

High Throughput Protein Purification and Data Management System for Structural Genomics of *Arabidopsis thaliana*

Won Bae Jeon¹, Adrian Hegeman¹, Brian Fox¹, Bryan Ramirez¹, Claudia Cornelescu¹, Craig Bingman¹, Craig Newman¹, Dave Aceti¹, Dave Dyer¹, Dave Hruby¹, David Smith¹, Dmitriy Vinaraov¹, Eldon Ulrich¹, Frank Vojtik¹, Gabriel Cornelescu¹, Gary Wesensberg¹, Hamid Eghbalnia¹, Hassan Sreenath¹, Heather Burch¹, Ivan Rayment¹, Jianhua Zhang¹, Jikui Song¹, Jing Ji¹, John Cao¹, John Markley¹, Jurgen Doreleijers¹, Kelly Kjer¹, Ken Johnson¹, Kory Seder¹, Michael Runnels¹, Mike Sussman¹, Milo Westler¹, Miron Livny¹, Paul Blommel¹, Peter Lee¹, Qin Zhao¹, Raj Arangarasan¹, Ramya Narayama¹, Rick Amasino¹, Ronnie Fredrick¹, Russell Wrobel¹, Sandy Thao¹, Scott Leisman¹, Todd Kimball¹, Zsolt Zolnai¹, Betsy Lytle², Brian Volkman², Francis Peterson², Chris Oldfield³, Keith Dunker³, Michal Linial⁴, Tatsuya Sawasaki⁵, Yaeta Endo⁵, Masatsune Kainosho⁶, Ray Salemme⁷, Ted Carver⁷ and George Phillips^{1*}

* Corresponding author

¹University of Wisconsin, Madison

²Medical College of Wisconsin, Milwaukee

³Molecular Kinetics, Pullman, Washington

⁴Hebrew University, Jerusalem, Israel

⁵Ehime University, Matsuyama, Japan

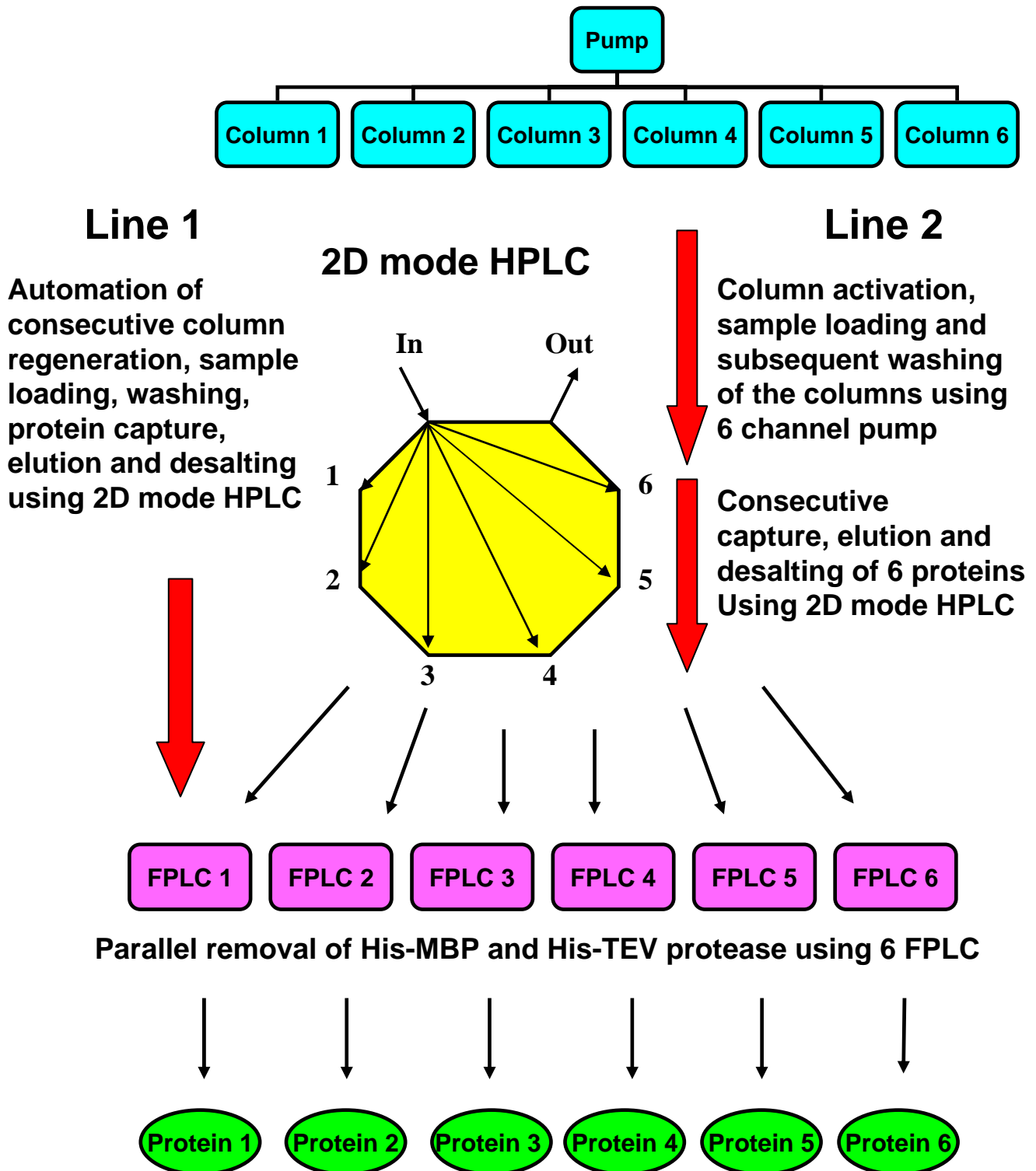
⁶Tokyo Metropolitan University, Tokyo, Japan

