HYDRA:

A HIGH-THROUGHPUT VIRTUAL SCREENING DATA VISUALIZATION AND ANALYSIS TOOL

Nara Institute of Science and Technology (NAIST)

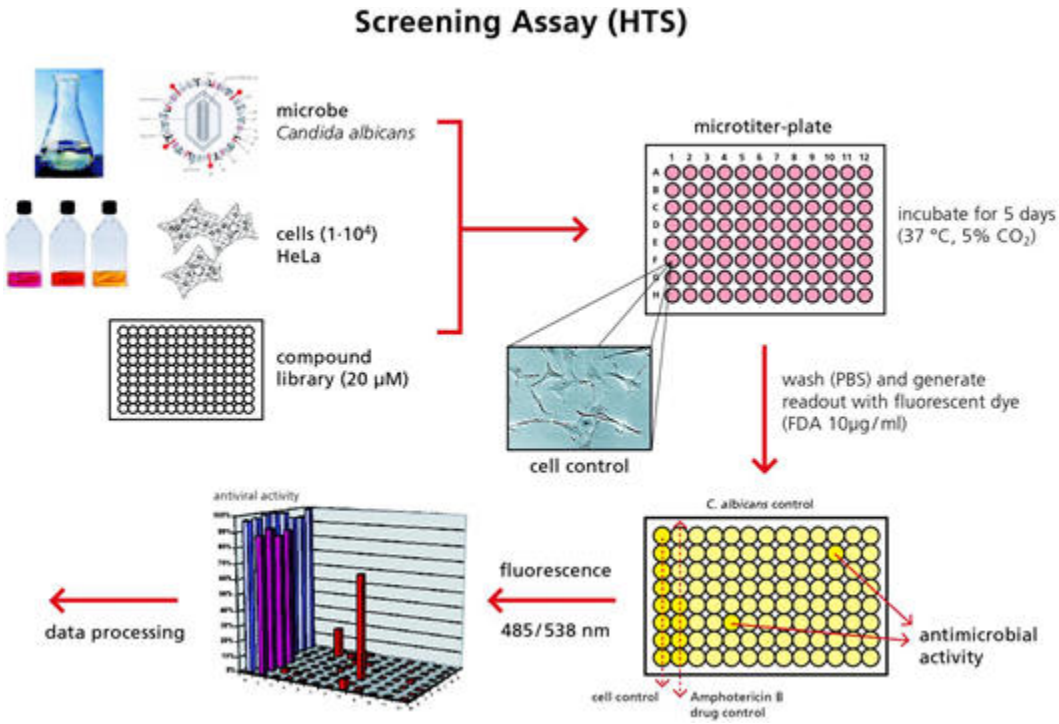
Nara, Japan

Curtis Sera, Kohei Ichikawa, Jason H. Haga

9 October 2015

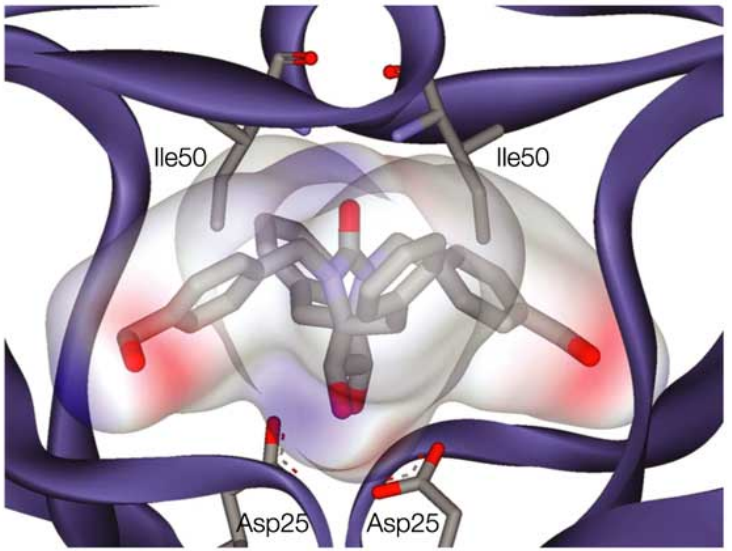
Background: High-throughput Screening

- Uses chemical libraries
  - From mid-1990s
  - Tests whole classes of compounds
    - Preliminary screening to chose parts of the library to test using basic chemical properties
  - 10,000-100,000/day
- Biology and drug discovery uses
- Problems:
  - Expensive
  - Resource intensive
  - False negatives



Background: Virtual Screening

- Simulates molecular interactions to predict compounds most likely to successfully interact
  - Uses crystal structures of compounds and target
  - Various methods
- Reduces necessary testing library size
  - Cost saving
  - Fewer false negatives
  - Improved “hit” rates



Nature Reviews | Drug Discovery

Dmp223 complexed with HIV protease with a surface showing relative electrostatic potential

