

Summary sheet of software and assay simulation information

Text taken from Dodson, Flower & Thomas (2023) A multi-disciplinary team-based classroom exercise for small molecule drug discovery *J Chem Ed* 100(9) p3320-3332

[10.1021/acs.jchemed.3c00066](https://doi.org/10.1021/acs.jchemed.3c00066) and [10.25416/NTR.21908502](https://doi.org/10.25416/NTR.21908502)

e-mail: c.a.dodson@bath.ac.uk, s.e.flower@bath.ac.uk

Software and database requirements

Students should have access to a number of specialist software packages and databases. ChemDraw, Excel, SciFinder-n and Reaxys are pay-for software packages or databases which are included in many University library and software subscriptions. Free alternatives to these software packages are listed below. PyMOL and SciDAVis are available free of charge.

- Excel (or other appropriate software if a different format is chosen for simulated data). Licensed copies of this software are widely available to students on University courses. Free alternatives include Google Sheets (www.google.com/docs/about/) and Calc from LibreOffice (www.libreoffice.org).
- PyMOL (www.pymol.org) or equivalent protein visualization and manipulation software. Licenses for the commercial build educational use are free and can be obtained by instructor registration at <https://pymol.org/edu/>. PyMOL is available for Microsoft Windows, macOS and Linux and students can download the software onto their own machines to work on in their own time. OpenSource PyMOL is available free of charge. (<https://github.com/schrodinger/pymol-open-source>) and installation instructions for Windows, MacOS and Linux were available on a number of independent websites in January 2023 (including www.pymolwiki.org).
- SciDAVis (www.sourceforge.net/projects/scidavis/) or other program suitable for non-linear least-squares fitting of data. Alternative licensed software includes SigmaPlot or GraphPad Prism. SciDAVis is available for Microsoft Windows, macOS and Linux operating systems and is licensed under GNU General Public License version 2.0.
- ChemDraw (<https://perkinelmerinformatics.com/products/research/chemdraw/>) or other software suitable for drawing chemical structures. Alternatives include Marvin Sketch (<https://chemaxon.com/marvin>; free for academic and personal use) and Chem Sketch (<https://www.acdlabs.com/resources/free-chemistry-software-apps/chemsketch-freeware/>). ChemDraw is available for Microsoft Windows and macOS.
- SciFinder-n (<https://scifinder-n.cas.org>). This curated database from the Chemical Abstracts Service (CAS) indexes the chemical literature and enables students to search for appropriate reactions to synthesize their target compound.
- Reaxys (<https://www.reaxys.com>). This index of the chemical literature is licensed by Elsevier and also enables students to search for appropriate reactions to synthesize their target compound.

At least one staff member requires access to GraphPad Prism and GOLD which are commercial pay for software (free alternatives listed below).

- GraphPad Prism (www.graphpad.com). Alternatives to Prism include programming environments such as Matlab, IDL or Python, and the expertise to write and/or run scripts generating data points from a mathematical equation with gaussian noise. An example Jupyter notebook (Python) is provided in Supplementary file S 21. Prism is available for Microsoft Windows and macOS.
- GOLD protein-ligand docking software (<https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/>)^{8,9} and the expertise to run simple docking. Alternative free software includes Autodock Vina (www.vina.scripps.edu) which is released under an

Apache License, Version 2.0 or other ligand / protein docking software. With the example kinase provided here – in which the protein is already prepared for docking – instructors should be able to self-teach using the instructions in Supplementary file S 1. As detailed in *Introduction of student docking*, an alternative to this would be to use the exercise in a format where students carry out their own dockings. In this case, students will require access to docking software.

There are no wet-lab requirements.

Preparing physicochemical and simulated data for students

Physicochemical data

This is calculated for each student-submitted compound by inputting a SMILES string for the compound into both SwissADME (<http://www.swissadme.ch/>)¹⁵ and the Chemicalize tool developed by ChemAxon (<https://www.chemicalize.com>; requires free registration). cLogP, LogS and tPSA values are the consensus cLogP, ESOL and TPSA values respectively from SwissADME. LogD_{7.4} is obtained from the Lipophilicity box in Chemicalize.