⁷3D Pharmaceuticals, Exton, PA, USA

ABSTRACT

A pipeline system has been developed that allows high throughput purification of *Arabidopsis thaliana* proteins expressed in *E. coli*. The key features of this system include (1) a compact 12-channel hi-performance peristaltic pump to load the protein samples onto Ni-IDA columns and subsequent washing of the columns; (2) an HPLC that uses a binary gradient pump to purify 6 proteins by applying gradient elution of buffers from 6 independent Ni-IDA columns; and (3) the parallel 6 HPLC systems for a one-step desalting and a subtractive Ni-IDA chromatography to remove His-tagged proteins from the target protein. By optimizing the purification protocols using this system, we have purified proteins with the MW range of 12 to 30 kDa with a purity of 97% that are suitable for structural determination by X-ray and NMR. We also utilize a highly ordered data storage system, Lamp module in Sesame software, to monitor the quality control of purification processes and to store the biochemical properties of purified protein such as mobility and purity on SDS-polyacrylamide gel, UV-visible spectra, MALDI-MS and ESI-MS.

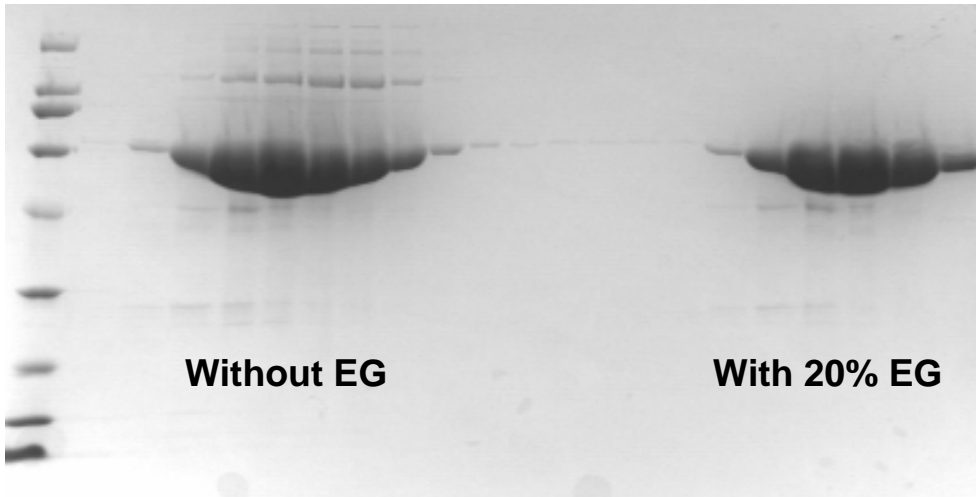
Schematic diagram of an effective and flexible purification pipeline



