

OOMMPPAA: A tool to aid directed synthesis using activity and structural data

Anthony Bradley^{1,2}, Ian Wall³, Darren Green³, Charlotte Deane², Brian Marsden^{1,4}

¹ SGC, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7DQ

² Oxford Protein Informatics Group, Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 1TG

³ Computational & Structural Chemistry, GlaxoSmithKline, Medicines Research Centre, Gunnels Wood Road, Stevenage, Hertfordshire, SG1 2NY

⁴ Kennedy Institute of Rheumatology, NDORMS, University of Oxford, Roosevelt Drive, Headington, Oxford, OX3 7FY

Aim: Develop tool for large activity and structural datasets to direct chemical synthesis and summarise data

Why?

- The pharmaceutical industry wants to make potent and selective inhibitors rapidly
- Companies frequently possess large activity and structural datasets
- Current automated methods cannot reliably predict activity and do not summarise this data well¹

OOMMPPAA was designed to:

- 1) Aid design of compounds by combining this data effectively
- 2) Visualise the distribution of available activity data

Method

1) Find matched molecular pairs



- MMPs: Molecules differing by one functional group
- Use method of Hussain *et al.*² to make a database to find MMPs efficiently

Molecule A:
Molecule B:



2) Generate conformations

- Use MMPs to generate coordinates for compounds with no complexed structure
- Energy minimise conformations
- Select best shape match

Find matched molecular pair

Overlay shared template

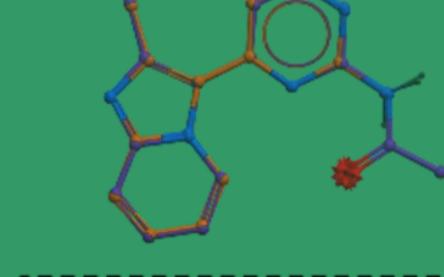
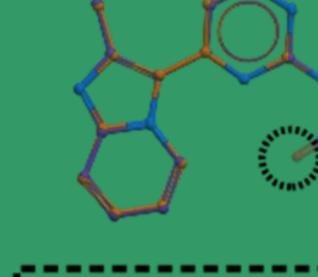
ChEMBL activity: 6.9
PDB ID: 3EJ1

ChEMBL activity: 8.7
No PDB structure

3) Find compound changes

Find pharmacophore differences

Attribute this to change in activity



- Find pharmacophoric differences between compounds
- Attribute these to activity changes
- View pharmacophoric trends in data to aid directed synthesis

Requirements



ChEMBL 3 4

Data

3D complexed structures

- Aligned based on protein
- Used to give activity data context

Activity data for the target

- IC_{50} and K_i data used for CDK2
- Tm shift data trialled too

Software

- Python source code freely available
- Windows installer for desktop application – no dependencies
- Web-based application to trial
- ActiveICM plugin for molecular visualisations



