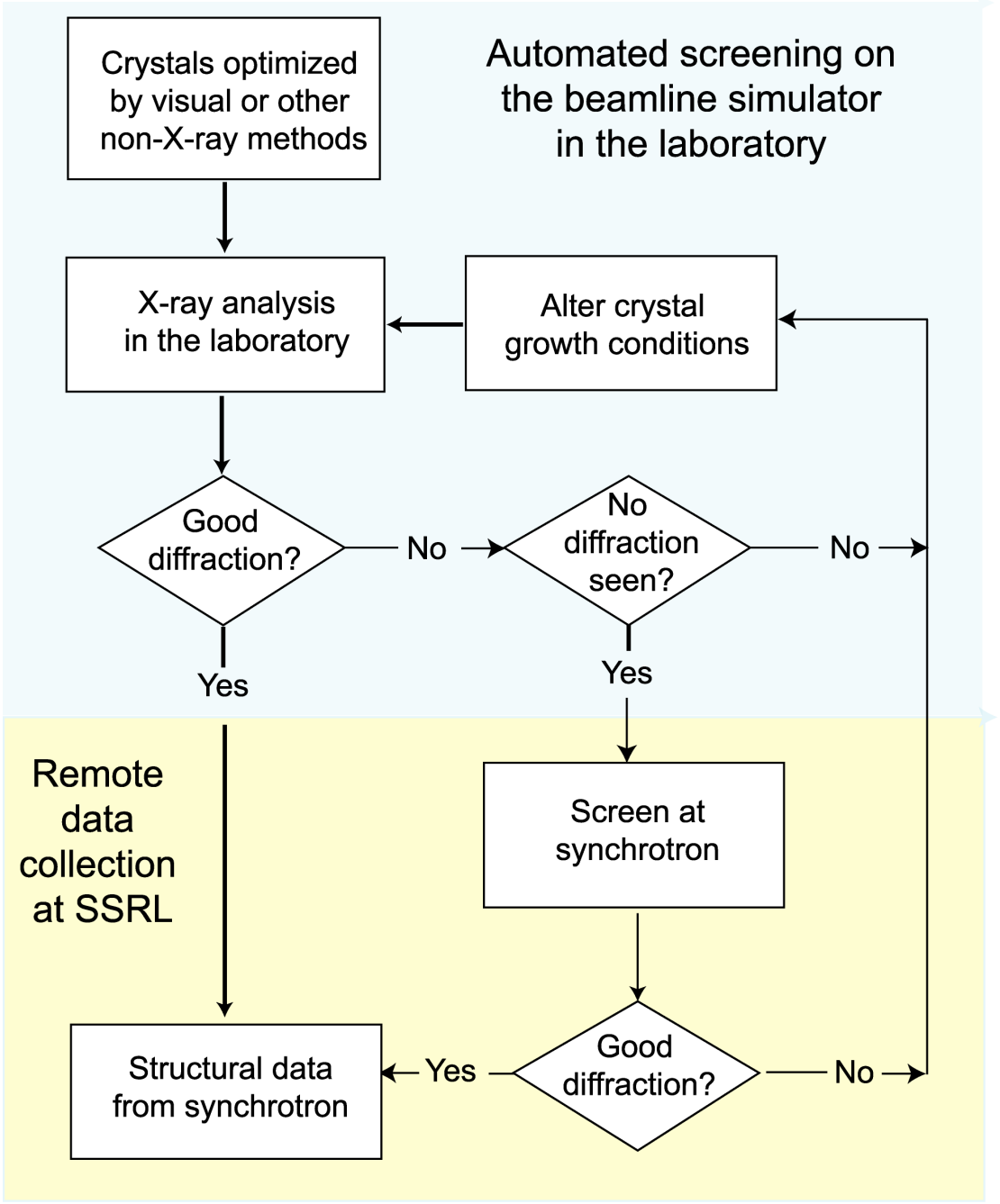
X-ray Data Collection: At HWI and Stanford Synchrotron Radiation Laboratory

The optimization and structural determination step

Introduction

Pipeline from crystal to structure



If the X-ray analysis shows good diffraction the crystal is placed in a cassette to ship to SSRL for the next available beamtime.

If no diffraction is seen the crystal is also placed in the cassette to ship to SSRL.

If diffraction is present but poor, e.g. problems such as twinning, cracking or very high mosaicity then the crystal growth conditions are further optimized. Poor resolution is not regarded as poor diffraction for this purpose.

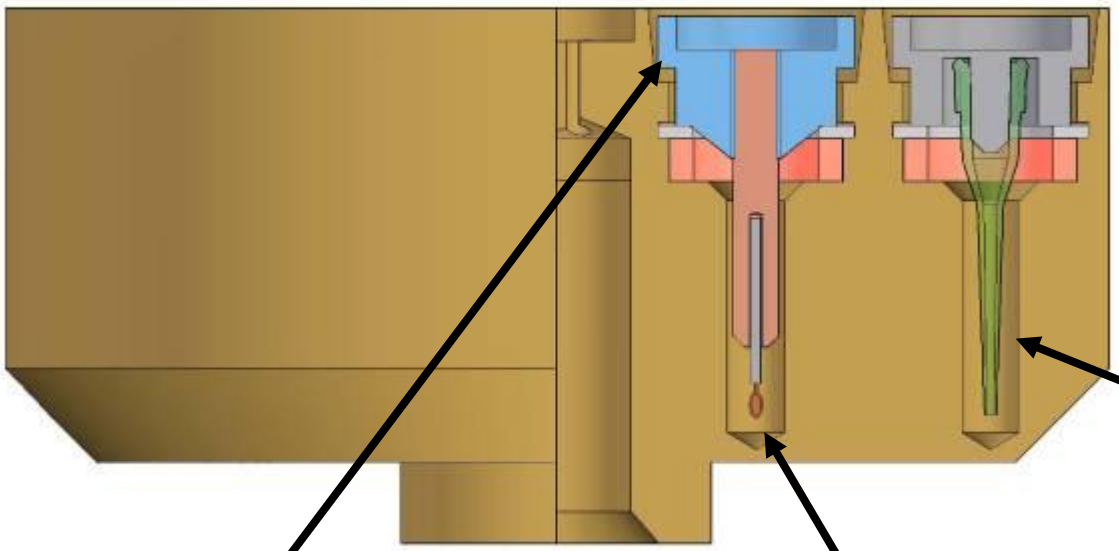
At the synchrotron structural data is collected from the good crystals and used for structural refinement.

For those with no diffraction at home screening takes place at the synchrotron. If good diffraction is present structural data is collected.

Visual Observation

Crystal observation in capillaries

Capillary mounting compared to cryoloop



Universal puck being developed by Stanford Synchrotron Radiation Laboratory (SSRL)

New implementation with the crystal cryoprotected in an X-ray transparent plastic capillary

Magnetic pin mount

Current use of cassette and puck with crystal cryocooled in a loop

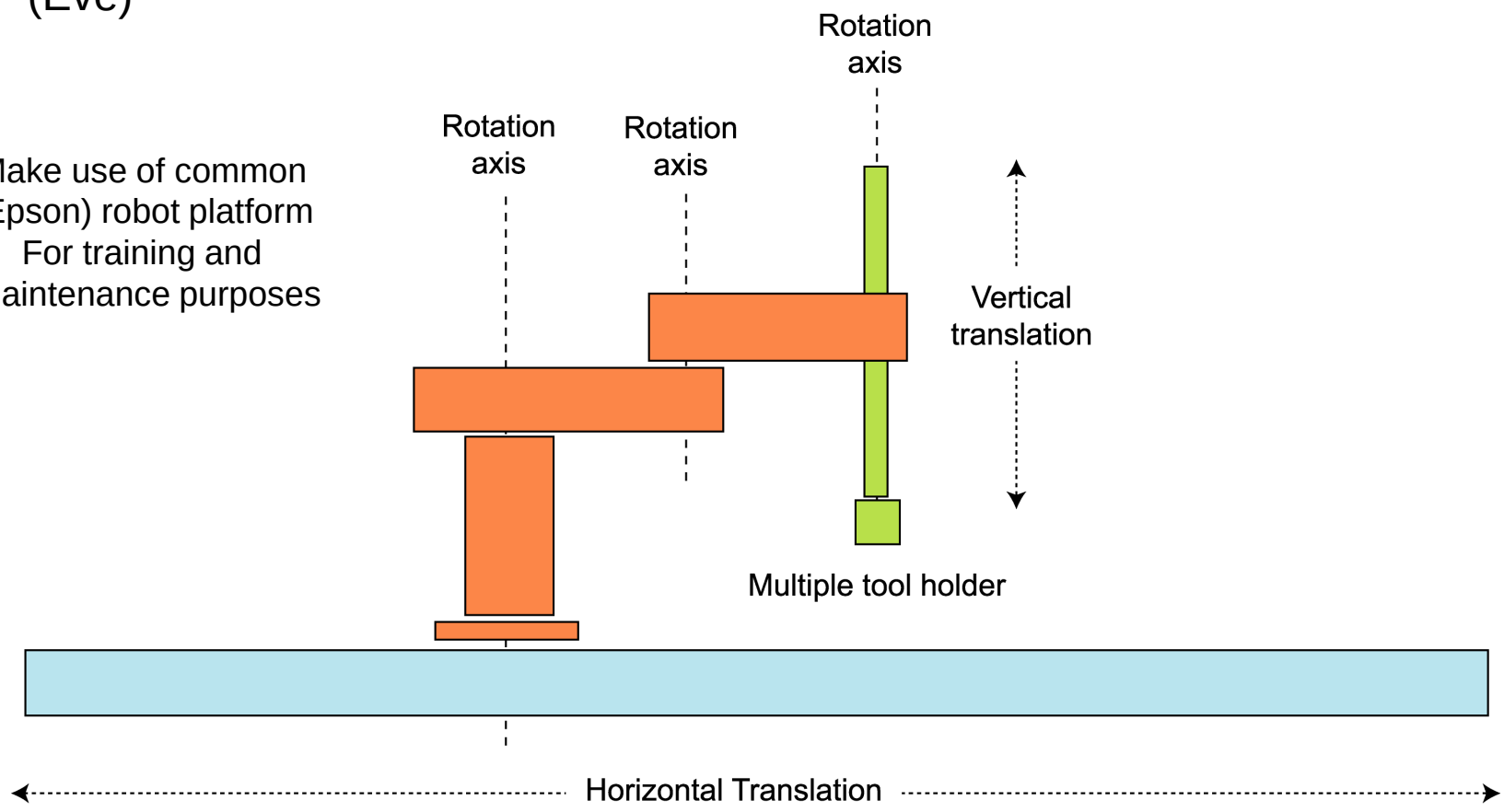
(Illustration from Dr. Aina Cohen, SSRL)