Optimization of purification protocols

Buffer condition: Buffering agent, pH and salt

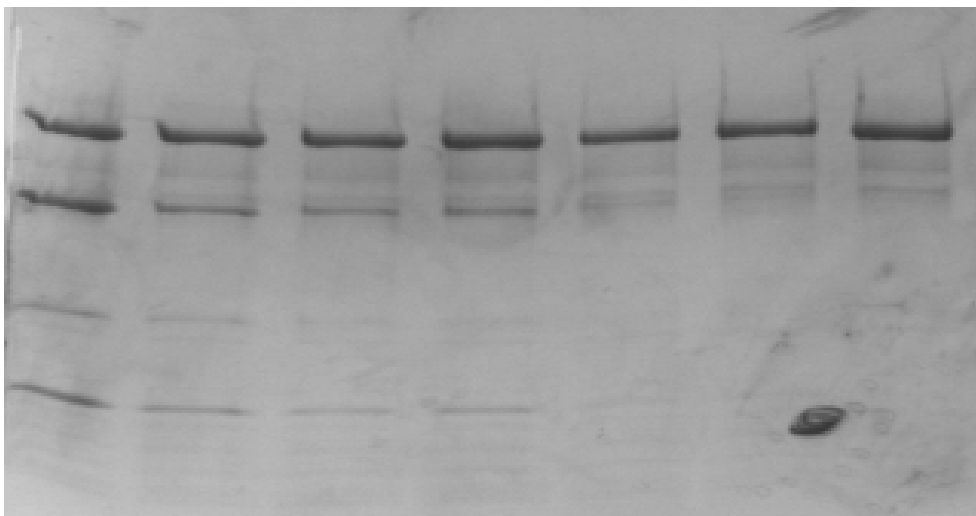
Effect of chaotropic agents: Ethylene glycol, glycerol, NaBr or CHAPS



Cleavage of His6-MBP-fusion protein by TEV protease

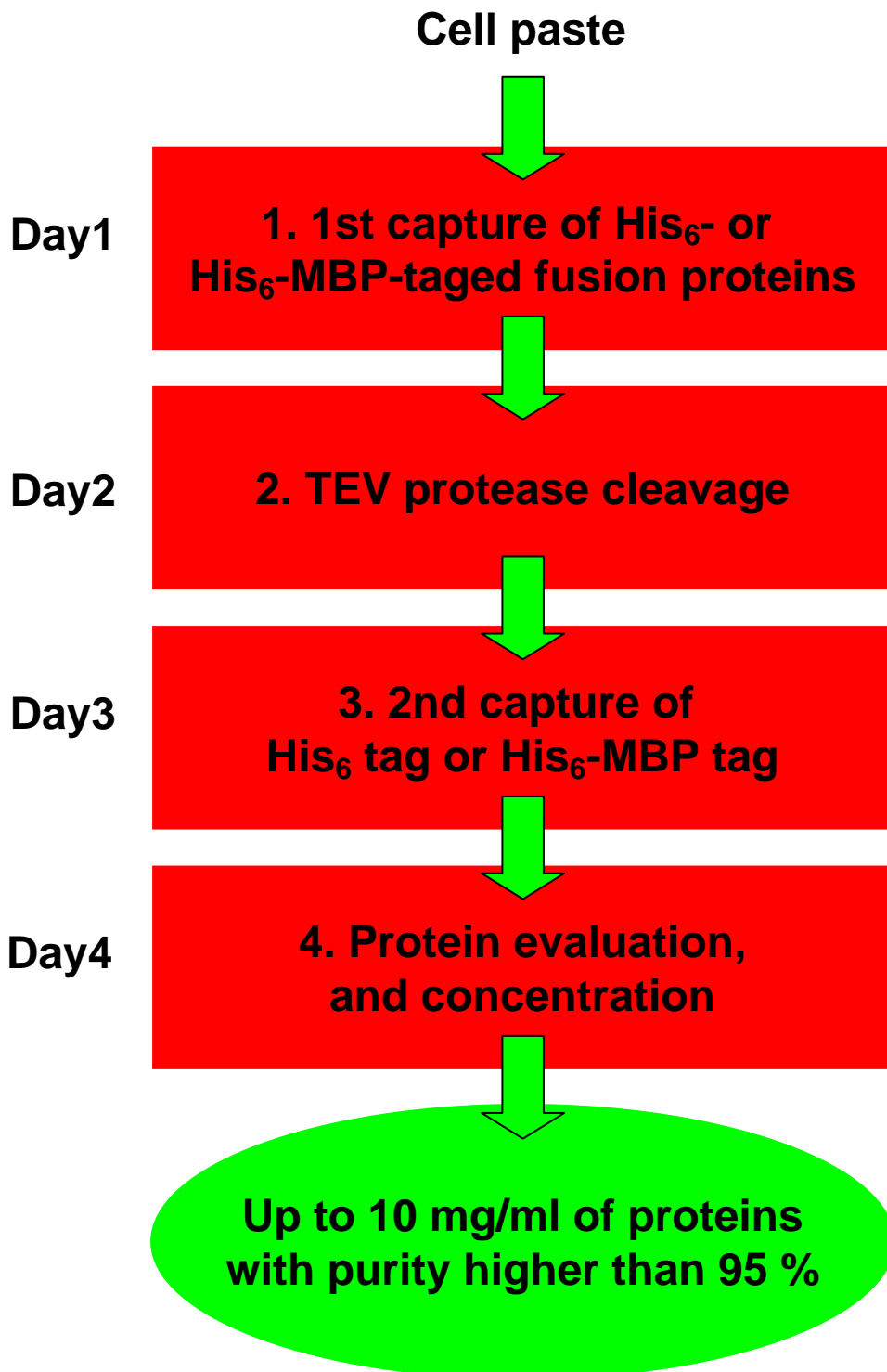
TEV to fusion protein ratio (w/w), 2 hrs at 20 °C