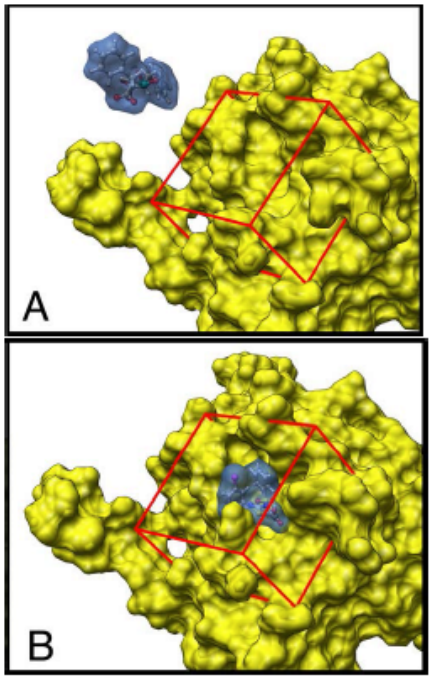
Motivation

- Virtual Screening is imperfect

		RANKINGS			SCORES	
Rank	ZINC ID	Total	Energy	AMBER	Energy	AMBER
1	4538484	1	1	-	-132.662	-
2	5553470	2	2	-	-127.161	-
3	3957200	4	4	-	-118.597	-
4	3193087	23	17	6	-92.150	-6.58E+11
5	5518012	33	24	9	-86.708	-4.96E+11
6	4410251	47	33	14	-82.004	-1.31E+11
7	4811922	150	145	5	-68.236	-6.71E+11
8	6645919	162	149	13	-68.067	-1.84E+11
9	5093622	178	157	21	-67.863	-1.96E+10
10	1532056	178	178	-	-67.280	-
11	6645916	183	172	11	-67.505	-2109
12	4892459	185	185	-	-67.120	-
13	4892457	219	219	-	-66.600	-
14	2306159	246	159	87	-67.837	-132.6018
15	2139774	320	68	252	-72.501	-105.0482
16	5093617	322	307	15	-65.064	-1.2E+11
17	2391170	362	350	12	-64.613	-1.94E+11
18	1874963	412	386	26	-64.220	-3.65E+09
19	6664179	563	316	247	-64.928	-105.7376
20	5413467	591	539	52	-63.027	-219.4566
21	5093619	636	607	29	-62.609	-2.41E+09
22	4089903	706	702	4	-62.083	-7.63E+11
23	3434768	714	420	294	-63.919	-102.3496
24	3399076	745	508	237	-63.247	-106.8366
25	4366621	761	392	369	-64.165	-99.03696

A: Rank 11

B: Rank 25



Target - protein molecule with characterized active site
Inhibitor - small compound that binds tightly to active site

Levesque MJ, Ichikawa K, Date S, Haga JH. Comput Methods Programs Biomed. 2009 Jan;93(1):73-82.

- Existing software solutions lack user-friendliness
  - Inhibits use in non-specialized biology labs

