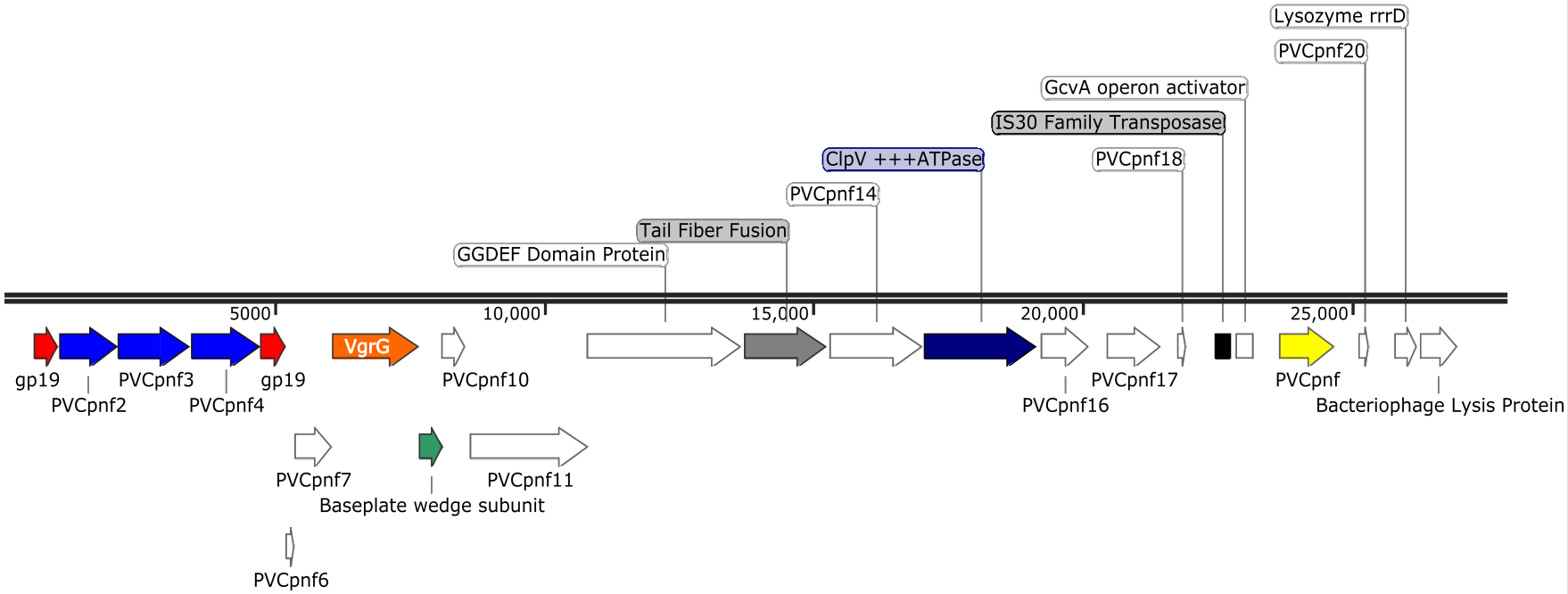
Quick Project Refresher

Quick Project Refresher: *Photorhabdus*

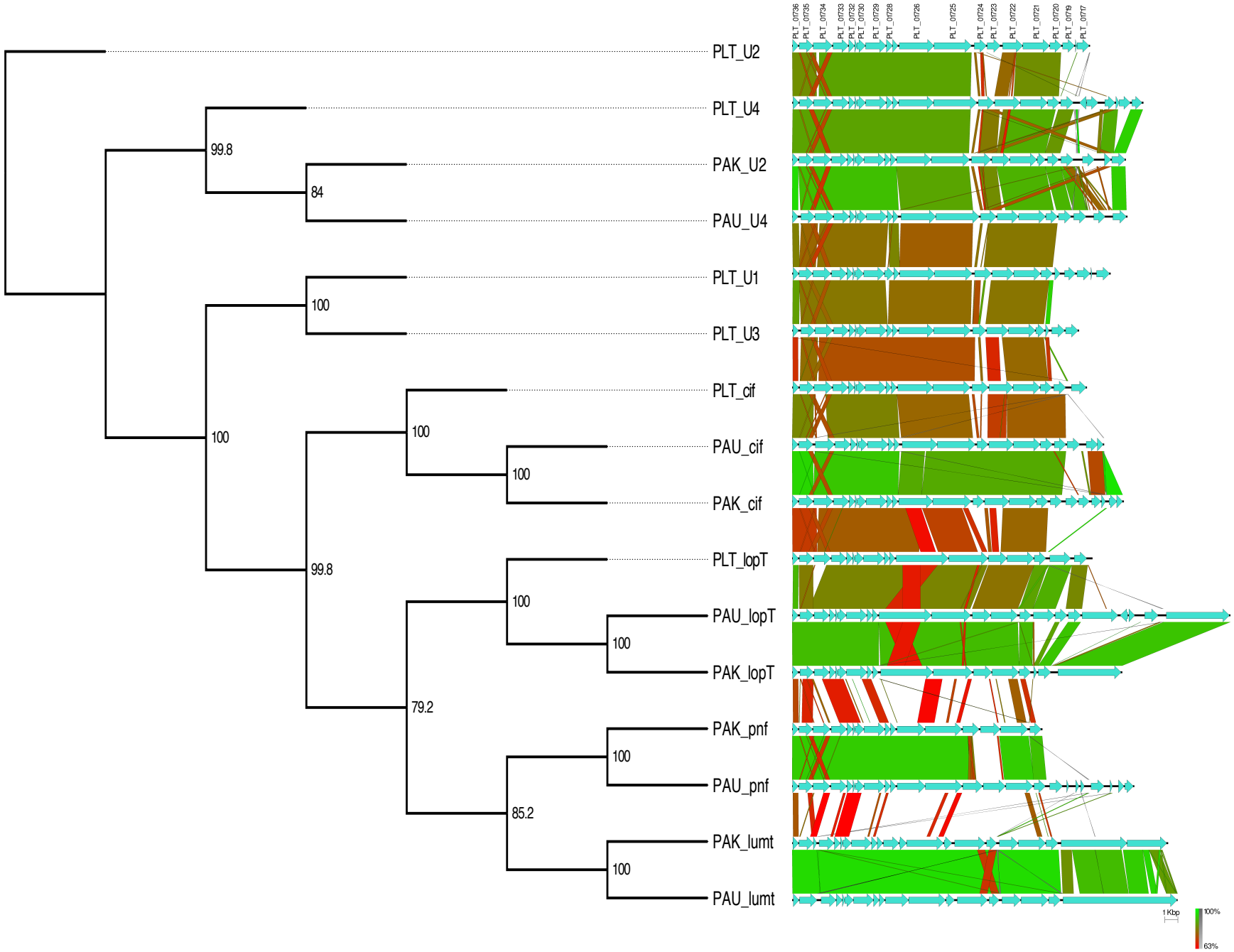
- Gram negative, bioluminescent, insect and (sometimes) human pathogen. Also a symbiont.
- Many unique/unusual metabolic/genetic characteristics.