Optimize by visual or non-X-ray methods

Schematic of the robotic solution picker and observation system

(Eve)

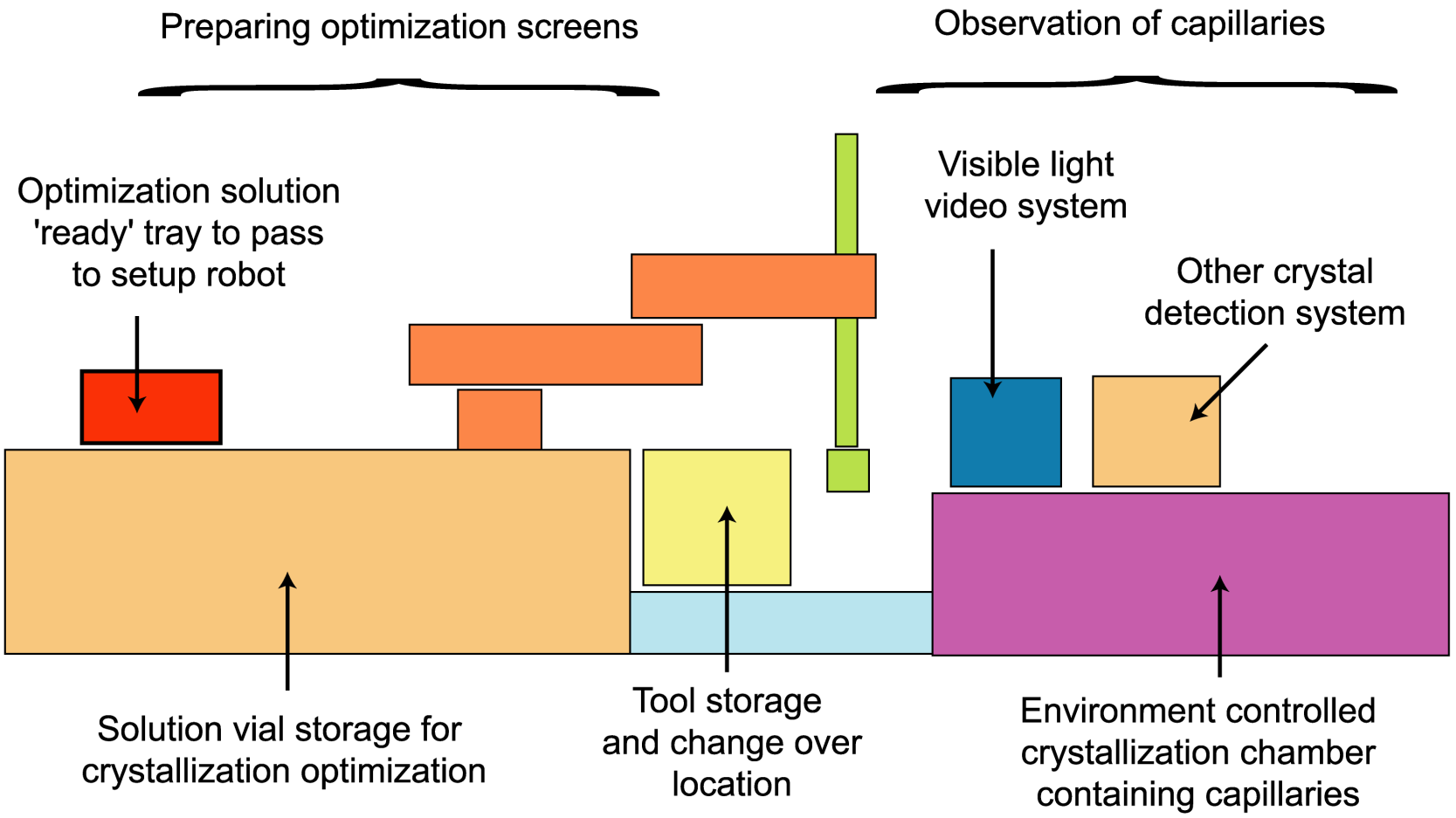
Make use of common (Epson) robot platform For training and maintenance purposes



Not to scale

Optimize by visual or non-X-ray methods

Schematic of equipment associated with the solution picker and observation system



X-ray analysis in the laboratory

Use Existing Developments

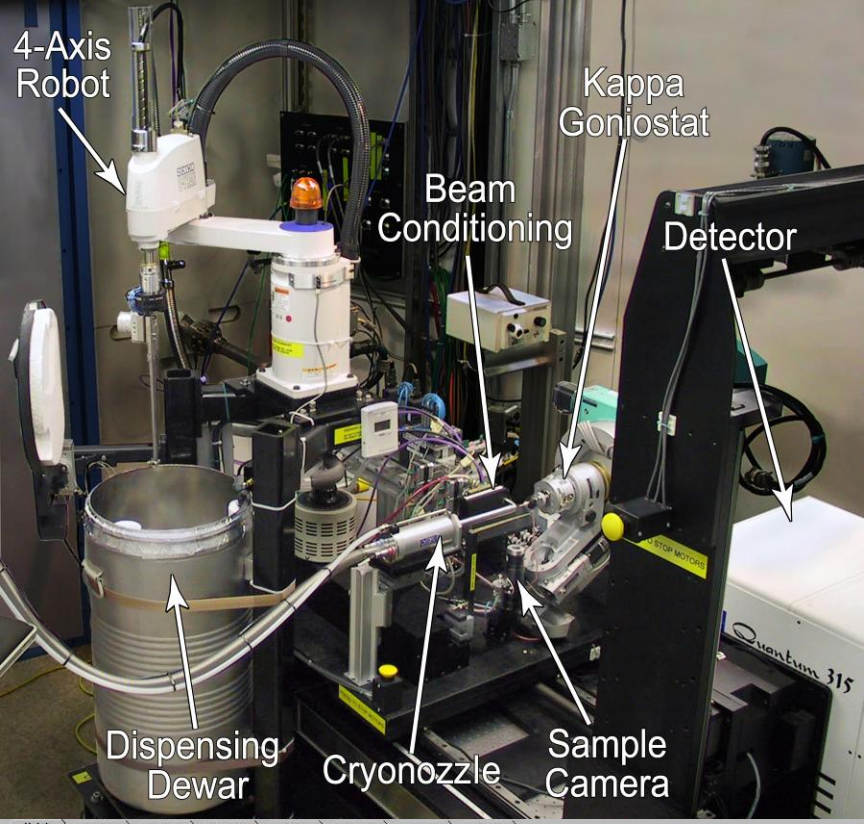
"If something works don't try and fix it"

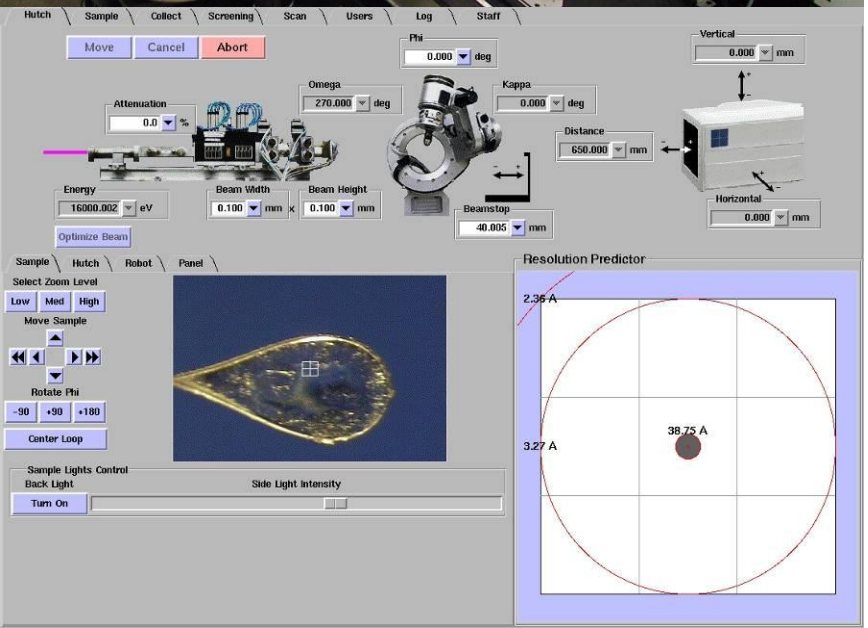
SSRL has developed a very efficient remote data collection system with:

- A reliable sample transport cassette system
- Remote unix clients to run software on SSRL computers
- A Robotic mounting system
- Automatic loop centering
- Integrated screening, data collection and control software
- Remote video systems
- Remote planning and data processing software
- X-ray based crystal centering
- Massively parallel data processing strategy

} Currently being tested

We will make use of these existing capabilities by endeavoring to ensure compatibility with the "tip" mount and the existing cassette and future puck sample systems.





Remote Data Collection Developments at SSRL

A common robotic mounting system, the Stanford Automated Mounting system (SAM).