Project Overview

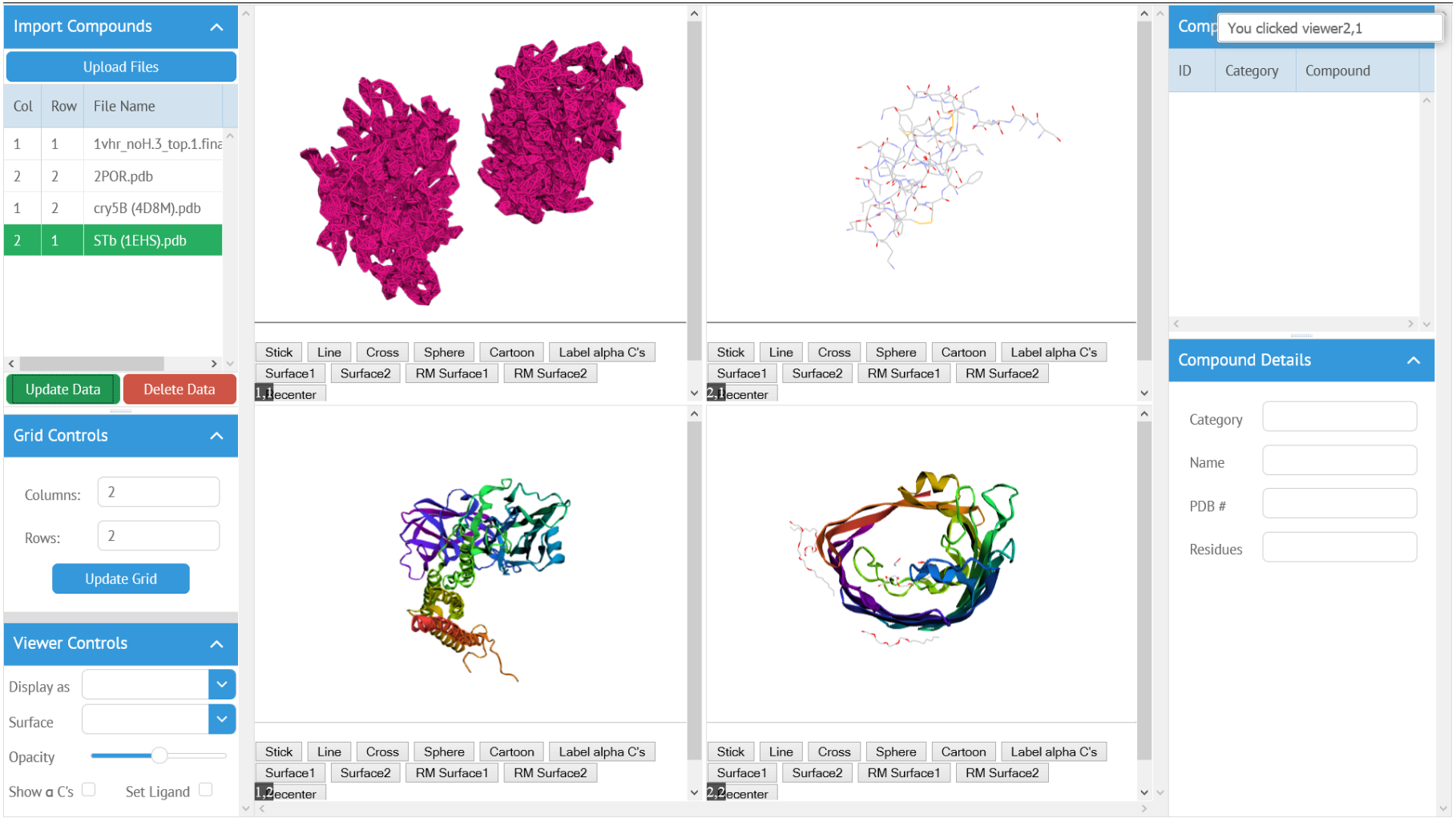
- Initial framework built by Yuan Zhao

- Primary goal
  - Creation of a *user-friendly* browser-based program to simultaneously display many molecular interactions in a dynamically sized grid of molecular viewers
    - Simulated interactions obtained from separate programs
  - Will enable usage on almost any device with an internet browser
- Secondary goals:
  - Improved efficiency
  - Unified control of model viewing
  - Ability to use raw data from virtual screening programs
  - Pull compound-specific information (properties and availability for purchase) into Hydra

Methods Overview

- Initial framework made about equal use of HTML5 and JS
- Made heavier use of JavaScript
  - Webix for GUI
  - Webix and jQuery for functionality
- Used WebGL-based molecular viewers
  - Initially used GLmol.js
    - Slow when manipulating large or complex models
    - Inadequate documentation
    - Did not support all necessary file types
  - Switched to 3Dmol.js
    - Improved efficiency: usage of web workers
    - Responsive developer, extensive comments, better documentation
    - Supported all required file types

Week 4



Dev. snapshot: Switched to 3Dmol.js. Upper left compound displaying incorrectly.

17 July 2015

Post-internship

- Back-end improvements:
  - Unified left and right panel data objects now synchronize to a single data object collection
    - Selection between the uploaded files table (left panel) and the compound list (right panel) is synchronized
  - Support for compound metadata in .mol2 files
  - Added vendor purchase information database
    - “Boutique Shards” chemical data set from ZINC database
  - Purchase information added to individual objects rather than in bulk to the vendor list
    - Vendor list now only displays information for the currently selected compound
  - Updated Webix library: bug fixes

Results

- All code is available on GitHub:
  - https://github.com/csera/Hydra

- Live demo

Conclusion

- The project's primary goal was successfully achieved
  - User-friendly program
    - No set-up required
    - Easy-to-use interface
    - Dynamically sized grid of molecular viewers
- Secondary goals also achieved
  - 3Dmol.js greatly increased the efficiency of the program
  - Single, unified control set for all viewers
  - Files outputted by the DOCK molecular simulation program can be processed within Hydra and sent directly to the main interface
  - Information on a compound's chemical properties and purchase availability is shown directly in Hydra

