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

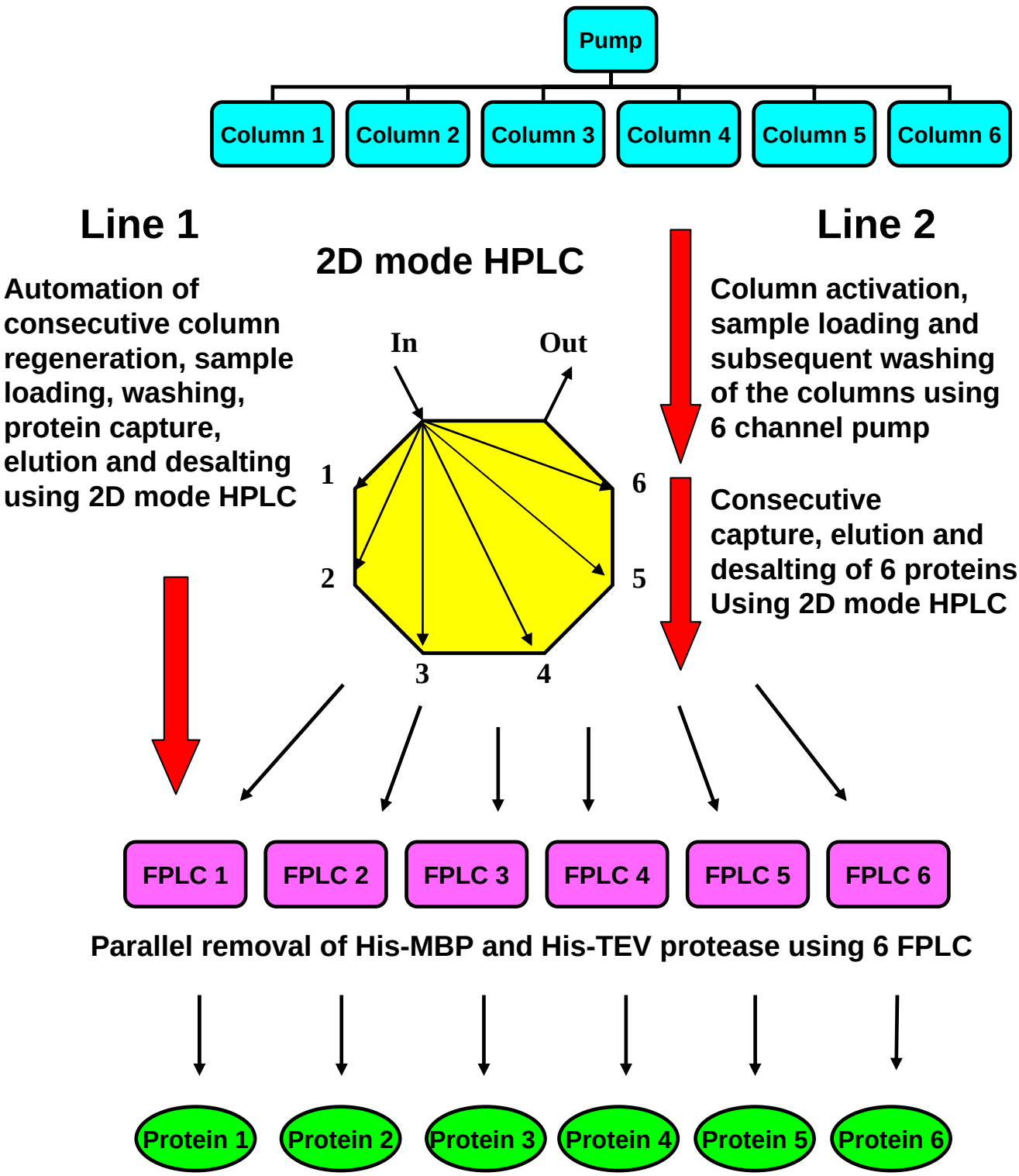
Won Bae Jeon[1], Adrian Hegeman[1], Brian Fox[1], Bryan Ramirez[1], Claudia Cornelescu[1], Craig Bingman[1], Craig Newman[1], Dave Aceti[1], Dave Dyer[1], Dave Hruby[1], David Smith[1], Dmitriy Vinaraov[1], Eldon Ulrich[1], Frank Vojtik[1], Gabriel Cornelescu[1], Gary Wesensberg[1], Hamid Eghbalnia[1], Hassan Sreenath[1], Heather Burch[1], Ivan Rayment[1], Jianhua Zhang[1], Jikui Song[1], Jing Ji[1], John Cao[1], John Markley[1], Jurgen Doreleijers[1], Kelly Kjer[1], Ken Johnson[1], Kory Seder[1], Michael Runnels[1], Mike Sussman[1], Milo Westler[1], Miron Livny[1], Paul Blommel[1], Peter Lee[1], Qin Zhao[1], Raj Arangarasan[1], Ramya Narayama[1], Rick Amasino[1], Ronnie Fredrick[1], Russell Wrobel[1], Sandy Thao[1], Scott Leisman[1], Todd Kimball[1], Zsolt Zolnai[1], Betsy Lytle[2], Brian Volkman[2], Francis Peterson[2], Chris Oldfield[3], Keith Dunker[3], Michal Linial[4], Tatsuya Sawasaki[5], Yaeta Endo[5], Masatsune Kainosho[6], Ray Salemme[7], Ted Carver[7] and George Phillips[1]*

