CCP CompMedChem Meeting - 11 Jan 2018

Executive Summary: A third meeting at the EBI to outline recent progress and desribe future plans. Three areas were discussed:

- The first CCP CompMedChem specific resource
- Development of potential partnering projects and useful methods from PDBe, ChEMBL, Sanger, GSK and CCP5
- The future shape of the group and proposed future plans

Outcomes:

- eCheminfo meeting / CCP CMC Training and exposure - September
- Building first CCP CMC workflows and technology at Diamond
- Next CCP CMC meeting - September - location TBD and organiser TBD (AB organise organiser)
    - Suggestion to be in London
    - Opportunity to discuss new methods.
    - Contribute ideas about what you'd like to see?
- Projects between ChEMBL and Diamond and Harry Jubb and Diamond / PDB

At EBI (21): Anthony Bradley - Oxford / DLS; Frank von Delft - SGC / DLS; Sam Hughes - AstraZeneca; Jason Cole - CCDC; Tim Dudgeon - Informatics Matters; Ben Allen - eTherapeutics; Ian Wall - GSK; Will Pitt - UCB; Andrew Leach - ChEMBL; Nathan Brown - BenevolentAI; Garrett Morris - Oxford; Steve Roughley - Vernalis; Wendy Warr; Aleksandras Gutmanas - PDBe; Chin Yong - STFC; Kuen Yeap - Charles Rivers; Kristian Birchall - LifeArc; Loeffler Hannes - Eli Lilly; Steve St-Gallay - Sygnature; John Liebeschuetz - Astex; Harry Jubb - Sanger;

http://www.ccp-cmc.org

Agenda

09:30 Arrivals & coffee
10:00 Introductions & overview of agenda (all)
10:20 Update on progress towards an SBDD pipeline and related activities (Anthony Bradley, Diamond)
11:00 Introduction to a newly initiated project on binding sites (Aleks Gutmanas, EBI)
11:30 Summary of recent experiences with simulation methods (Ian Wall, GSK)
12:00 Recent changes to ChEMBL and future plans (Andrew Leach, EBI)
12:30 COSMIC-3D (Harry Jubb, Sanger)
12:45 CCP5 update (Chin Yong, STFC)
1:00 lunch, informal discussions, demos etc.
1:45 Discussion on future directions of this group (Andrew Leach, all)
2:45 AOB
3:00 departures

Update on progress towards an SBDD pipeline and related activities (Anthony Bradley, Diamond)

Can fund formal CCP-CMC development starting now Jan 2018
Goal 1: Standard workflows for normal fragment follow-up; — just login and start doing science.
Goal 2: much smarter scoring and enumeration. — Deploy latest algorithms; capture expt outcomes; prepare ground for AI approaches.

- Outlined stack at Diamond of GitHub -> TravisCI -> Docker -> Luigi and other tools to run
- Stack - offers easy addition by others
- Eg1: Simple tools e.g. docking - easy addition of multiple components
- Eg2: Graph Network - with Docker-Compose - scalable instant tools
- Eg3: Scientifically Complex Tools (e.g. complex projects like Dynamic Undocking) requires co-ordination and incentive
- ~2 FTE Team at Diamond to implement resource
    - Anthony will lead it (WT funding it);
    - Rachel Skyner
    - Recruiting compchem engineer

11:00 Introduction to a newly initiated project on binding sites (Aleks Gutmanas, EBI)

FunPDBe - (UK) community-driven annotations: (1) functional sites and biological assemblies (Thornton) annotations mapped onto PDB structures; (2) enzyme mapping; (3) genetic variant (Sternberg) mapping; pharmacophore and ligand binding (CREDO) annotation (Blundell)
PDB (43,000 non Ab proteins);
1.5M proteins with 90% sequence ID & 70% coverage.

- Will have APIs for input and output of data
- Github project with issue tracking
- http://funpdbe.org - registered domain for resource
- Agreed draft schema for data exchange (res-level; evidence and sequence ontologies; confidence.
- API & Neo4j graph db that can be downloaded for advanced queries
- LiteMol-based interactive molecule viewers
- New mmCIF format with better chemical data

11:30 Summary of recent experiences with simulation methods (Ian Wall, GSK)

Ian Wall discussed FEP and MM-PB(GB)-SA methods, including Schroedinger FEP and Peter Coveney ESMACS method and QM-based scoring functions (e.g. FMO). Automated parameter generation a major advantage of Schrödinger.

Other gaps:

- Absolute binding free energy methods for unrelated molecules;
- High-throughput methods for assessing VS results.

