Future Protein-Ligand Challenges:
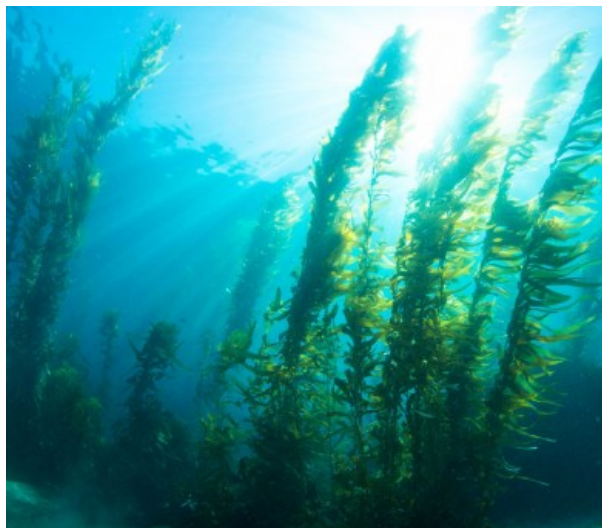# *CELPP & New Datasets*

Vicki Feher

CELPP

Weekly automated cross-docking challenges enabled by PDB release of small molecule & protein sequences
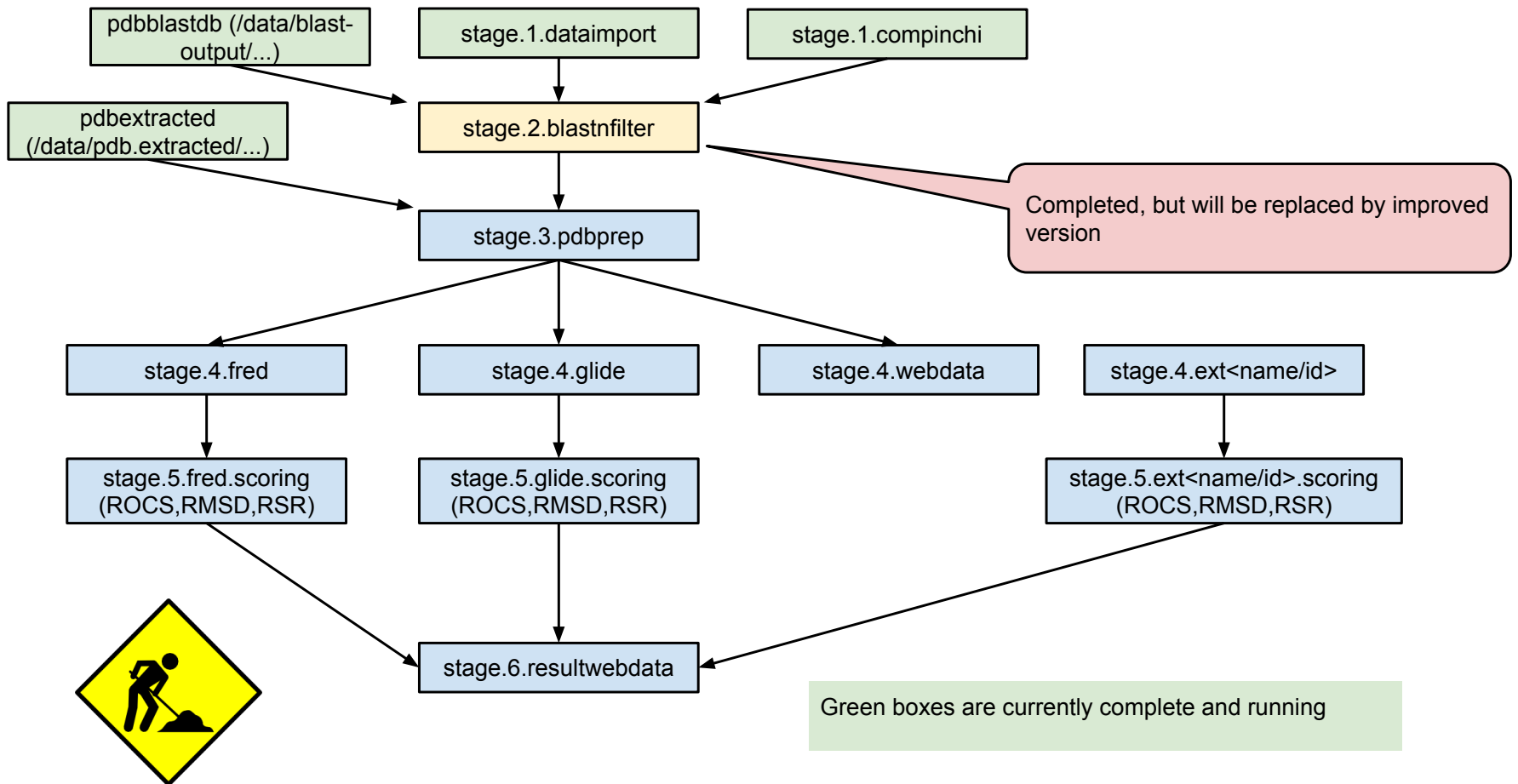
We envision three possible ways to participate:

1. Docking program hosted on our internal server

2. Automated link between our webservices and an external automated docking websites

3. Downloading our weekly posts & uploading results

Can we move beyond reproducing PDB structure poses?

1. Work with the journals/PDB to have notification of "accepted" papers and "HOLD" structures with affinities.

2. SGC - Toronto will provide ~1-2 month notice of datasets being prepared for publication
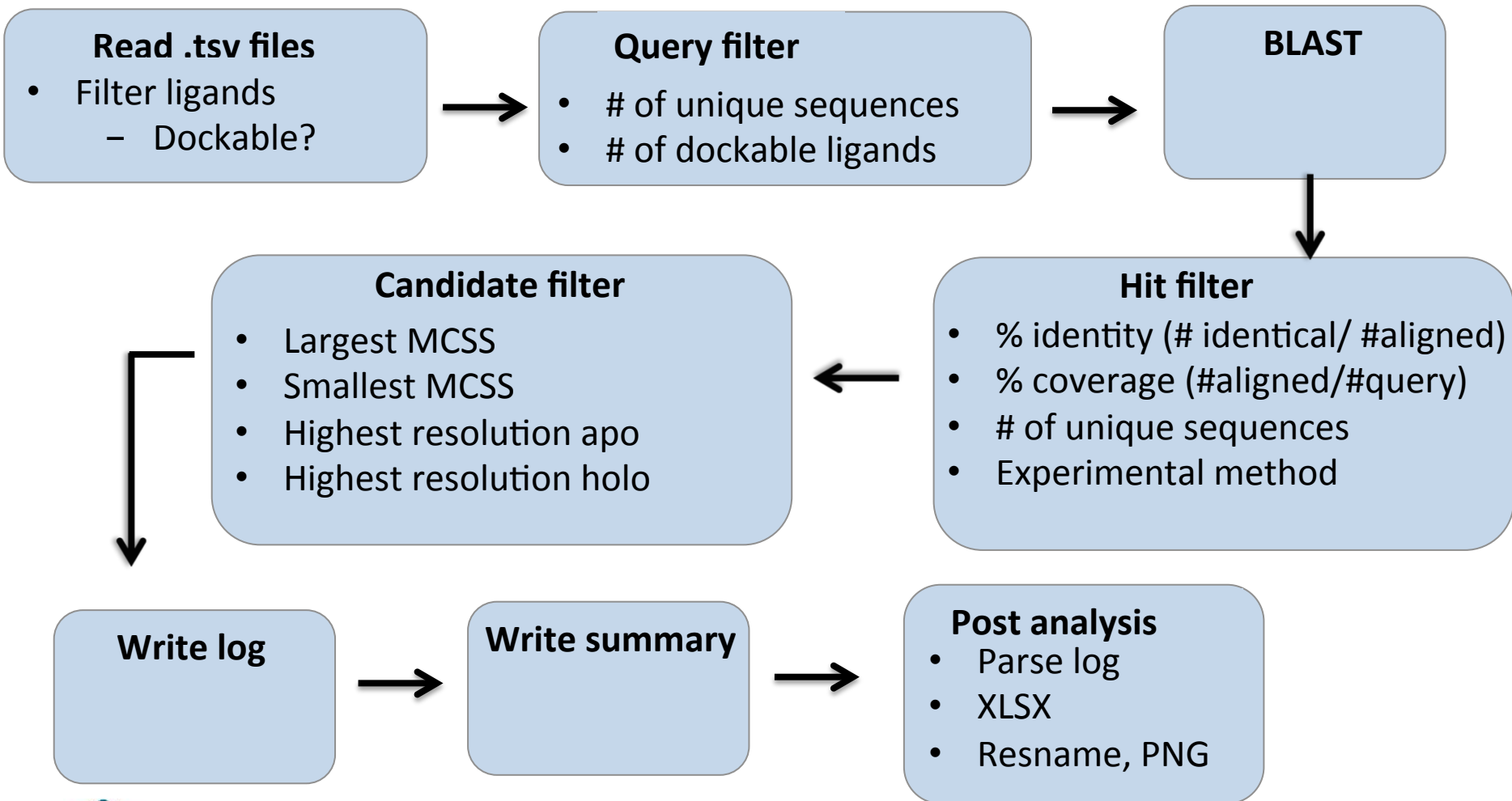
Drug Design Data Resource