* Corresponding author
[1]University of Wisconsin, Madison
[2]Medical College of Wisconsin, Milwaukee
[3]Molecular Kinetics, Pullman, Washington
[4]Hebrew University, Jerusalem, Israel
[5]Ehime University, Matsuyama, Japan
[6]Tokyo Metropolitan University, Tokyo, Japan
[7]3D 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-polyacylamide gel, UV-visible spectra, MALDI-MS and ESI-MS.

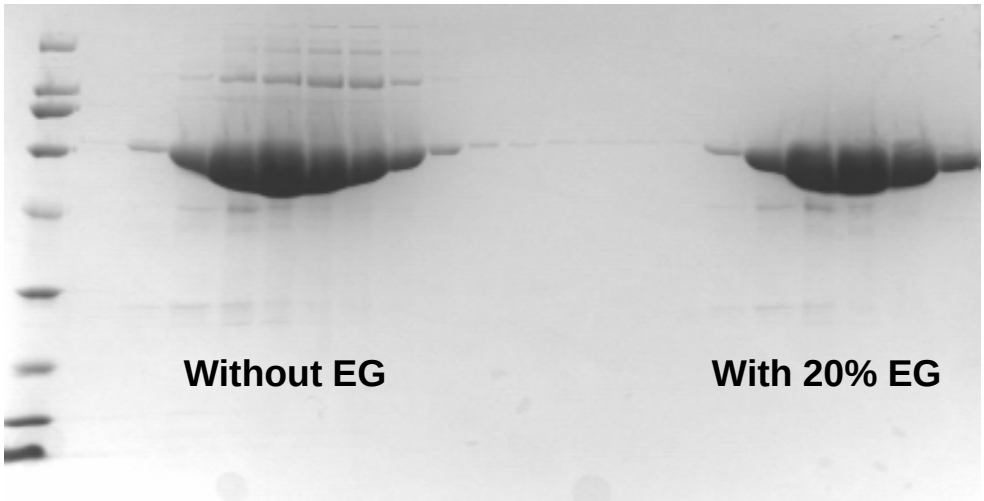
Schematic diagram of an effective and flexible purification pipeline



Optimization of purification protocols

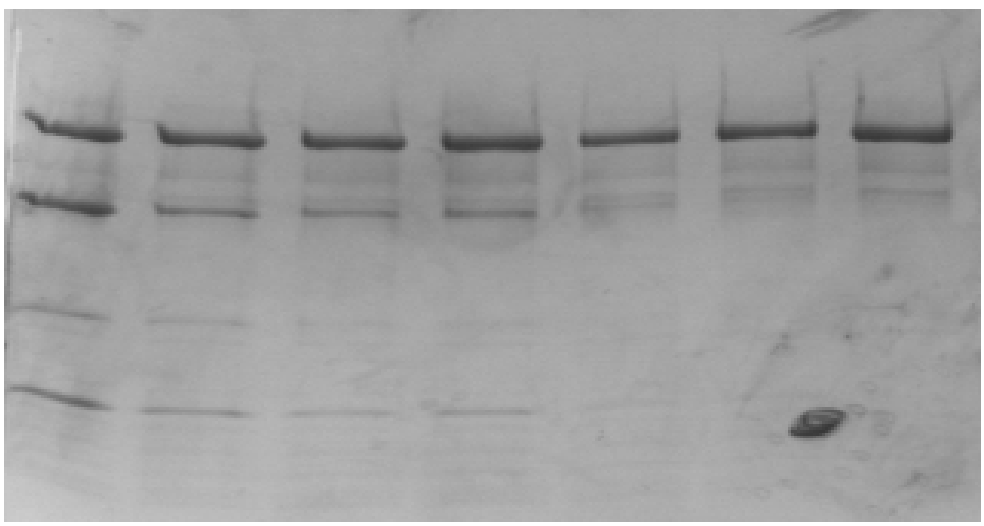
Buffer condition: Buffering agent, pH and salt