Other pertinent points

- MD - sporadic use with anecdotal successes;
- Protein conformational analysis, cryptic pockets, *etc*.;
- Key requirement is to be able to translate output into robust designs
- FEP failures involved saturated 7-rings (started in equatorial solution-phase but axial position is needed to bind...) ΔG overestimated.
- Tautomers need careful consideration in FEP preparation.
- MM/PBSA or MM/GBSA might be good enough/on the border to being usable as an automated triage/filtering tool before FEP is run.

12:00 Recent changes to ChEMBL and future plans (Andrew Leach, EBI)

Andrew Leach outlined Open Data & Resources @ EMBL-EBI.

- ChEMBL - SAR Data 1.7 m cpds; 15 m activities
- SureChEMBL - patent data

He then discussed upcoming technical projects

- Enhancing release process
- New DB schema
- New ChEMBL web interface https://chembl-glados.herokuapp.com
- "Make ChEMBL" green button (is AL's goal)

Future plans (Ideas):

- Drug development pipelines;
- New sources (ChEMBL is based on core journals); New modalities;
- Maximizing value from patents;
- Deposited datasets (including negative data)
- Data publishing workflows. (PDB model of data publishing required when paper published.)

Future plans (Infrastructure):

- UniChem enhancements (based on InChI, substructure searching, layers);
- New loader and schema;
- Complex data and ools and models.

COSMIC-3D: Exploring cancer driver mutations and druggability in three dimensions (Harry Jubb, Sanger / Astex Pharmaceuticals)

Harry Jubb presented COSMIC3D. COSMIC (Catalogue of Somatic Mutations In Cancer) is the world's largest resource for cancer mutations.
His work is mapping from genome to structural proteome for cancer.
Several examples showing how this helps direct design on the basis

CCP5 Software Tools: DL_FIELD and DL_ANALYSER (Chin Yong, STFC)

Chin Yong presented CCP5's work on creating single, unified format for all forcefileds.

AMBER; CHARMM; OPLS-AA/2005 (OPLS is available in Tinker); DREIDING; PCFF; CVFF; Gromos-G54A7
Yong (2016) JCIM 56: 1405

DL_FIELD uses atom keys to map from different force fields.

New standard expression system to annotate specific atomic interactions:

Yong & Toderov, (2018) *Molecules* 23:36.

Eg ethanoic acid liquid interactions with a H-bond network sponge-like with methyls in holes.

Group discussion on the future of CCP CMC

Andrew Leach presented an overview:

- What is the purpose of this group?

- What is it aiming to deliver?
- Are its functions already covered by other groups?
- What needs to happen next?

Computational methods in drug discovery: from *post hoc* “rationalization” to prediction.

Reasons to believe?

FEP
Predicting cardiovascular safety (in silico and exptl)
Cell / systems biology modelling by guy at Stanford

A vision?
True in silico drug discovery
Drug design with in vitro and in silico assays
“From disease to hypothesis to the clinic in one year”

High Level Goals / Mission

- to fully exploit uk/european expertise in the application of computational sources to drug discovery to deliver:
- Maximize investments in methods dev & avoid duplication of effort
- Achieve critical mass & greater visibility
- Create community
- Deliver tools, capabilities, expertise of practical value
- Demonstrate the value of out science to drive further investment

Where do we fit in → hands-on tools for drug discovery.

<u>What are CCPs?</u>
The main activities of the CCPs are to:
Carry out flagship code development projects;
Maintain and distribute code libraries;
Perhaps they appeal to a broader church?

What can we learn feom other disciplines?
List of free and open-sourxe software packages
“How bioinformatics tools are bringing genetic analysis to the massess”
Clustal Omega

<u>Strawman Proposal</u>

<u>Focus on the practical</u>

- Making “academic” software/methodsbfor drug discovery more widely available
- Creating a community
- Sharing expertise and experiences
- Community projects

<u>Operating Model</u>

- Meetings
- Joint projects
- Web presence
- Tools and workflows
- Sustainability and funding

Making academic methods applicable to the real world

- Evaluation on well-curated (published) public domain data sets
- Evaluate on unseen public or industry data sets
- Sensitivity analysis
- Code is “stable” and “useable”

Next steps?

- Decide to continue
    - Yes
- Future meeting(s) / frequency
    - Yes - 2-3 / year
- An organizational structure (or keep “ad hoc”)
    - Discussion and decision from FvD and ARL and others
- Define areas of interest / projects
    - FEP and MMGB(PB)SA pharma interest
    - 3rd party integration (Cosmic3D onto XChem data)
    - ChEMBL and XChem integration