Interactive Visualisation tool

Demo version

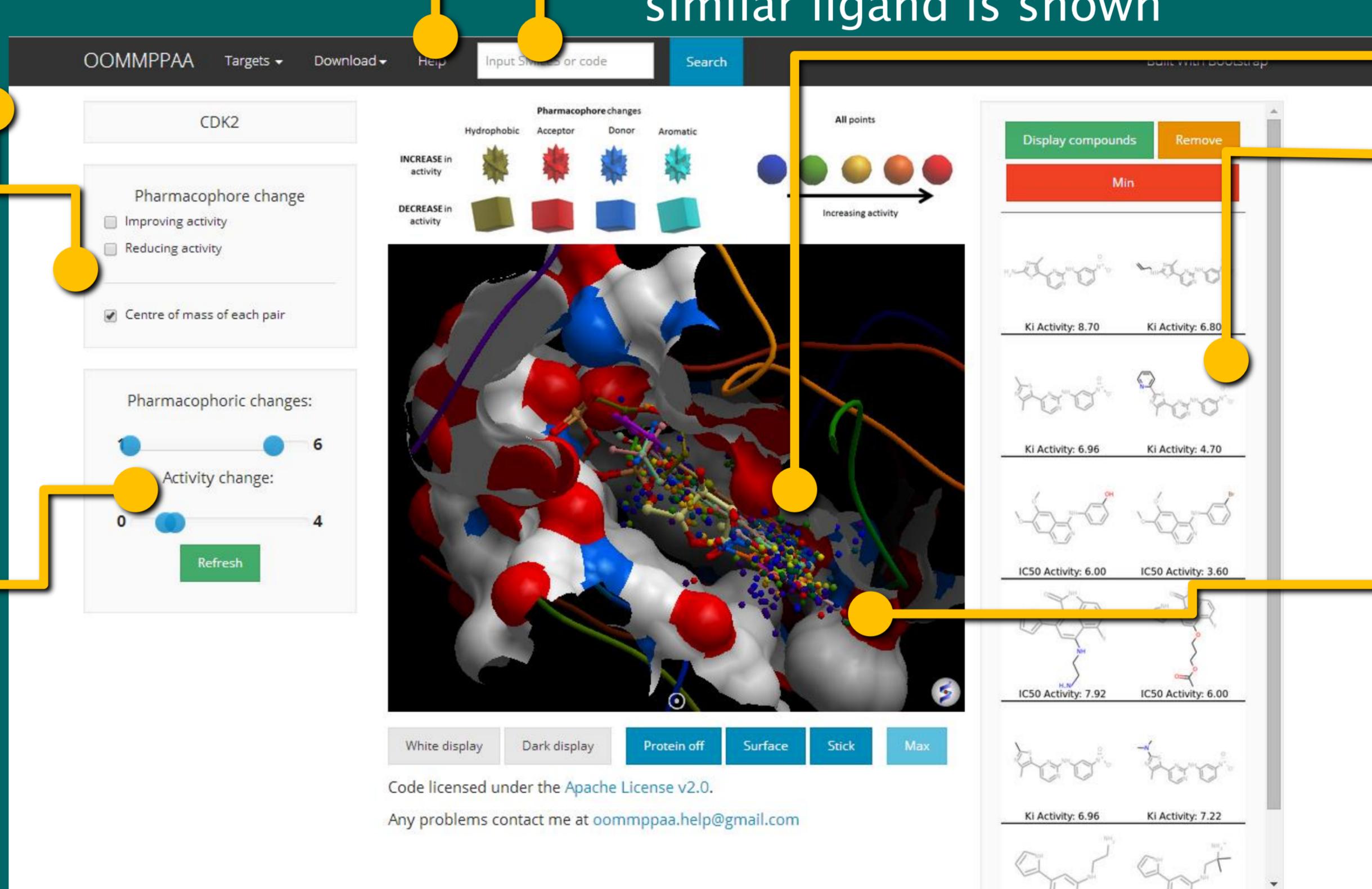
- Tutorial version available for new users
- Shows all the elements and how to use them interactively

Choose data

- Show points that improve or reduce activity, or simply view all changes

Filter data

- Filter the points shown on:
 - Activity change
 - Number of pharmacophoric changes



Show ligand

- Input a smiles string or PDB code to show complexed ligand in the 3D display
- If no exact match, chemically similar ligand is shown

Interactive 3D display

- Show points in 3D display
- Select points and show the underlying activity data