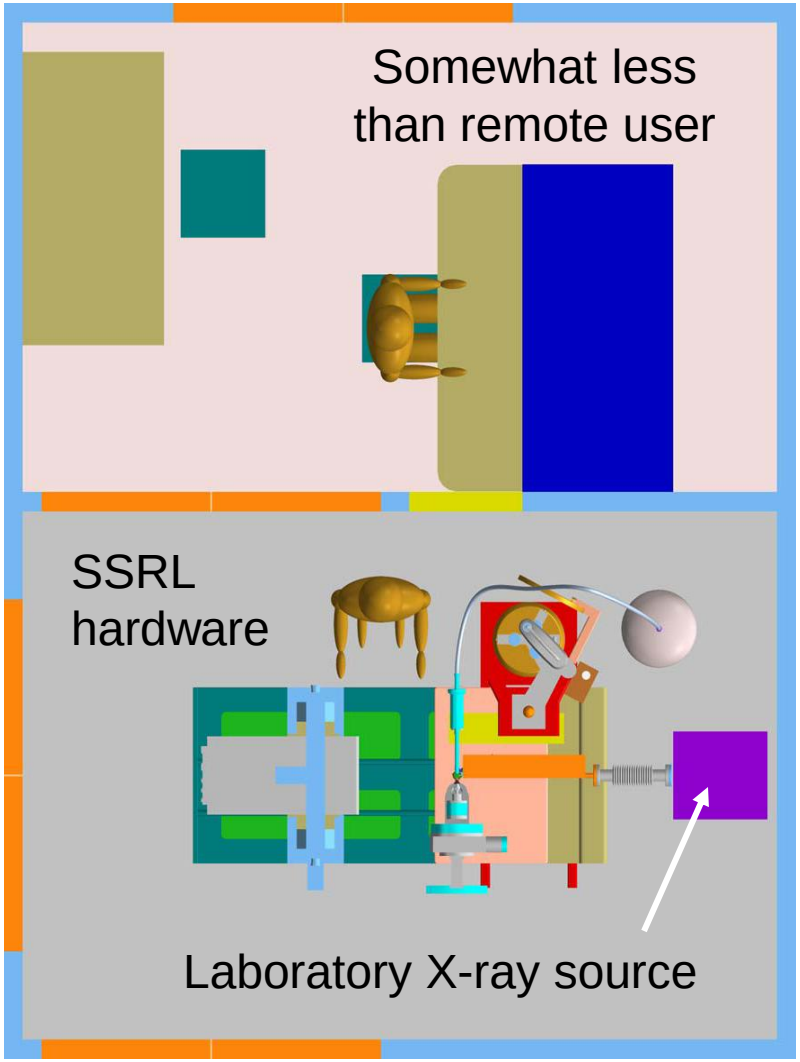
Blu-Ice control, screening and data collection software.

Automated analysis of images as screening takes place.

Web-Ice, a web based interface to the screening and data collection results which includes strategy, experiment optimization and radiation damage determination procedures.

To be covered by Dr. Aina Cohen in the next presentation

The Beamline Simulator (Son of SAM)



(Illustration from Dr. Aina Cohen, SSRL)

Replicate the beamline hardware and software of an SSRL beamline system into the laboratory.

Ensure compatibility at all steps of the process.

Avoid unnecessary development of control, mounting, screening and data collection software.

Train users in efficient beamline use.

See the next presentation by Dr. Aina Cohen.

Thought Experiment

Estimated cycle time

- Robot time is 3 minutes to dismount/mount/center sample.
- Exposure time is longer in the home laboratory, 10-20 minutes per image.
- Typically use 2 images, 90 degrees apart.
- Total maximum time per crystal sample, 43 minutes.
- To screen a single cassette of 96 crystals total time is 68 hours. With a single image this could be cut to 37 hours.
- Initially the system would be low throughput.
- A similar throughput could be achieved for short periods by non-automated means however the robotic system will operate continously for the total screening period.
- Synchrotron time occurs approximately every six weeks – about 14 cassettes worth (~1400 samples) of screening at home.
- Cycle time for a single system is suitable for optimization and the current synchrotron beamtime available.
- From previous experience synchrotron screening time totals about 4 minutes per sample allowing a cassette of 96 samples to be screened in under 7 hours.

Going from development to high-throughput

- A key consideration of technology development is the application of the technology to high-throughput and how to accomplish this.
- For the laboratory screening system we need to decrease cycle time, increase the number of cycles or both.
- The planned development system is based on an RU-200 rotating anode generator (use as much existing infrastructure as possible).
- Microsource generators with new optics increase the flux on the crystal compared to our existing system, offer a smaller footprint and have a reduced maintenance overhead.
- Taking it to high-throughput: Use several microsources and detectors with a single mounting robot (not part of this development but a direction that could be taken).
- With two microsources and detectors a 96 sample cassette can be screened in 34 hours, with three, 23 hours (if we assume a single image, 19 and 13 hours respectively). With three sources and detectors this allows us to study ~2800 (or ~4200) samples between synchrotron runs.
- Can we increase throughput further, how much exposure time is necessary for optimization?

Optimization of cycle time

- Our calculations assume a 20 minute exposure per image.
- If we can reduce this we can increase the number of samples studied.
- To do this we can improve the flux but we also have a number of other possibilities including:

1. An initial image to check that the crystal is not salt.

   Mar detector has 34 second cycle time in fastest mode. Use a 1 minute exposure, several degree oscillation and rule out salt. This decreases the time taken to screen the crystal from 43 minutes to 5 minutes for a salt crystal (including sample mounting and dismounting). Once confirmed as protein this step is not used.

2. Use of a shorter exposure time.

   It may be possible to make use of a shorter exposure time remembering that crystals that do not show diffraction still end up at the synchrotron. There will have to be a balance between the available synchrotron time and the exposure time cutoff used at home. This will be determined empirically.

3. Feedback of exposure time from previous results (a database function).

   Crystal optimization is a multistage process whereby the data from one crystal of a sample is ranked against other crystals of the same sample and the best crystals sent to the synchrotron. Once X-ray data from one crystal is recorded we will know if it is a strong diffractor or not and the exposure time for corresponding samples can be adjusted accordingly. Similarly other experimental parameters could be optimized, i.e. oscillation angle and crystal to detector distance. This has the highest potential but is the most complex solution.

Possible problems

- Centering the crystal – The Stanford Case
  - It is not known where the crystal is in the loop but loops are chosen to match crystal size.
  - The system uses automated loop centering, the crystal is not centered.
  - Variables include illumination and magnification.
  - Recent developments include the ability to perform on demand X-ray based centering. This is typically used after loop centering has taken place and is more time consuming.
- Possible problems
  - Our crystal is in a tip. It can be anywhere along that tip.
- Solutions
  - The tip has already been imaged during crystal growth so we know approximately where the crystal is.
  - We also know that we are looking for a crystal not a ‘hit condition’.
  - Visual edge detection becomes possible.
  - Offline light absorption or in situ X-ray absorption measurements are also possible and easily integrated into the control software.

Current state

- Eve, the solution picker and observation system
  - Eve is here, the robot hardware is being installed.
  - The barcode reader for optimization solution tracking and the visual crystal imaging system are also here ready to be installed.
  - Tools are still in the design stage.
  - Software control and database interaction have still to be developed.
  - The overall scheme of operation is developed but the specific details are not.
- Son of SAM, the beamline simulator
  - Detail in the next talk
  - HWI implementation will be placed on an existing rotating anode generator (non-ideal).
  - Improved optics will be ordered after this meeting.
  - A MAR345 detector on loan from SSRL is ready to be installed.
  - The detector table will be modified as not as many motions are needed.

X-ray analysis at the synchrotron

Results from remote data collection at SSRL

Preparation: Training of Staff

- A remote data collection workshop was held at HWI on August 4th.
- Workshop consisted of lectures on preparation of samples, beamline control, screening, data collection and data processing.
- In the afternoon practical sessions covered sample loading and actual control and data collection from two beamlines at SSRL.
- The workshop was open to the community and attendees included 16 members of HWI scientific staff and 14 from other institutions.
- At SSRL two representatives from the European Synchrotron Radiation Facility attended to watch the process from the beamline end.
- During each remote run experienced staff at HWI are on call as a filter to SSRL user support for new and inexperienced HWI users.

Use of Robotic Data Collection to Date

Date	Cassettes, samples and beamline.
May 21st - 23rd 2004	Beta testing of robot at SSRL, 8 crystals* (Beamline 11-1).
Jul 12th - 13th 2005	1st HWI robotic collection – 2 cassettes, 160 samples (Beamline 9-2)
Jan 27th - 28th 2006	1st Center Robotic Data Collection, 1 cassette, 86 samples.
Mar 1st - 3rd 2006	2 cassettes, 162 samples (Beamline 9-2)
Apr 12th - 13th 2006	1 cassette, 76 samples (Beamline 9-1)
Apr 19th - 21st 2006	1st pipeline sample, 2 cassettes, 171 samples (Beamline 11-1)
May 9th - 11th 2006	1 cassette, 91 samples (Beamline 11-1)
Jun 5th - 7th 2006	3 cassettes#, 141 samples (Beamline 11-1)
Jun 30th - Jul 3rd 2006	3 cassettes, 194 samples (Beamline 9-1)
Jul 19th - 21st 2006	1 cassette, 69 samples (Beamline 9-1)
	Total cassettes 16
	Total samples 1150

* Testing of robot during non-robotic time. Early problems were discovered.

Includes a cassette left from the previous run for further study

Testing during regular, non-robotic beamtime. Beamtime as part of HWI proposal and as part of the NIH center grant

Result Highlights

- *Pneumocystis carinii* Dihydrofolate Reductase
  - The causative agent *of Pneumocystis carinii* pneumonia in patients who are immune compromised. Inhibitor bound complexes will help develop more selective inhibitors for treatment of AIDS-related pneumonia.
  - Five data sets of pcDHFR inhibitor complexes were collected from 8 of the crystals screened. Preliminary electron density maps revealed density for the inhibitors and cofactor, NADPH. Refinement is in progress.
- Beta Sliding Clamp Protein
  - The beta subunit of *E. coli* replicative DNA polymerase III holoenzyme is a sliding clamp that interacts with the alpha polymerase to maintain its high processibility of the enzyme.
  - Diffraction data were collected to 1.76 Å resolution. The final stages of refinement are in progress. Preliminary reports of these data were presented at the ACA meeting in Hawaii.
- Histidine Triade Nucleotide Binding Protein (Hint)
  - Histidine triad enzymes are a superfamily of nucleoside monophosphate hydrolases and transferases containing an active site motif related to His-X-His-X-His-X-X where X is a hydrophobic residue. *E. coli* Hint proteins have been shown to have purine nucleoside phosphoramide substrate activity, unlike the homologous protein from rabbit. Structural data are needed to understand differences in their mechanistic properties.
  - Crystals of ecHint-GMP complexes were screened and obtained higher resolution data than were previously collected. Crystals of a series of mutant complexes were also screened at the June 2006 run that resulted in the collection of two data sets with resolutions from 1.7–1.5 Å. Refinement is in progress. In the case of the mutant complex, the lattice is half the size of the native data set.