1:30 1:50 1:100 1/200 1:300 1:500 1:1000



Purification of ^{13}C -, ^{15}N -, ^{13}C and ^{15}N -, and Se-Met-labeled protein

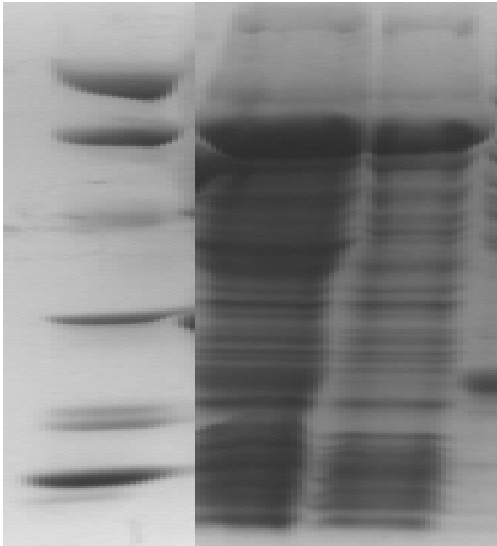
Flow diagram of purification steps



Behaviors of protein during purification

Didn't bind Ni-column

At5g17090

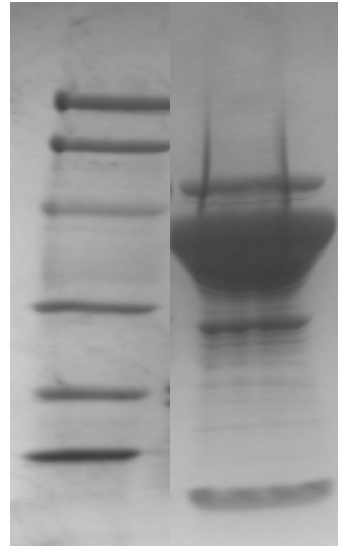


Small insert size

At1g75090

MW : 23 kDa

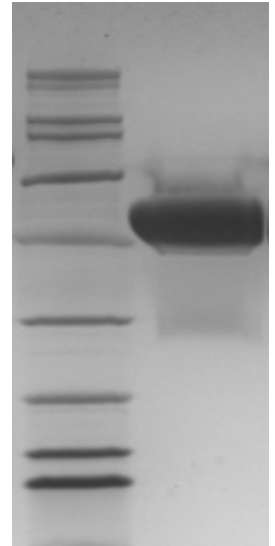