Show activity data

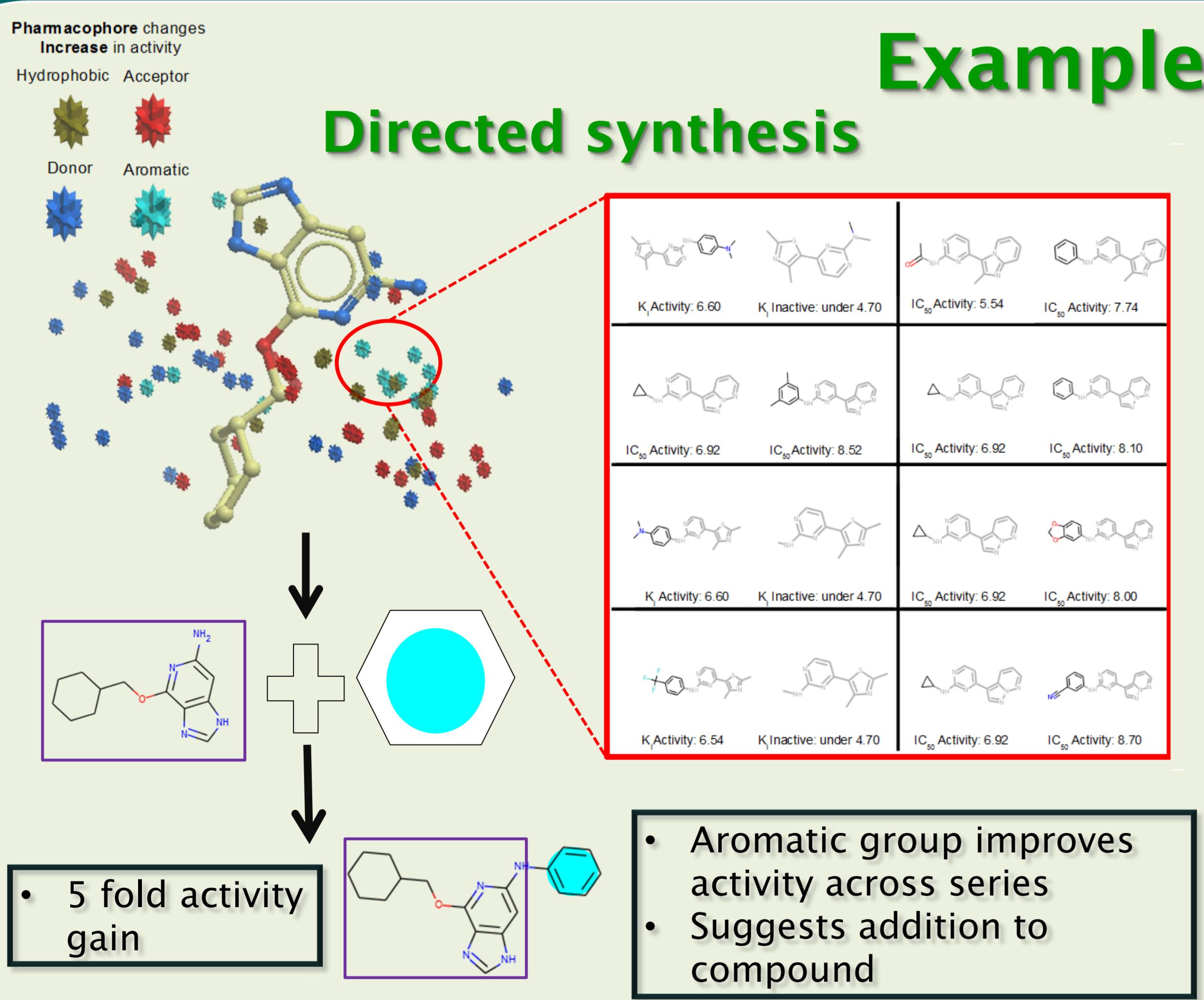
- Points are selected in the graphics window
- The relevant compound changes are shown in 2D with activity data