Vivian Cody

Summary:

78 crystals screened resulting in 12 data sets. 10 structures are under refinement representing two new protein families and one continuing project

- Barrier-to-autointegration factor
  - Barrier-to-autointegration factor (BAF) is a host cell protein that plays a crucial role in retroviral integration. Preintegration complexes (PICs) stripped of BAF lose their normal integration activity, which can be restored by incubation with purified BAF. BAF bridges double-stranded DNA both intra- and intermolecularly in a non-sequence-specific manner, leading to the formation of a nucleoprotein network. BAF also binds to the nuclear protein lamina-associated polypeptide 2 (LAP2), and is localized with chromatin during interphase and mitosis.
  - Diffraction data extended resolution to 1.05 Å in order to resolve differences between NMR and X-ray studies. Refinement in progress.
- Short Chain Oxidoreductase (SCOR) proteins
  - The short-chain oxidoreductase (SCOR) family of enzymes includes over 6,000 members identified in sequenced genomes. Of these enzymes, 300 have been characterized functionally, and the three-dimensional crystal structures of 40 have been reported. Since some SCOR enzymes are steroid dehydrogenases involved in hypertension, diabetes, breast cancer, and polycystic kidney disease, it is important to characterize the other members of the family for which the biological functions are currently unknown and to determine their three-dimensional structure and mechanism of action.
  - Four new structures determined including apo and co-factor bound enzymes.
  - Diffraction data from two other SCOR proteins collected which confirmed that the crystals were protein and which will lead to new structure determinations
    - Complete data to 2.7 Å for one, and a preliminary 4.5 Å data set for the other protein.

Tim Umland

Summary:

1 new high resolution structure, 4 new structures determined, and 2 new structures are in the pipeline.

- *Escherichia coli* enterobactin synthetic cluster (EntA-F)
  - The Escherichia coli enterobactin synthetic cluster is composed of six proteins, EntA-EntF, that form the enterobactin molecule from three serine molecules and three molecules of 2,3-dihydroxybenzoic acid (DHB). EntC, EntB and EntA catalyze the three-step synthesis of DHB from chorismate. Structures of three of these proteins have been solved.
  - 51 crystals have been screened. MAD Structure (3.0Å) of one protein recently determined and native data to 1.8Å.
- PvcA and PvcB
  - 23 crystals have been screened. The MAD structure was solved to 2.6 Å from one of the proteins under study. Good native data from both.
- Luciferase
  - 11 crystals were screened. Structural data was collected to 2.8 Å showing a new inhibitor bound.
- CBAL
  - 14 crystals screened, structural data of the protein bound to new ligand resulted.
- AAE
  - 2.0 Å native data set collected
- PA4078
  - Low resolution diffraction confirmed crystals are protein.

Andrew Gulick

Summary:

2 New MAD Structures

2 As yet unphased native data sets

1 Structure of New (to our lab) protein-ligand complex

4 Structures of Protein-ligand complexes (previously studied in our group)

Results Summary

- Structural results
  - New structures in refinement: 12
  - Extended existing resolution resulting in publishable result: 5
  - Structures with new ligands bound: 16

- Results enabling further study
  - New crystal samples determined to be protein*: 5
  - Screening of samples to optimize conditions*: 6
  - Extended resolution*: Most samples

- Preparing for the pipeline
  - Proteins from the NIH Center Pipeline: 1(2)

- Methodology
  - Free radical scavengers studied: 4
  - Cryoprotectants studied: 4

*Specific protein samples, not individual crystals.

Key Center Result

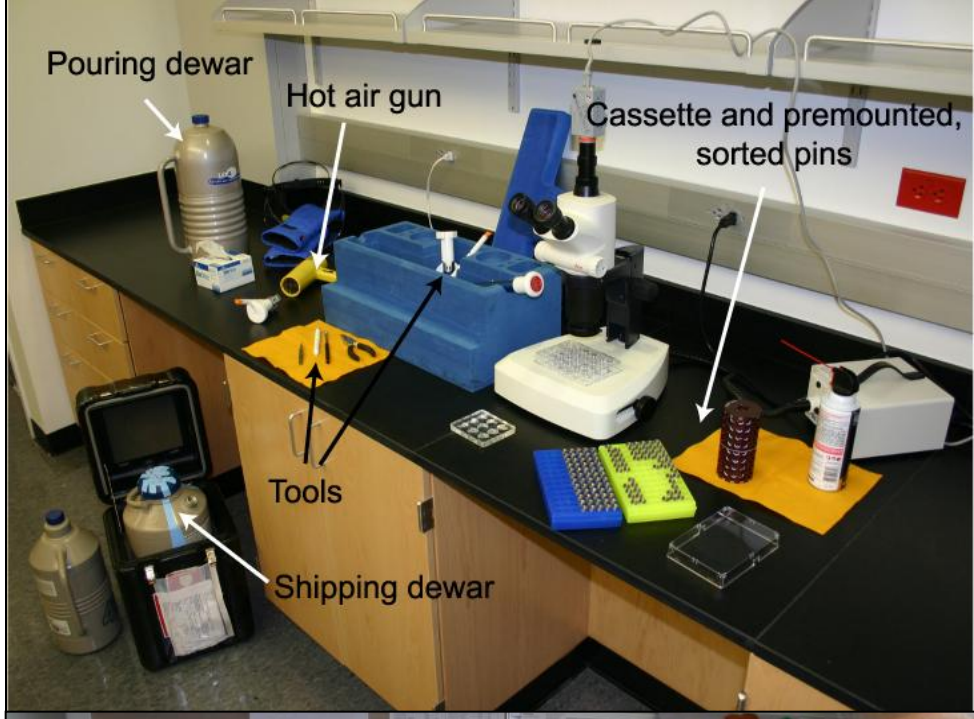
- First structure from the pipeline, WRS1, demonstrating protein production, SeMet incorporation, crystal hit identification, crystal optimization and production, remote data collection and structural solution. See the presentation by Dr. Mike Malkowski.

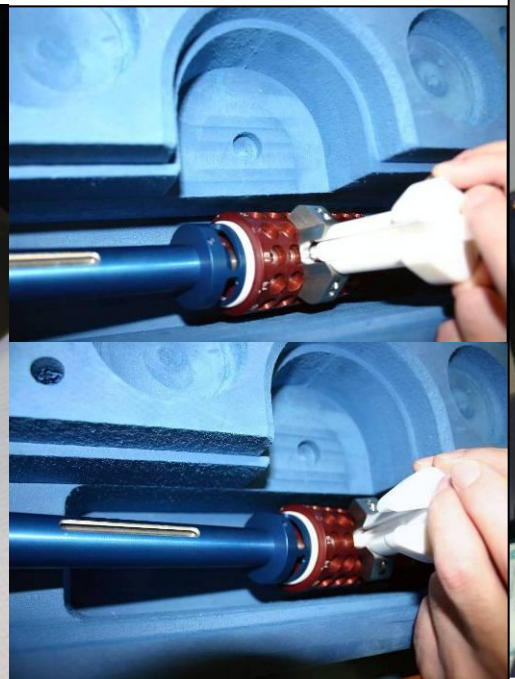
Additional Results

- New crystals of Bovine interphotopeceptor retinoid-binding protein.
- Estrone sulfatase complex extended from 2.60 to 2.10 Å
- COX-2 studies to screen many crystals for optimization of crystallization conditions.
- Similar studies with PIOX

Practical Tips Developed

- Loading in fume hood
- Avoiding icing issues
- Tools and techniques
- Computer setup and communication
- Reusing supplies effectively





Cryopreservation 101

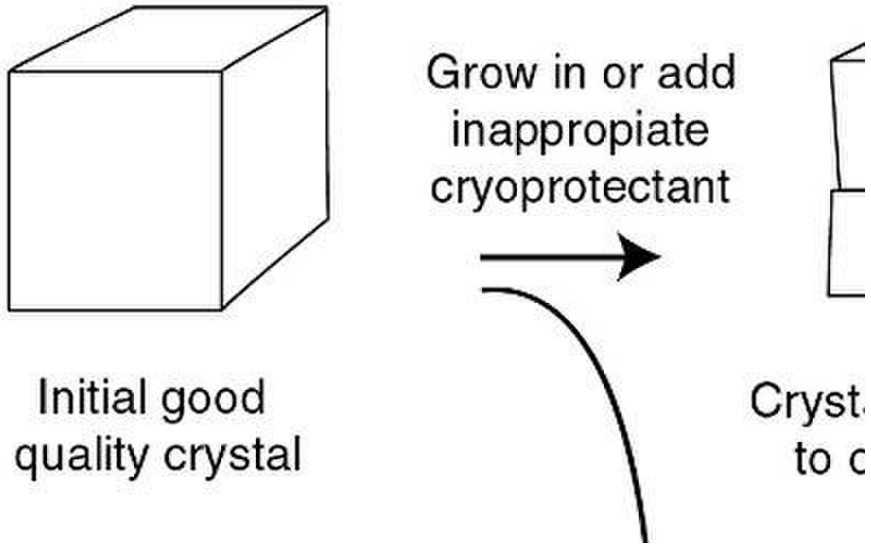
- X-rays are really, really nasty to proteins.
- Primary X-ray damage releases electrons which cause secondary damage through the migration of these electrons and the formation of mobile free radicals.
- The most important samples tend to be the weakest scatters and therefore require the most intense X-ray sources, synchrotrons.
- Synchrotron radiation is 10^9 times more brilliant than the sun and about 100 million miles closer
- X-rays from synchrotrons are really, really, really, really, nasty to proteins.
- We mitigate the secondary damage by cooling the sample to 100K where the free radical mobility is significantly decreased or eliminated.
- This also makes the samples easier to handle and ship than manipulating them at room temperature (?)
- Cryocooling must prevent crystalline ice formation – the crystals must be vitrified.
- Cryoprotectants (e.g. glycerol, propanediol, ethylene glycol etc.) are used to accomplish this