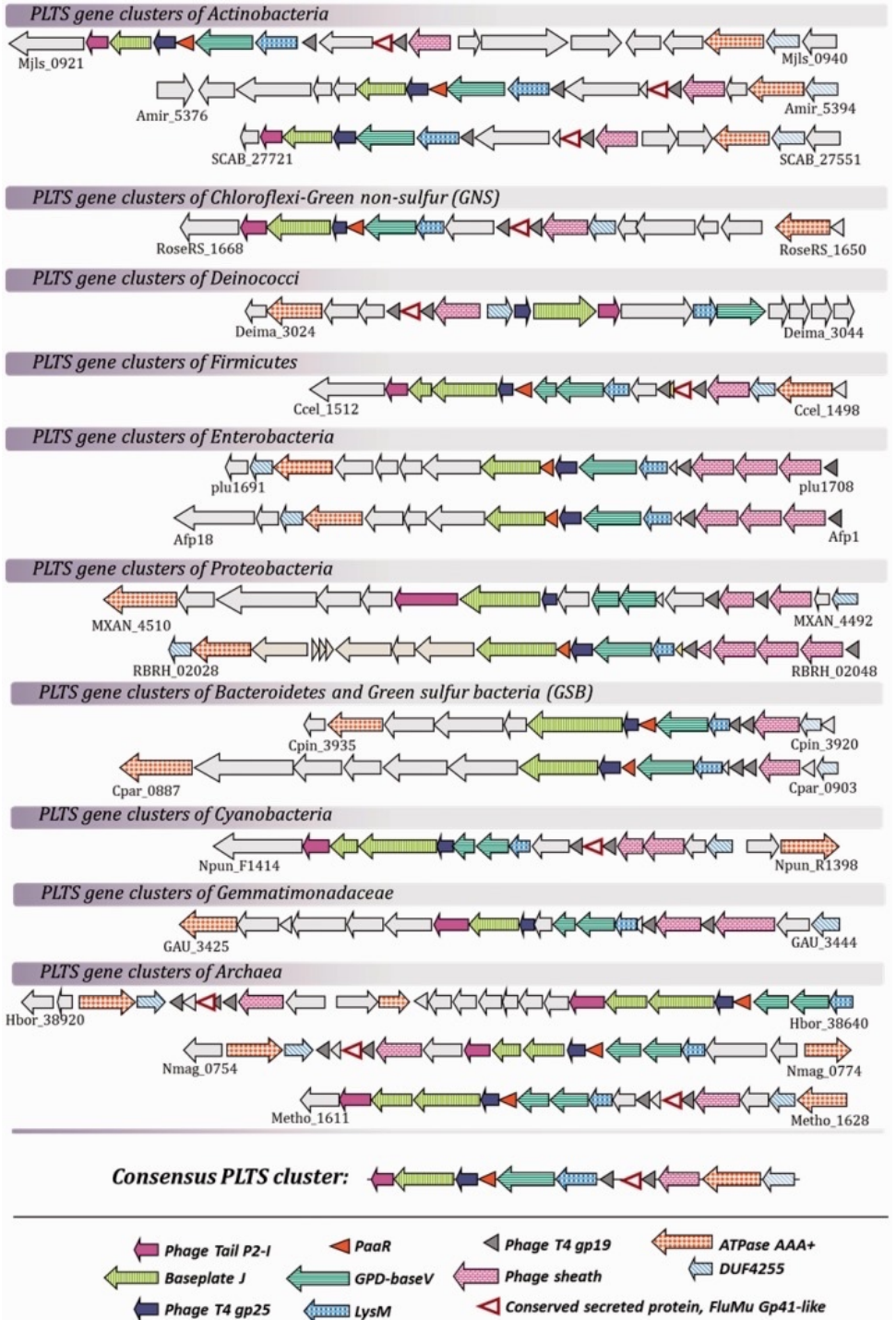
Quick Project Refresher: PVCs



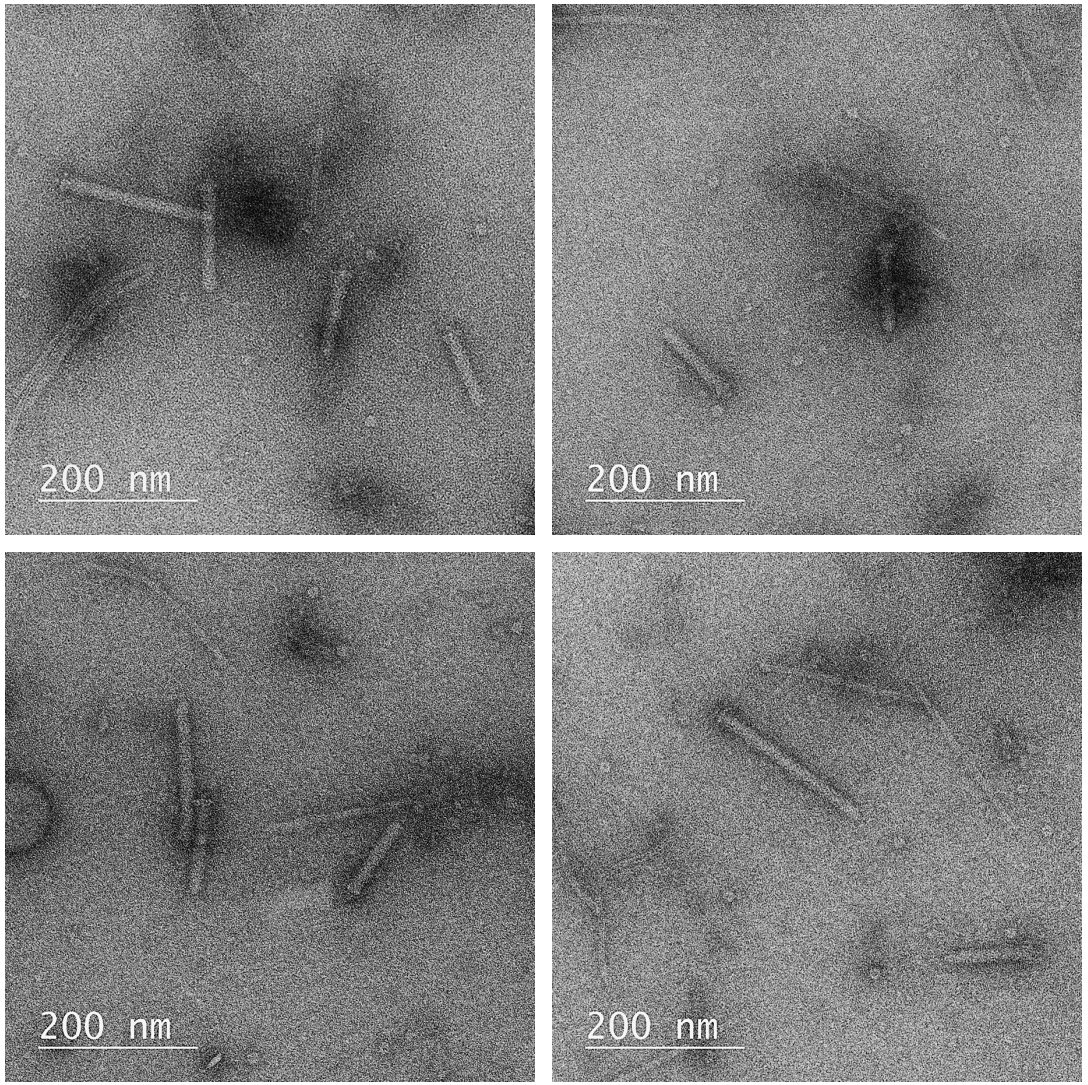
Quick Project Refresher: PVC Genetics

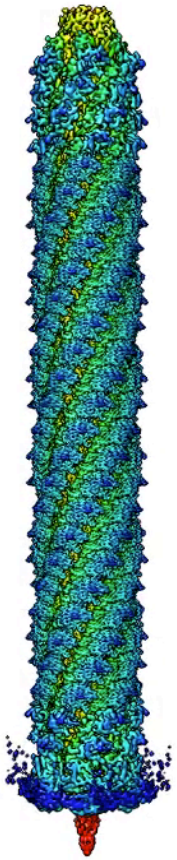


Quick Project Refresher: PVC Genetics



Quick Project Refresher: PVC Structure



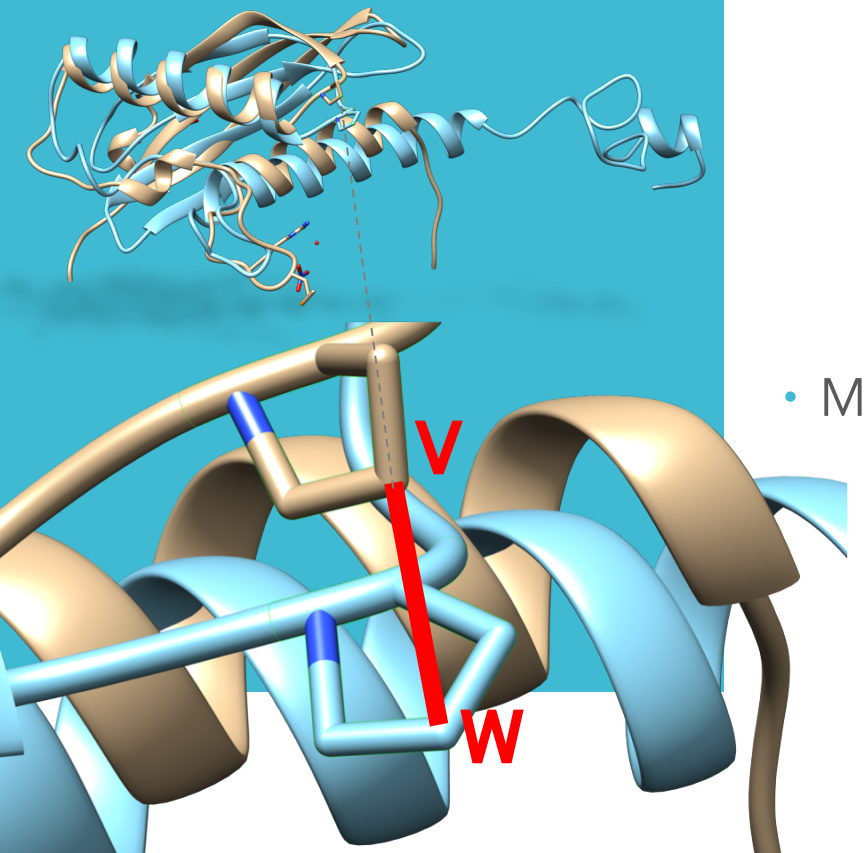


In silico Structural Work

Simulations

- 11 MRC CLIMB VMs calculated theoretical structures in parallel.
  - 352 vCPUs, >1TB RAM
- Simulated (almost) every protein from 16 different PVCs + one XVC
- Of all simulations:
  - 332/344 structures completed successfully.