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



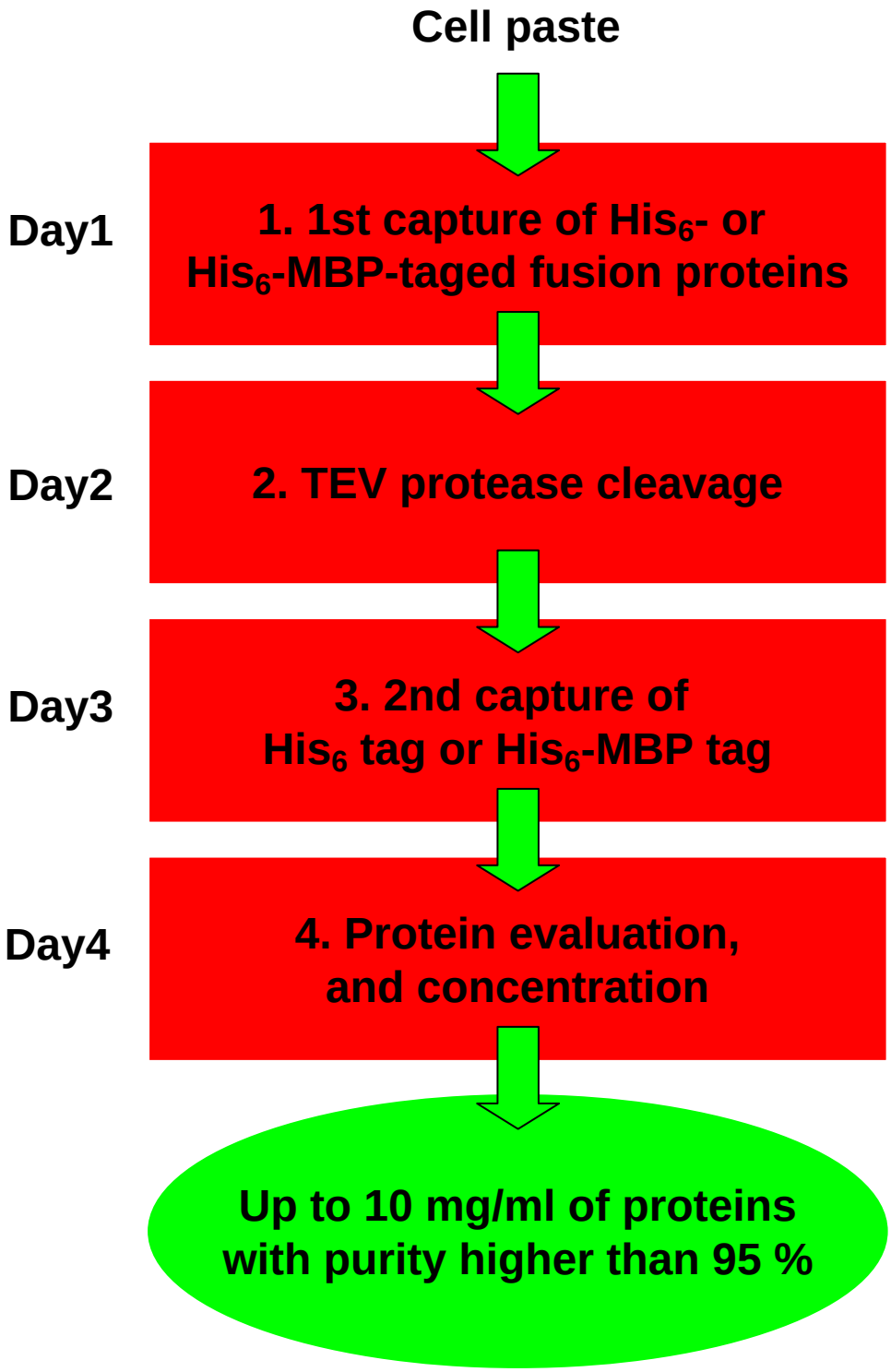
Cleavage of His6-MBP-fusion protein by TEV protease
TEV to fusion protein ratio (w/w), 2 hrs at 20 ^{o}C