IC₅₀ values

Unless explicitly requested otherwise, we assume that we have run an ADPglo assay (Promega)¹⁶ against recombinant Chk1 *in vitro* following the manufacturer's instructions. This commercially available assay measures [ATP], and reports [ADP] present in solution at the end of the 1 hr assay run-time. A standard IC₅₀ graph will therefore plot [ADP] against [compound] (using a log scale on the x-axis) and give a sigmoidal plot which decreases in y-value with increasing x.

Experimental data is simulated using the 'simulate data' function of GraphPad Prism. Briefly, we create a blank XY data table with y-values in triplicate. This table is then 'analyzed' using the 'Simulate XY data' option in the 'Simulate Data' family of analyses with the parameters given in Table V. The values in the resulting data table are cut/pasted into Excel and uploaded to the VLE as experimental assay results for fitting by the students. A Prism file with example data sets is provided in Supplementary file S 20. Alternatively, a python Jupyter notebook with similar functionality is provided in Supplementary file S 21.

Cell viability assays

Our students have requested either MTT¹⁷ or CCK8¹⁸ (Abcam) assays¹⁹. Both are colorimetric assays in which a tetrazolium salt is reduced by NAD(P)H in the cell to form a colored formazan product. The raw data is therefore a measured absorbance value at 460 nm, indicating the amount of formazan product produced. A low absorbance value indicates dead cells (low conversion to formazan) and a high absorbance indicates live cells (metabolizing cells giving high conversion to formazan).

Cell viability plots show percentage of live cells (*ie* a percentage of an untreated control) against [compound], so the data we provide to students are simulated absorbance values for both a compound dilution series and control cells (zero compound). In practice, we generate data for the dilution series in Prism using the parameters in Table V then re-run the calculation and copy an additional set of low-concentration values to act the values for control cells. A Prism file with example data sets is provided in Supplementary file S 20.

When choosing an EC₅₀ value for cell viability assays, the cLogP value²⁰ of each compound is used to determine whether or not the compound is likely to cross the cell membrane (or – occasionally – whether the compound is so hydrophobic it will remain in the membrane). If a compound is likely to enter the cell ($1 < \text{cLogP} < 5$)²¹, the EC₅₀ value is set to be 2-5 times larger than the IC₅₀ value (since the midpoints for cellular and *in vitro* assays are not directly comparable, likely due to different concentrations of ATP or physiological binding partners).

Table V: Graphpad Prism simulation parameters

Parameter	Value for IC_{50}	Value for MTT / CCK8 assay	Value for FP binding assay	Value for CYP inhibition assay	Value for Comet assay	Notes
Generate a series of X-values Start at X=	0.005	0.005	0.005	0.001	0.005	This will depend on the IC_{50} (so that the data points spread across the whole sigmoidal transition). Units should match those used for the IC_{50} (we use M throughout).
Each value equals the previous value	Times 0.2	Times 0.2	Times 0.2	Times 0.2	Times 0.2	<i>ie</i> a 5x dilution series
Generate ... values	8	8	8	8	8	Equivalent to #rows on a 96-well plate
Equation	Generate Y values from equation list -> Dose-response – Inhibition -> [inhibitor] vs. response (three parameters)	Generate Y values from equation list -> Dose-response – Inhibition -> [inhibitor] vs. response (three parameters)	Generate Y values from equation list -> Dose-response – Inhibition -> [inhibitor] vs. response (three parameters)	Generate Y values from equation list -> Dose-response – Inhibition -> [inhibitor] vs. response (three parameters)	Generate Y values from equation list -> Lines -> Horizontal line	This is a built-in equation in GraphPad Prism. The mathematical form of the inhibition equation is $Y = \text{Bottom} + (\text{Top} - \text{Bottom}) / (1 + (X / IC_{50}))$. The mathematical form of the Horizontal line is $Y = \text{Mean} + 0 * X$ (all equations in Prism must contain X).
Parameter values	4	4	4	4	4	Or number of data sets required – one per compound
Simulate ... data sets	3	3	3	3	3	<i>ie</i> assay carried out in triplicate
For each data set generate ... replicate values						
Bottom	50	0.01	0	0	N/A	Arbitrary value held constant across all compounds to simulate high inhibition control
Top	200	0.7	200	100	N/A	Arbitrary value held constant across all compounds to simulate low / no inhibition control
Midpoint (IC_{50} , EC_{50} , K_d)	As chosen for compound – see main text	As chosen for compound – see main text	As chosen for compound – see main text	As chosen for compound – see Table VI	N/A	Units should match units used for X
Mean	N/A	N/A	N/A	N/A	20	Same as the value for zero compound.
Random error	Gaussian, absolute	Gaussian, absolute	Gaussian, absolute	Gaussian, absolute	Gaussian, absolute	Adding random error to each data point to simulate experimental scatter
SD	15	0.03	5	8	8	Arbitrary value chosen so that the simulated data resembles plausible experimental values.