  - Of the failures:
    - 9 had ambiguous amino acids (could possibly be manually fixed in future)
    - 3 were too long (>1500AA – 2x LUMT, 1x ssel toxins)
- Begun various analyses of these structures to decide how useful they are (there will be a whole spectrum of validity)

Simulations: RMSD Calculation



- A method for broadly assessing the 'sensibleness' of each model.
  - Calculates 3D Euclidean distance between 2 points (atoms)
- Used a UCSF Chimera python wrapper ("Pychimera") written by a (super helpful) student in Spain (https://github.com/insilichem/pychimera).
- Python script of my own:
  - Calls HHPred to HMM profile the amino acid fastas
  - Retrieves the closest known published structure from PDB (NB, closest ≠ good)
  - PDB Reference structure and 5 models from each simulation loaded in to Chimera via pychimera
  - Calculate the RMSD of each model vs. reference. Lower the RMSD, better the 3D fit (though RMSD's usefulness is debated).
- May test the fits further in future with COFACTOR and TM-align.

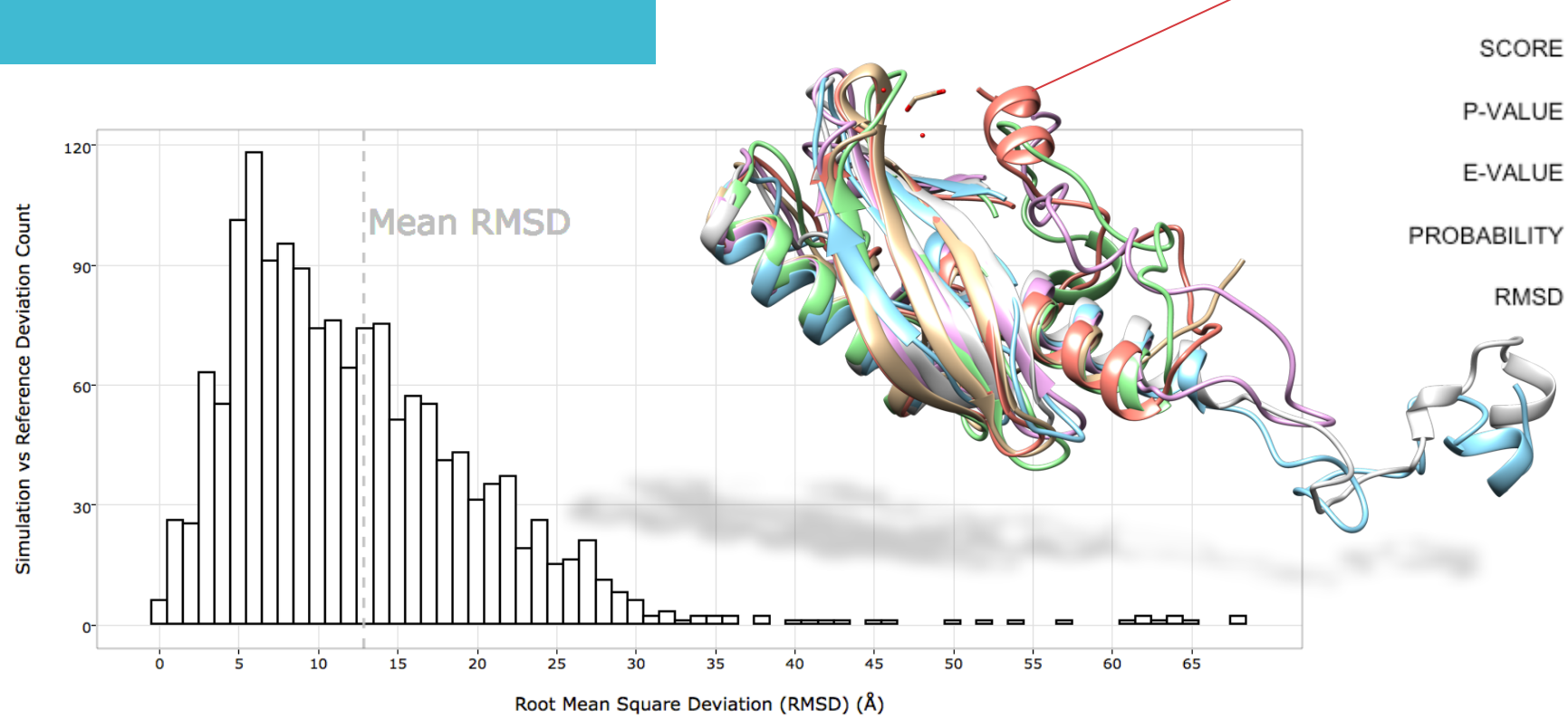
$$\mathrm{RMSD}(\mathbf{v},\mathbf{w}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|v_i - w_i\|^2}$$