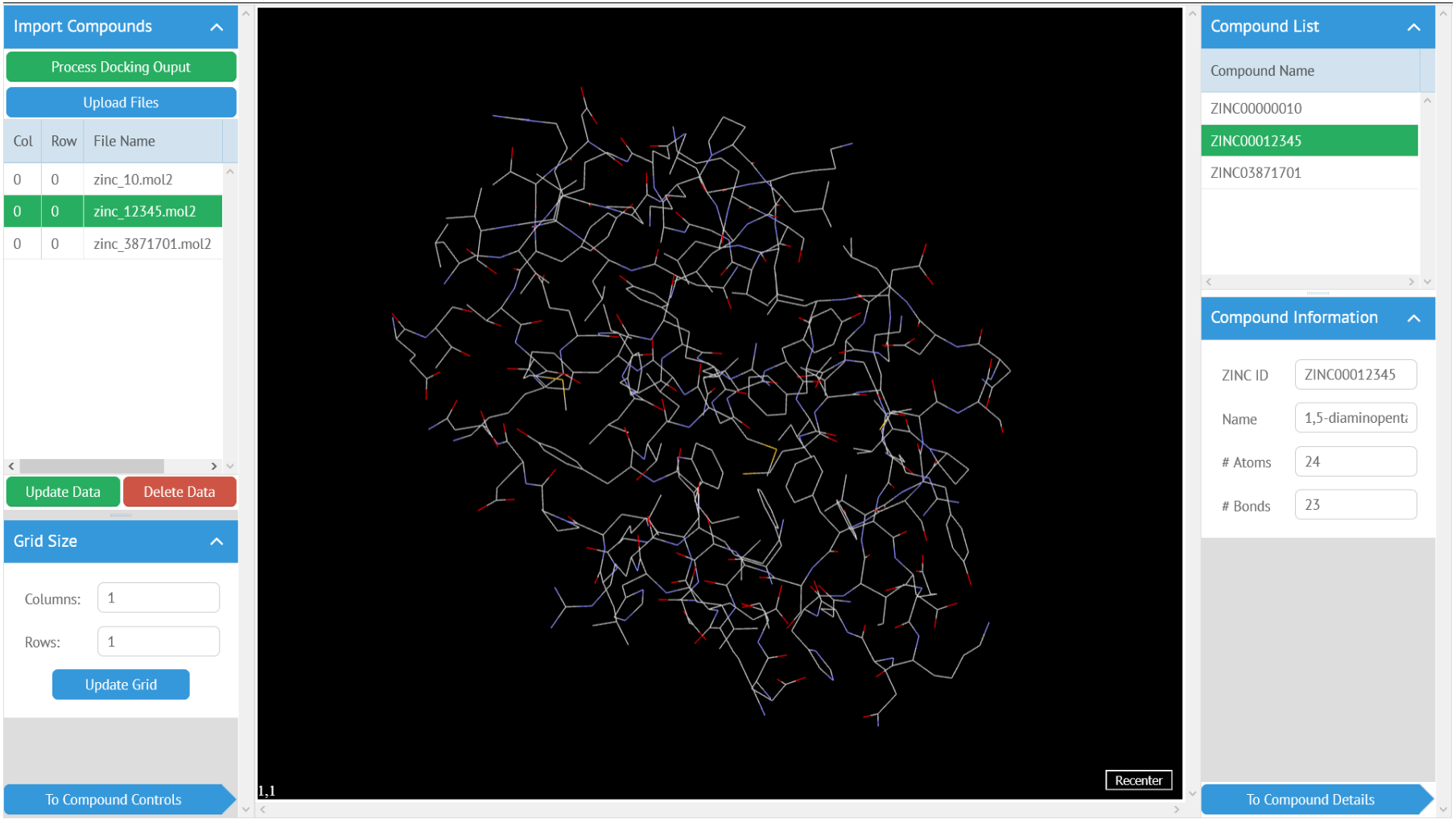
Future Work

- Improved algorithm for searching databases
  - Current method can cause noticeable lag
- Add markup to vendor (purchase) information popup
  - Add hyperlinks, mailto, etc
- Fix synchronized model manipulations
  - Currently works but does not transmit events until after initial mouse release

Acknowledgements

- I would like to thank all those involved in making this possible:
  - Mentors
    - Dr Jason Haga & (Kohei) Ichikawa-sensei
  - UCSD Pacific Rim Experiences for Undergraduates (PRIME)
    - Dr Gabriele Wienhausen, Teri Simas, Madhvi Acharya, & Jim Galvin
  - NAIST's Software Design & Analysis Lab (SD Lab)
    - (Hajimu) Īda-sensei, (Yasuhiro) Watashiba-sensei, & (Tomoko) Arai-sensei
  - Financial support
    - Julia Brown
      - UCSD Undergraduate Research Scholarship coordinator Dr Sophia Tsai
    - NAIST & Japan Student Services Organization (JASSO)
      - NAIST coordinator Nao Terada
    - National Science Foundation
  - Additional thanks to all the SD Lab graduate students, the PRIME alumni, Yuan Zhao, and the developers of Webix & GLmol
    - Special thanks to 3Dmol developer Dr David Koes for his great helpfulness and communicativeness

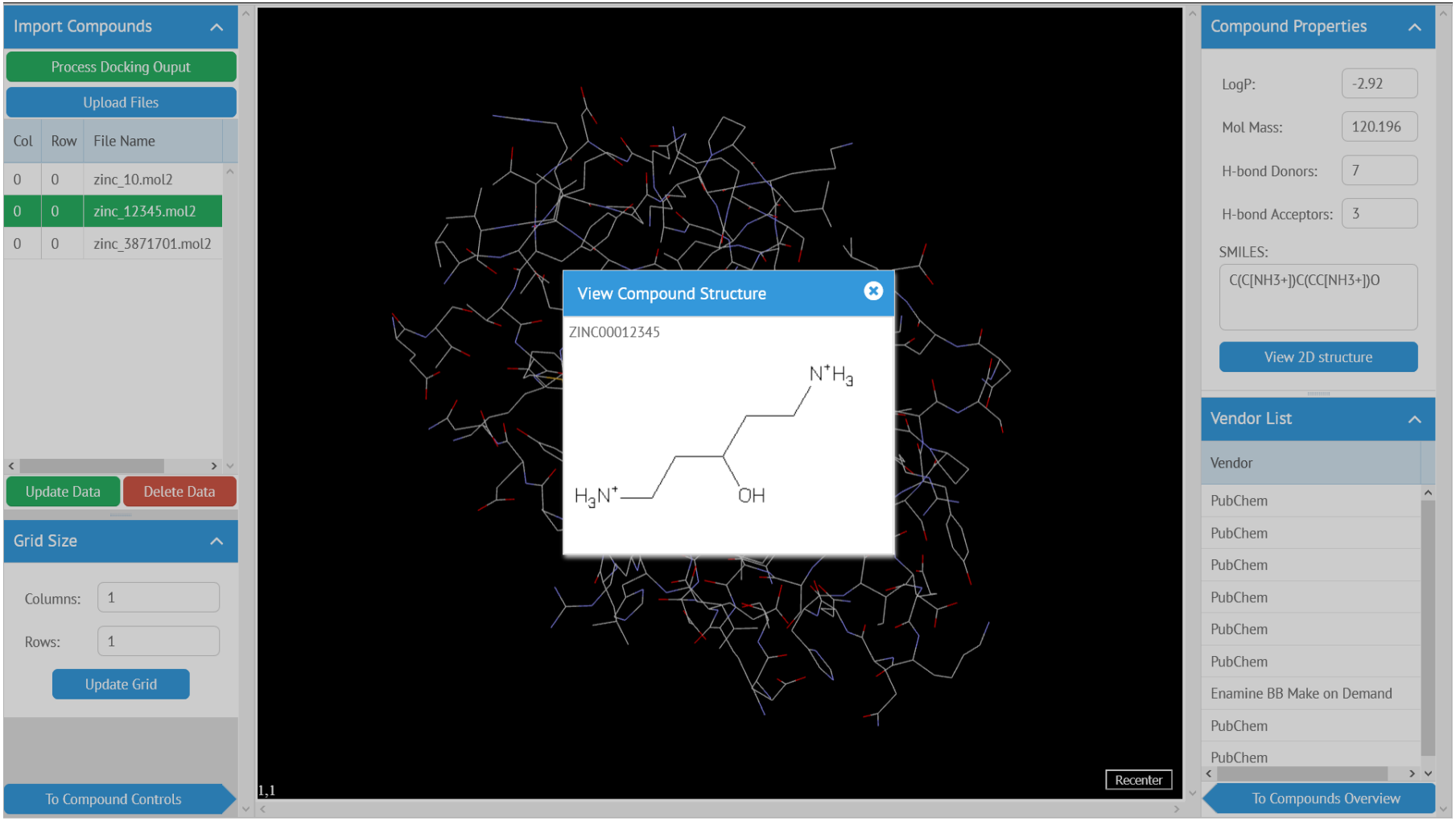
Post-internship (cont)



