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 Phillips¹*

* 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
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

**Line 1**
Automation of consecutive column regeneration, sample loading, washing, protein capture, elution and desalting using 2D mode HPLC

**Line 2**
Column activation, sample loading and subsequent washing of the columns using 6 channel pump
Consecutive capture, elution and desalting of 6 proteins Using 2D mode HPLC

Parallel removal of His-MBP and His-TEV protease using 6 FPLC

Protein 1  Protein 2  Protein 3  Protein 4  Protein 5  Protein 6
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
1. 1st capture of His$_6$- or His$_6$-MBP-tagged fusion proteins

2. TEV protease cleavage

3. 2nd capture of His$_6$ tag or His$_6$-MBP tag

4. Protein evaluation, and concentration

Up to 10 mg/ml of proteins with purity higher than 95%
### Behaviors of protein during purification

<table>
<thead>
<tr>
<th>Protein</th>
<th>Did’t bind Ni-column</th>
<th>Small insert size</th>
</tr>
</thead>
<tbody>
<tr>
<td>At5g17090</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MW : 23 kDa</td>
<td>At1g75090</td>
</tr>
<tr>
<td></td>
<td>Insert : 6 kDa</td>
<td>25 kDa</td>
</tr>
<tr>
<td></td>
<td></td>
<td>At3g51890</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10 kDa</td>
</tr>
<tr>
<td>At5g62290</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protein degradation</td>
<td></td>
</tr>
<tr>
<td>At1g80940</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 its 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.