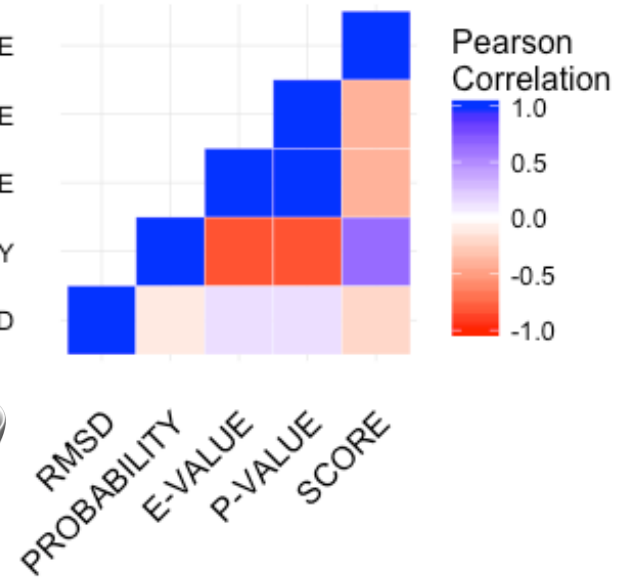
$$= \sqrt{\frac{1}{n}\sum_{i=1}^{n}((v_{ix}-w_{ix})^2 + (v_{iy}-w_{iy})^2 + (v_{iz}-w_{iz})^2)}$$

Simulations: RMSD Calculation

- End result is a list of reference PDBs, models, RMSD and other stats:

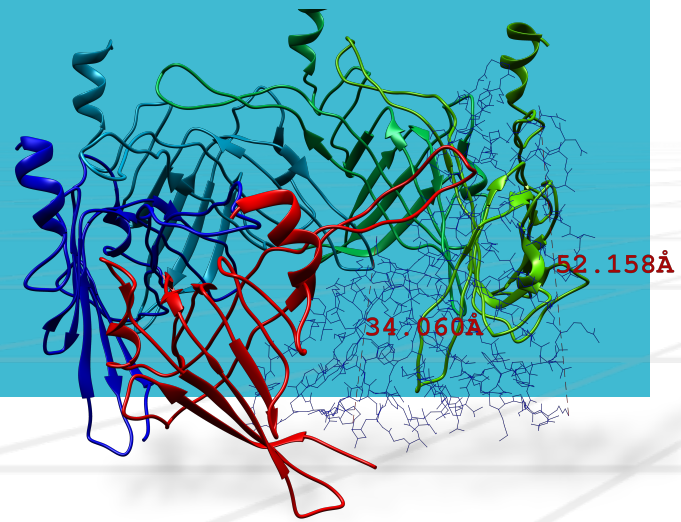
BEST HIT	LOCUS/MODEL	RMSD	PROB	E-VAL	P-VAL	SCORE
...						
3j9q	PAK_03201_model5.pdb	20.6238008807	99.9	8.7e-33	2.3e-37	239.9
3j9q	PAK_03201_model2.pdb	7.06613161922	99.9	8.7e-33	2.3e-37	239.9
2ia7	PAU_01969_model3.pdb	3.09411432359	99.7	8.1e-22	2.1e-26	132.1
2ia7	PAU_01969_model4.pdb	1.57304926279	99.7	8.1e-22	2.1e-26	132.1
2ia7	PAU_01969_model1.pdb	5.19152804989	99.7	8.1e-22	2.1e-26	132.1
2ia7	PAU_01969_model5.pdb	4.7178320049	99.7	8.1e-22	2.1e-26	132.1
2ia7	PAU_01969_model2.pdb	3.84606090378	99.7	8.1e-22	2.1e-26	132.1
3gkn	PAU_01980_model5.pdb	5.70481991885	37.8	13	0.00034	22.7
3gkn	PAU_01980_model2.pdb	6.01218642657	37.8	13	0.00034	22.7
...						

