

An open-source molecular builder and free energy preparation workflow

Mateusz K. Bieniek ^{1,3}, Ben Cree^{1,3}, Rachael Pirie ¹, Joshua T. Horton ¹, Natalie J. Tatum ² & Daniel J. Cole ¹✉

Automated free energy calculations for the prediction of binding free energies of congeneric series of ligands to a protein target are growing in popularity, but building reliable initial binding poses for the ligands is challenging. Here, we introduce the open-source FEgrow workflow for building user-defined congeneric series of ligands in protein binding pockets for input to free energy calculations. For a given ligand core and receptor structure, FEgrow enumerates and optimises the bioactive conformations of the grown functional group(s), making use of hybrid machine learning/molecular mechanics potential energy functions where possible. Low energy structures are optionally scored using the gnina convolutional neural network scoring function, and output for more rigorous protein-ligand binding free energy predictions. We illustrate use of the workflow by building and scoring binding poses for ten congeneric series of ligands bound to targets from a standard, high quality dataset of protein-ligand complexes. Furthermore, we build a set of 13 inhibitors of the SARS-CoV-2 main protease from the literature, and use free energy calculations to retrospectively compute their relative binding free energies. FEgrow is freely available at <https://github.com/cole-group/FEgrow>, along with a tutorial.

¹School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne NE1 7RU, UK. ²Newcastle University Centre for Cancer, Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. ³These authors contributed equally: Mateusz K. Bieniek and Ben Cree. ✉email: daniel.cole@ncl.ac.uk

Computational structure-based molecular design, in particular aiding the discovery of novel chemicals with desired biological activity, plays a crucial role in the modern drug discovery pipeline. High-throughput virtual screening is widely used in hit discovery¹, but relies on pre-defined libraries of compounds. De novo design software packages aim to construct a model of a ligand in a target binding pocket using growth algorithms, either starting from a scaffold of a known hit compound or entirely from scratch. Such approaches can be beneficial as they do not rely on a (physical or virtual) library, and molecules can be tailored specifically to the problem at hand. Advances in de novo design software have been extensively reviewed², and examples include both rule-based generation methods such as OpenGrowth³, AutoGrow⁴, and LigBuilder⁵, and recently deep generative methods for molecule design⁶.

With advances like these described above, much progress has been made in the important problem of optimising a molecular design within the context of a pre-defined scoring function and binding pocket. However, whether the designed molecule indeed has high biological activity is crucially reliant on the accuracy of the methods that are used to generate and score poses of the designed molecules, as well as other assumptions, such as a rigid receptor, that might be employed. Furthermore, the generated molecules can be quite esoteric, which may be advantageous with regards to arriving at novel intellectual property, but may not be ideal from a synthetic tractability viewpoint⁷. More commonly, a drug discovery effort may have identified a hit compound with a well-defined binding mode and wish to explore structure-activity relationships amongst a small library of synthetically accessible analogues. In this case, it would be beneficial to make use of prior knowledge about the binding mode when generating poses of designed compounds. One example of this approach is the E-novo workflow⁸, which was made available in Pipeline Pilot or Discovery Studio. The available conformations of added chemical functional groups (R-groups) were enumerated with a rigid core, and scored using a CHARMM-based docking method. The physics-based molecular mechanics-generalised Born with surface area (MM-GBSA) was then used to provide a more accurate score. Further, more recent, examples include FragExplorer⁹, which aims to grow or replace fragments to optimise molecular interaction fields generated by the GRID software¹⁰, DeepFrag¹¹, which predicts appropriate fragment additions using a deep convolutional neural network trained on thousands of known protein–ligand complexes, and DEVELOP¹², which uses deep generative models to output 3D molecules conditional on provided pharmacophoric features of the binding site. However, the employed approximate physics- or knowledge-based approaches to scoring the designs will limit to some extent their ability to predict and optimise binding affinity.

On the other hand, free energy methods are much more computationally expensive approaches to molecular design that employ rigorous thermodynamics and carefully parameterised force fields to compute (relative or absolute) protein–ligand binding free energies. As such, they overcome many of the accuracy limitations of de novo design workflows, and are commonly employed in prospective design efforts to explore and prioritise relatively small perturbations in the hit-to-lead stage¹³. Many excellent tutorials and best practice documentation are available^{14–18}, but most start from the assumption that the user has already built initial poses of the ligands in the binding pocket. For simple R-group additions, input coordinates may be built from maximum common substructure alignment, for example, but it may be difficult to resolve steric clashes or decide between two alternative 3D poses in more complicated cases¹⁶. Some widely-used graphical user interfaces, such as Maestro¹⁹ and Chimera²⁰, are also available for building R-groups, but these can

be proprietary and/or difficult to build into automated workflows and modify according to user needs.

Notable successful computer-aided design efforts have used free energy calculations in conjunction with de novo design tools to build (and maybe score) new molecules. Jorgensen and co-workers¹³ have pioneered this approach for many years, linking de novo design through the biochemical and organic model builder (BOMB) software with free energy perturbation (FEP) through the MCPRO software. BOMB builds ligands into a binding pocket by linking user-defined R-groups to an existing core. Functionality is available for conformer searching, structural optimisation, and scoring, using a custom scoring function trained via linear regression on >300 experimental activity data points²¹. Once hits have been built and scored, hit-to-lead optimisation may be further refined through free energy calculations. Such an approach has yielded extremely potent series of leads against HIV reverse transcriptase²², macrophage migration inhibitory factor²³, and the SARS-CoV-2 main protease²⁴. In other drug discovery programmes, as part of the recent COVID Moonshot open science effort to crowd source design of SARS-CoV-2 main protease inhibitors²⁵, the Omega toolkit by OpenEye²⁶ is used for constrained conformer generation, and optimal binding poses are then taken through to free energy calculations using the perses software²⁷. The evident importance of input structure to the reliability of free energy calculations¹⁶ means that open-source tools to automate this step are crucial.