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

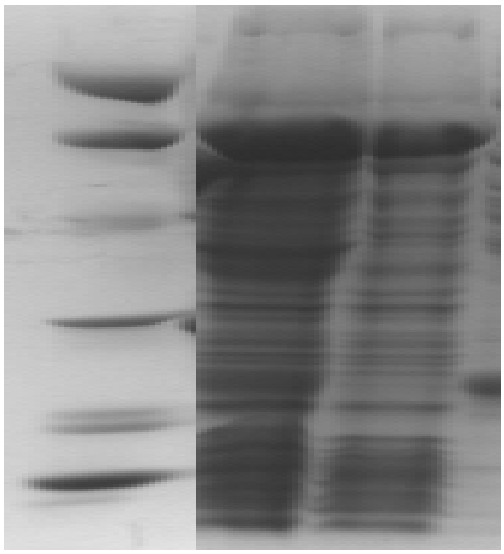
Flow diagram of purification steps



Behaviors of protein during purification

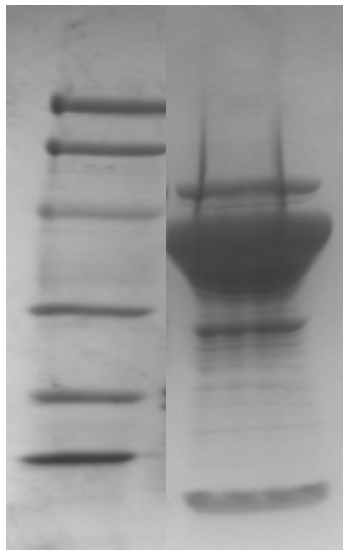
Didn’t bind Ni-column

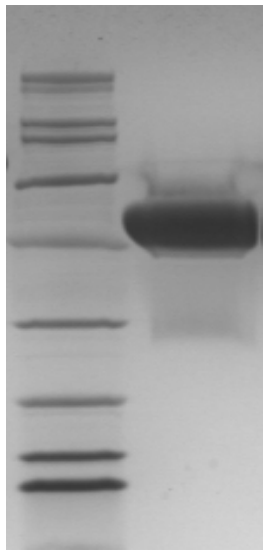
At5g17090



Small insert size

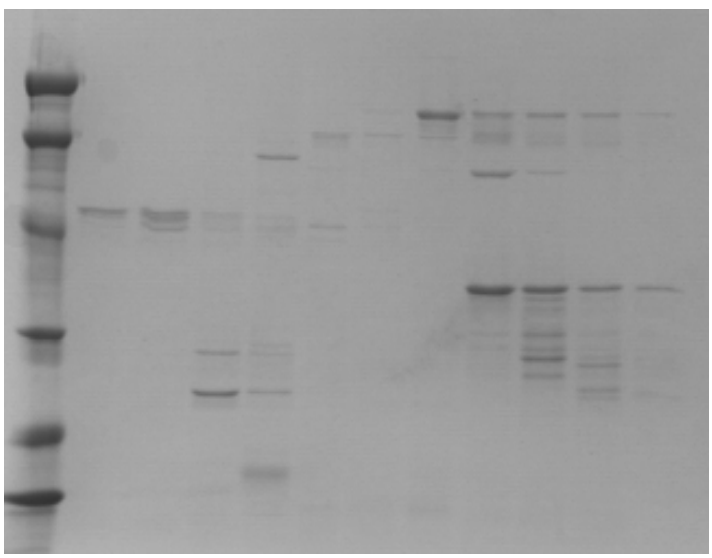
	At1g75090	At3g51890
MW :	23 kDa	25 kDa
Insert :	6 kDa	10 kDa