Observation

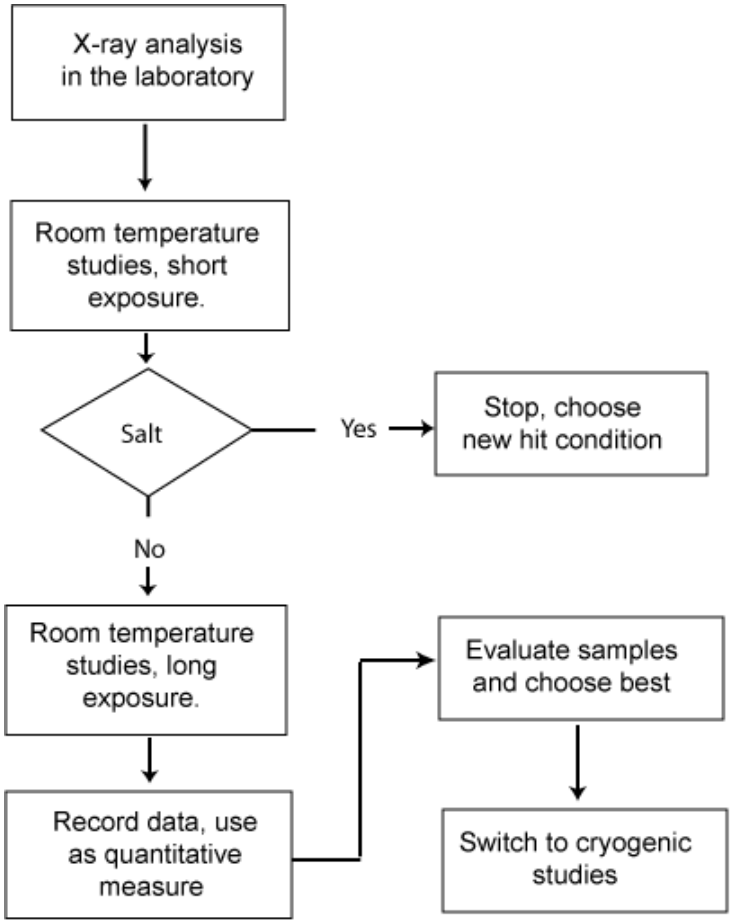
- In a significant number of cases cryoprotectant was used in very high concentrations or added to a system that was already a cryoprotectant.

Assumption (always dangerous)

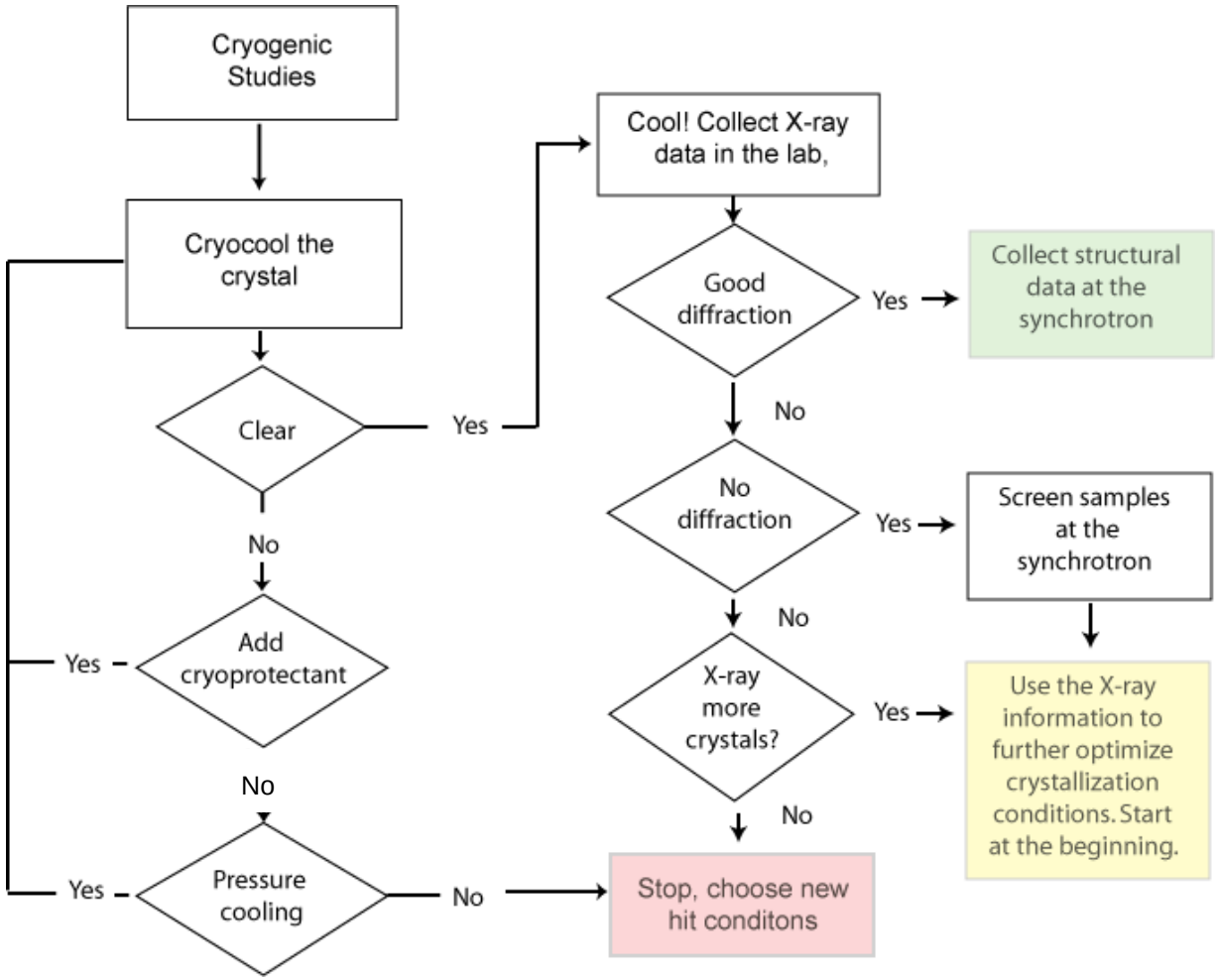
- The best cryoprotectant is the one that works and does not significantly impact the crystallization system, i.e. it is present in the smallest concentration to prevent crystalline ice formation.