Show protein

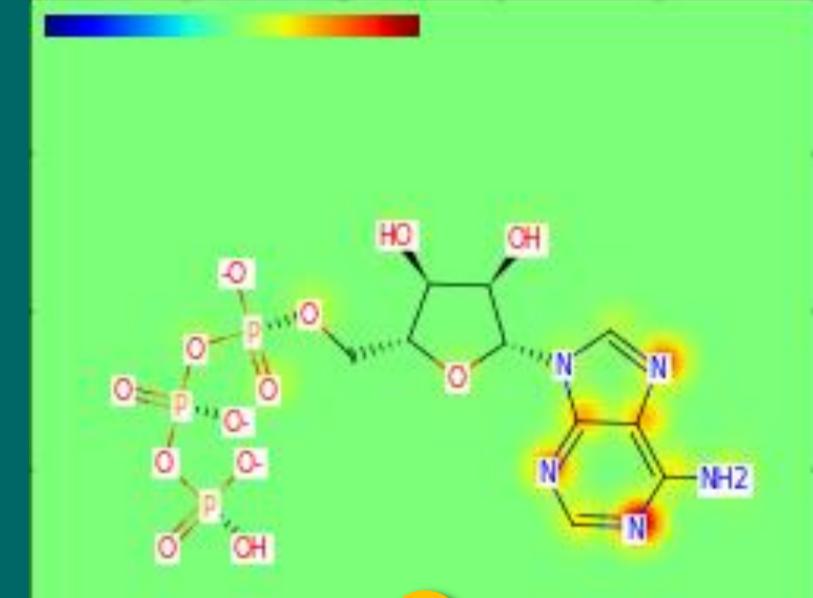
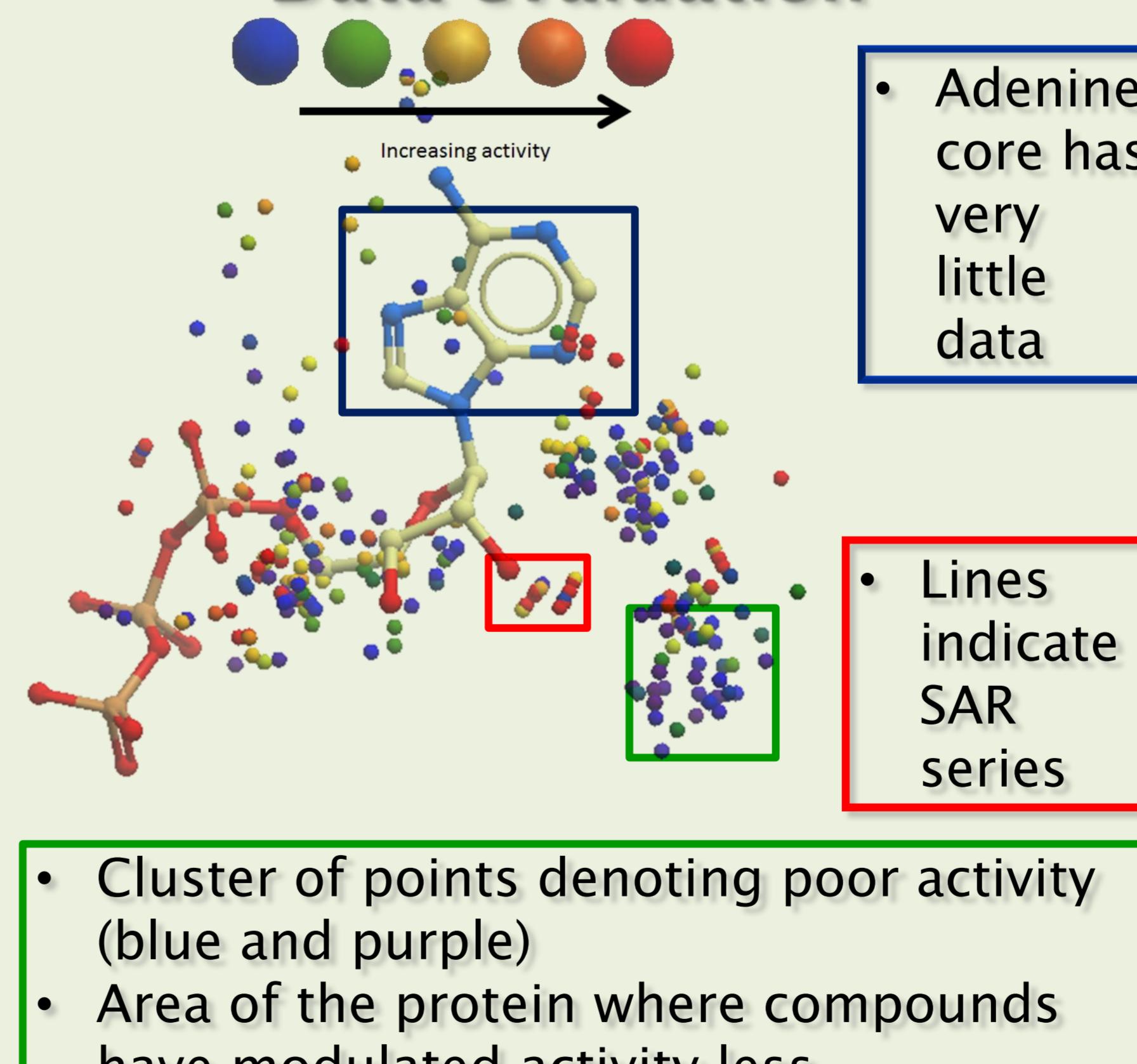
- Show activity data in the context of the protein
- Find potential interactions