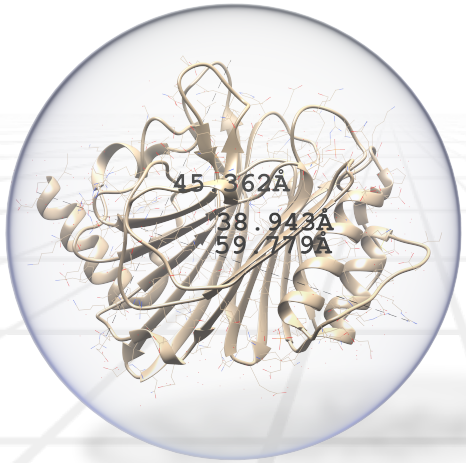


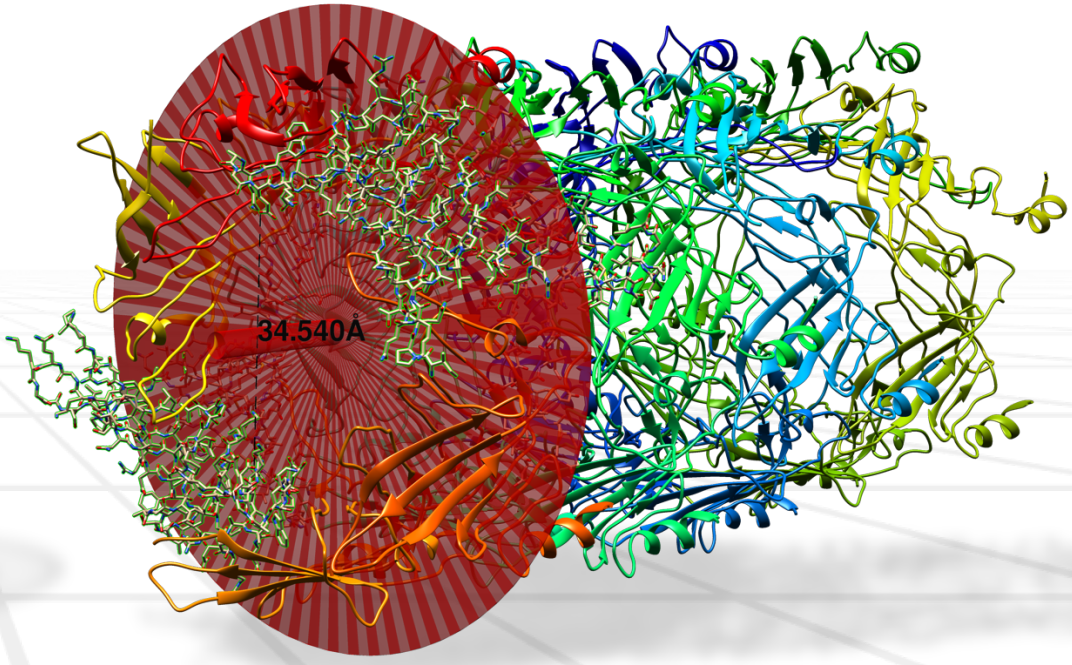


Simulations: Atomic Measurements

- Slightly academic until we can obtain a reliable EM density map of our own but…
- HHPred identified some very good solved structures (Ge et al's Pyocin).
- For FutuRx (more later possibly), we keep being asked about the dimensions.
- Need to make as accurate as possible measurements from EM images (tube length, total capacity etc).
  - Likely that effector molecules are present inside the tube semi- or unfolded







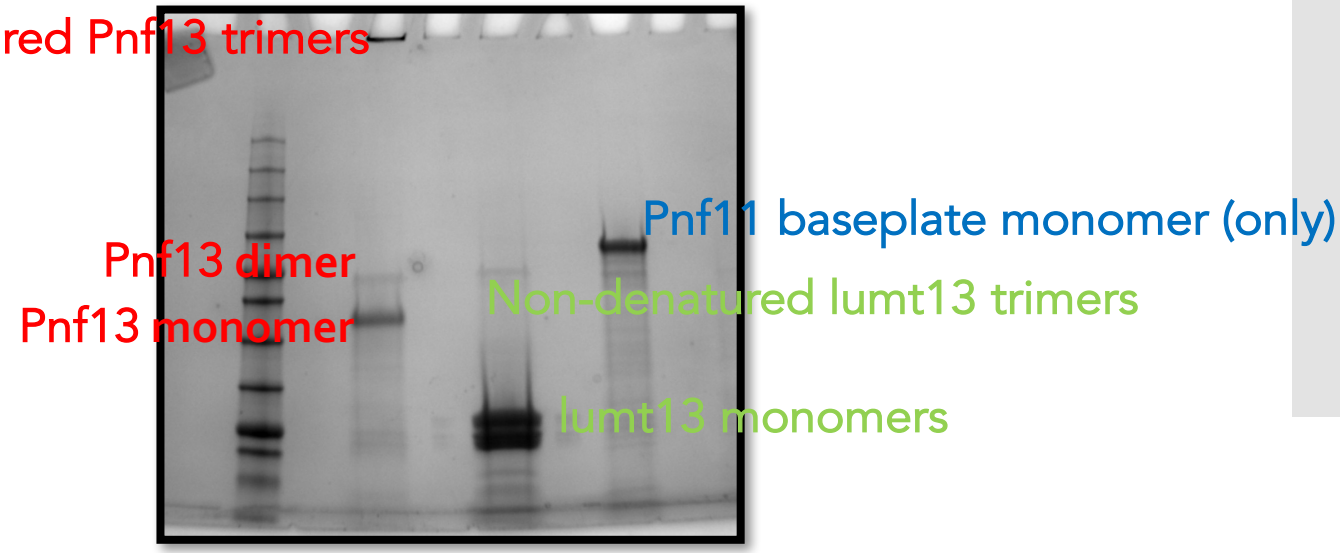
Simulations: Next Steps

- Repeat some analysis once a few simulations have re-run. Run RMSDs for effector molecule simulations
- Continue to work out molecular measurements.
- Plan to do CD-HIT clustering of the effector sequences, to see if they fall in to separate classes.
  - Maybe do the same for effector N50 sequences which seem to be variable/conserved differentially – loading signal?