The crystal to cryocooling path

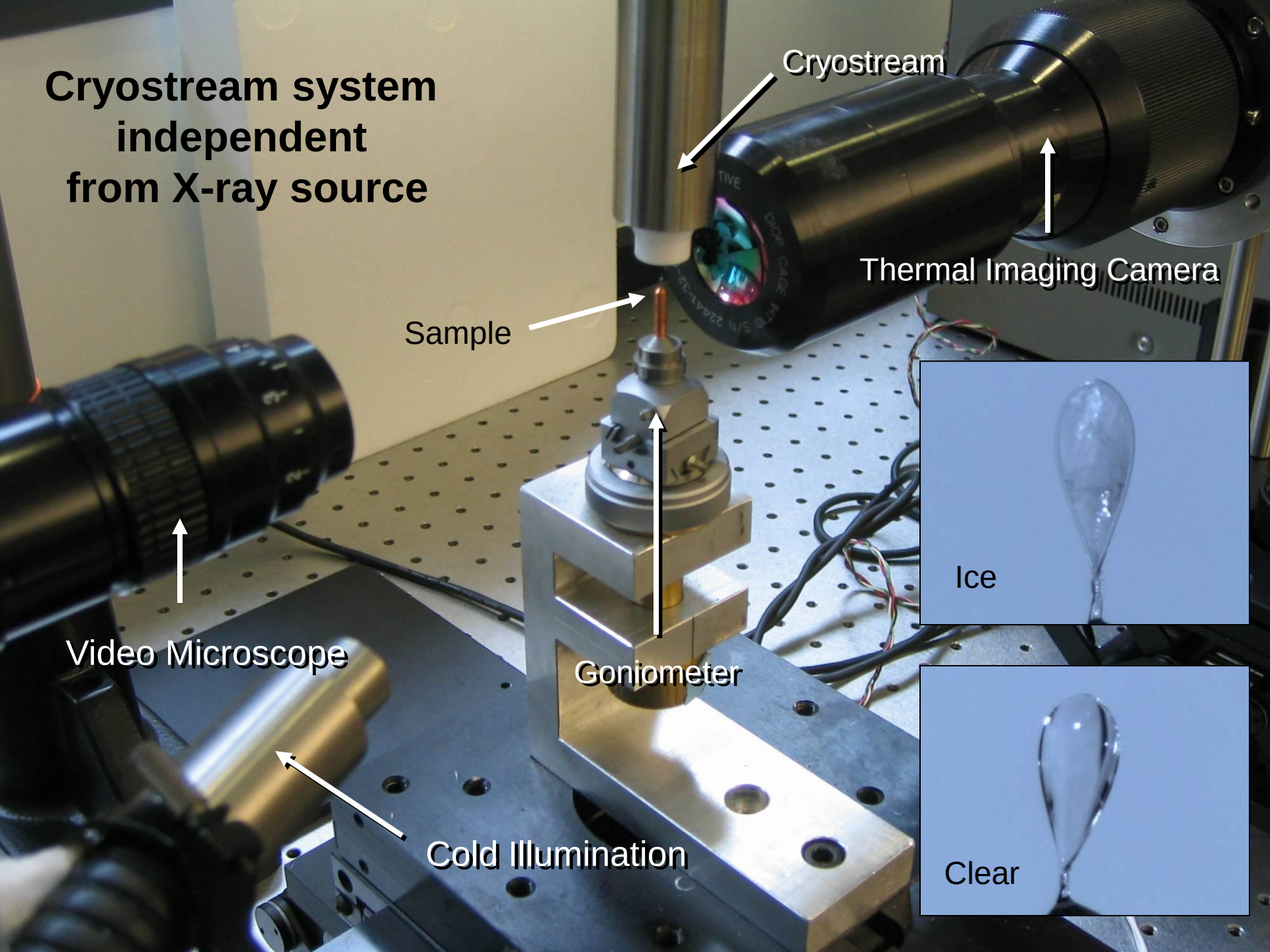


The crystal to cryocooling path



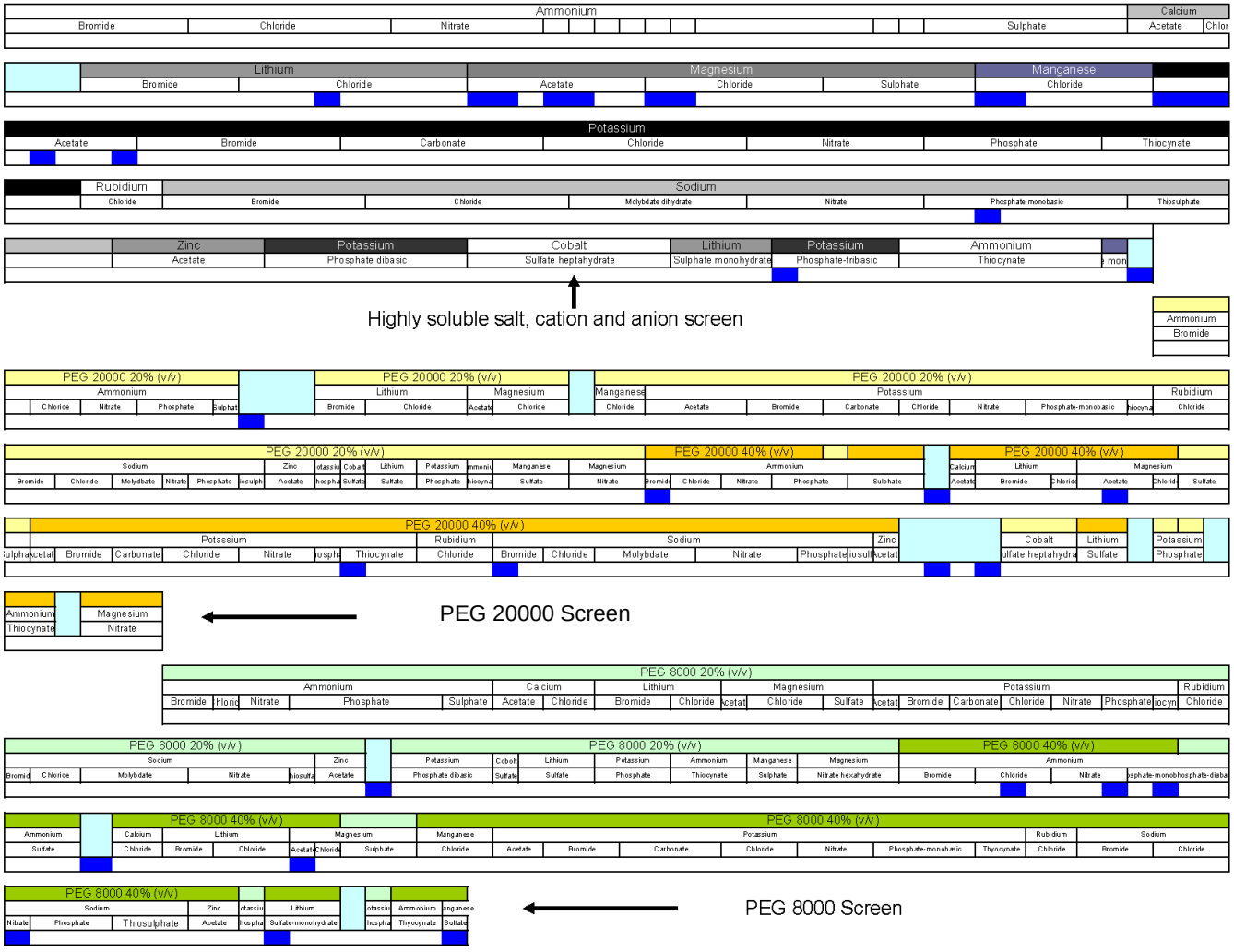
Simple Experiments

- Simple question 1; how many of the 1536 screening conditions are already cryoprotectants?
- Simple question 2; how many can be made into cryoprotectants with a minimal addition of glycerol (as a first example)?
- Previous research has looked at the concentration of glycerol required to freeze 50 standard Hampton Research crystallization conditions (Mitchell & Garman, J. Appl. Cryst., 584-587, 1996)
- Other research examined 4 other cryoprotectants with 98 standard Hampton Research crystallization conditions (McFerrin & Snell, J. Appl. Cryst. 538-545, 2002)
- Both studies used visual observation confirmed by X-ray observation to detect crystalline ice formation.
- In both studies simple visual observation was sufficient to detect crystalline ice in the vast majority of cases studied.
- Simple experiment; visual observation of screening condition mixed 1:1 with cryoprotectant solution to answer questions 1 and 2.

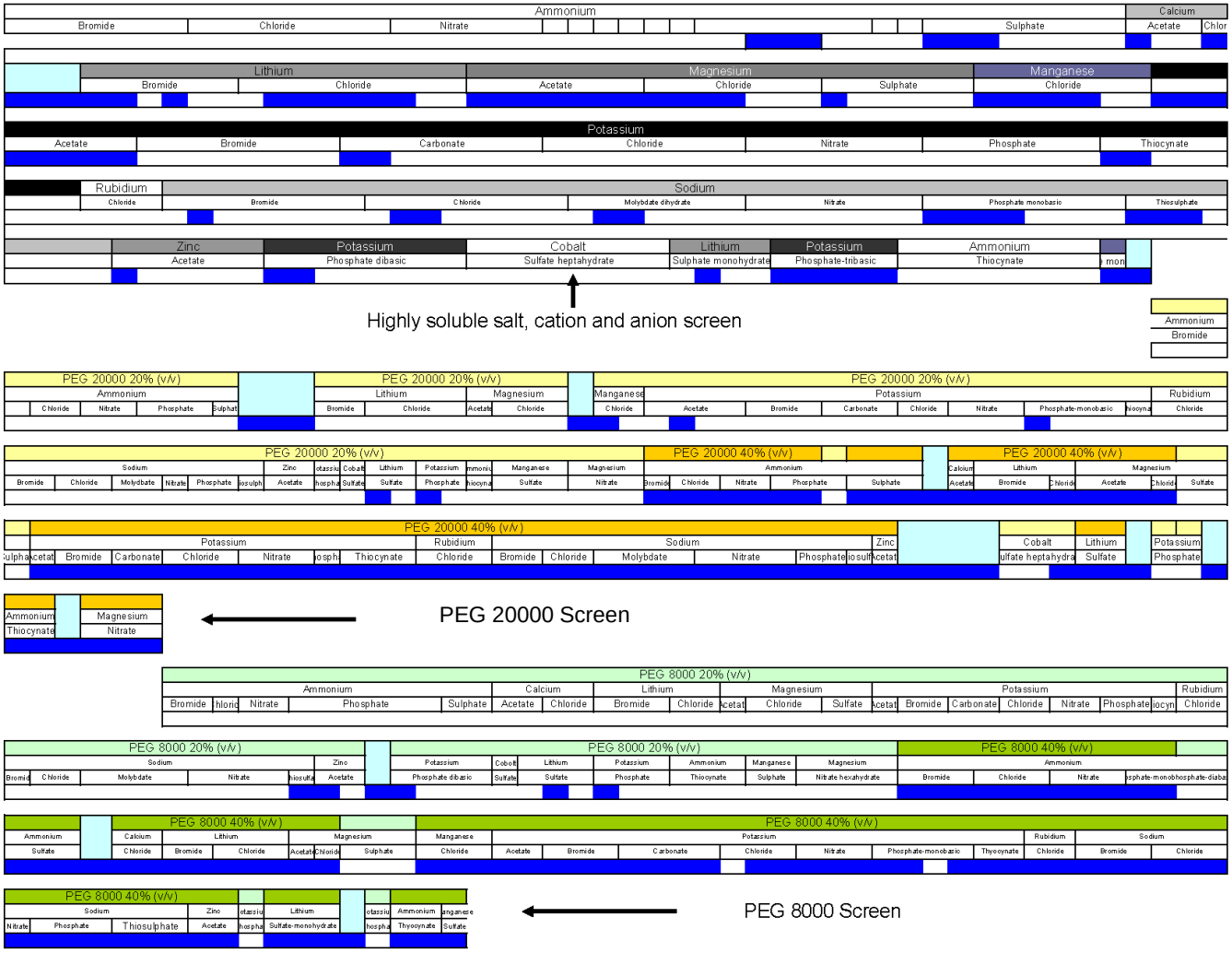

Cryostream system independent from X-ray source
Cryostream
Thermal Imaging Camera
Sample
Ice
Video Microscope
Goniometer
Cold Illumination
Clear