Examples – CDK2

Directed synthesis

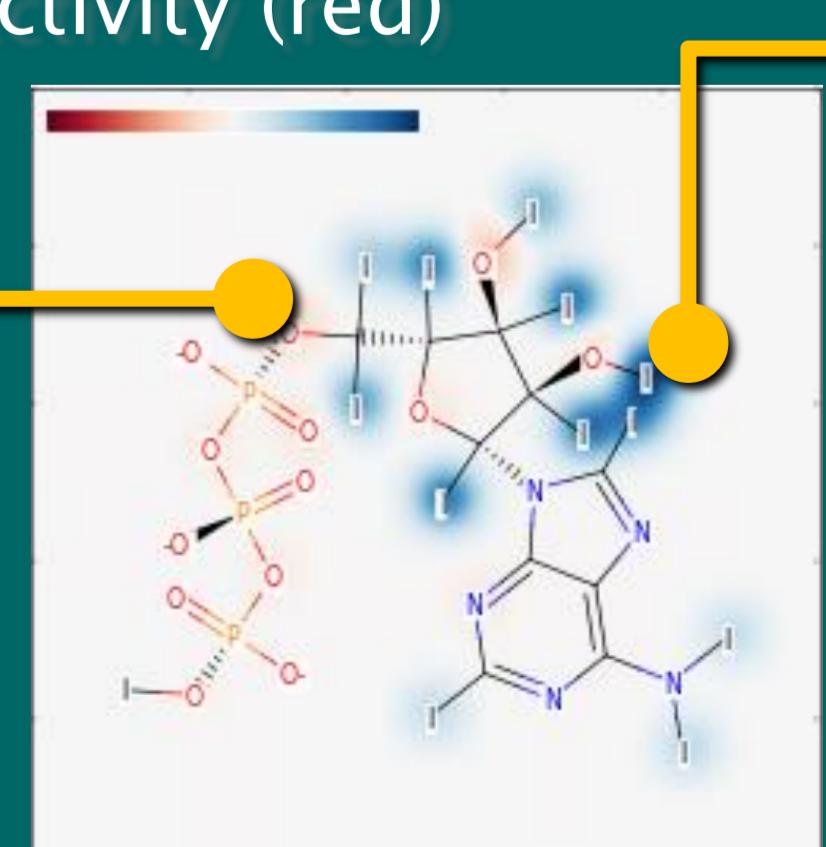


Data evaluation



Feature maps

- Pharmacophoric conservation across the target's co-complexed ligands
- Areas that have potential to be extended from (blue)
- Features that have previously reduced activity (red)



What's next?

- Prospective experimental validation will be carried out:
 - Compounds predicted to increase potency
 - Compounds predicted to improve knowledge of binding
 - Consider protein-ligand interactions and clashes

- How to Recognize and Work around Pitfalls in QSAR Studies: A Critical Review, *Curr. Med. Chem.*, 2009
- Computationally efficient algorithm to identify matched molecular pairs (MMPs) in large data sets. *J. Chem. Inf. Model.*, 2010
- The Protein Data Bank: A Computer-based Archival File For Macromolecular Structures, *J. of. Mol. Biol.*, 1977
- ChEMBL: a large-scale bioactivity database for drug discovery, *Nucleic Acids Res.*, 2012