CELPP Status

pdbblastdb (/data/blast-output/...)

stage.1.dataimport

stage.1.compinchi

pdbextracted (/data/pdb.extracted/...)

stage.2.blastnfilter

Completed, but will be replaced by improved version

stage.3.pdbprep

stage.4.fred

stage.4.glide

stage.4.webdata

stage.4.ext<name/id>

stage.5.fred.scoring (ROCS,RMSD,RSR)

stage.5.glide.scoring (ROCS,RMSD,RSR)

stage.5.ext<name/id>.scoring (ROCS,RMSD,RSR)

stage.6.resultwebdata

Green boxes are currently complete and running

Drug Design Data Resource

CELPP Week

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday

Tuesday:
8:01 pm
1.pdbdownload

11:01 pm
5.*.evaluation
6.resultwebdata

Friday:
8:01 pm
1.makeblastdb
1.dataimport
2.blastnfilter
3.proteinligprep
4.webdata
4.fred
4.glide

12:01 am Saturday to 2:59 pm Tuesday
User submissions accepted

3:00 pm Tuesday to 12:00 am Saturday
No user submissions allowed during this time 🚫

12:01 am Saturday to 2:59 pm Tuesday
User submissions accepted

Drug Design Data Resource

Stage 2: BLAST-n-Filter Workflow

Read .tsv files
- Filter ligands
  - Dockable?

Query filter
- # of unique sequences
- # of dockable ligands

BLAST

Candidate filter
- Largest MCSS
- Smallest MCSS
- Highest resolution apo
- Highest resolution holo

Hit filter
- % identity (# identical/ #aligned)
- % coverage (#aligned/#query)
- # of unique sequences
- Experimental method

Write log

Write summary

Post analysis
- Parse log
- XLSX
- Resname, PNG

Drug Design Data Resource

Numbers by week

OUTPUT SUMMARY
Targets found: 66
Target: 5i3r|Sequences: 1|Hits: 336|Candidates: 2|Elected:2|PDBids: 4w9x,4w9w
Target: 5aal|Sequences: 1|Hits: 40|Candidates: 37|Elected:7|PDBids: 3mcy,4x5p,4xob,4xo9,4ca4,1kiu,4xod
Target: 4zg3|Sequences: 1|Hits: 102|Candidates: 82|Elected:4|PDBids: 2vu6,3wxs,5avg,2vhk
Target: 4yyy|Sequences: 1|Hits: 20|Candidates: 10|Elected:5|PDBids: 4yek,4lhm,4eaf,4ead,1tpt
Target: 5ajg|Sequences: 1|Hits: 44|Candidates: 15|Elected:5|PDBids: 4cqh,1ztu,4q0j,4o0p,4o01
Target: 5eh8|Sequences: 1|Hits: 492|Candidates: 440|Elected:4|PDBids: 5fnm,3ks3,3k34,2hnc
Target: 5i3o|Sequences: 1|Hits: 336|Candidates: 2|Elected:2|PDBids: 4w9x,4w9w
Target: 5hk7|Sequences: 1|Hits: 40|Candidates: 4|Elected:1|PDBids: 4lto
Target: 5ehv|Sequences: 1|Hits: 492|Candidates: 440|Elected:4|PDBids: 3ks3,3k34,2hnc,2f14
Target: 5ehw|Sequences: 1|Hits: 492|Candidates: 440|Elected:4|PDBids: 5fnk,3ks3,3k34,2hnc
Target: 5dro|Sequences: 1|Hits: 39|Candidates: 14|Elected:4|PDBids: 3p3c,4u3d,1yhc,1p42
Target: 5i91|Sequences: 1|Hits: 6|Candidates: 1|Elected:1|PDBids: 1vr3