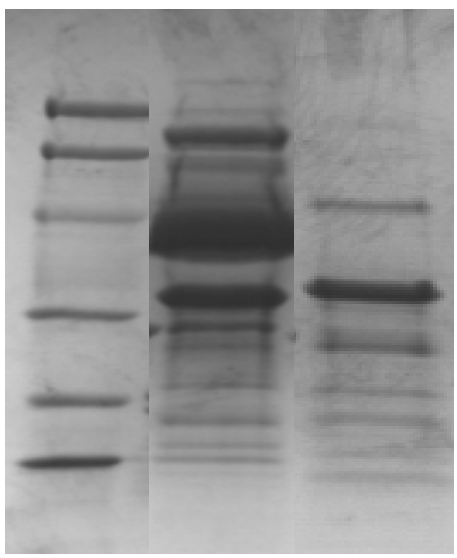


Protein degradation

At5g62290

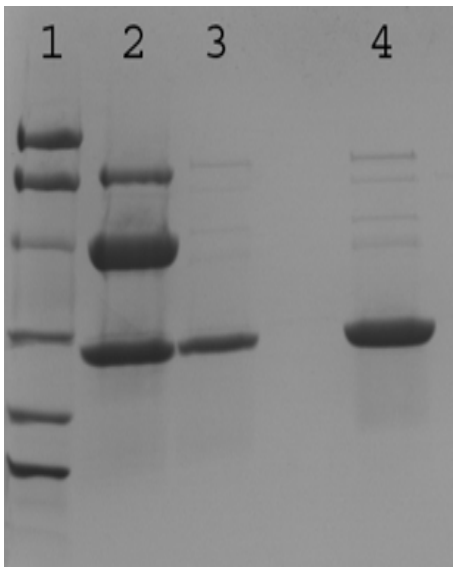


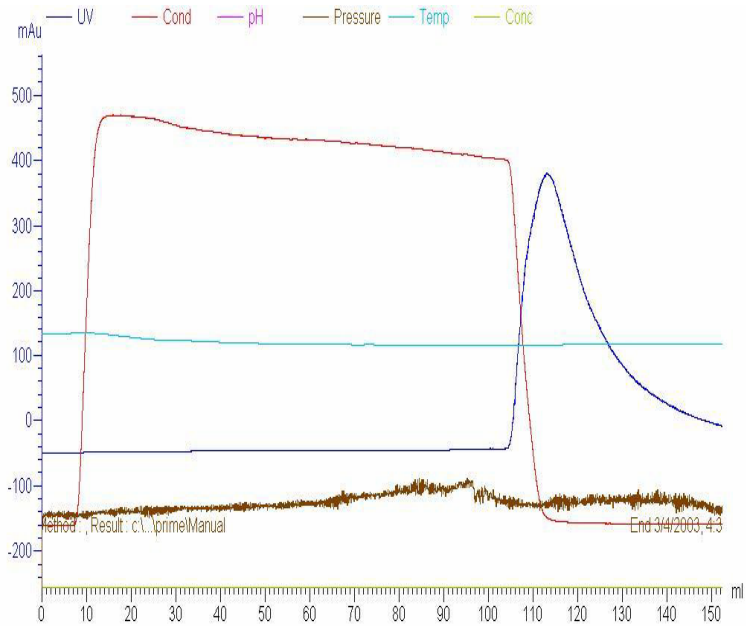
At1g80940



Typical examples of pipeline products

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

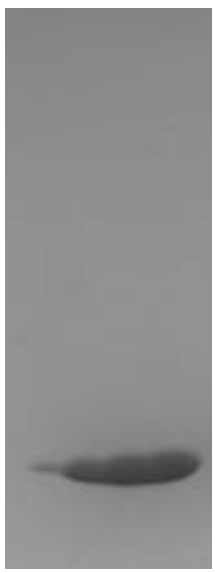




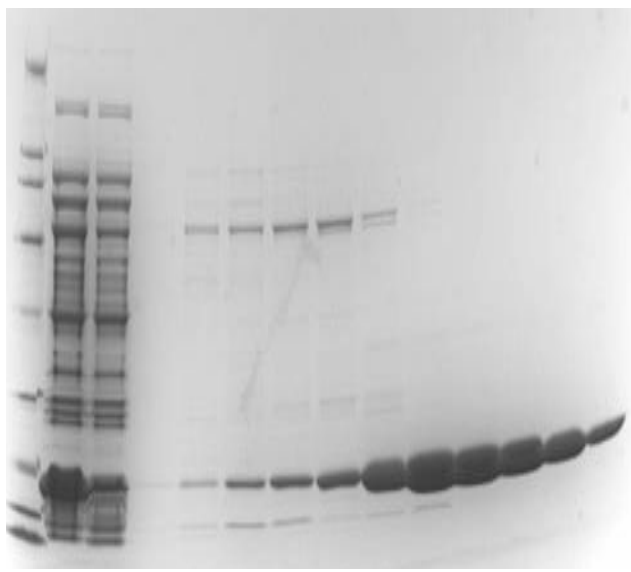
Purification of small size proteins for NMR study