Insert : 6 kDa



At3g51890

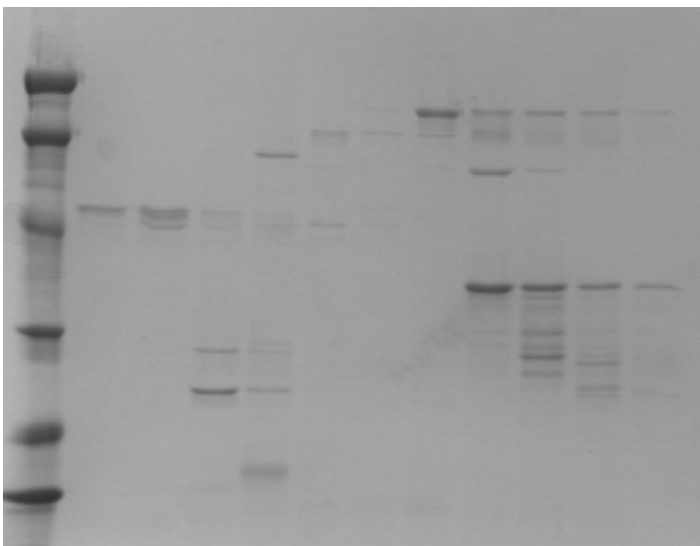
25 kDa

10 kDa

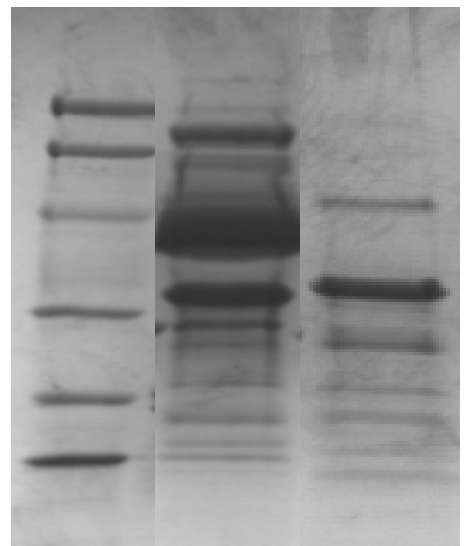


Protein degradation

At5g62290

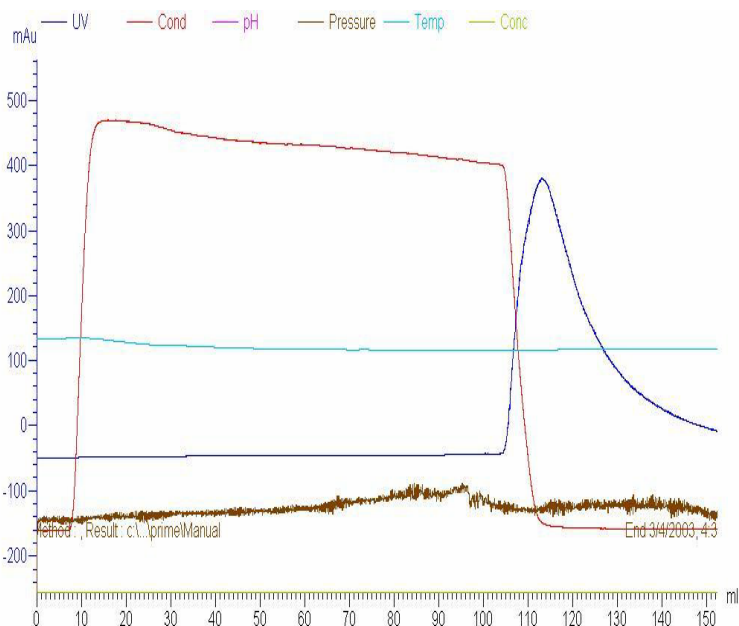
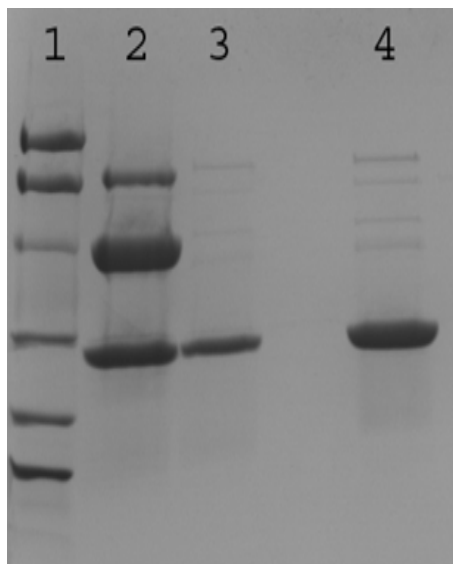


At1g80940



Typical examples of pipeline products

SDS-PAGE and elution profile of Se-Met-labeled At3g16990 protein from 2nd Ni-IDA column