Experimental

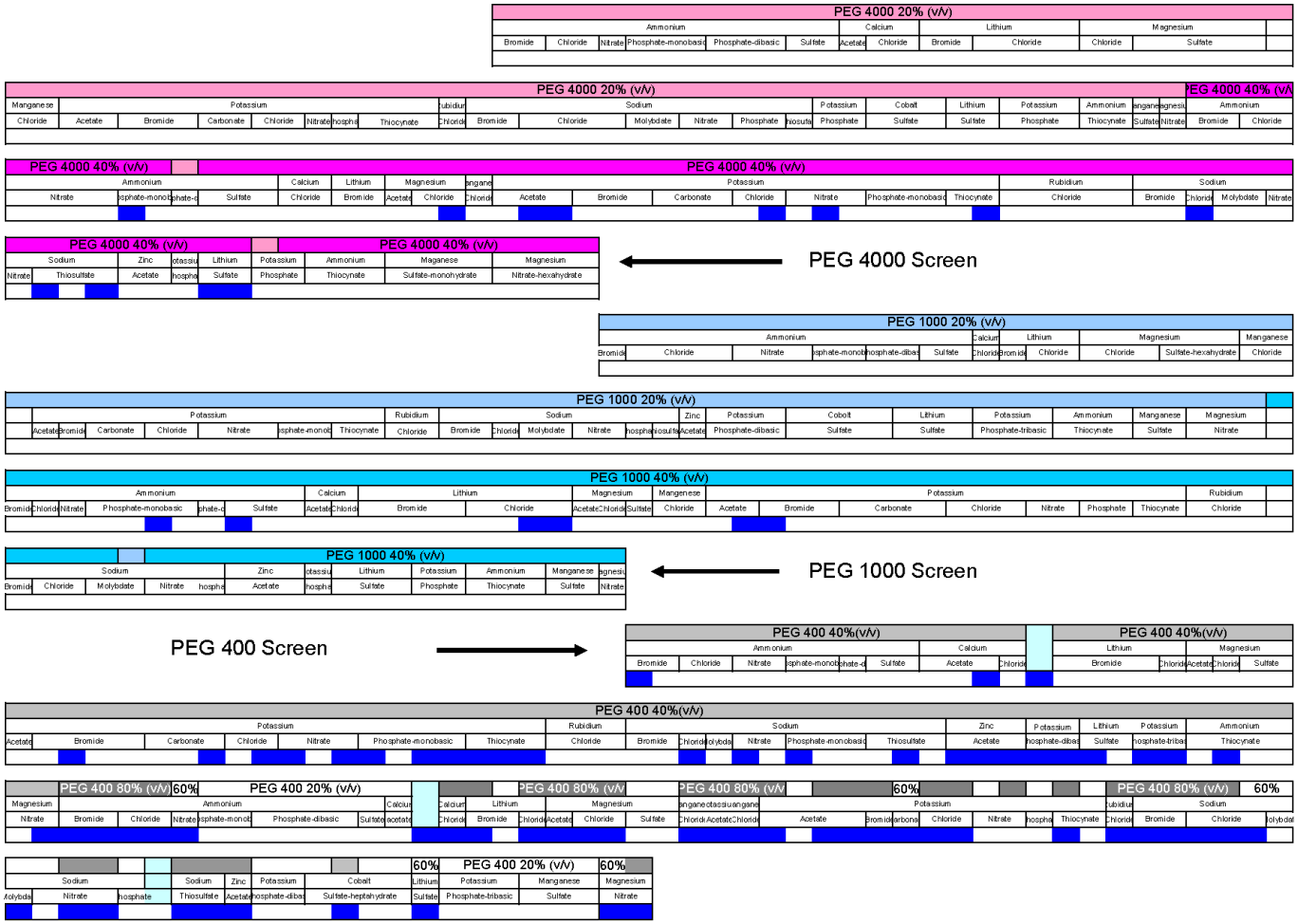
- Off line Oxford 700 cryostream system with video microscope.
- Standard goniometer head with a 0.5-0.7 mm Hampton Research cryoloop.
- Several loops used, washed after use in water followed by ethanol and allowed to dry. Hampton Research recommend buying a new loop for each use ☺
- Look at screening condition alone to determine if it is a cryoprotectant.
  - Use 10μl drop on glass slide and pick up with loop in same manner for each case.
  - Block stream, put loop in place, unblock stream observe and record photograph.
- Repeat experiment for conditions that were not cryoprotectants.
  - Add 12.5% glycerol to system, if clear reduce glycerol concentration by 2.5% each time until ice is seen. If not clear increase glycerol concentration to 20% and determine if clear.
- Give the job to two summer students, and make use of the Hampton research cryoscreen as control points to check the methodology throughout.
- Treat the students nicely to see if you can get them back next year!

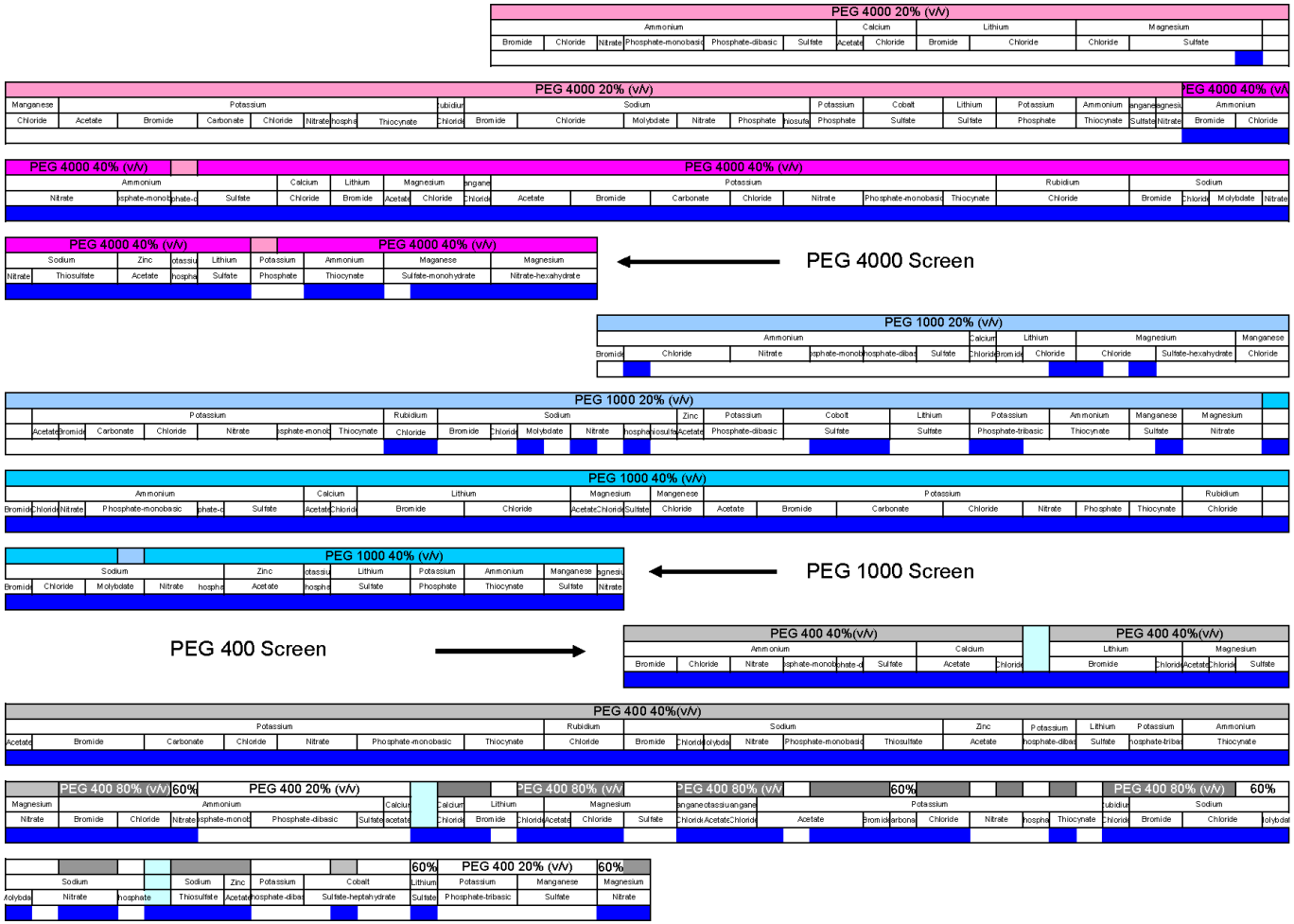


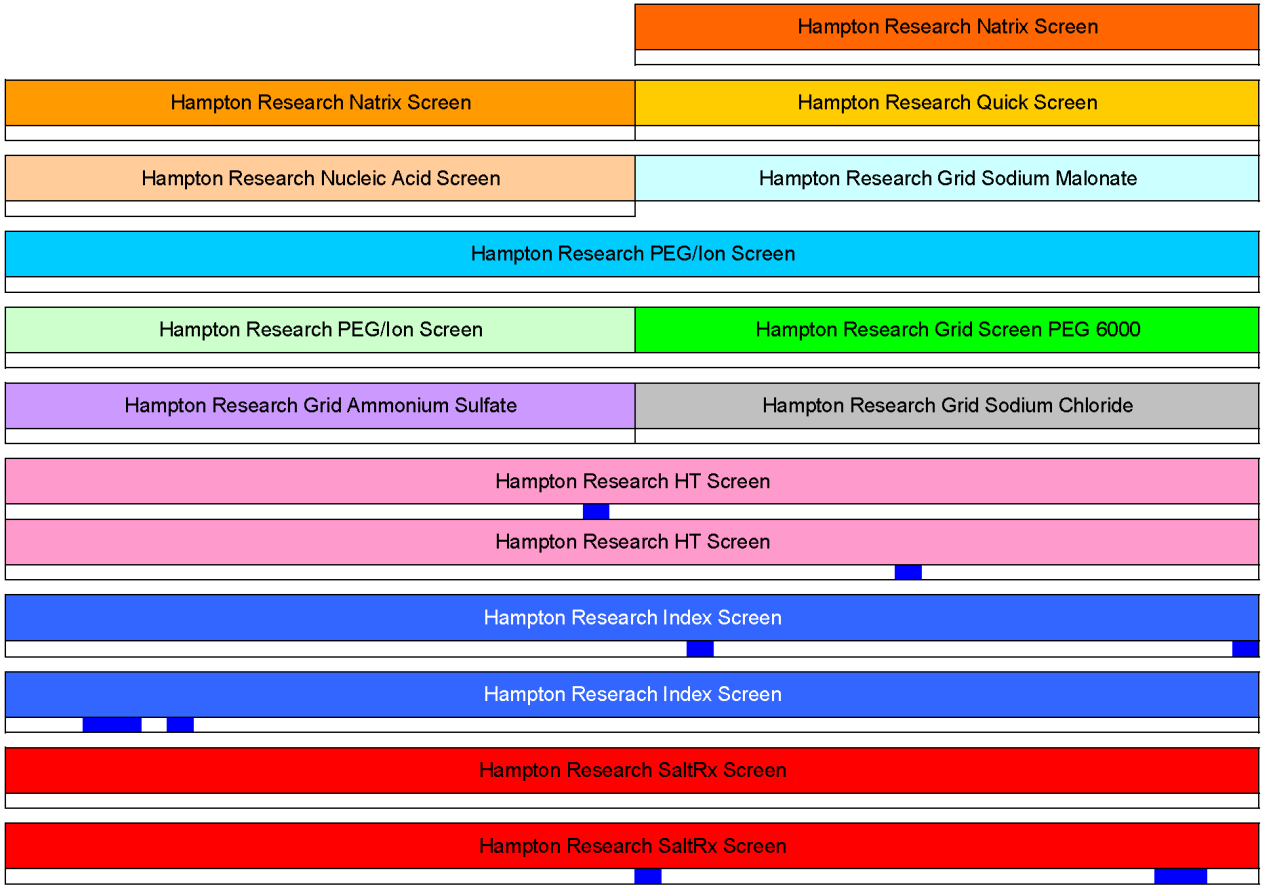
1536 Cocktails described in “Macromolecular crystallization in a high throughput laboratory – the search phase”, Luft, Wolfley, Jurisica, Glasgow, Fortier and DeTitta, J. Crystal Growth 232, 591-595 (2001).