At1g29250

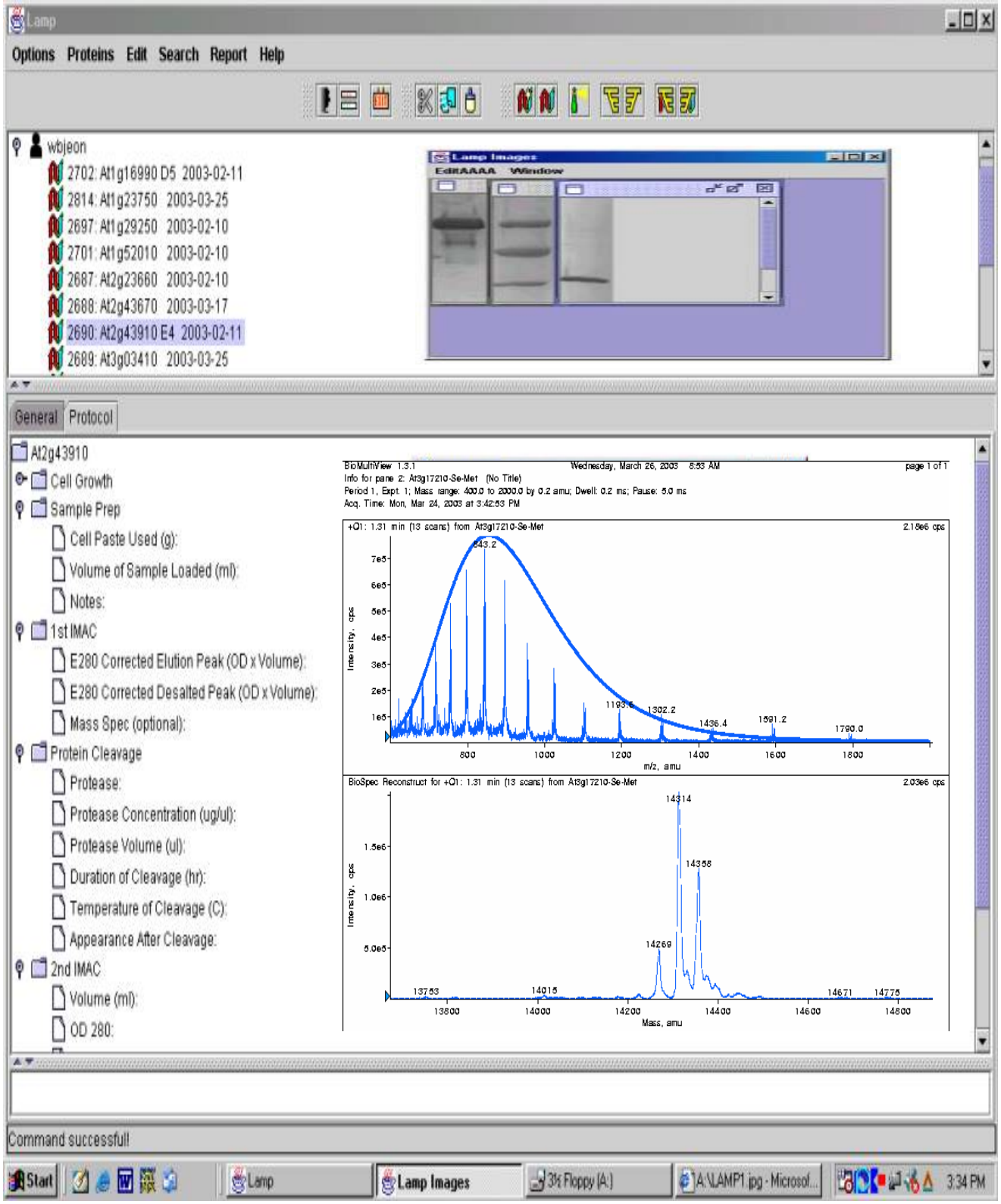
At5g22580



At1g52010

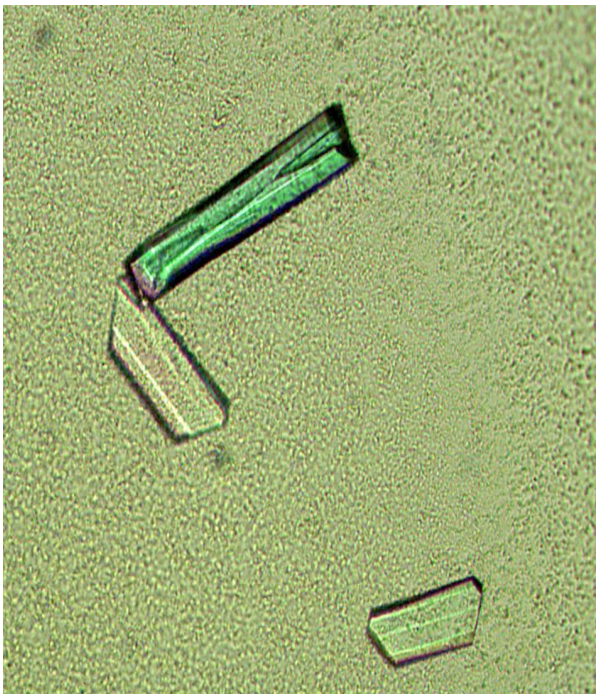


Lamp Module in Genie Software: Highly Ordered Protein Data Management System



