Second CCP CMC - Workshop Cambridge - 21st June 2017

Attendees

Oxford / DLS	**Anthony Bradley**
SGC / DLS	**Frank von Delft**
AstraZeneca	**Sam Hughes**
Charles Rivers	**Dave Clark**
CCDC	**Jason Cole**
Informatics Matters	**Tim Dudgeon**
eTherapeutics	**Ben Allen**
GSK	**Ian Wall**
UCB	**Will Pitt**
ChEMBL	**Andrew Leach**
ICR	**Nathan Brown**
Oxford	**Garrett Morris**
Sheffield	**Val Gillet**
Vernalis	**Steve Roughley**
Suffolk/Elixir	**John Hancock**
Cambridge	**Andreas Bender**
Oxford	**Brian Marsden**
Coot/CCP4	**Paul Emsley**
MD Catapult	**John Overington**
IOTA	**David Bailey**
Cambridge	**Fredrik Svensson**

Agenda

10.30-11 Arrive (coffee and tea available)

11-12 AM Introduction and MD Catapult (Frank von Delft DLS and John Overington MD Catapult)

- Frank will outline the InnovateUK Proposal and reintroduce CCP CMC
- John Overington will describe the MD Catapult

12-12.30 PM SQUONK progress and examples (Anthony Bradley DLS/Oxford and Tim Dudgeon IM)

- Anthony and Tim will outline recent developments for SQUONK (https://squonk.it/)
- Video - https://www.youtube.com/watch?v=Ek1ojZiIAG0&feature=youtu.be
- Reaction enumeration, docking, molecule standardisation, molecule scoring workflows

12.30-1.15 PM lunch

1.15-2 PM Proposal for Workshop / Study Weekend (Garrett Morris Oxford and Nathan Brown ICR)

- Garrett and Nathan to outline training initiative similar to CCP4 study weekend (20 mins)
    - Aimed at enlisting and supporting naive users
    - Reach a broad agreement about how to progress this in CCP-CMC.
- Group discussion on workshop / study weekend - what each want to achieve (25 mins)

2-3.45 PM Crowdsourced challenge: brainstorm on XChem dataset / issues (whole room):
Format - proposal/summary by expert and then roundtable discussion

- Shared community approach
- Infrastructure gaps and scientific gaps
- Funding opportunities

From this we'd like to establish three things:

1) What is current best-practice for these tasks?
2) What are the specific challenges that could be addressed?
3) How can people be involved - what can people contribute - to resolve these problems?

We will be focussing on three distinct tasks

1) Hit-triaging from an ensemble of fragment structures (Ian Wall GSK) 10 minutes of proposal + 1-2 mins each of thoughts
2) Follow-up workflow on a given hit (Fredrik Svensson - IOTA / Cambridge) 10 mins proposal + 1-2 mins each of thoughts
3) Compound acquisition (Tim Dudgeon - IM) 5 minutes proposal + 10 minutes of general thoughts

3.45-4.15 PM Wrap-up and future planning

- Actions for achieving ambitions
- Schedule and purpose of next meeting(s)
    - November at EBI - Lead by ChEMBL / Elixir?
    - Every 3-6 months? Split into different interest groups? (e.g. Comp-chem / cheminformatics)

4.30 PM Close

CCP CMC Challenges

Data for challenges one and two is here:
https://drive.google.com/open?id=0BzjYwKmOBCrqX0pYc0NxY0wtcWM

Anthony Bradley thoughts on these topics

Challenge one - hit triaging

Issue One - Providing analysis to structural data

a) Generic APIs for atom-based scoring data
b) Generic visualisers for considering that data (e.g. JavaScript components)
    i) Incorporation into Proasis, 3Decision and CCP4 tools
c) Can we provide value added analysis to the community for openly-available data on (e.g. Proasis / 3Decision)

Issue Two - Once all data / analysis available - how do I decide?

a) Docking scores - and removal of crud
b) Semi-intelligent selection and prioritisation (e.g. of potential merging projects)

Input data: an ensemble of protein-ligand complexes
NUDT7A-x0159_event2.pdb, NUDT7A-x0254_event1.pdb, NUDT7A-x0452_event1.pdb, NUDT7A-x0132_event1.pdb,NUDT7A-x0140_event1.pdb

Challenge two - hit follow-up

Issue One - Availability of tools / best practice for such workflows

a) Reaction nodes - how to construct nodes - e.g. (reaction SMARTS / vectors)
b) Filtering and curation (tautomer standardisation / enumeration)
c) Scoring of available data (constrained docking) - integration with D3R and similar concepts

Issue Two - Connection of available tools

a) Common workflow language integration - Chemistry specific APIs

Input data: an individual hit series (merging and individual series)
NUDT7A-x0452_event1.pdb and NUDT7A-x0132_event1.pdb

Challenge three - compound acquisition

Providing “useful” systems for finding purchasable compounds

a) Available - and queryable dataset - focus on Enamine Real Space Database
b) Chemically sensible ways of viewing (e.g. Astex Fragment Network like)
c) Possibility to serve and maintain as a community resource

Input data: a large (millions of compound) library - e.g. from Enamine