Dev. snapshot 1: Synchronized tables and new right panel

7 October 2015

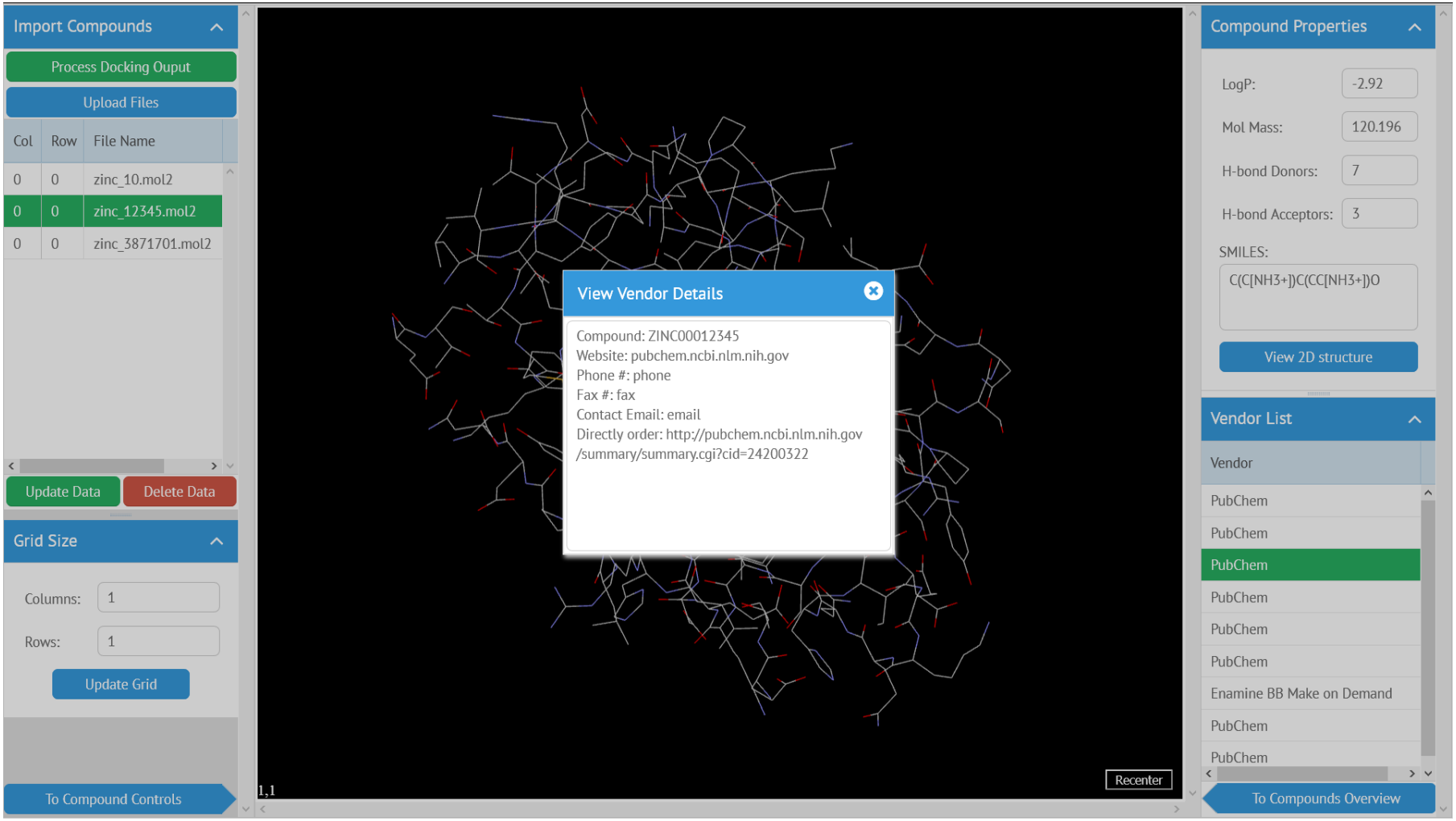
Post-internship (cont)



Dev. snapshot 2: Updated “Vendor List” and new properties pane with structure popup