Stage 4: Docking

- Current on our server
  - OEDocking
  - CCDC: GOLD
  - Schrodinger: Glide
  - AutoDock

- Outreach to
  - Shoichet, DOCK 3.7
  - Robert Rizzo Dock 6.7
  - Jain, Surflex

Stage 5: Evaluation

- Evaluations will be same as those for Grand Challenges and automated in the CELPP workflow; RMSD, RSR, ROCS (or some type of interaction fingerprint)

Drug Design Data Resource

Feedback

- Ideas for how best to post results?

- What sort of results should we post?

- Would it be useful to post results relative to using various ligand preparation methods?
    - Ie. tautomer prediction, pKa predictions

Next Challenges

ITK – Genentech donation

FXR – Roche donation

FBPase – Roche donation

**Drug Design
Data Resource**

Farnesoid X receptor (FXR)
Roche

A nuclear receptor involved in cholesterol and bile acid homeostasis

Antagonists are potential dyslipidemia lowering agents and potential Type II diabetes treatments

The binding site is highly lipophilic and flexible

4 distinct chemotypes + 6 miscellaneous compounds

There are ~ 25+ published crystal structures, including examples of these chemotype crystal structures published

Chemical series	No. of compounds	Activity range for active compounds IC_{50} (µM)	No. of blinded experimental structures	Resolution(Å) range of blinded crystal structures
4 main	102	0.008-263	36+1 apo	1.8 – 2.6

Drug Design Data Resource

Fructose-1,6-bisphosphatase
Roche

A protein involved in gluconeogenesis, converts fructose-1,6-bisphosphate to fructose-2,6-bisphosphate

Target for decreasing glucose production for TypeII diabetes

Allosterically regulated; the compounds are targeting the allosteric site

2 chemotypes + some miscellaneous compounds

There are ~ 20+ published crystal structures, including examples of these chemotype crystal structures published

Chemical series	No. of compounds	Activity range for active compounds IC_{50} (μM)	No. of blinded experimental structures	Resolution(Å) range of blinded crystal structures
1 (2)	82	0.017-21	23+1 apo	1.9 – 2.8

Drug Design Data Resource

D3R Team

Rommie Amaro   Vicki Feher   Mike Gilson
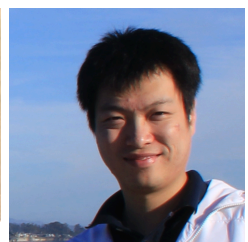
Symon Gathiaka   Shuai Liu   Jeff Wagner

Chris Churas   Jeff Grethe   Mike Chiu

Stephen Burley
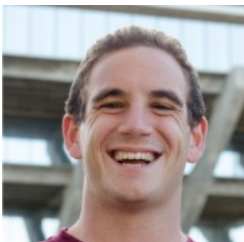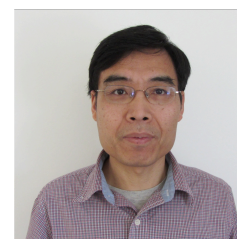
Huanwang Yang   Jasmine Young

Thanks to Previous Dataset Donors

- Roche

- Genentech

- GSK

- Abbvie

- Vertex

**Drug Design
Data Resource**