Fluorescence polarization (FP) binding assay

When students decide to measure the binding affinity of the compounds to the kinase directly, we assume that a fluorescence polarization competition assay²² has been carried out. The physical basis of this assay is the different tumbling times of a fluorescently labelled Chk1 ligand free in solution and bound to Chk1: free ligand will tumble rapidly in solution, bound inhibitor will tumble more slowly. When the solution is illuminated with plane polarized light, the emitted fluorescence will also be polarized. Polarization of the emitted light will be lower for unbound ligand molecules compared with bound ones since the rapid tumbling of the unbound inhibitor will remove the direction dependence of the emission more quickly.

We assume that the FP assay has been performed in a competition format, where unlabeled test compound is titrated into a mixture of Chk1 and a tight-binding fluorescent tracer compound (at fixed concentration) under non-stoichiometric conditions. Unlabeled test compound competes with the fluorescent tracer and the measured fluorescence polarization signal gives a traditional sigmoidal IC₅₀ curve of signal against [compound] (using a log scale on the x-axis), decreasing in intensity with increasing x. Under these conditions, IC₅₀ is mathematically equivalent to the dissociation constant of the compound for the enzyme (K_d)²³, and so the binding affinity of the compound for Chk1 has been determined experimentally.

In practice, experimental data are simulated using the 'simulate data' function of GraphPad Prism to provide triplicate fluorescence polarization values for a test compound titration using the parameters in Table V. A Prism file with example data sets is provided in Supplementary file S 20.

The following DMPK assays are 'outsourced' to a fictional contract research organization *ADME Express Ltd* for the practical reason that these assays are well beyond our personal expertise. As such, *ADME Express* provides the results of the assays rather than the raw data for data plotting. Students tend to request these assays in the final rounds of the exercise, when ensuring that they have a balance of assay data to support further development of their compound. Therefore the intended learning outcomes here relate to assay selection rather than data handling.

Caco-2 permeability assays

In vitro Caco-2 permeability assays monitor the extent to which compounds cross a polarised monolayer of Caco-2 cells and act as a model of intestinal permeability (for the uptake of orally administered compounds)²⁴. When students request that a Caco-2 assay is carried out on their compound, we provide values for two commonly reported parameters: the apparent permeability of the compound, P_{app} (apical-basolateral), and the efflux ratio, P_{app} (basolateral-apical) / P_{app} (apical-basolateral).