Lab Structural Work: PVC tail fibres

Tail Fibre Work

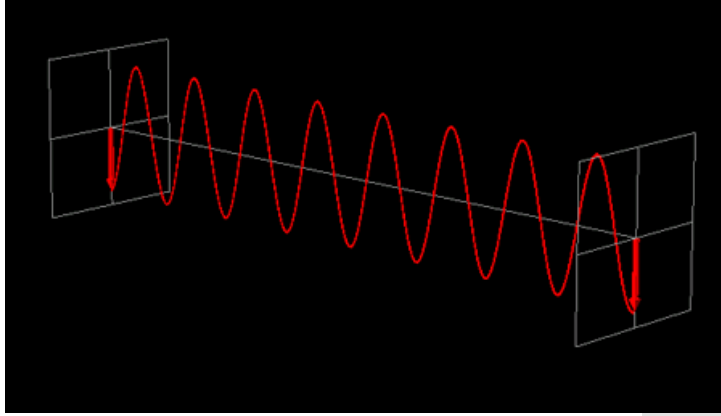
- RMSDs for tail fibre simulations is unsurprisingly poor (avg=28Å, range=61Å)
  - Mainly because they're trimers, and only monomers can be simulated (though there's still some useful info in the simulations)
  - Only recourse is to try and do the structure!
  - Useful because they can't be resolved in EM due to mobility (Ge structure for example (AKA good thesis narrative…)
- Tail fibres are clones and 6xHIS tagged. They express and can be purified reasonably well (tested IMAC and AmSO4 Precipitation).
  - C-terminal HIS tags don't express interestingly...



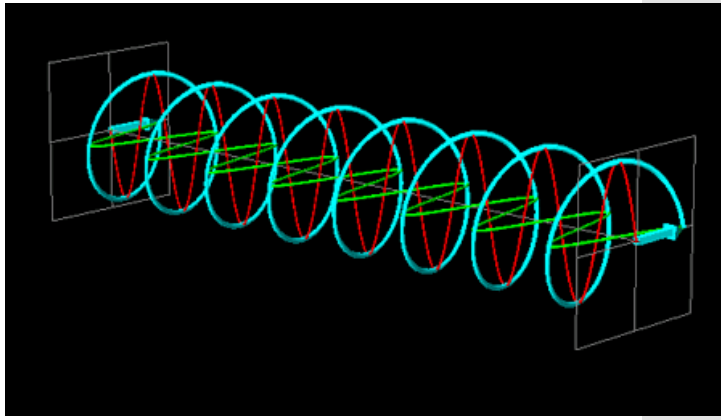
Tail Fibre Work: CD

- Noticed when trying to do SDS-PAGE they are staggeringly stable (8M Urea and boiling in SDS buffer insufficient for full denaturing).
  - Decided to quantify this further with Circular Dichroism
- Circular Dichroism is a spectroscopic technique which measures differences in absorbance of differentially polarised light

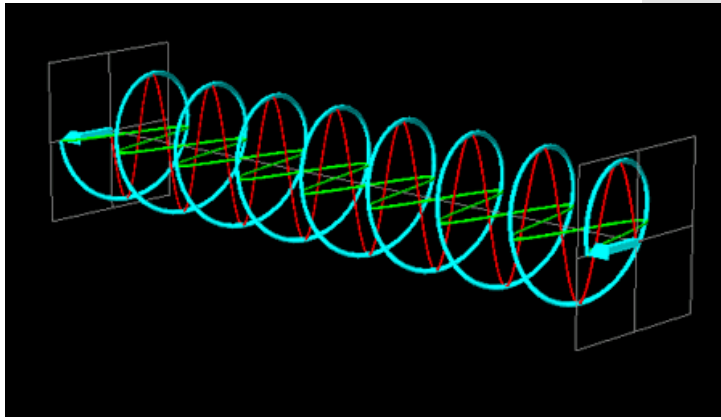
$$\Delta A = A_L - A_R$$



Vertically Polarised (Linear)



Circularly Polarised (Left)



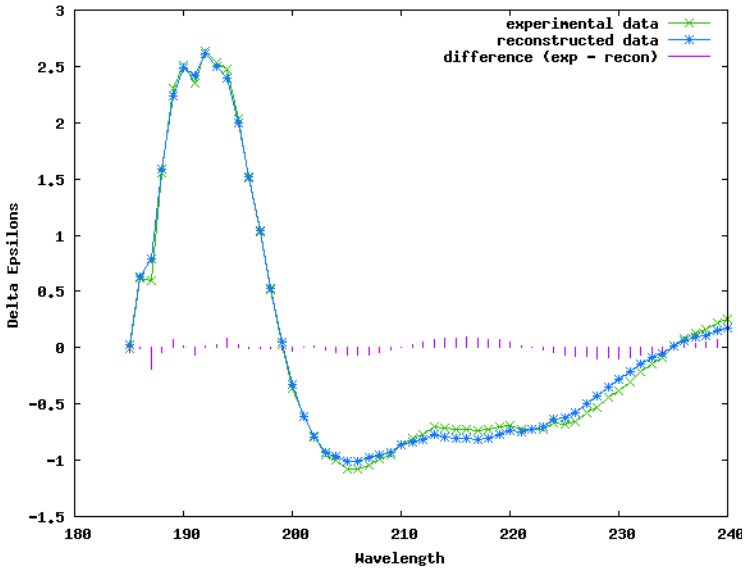
Circularly Polarised (Right)