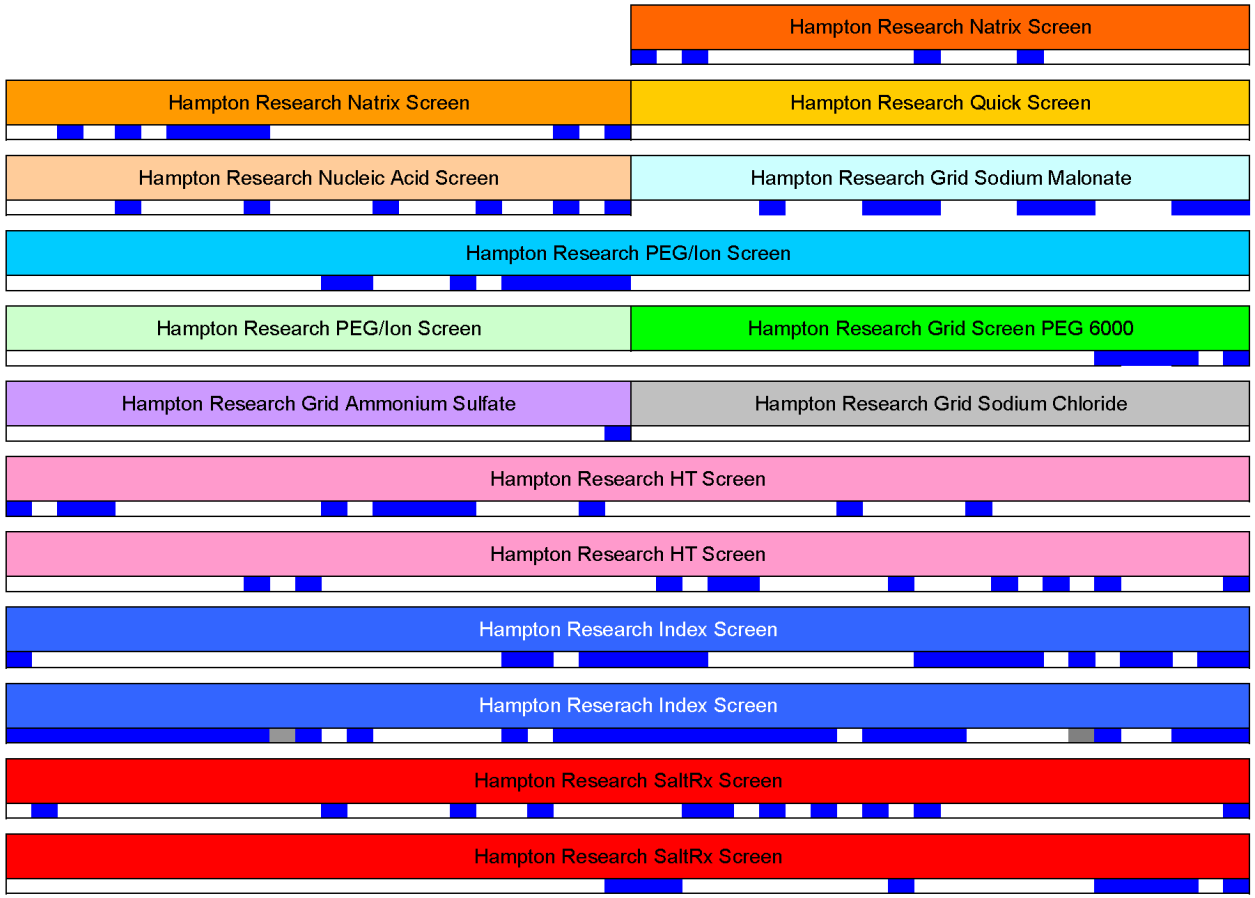


The dark blue shading indicates conditions that are suitable cryoprotectants by the addition of a maximum of 20% glycerol.


PEG 4000 20% (v/v)
PEG 4000 40% (v/v)
PEG 4000 Screen
PEG 1000 20% (v/v)
PEG 1000 40% (v/v)
PEG 1000 Screen
PEG 400 Screen
PEG 400 40%(v/v)
PEG 400 80% (v/v)
PEG 400 20% (v/v)
60%


PEG 4000 20% (v/v)
PEG 4000 40% (v/v)
PEG 4000 Screen
PEG 1000 20% (v/v)
PEG 1000 40% (v/v)
PEG 1000 Screen
PEG 400 Screen
PEG 400 40%(v/v)
PEG 400 80% (v/v)
PEG 400 20% (v/v)
60%


Hampton Research Natrix Screen
Hampton Research Natrix Screen
Hampton Research Quick Screen
Hampton Research Nucleic Acid Screen
Hampton Research Grid Sodium Malonate
Hampton Research PEG/Ion Screen
Hampton Research PEG/Ion Screen
Hampton Research Grid Screen PEG 6000
Hampton Research Grid Ammonium Sulfate
Hampton Research Grid Sodium Chloride
Hampton Research HT Screen
Hampton Research HT Screen
Hampton Research Index Screen
Hampton Reserach Index Screen
Hampton Research SaltRx Screen
Hampton Research SaltRx Screen


Hampton Research Natrix Screen
Hampton Research Natrix Screen
Hampton Research Quick Screen
Hampton Research Nucleic Acid Screen
Hampton Research Grid Sodium Malonate
Hampton Research PEG/Ion Screen
Hampton Research PEG/Ion Screen
Hampton Research Grid Screen PEG 6000
Hampton Research Grid Ammonium Sulfate
Hampton Research Grid Sodium Chloride
Hampton Research HT Screen
Hampton Research HT Screen
Hampton Research Index Screen
Hampton Reserach Index Screen
Hampton Research SaltRx Screen
Hampton Research SaltRx Screen

Results

- 40% of the screening conditions can be made into a suitable cryoprotectant with a maximum addition of only 20% glycerol.
- 30% of the screening conditions at100% concentration are cryoprotectants already.
- 8% are still cryoprotectants at 50% concentration.

What does this mean?

- With multiple screening hits we can deliberately choose those that are cryoprotectant already or can easily be made into a cryoprotectant.

Does it work?

- Choose appropriate pH for buffer based on calculated pI (Kantardjieff & Rupp, Bioinfomatics, 20, 2162-2168, 2004).
- Send sample for screening service.
- Buy chocolates for the staff.
- Wait, get hit, optimize.

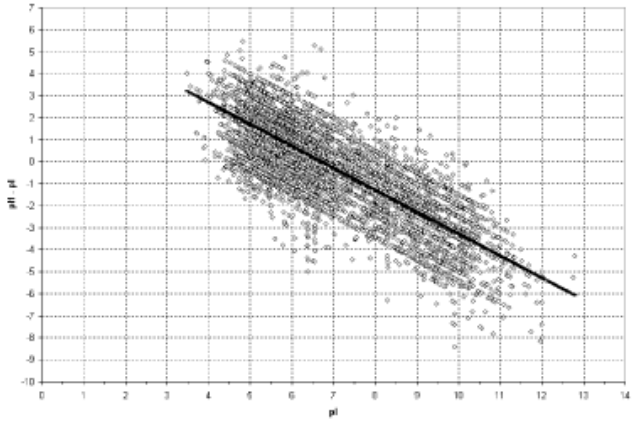
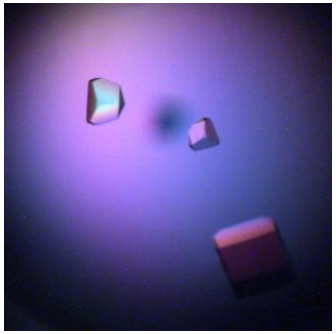


Fig. 2. Correlation between pI and pH. Correlation between calculated pI of successfully crystallized protein and difference between reported crystallization pH and pI. $R^2 = 0.62$, P-value $< 10^{-7}$. Not shown, protein–nucleic acid complexes ($R^2 = 0.77$, P-value $< 10^{-7}$).

- Test case 1, Insulin, over 500 hits, crystals in 10 biochemically different cryoprotectants (study on cryoprotectant action). X-ray data to come
- Test case 2, Ribonuclease A, sub Angstrom resolution, 3 different cryoprotectants. X-ray data instrument limited.
- Human CuZn superoxide dismutase (SOD-1). X-ray data to come.
- Human catalase. X-ray data to come.



To come….

- Look at other cryoprotectants.
- Make use of cryoprotectants already in solutions, e.g. increasing PEG concentrations where PEG is already available.
- Add cryoprotectant information to the data base. Allow the data base to recommend how to make something a good cryoprotectant condition.
- Biochemical and physical analysis of solution properties to try and both understand and predict suitable cryoprotectants.

- This time next year the hardware described will be in operational testing, e.g. the beamline simulator and the solution picking and crystallization optimization robot. Look for some nice movies.

- Integrate X-ray screening data and finally the X-ray images and resulting structural information into the laboratory database.
- More headaches for the computer support group.

Summary

- Efficient use of crystal growth predisposes efficient use of data collection. This is now the case.
- Robotic data collection works well. Our user group is well trained and has produced a significant number of structures as part of that training process.
- The structure of first protein through the pipeline is solved.
- The same X-ray techniques and control software that work well remotely will be adopted into a laboratory system.
- We have several paths to cryoprotect samples so there is no single point failure.

Acknowledgements

- Jen Wilson, Robin Kempkes, Kap Lam and Beth Stofko.
- Joe Luft and the high-throughput crystallization lab.
- The staff at SSRL.
- All the investigators at HWI who put their samples in the hands of FedEx.

- NIH for funding.
- Other support from the John R. Oishei Foundation, NASA and Erie County.
- FedEx (for not losing samples for too long).