To select values for these parameters, we input the student compound into the online server SwissADME (<http://www.swissadme.ch/>)¹⁵ using a SMILES string (easily output from ChemDraw). The server reports a parameter 'GI permeability' (based on a prediction by²⁵) and the output from this (eg 'high') is used to choose a value for P_{app} according to the guidance in Table VI. In order to inform the value that is selected for the efflux ratio, we use the SwissADME prediction for whether the compound is a substrate for the P-glycoprotein 1 efflux pump P_{gp} . Compounds which are P_{gp} substrates will be pumped out of intestinal epithelial cells back into the intestinal lumen in an ATP-dependent manner. An efflux ratio greater than one indicates that efflux occurs, an efflux ratio less than one indicates that efflux does not occur and we have selected values of around 0.92 for compounds with good efflux ratios. In practice, efflux ratios less than two are acceptable, so for those compounds which are predicted to be P_{gp} substrates we have chosen efflux ratios close to this upper limit. This is for the practical reason that teams tend to request ADME assays in rounds 4 or 5 (ie late in the development of their compounds), and we would like to enable students to write up a reasonably successful drug discovery story, rather than one which is fully realistic.

Table VI: Guide values for ADME parameters

Property	P_{app} /nm s ⁻¹ Caco-2 and PAMPA assays	CL_{int} / μ l min ⁻¹ mg protein ⁻¹ Microsomal stability	IC_{50} / μ M CYP inhibition
High permeability ²⁶	> 100		
Medium permeability ²⁶	34-100		
Low permeability ²⁶	< 34		
High metabolism ^a		> 47.0	
Low metabolism ^a		< 8.6	
Potent inhibition ^b			< 1
Moderate inhibition ^b			1-10
No or weak inhibition ^b			> 10

^a Values calculated as described in the main text.

^b Inhibition bands from <https://www.cyprotex.com/admepk/in-vitro-metabolism/cytochrome-p450-inhibition>

PAMPA assays

The *in vitro* parallel artificial membrane permeability assay (PAMPA) measures the extent to which compounds cross an artificial hexadecane or lipid membrane ²⁴ and acts as a model of passive compound absorption across the gut wall. Once more, PAMPA reports an apparent permeability of the compound (P_{app}) and so the SwissADME prediction of GI permeability ¹⁵ is used to inform the value that we select for student compounds (guided by the values Table VI).

Liver microsomal stability assays

The majority of drug metabolism in the human body occurs in the liver, and estimates of the metabolic stability of a compound are often obtained by *in vitro* experiments on different subcellular fractions of liver cells. Assays using liver microsomes are popular since microsomes contain many of the relevant oxidative enzymes (including CYP enzymes), and these assays report an intrinsic clearance value, CL_{int} ²⁴.

When students request a liver microsomal stability assay, the online prediction tool PredMS (available at <https://predms.netlify.app/>) ²⁷ is used to obtain a probability of compound stability, P_{stable} . Compounds for which $P_{stable} < 0.5$ are likely to be unstable, compounds for which $P_{stable} > 0.5$ are likely to be stable. Based on the PredMS output we then choose a value of CL_{int} to report to the students using the guide values in Table VI (calculated as detailed below).

The relationship between CL_{int} , extraction ratio (E), liver blood flow (Q_H), microsomal protein per gram of liver (MPPGL), liver mass (m_{liver}) and fraction of drug unbound in blood plasma (f_u) is given in equation (3) ²⁸. Using the approach taken by the Cyprotex (<https://www.cyprotex.com/admepk/in-vitro-metabolism/microsomal-stability>) and given in ref ²⁹ and assuming liver blood flow of 1450 ml min⁻¹ ³⁰, liver mass of 1800 g ³⁰ and completely free drug (*ie* $f_u = 1$), boundary values of $E = 0.3$ and $E = 0.7$ were used to calculate clearance values for low and high clearance respectively.

$$CL_{int} = \frac{E \cdot Q_H \cdot MPPGL}{f_u \cdot (1 - E) \cdot m_{liver}} \quad (3)$$

CYP inhibition

Cytochrome P450 (CYP) superfamily enzymes are monooxygenases, a handful of which are responsible for the majority of oxidative drug metabolism and elimination ³¹. When students request that assays are carried out to determine CYP inhibition, we assume that an IC_{50} value has

been determined for the inhibition of the CYP enzymes in liver microsomes by the compound of interest. In this assay, the amount of a defined CYP substrate remaining after the assay duration (expressed as a percentage) is determined using LC-MS/MS.