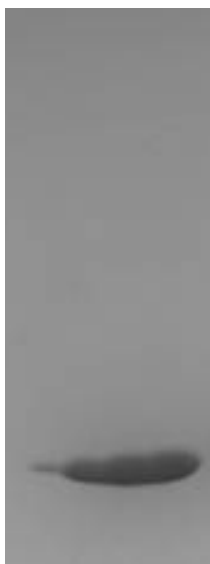


Purification of small size proteins for NMR study

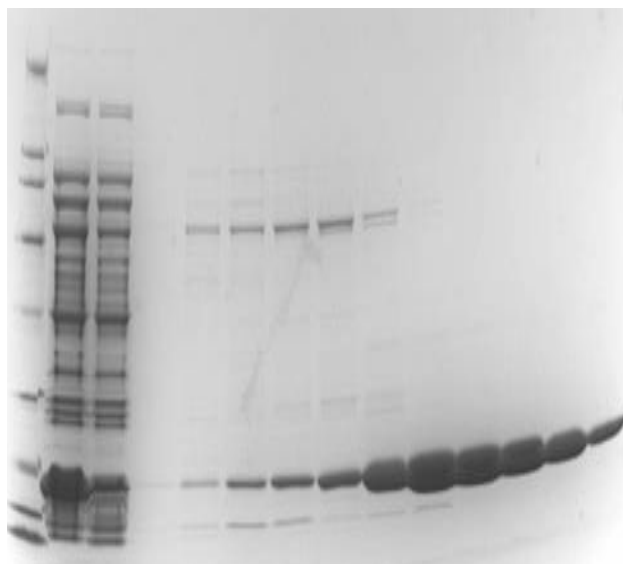
At1g29250



At5g22580



At1g52010



Lamp Module in Genie Software: Highly Ordered Protein Data Management System

The screenshot displays the Lamp software interface. At the top, there is a menu bar with 'Options', 'Proteins', 'Edit', 'Search', 'Report', and 'Help'. Below the menu bar is a toolbar with various icons for file operations and data management. The main window is divided into several sections:

- Left Panel:** A list of protein samples with columns for ID, Name, and Date. The selected entry is '2690: At2g43910 E4 2003-02-11'.
 - 2702: At1g16990 D5 2003-02-11
 - 2814: At1g23750 2003-03-25
 - 2697: At1g29250 2003-02-10
 - 2701: At1g52010 2003-02-10
 - 2687: At2g23660 2003-02-10
 - 2688: At2g43670 2003-03-17
 - 2690: At2g43910 E4 2003-02-11
 - 2689: At3g03410 2003-03-25
- Top Right Panel:** A window titled 'Lamp Images' showing a grid of images, with the 'EdrAAAA' window selected.
- Bottom Panel:** A detailed view for the selected protein 'At2g43910'. It includes a 'General' tab and a 'Protocol' tab. The 'Protocol' tab shows a list of steps: 'Cell Growth', 'Sample Prep', '1st IMAC', 'Protein Cleavage', and '2nd IMAC'. Each step has associated input fields for parameters like volume, concentration, and duration.
- Mass Spectrometry Data:** Two plots are shown on the right side of the bottom panel.
 - Top Plot:** A mass spectrum plot titled '+01: 1.31 min (13 scans) from At3g17210-Se-Met'. The x-axis is 'm/z, amu' (ranging from 600 to 1800) and the y-axis is 'Intensity, cps' (ranging from 0 to 7e5). A prominent peak is labeled at 1431.2. Other labeled peaks include 1193.4, 1302.2, 1436.4, 1591.2, and 1790.0. The total intensity is 2.186e cps.
 - Bottom Plot:** A 'BioSpec Reconstruct' plot for the same sample. The x-axis is 'Mass, amu' (ranging from 13600 to 14800) and the y-axis is 'Intensity, cps' (ranging from 0 to 1.5e6). The main peak is at 14314. Other labeled peaks include 13753, 14015, 14269, 14358, 14671, and 14775. The total intensity is 2.036e cps.