7 October 2015

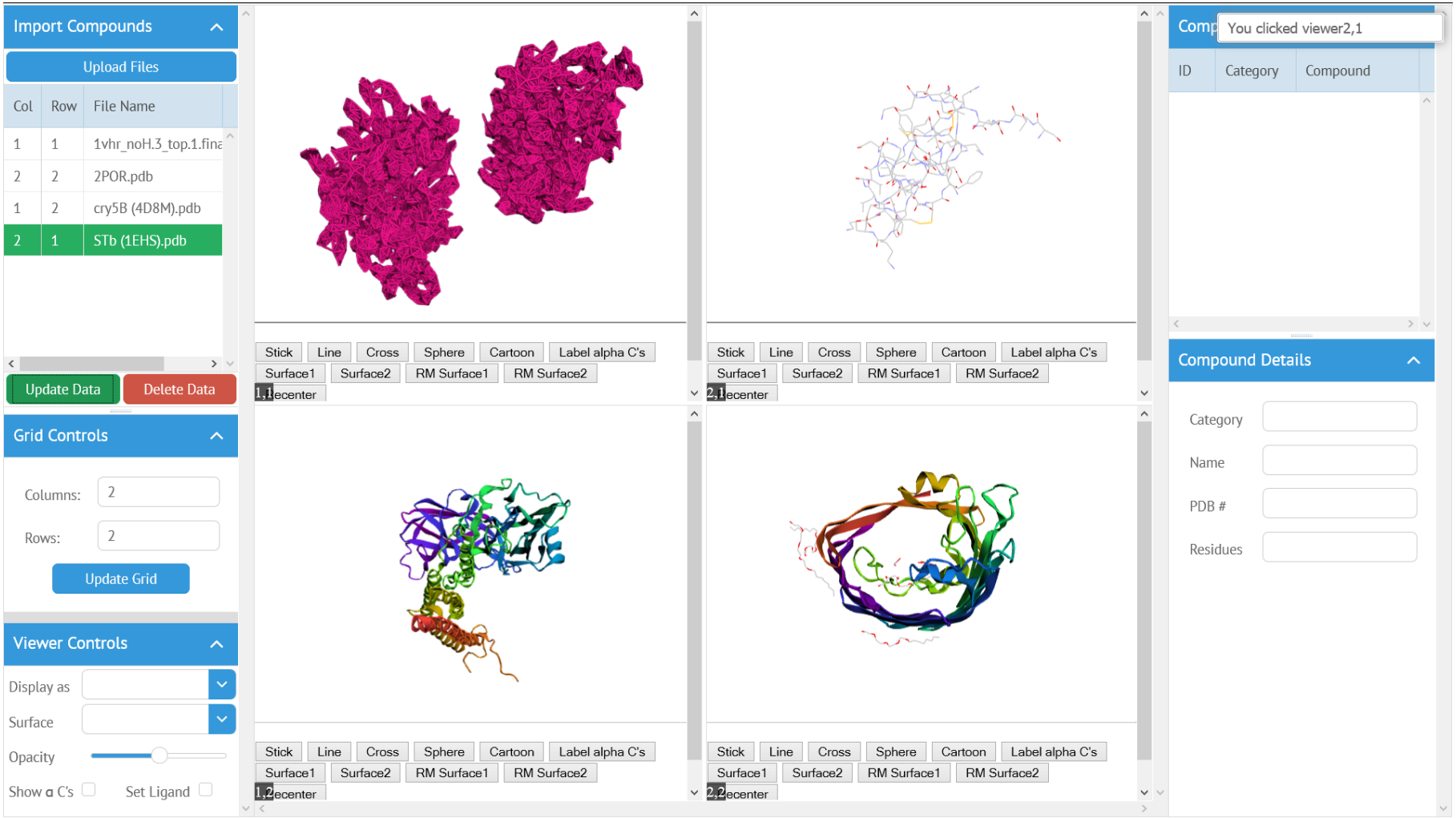
Post-internship (cont)