We use SwissADME to predict whether the student compound is likely to be a substrate for five different CYP enzymes (CYP1A2, CYP2C19, CYP2C9, CYP2D6 and CYP3A4) and use this to choose an overall IC₅₀ value for the *in vitro* assay based on the guide values in Table VI. We then simulate IC₅₀ data in GraphPad Prism using the parameters in Table V.

Plasma protein binding

Small molecule drugs may bind to proteins in human blood (eg human serum albumin, α_1 -acid glycoprotein). When complexed with blood proteins, compounds will not be taken up into tissues and so will be pharmacologically inactive ²⁴.

We have assumed that equilibrium dialysis against blood plasma proteins has been carried out. In this assay, the concentration of free drug either side of an equilibrated dialysis membrane is measured and the fraction of compound unbound is reported ²⁴. This assay is requested infrequently, but when it is we submit student compounds to the PreADMET server (accessed via <https://preadmet.qsarhub.com/>) ³² and report the fraction unbound (f_u) (calculated as (100-Plasma_Protein_Binding parameter)/100).

Comet genotoxicity assay

The Comet assay is a conceptually simple genotoxicity assay in which single cells are immobilized in agarose, lysed and undergo electrophoresis ³³. Expansion of cellular DNA away from a compact nucleoid attached to the nuclear matrix towards the anode occurs most rapidly when the DNA has multiple sites of breakage (and thus relaxation of supercoiling). When visualized using fluorescence microscopy, this expansion resembles the tail of a comet (hence the assay name).

The degree of tail expansion is quantified as a relative intensity value (% of tail DNA) and plotted against compound concentration to visualize toxicity: a flat line (no change with compound concentration) indicates no toxicity while a linear increasing relationship between % tail DNA and compound concentration indicates toxicity. At zero compound, the value of % tail DNA is around 20%.

Students tend to request a Comet assay towards the close of the project, so if there are no obvious flags to the contrary (which should have been corrected as a result of earlier discussions), results for a negative Comet assay are simulated in Prism by creating a set of compound dilution data for a flat line using the parameters in Table V.

Compound docking

Molecular docking is a computational technique which approximates the process of a small molecule ligand binding to a protein active site. It uses iterative rounds of ligand pose prediction and scoring to optimize protein-ligand interactions and provide physically-informed models of the 3D structure of the final ligand-bound complex. The magnitude of the docking score is used as a qualitative estimate of the strength of the protein-ligand interaction and – in the absence of information to the contrary – the model with the highest docking score is accepted as a model of the likely ligand binding pose ^{8, 34}.

Instructions to enable generalist staff to dock student-drawn compounds into Chk1 are included in Supplementary file S 1. Briefly, compounds are designed by the students and uploaded as ChemDraw files. Students are reminded to specify stereochemistry around chiral centers and staff check for unspecified stereocenters prior to carrying out compound docking. Student-designed molecules are copy/pasted into Chem3D by staff members and a short round of MM2 energy minimization is carried out. The energy-minimized compound is saved as a .sdf file. Ten rounds of docking into the ATP binding pocket of Chk1 are carried out for each compound using GOLD software and the highest scoring pose is used as a model of the final structure. A pdb file of Chk1

prepared for docking, a file defining the ATP binding cavity and a configuration file for docking in GOLD are included in the supplementary information (Supplementary file S 17; Supplementary file S 18, Supplementary file S 19).

List of summary files

Supplementary file S 1: Brief instructions for docking ligands into Chk1.docx, Brief instructions for docking ligands into Chk1.pdf

Supplementary file S 17 (Chk1 kinase prepared for docking using GOLD): Chk1.pdb

Supplementary file S 18 (Configuration file for compound docking in GOLD): gold.conf

Supplementary file S 19 (Cavity file for compound docking in GOLD): cavity.atoms

Supplementary file S 20: Template Prism file for data simulation.pzf

Supplementary file S 21: (Jupyter notebook (python) and tools file as an alternative to GraphPad Prism): Python code for simulating experimental data.zip