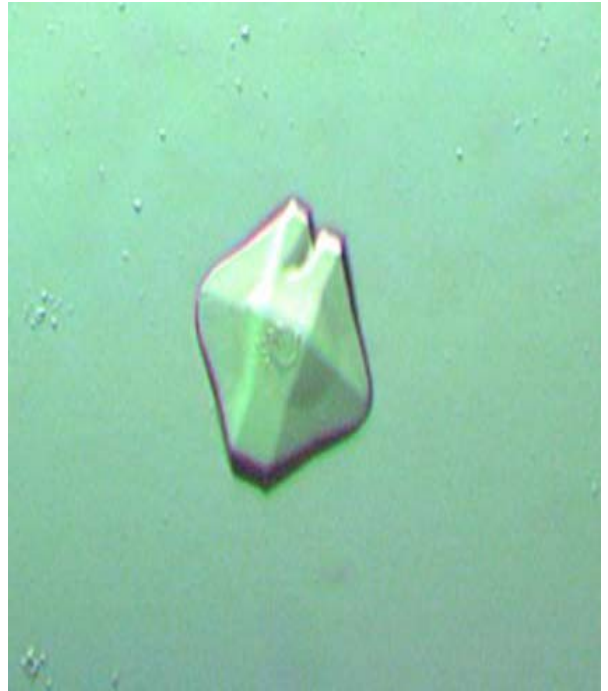
At the bottom of the window, a status bar shows 'Command successful!'. The Windows taskbar at the very bottom includes the Start button, several application icons, and the system tray showing the time as 3:34 PM.

Crystal screening

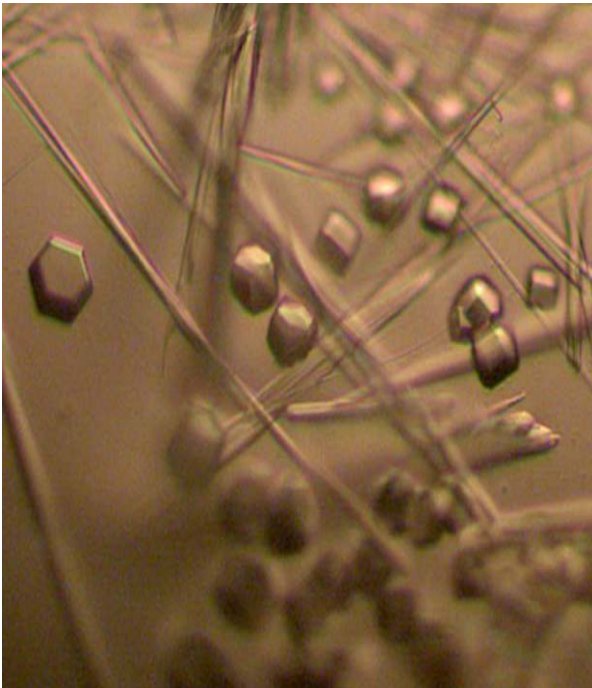
At3g03410



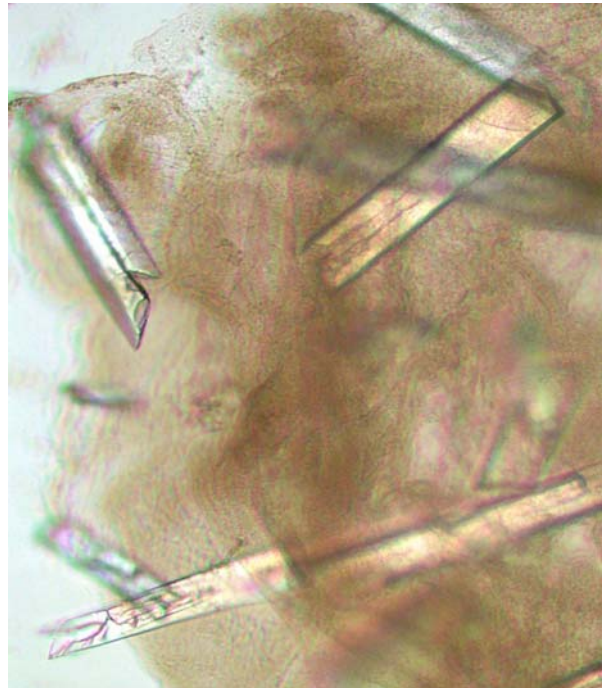
At3g16990



At3g20970



Se-Met-labeled At3g16990



CONCLUSION

- 1. A purification pipeline that consists of a pump, a mode HPLC and 6 simple FPLC has been developed and tested.**
- 2. Optimization of protocols together with high speed automation allows us to purify up to 50 mg of 6 different fusion proteins with a purity higher than 95% in a day.**
- 3. Each protein is being analyzed it's identity by MALDI- and ESI-MS before it is passed on for crystallization screening.**
- 4. The Lamp module in Genie software is being used to store chromatographic elution profiles and gel images.**
- 5. The system provides data that will be eventually used to further improve current high throughput protein production.**