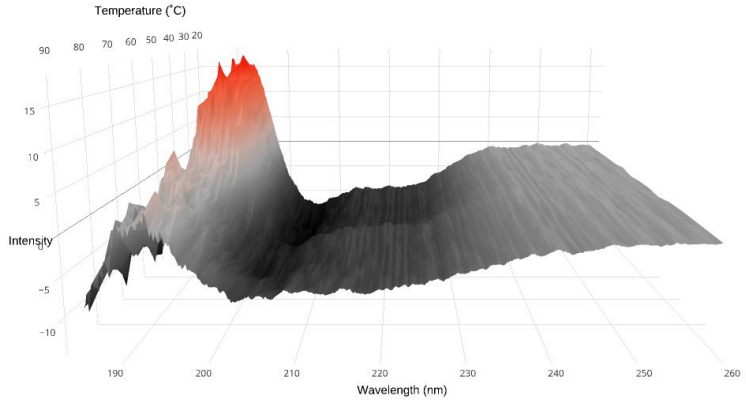
Tail Fibre Work: CD

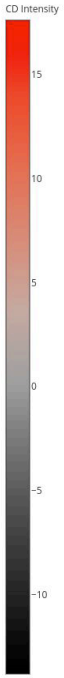
- Then you end up with a spectrum which can be analysed for 2 things:
- A web service called DichroWeb gives you secondary structure prediction (fibres are low in helix consistent with expectation)

Helix1	Helix2	Strand1	Strand2	Turns	Unordered	Total
0.01	0.04	0.23	0.13	0.21	0.36	0.98

- And the spectra (plus how well it fits known reference sets)*:



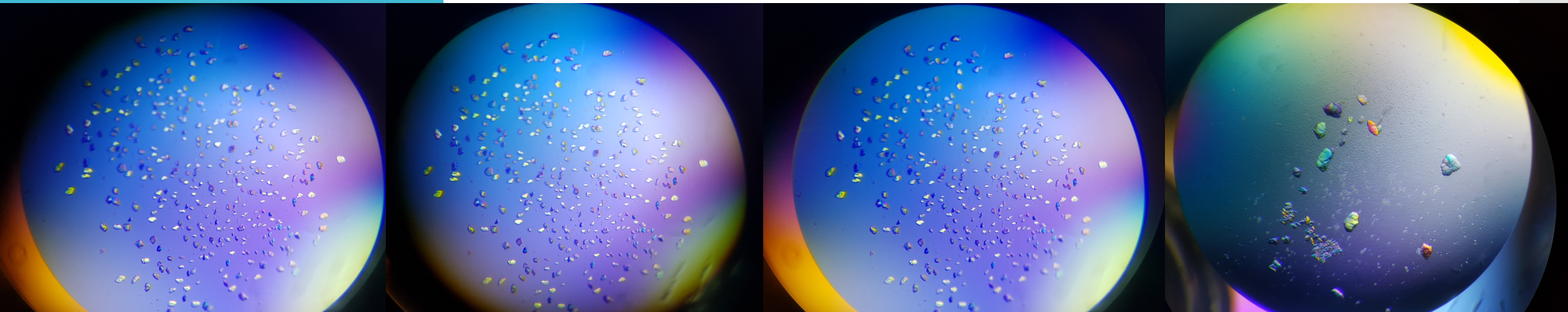




- And you can run automated temperature ramps to assess stability

Tail Fibre Work: Crystallography

- Obtained crystals of Lumt13 surprisingly easily.
  - Crystals in 12 out of ~300 conditions screened (common to not get even 1!)
  - Crystals reached a good size in less than a week (can take months).
- But that run of luck is far too much for lab work.
  - They are quasi-crystalline (jelly like) and don't diffract sufficiently.
- There are some further optimisations to make.



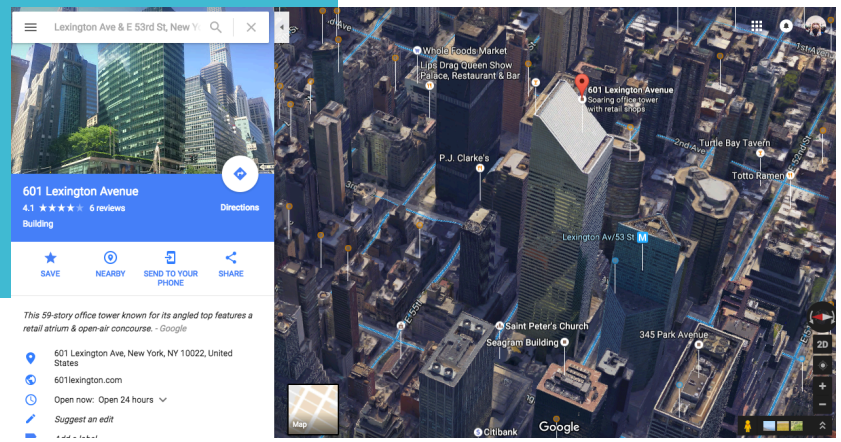
Tail Fibre Work: Next Steps

- Actually need to re-export data for dichroweb as it's not baseline subtracted (not an issue for my own plots)
- Repurify PNF13 & 11 and perform CD
- Repurify everything and do gel-filtration with *in-situ* proteolysis to try and get diffracting crystals
- Structural biology are also buying a nanothermophoresis machine which gives more resolution to unfolding studies, so could test out that methodology (corroborate the CD data).

Biotech Busy Work

Biotech

- Last time I spoke I mentioned us coming in as finalists in BioStars (though not in depth)
- Since then we had a discussion with Oxford Technologies – though that's all gone unexpectedly quiet.
- We've recently been talking with FutuRx
  - Weird set up – Israeli gov't tender by 3 US biopharma VCs (OrbiMed, Takeda, J&J)
  - If successful, $2million, they hire staff and we license the IP to them and they 'get on with it'.
  - Company is incubated in Israel with us as SAB/Directors etc.
  - 2 conference calls and they're still interested.
  - Trip to NYC in Nov to discuss it with their SAB – last hurdle?
- £30K Warwick Impact Fund



FutuRx
The biotech accelerator