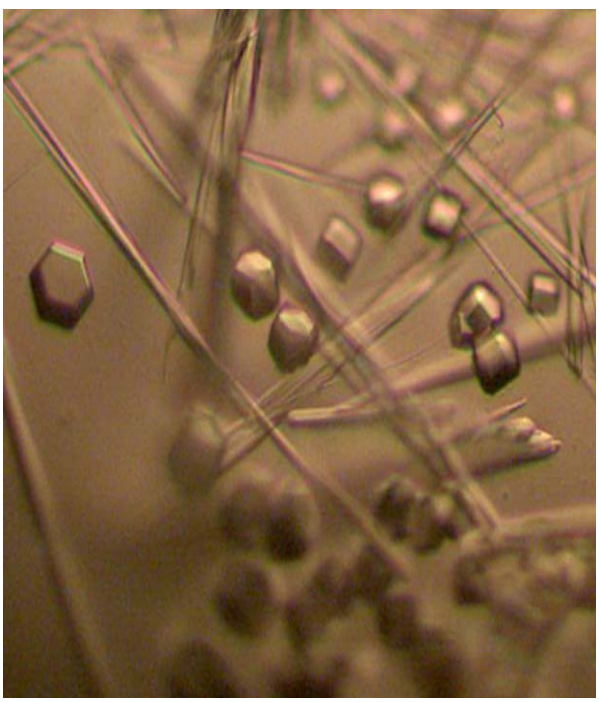
Crystal screening

At3g03410

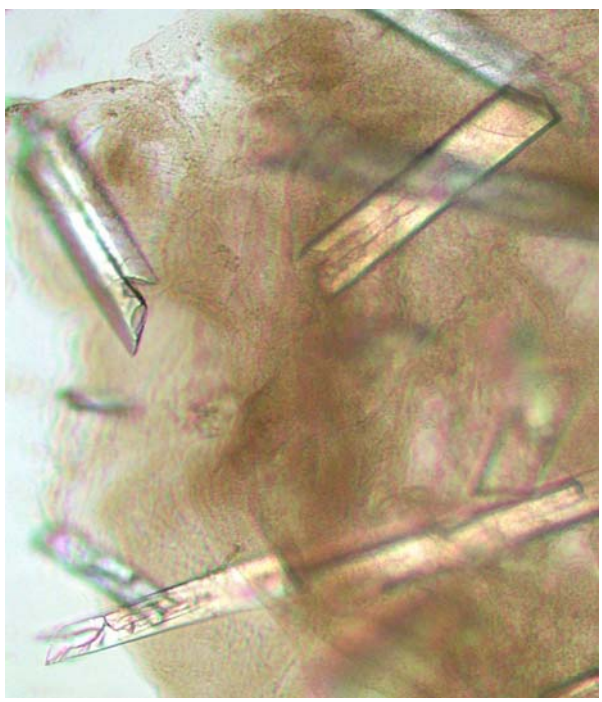


At3g16990

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.