Dev. snapshot 3: Improved vendor popup

7 October 2015

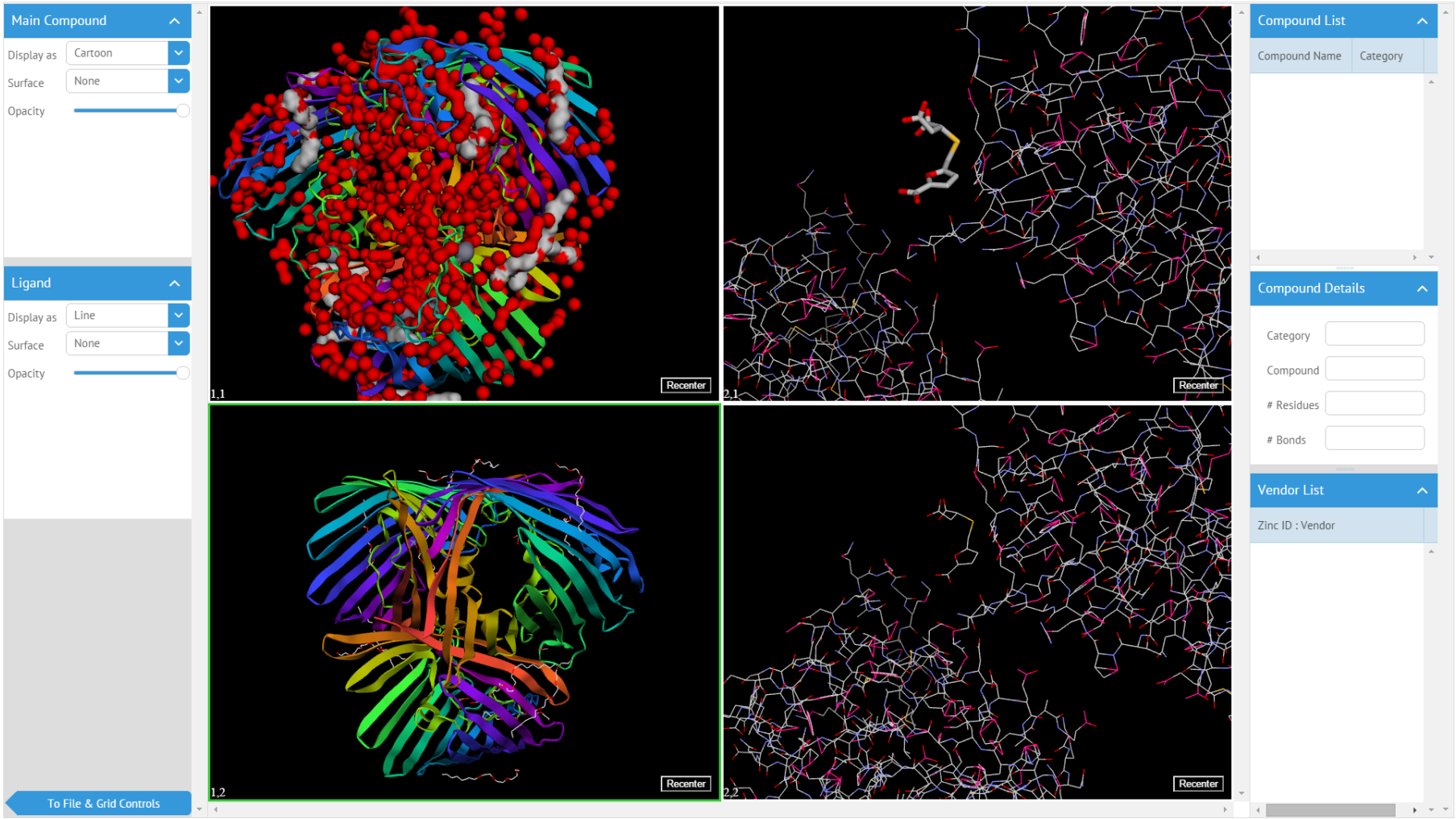
Week 4



Dev. snapshot: Switched to 3Dmol.js. Upper left compound displaying incorrectly.

17 July 2015

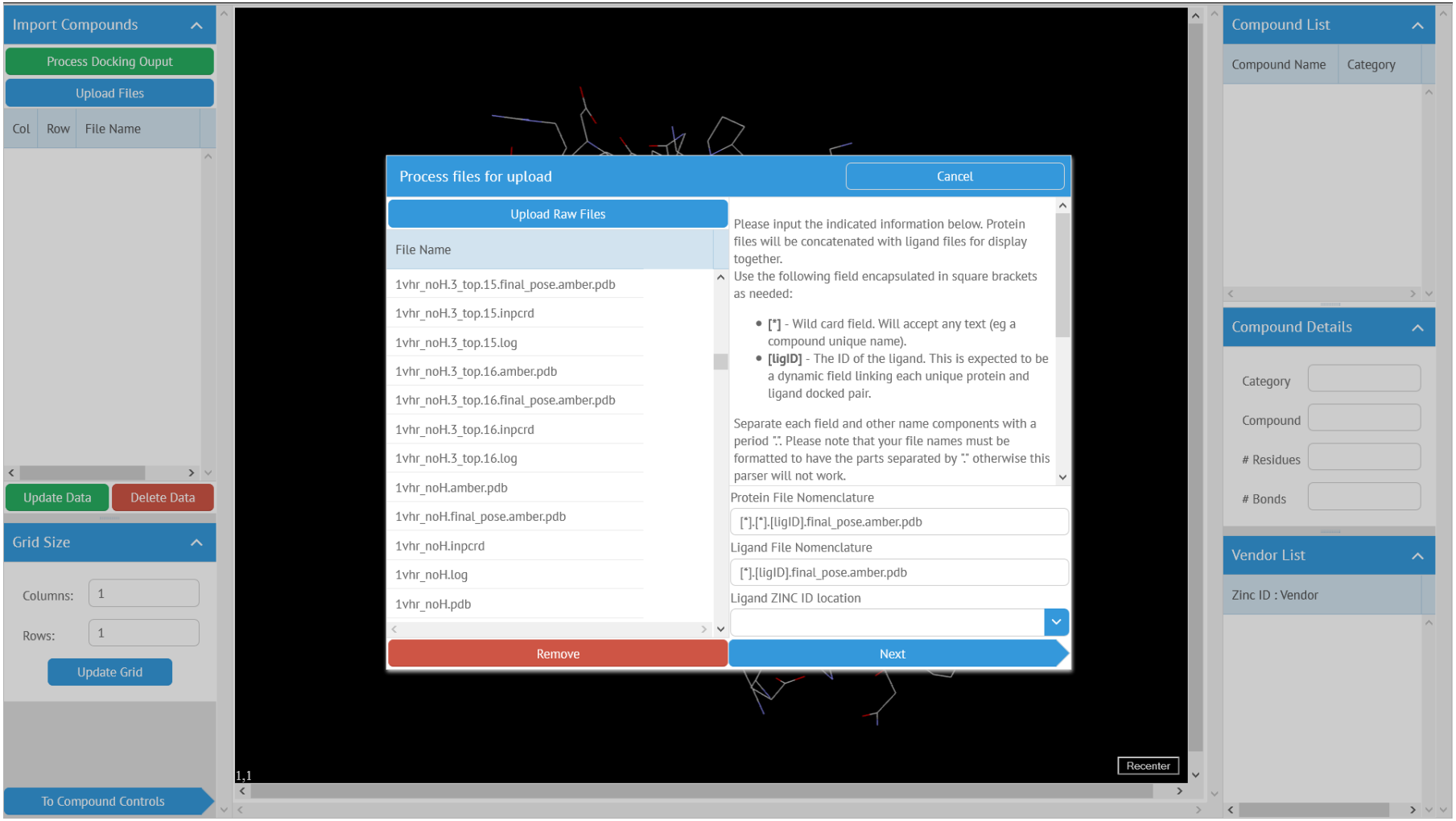
Week 6



Dev. snapshot: The same files are loaded on the top and bottom with different ligand settings. As marked by the green border, the bottom-left viewer is currently selected.

31 July 2015

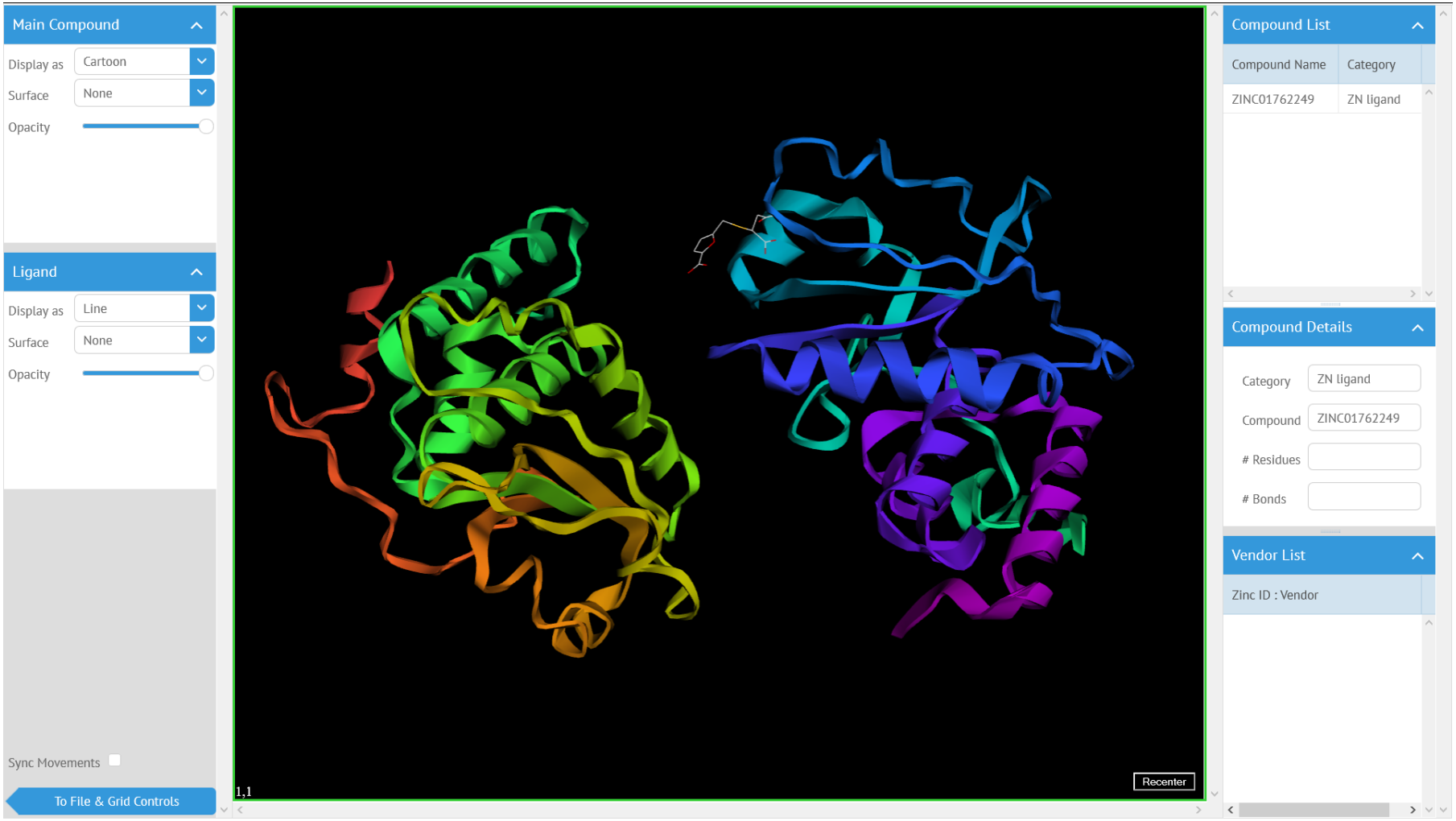
Week 7



Dev. snapshot: New file processing GUI with test files loaded and inputted filters.

7 August 2015

Week 8



Dev. snapshot: Successfully processed file set loaded into a viewer

14 August 2015