Inspired by the BOMB/MCPRO approach to molecular design¹³, we introduce here the FEgrow open-source workflow for growing functional groups, chosen by the user, from a defined position on a core compound. To account for the multi-objective nature of molecular design, we output simple rule-of-five indicators of oral bioavailability, as well as flags for undesirable substructures and synthetic accessibility estimates. For the designed ligands, we enumerate 3D conformers of the added R-group, with options for additional flexibility if desired, within the context of the protein (discarding conformers with steric clashes). A common issue with generating docked poses is inaccuracy in the molecular mechanics force fields used to refine them, particularly for uncommon chemistries. To overcome this, we employ a hybrid machine learning/molecular mechanics (ML/MM) approach to optimisation, whereby the ligand is (optionally) described by the ANI neural network potential^{28,29}, and non-bonded interactions with the static protein are described by traditional force fields. The binding affinities of low energy poses are predicted using a deep learning-based scoring function. Finally, FEgrow outputs binding poses in a form suitable for input to free energy calculations, and we illustrate this process with a case study, using the SOMD software³⁰ to retrospectively compute relative binding free energies of several inhibitors of the SARS-CoV-2 main protease²⁴.

In this way, we hope to integrate medicinal chemistry expertise in the FEgrow design workflow, with state-of-the-art methods for pose prediction, scoring, and free energy calculation. By building ligands from the constrained core of a known hit, we maximise the use of input from structural biology, and reduce reliance on docking algorithms. We aim for an open-source, customisable, fast, and easy-to-use (accessed through Jupyter notebooks) workflow that can adapt to community needs and advances in molecular design.

Results

Workflow design. The FEgrow package is written in Python, and supports Jupyter visualisation at each stage using py3Dmol³¹. Underneath, the main unit in the package is RMol which extends the RDKit class `rdkit.Chem.rdchem.Mol`³² with additional

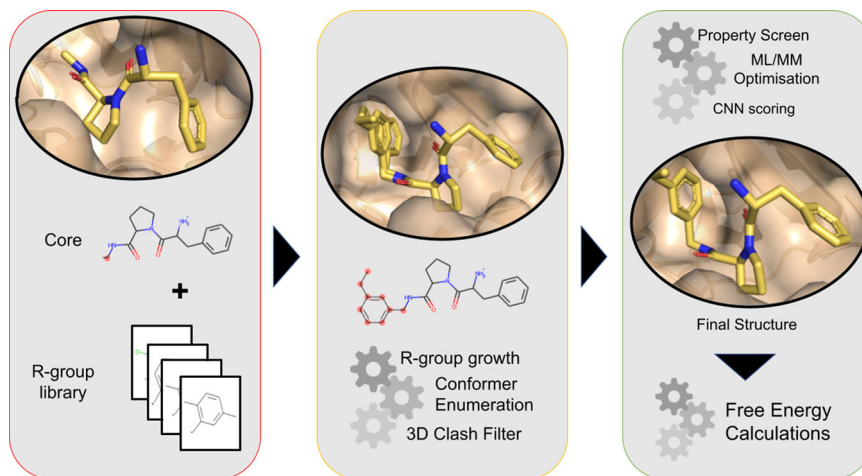


Fig. 1 Overview of the FEgrow workflow. (left) The user specifies the receptor, ligand core, and a list of functional groups, along with their attachment points. (centre) RDKit³² is used to attach the selected R-group(s) and enumerate the available conformers with a rigid core. (right) Possible bioactive conformers undergo structural optimisation using a hybrid ML/MM potential energy function. The binding affinity is predicted using a convolutional neural network scoring function⁴⁷ and molecular properties are assessed. Final structures are output for further free energy based binding affinity assessment.

functionalities, such as visualisation, molecule merging, conformer generation, as well as storage of energies and other metadata. A convenience class RList is provided with the same functions for operating on a set of molecules, which allows also for future parallelisation. A modular workflow allows for addition/removal of functionality, such as new scoring functions or optimisation algorithms. FEgrow is freely available at <https://github.com/cole-group/FEgrow>, along with a tutorial. Figure 1 shows the overall design of the FEgrow workflow, and the component methods are described in the following sections.

Input and constrained conformer generation. The first task is to define the receptor and the ligand core, along with an attachment point for growth (currently only growth from hydrogen atoms is supported). Users may download receptor and ligand structures directly from the protein databank (PDB), or upload pre-prepared structures. In this study, we used the Open Babel software³³ for parsing input structure files and ligand protonation (at pH 7).

Merging the ligand core with a new R-group requires that both the linking atom on the template core and on the attachable R-group are specified. The merging is carried out with the RDKit editable molecule³². RDKit is further used to generate 3D conformers using the ETKDG method³⁴. The generated conformers are aligned, and energy minimised using the Universal Force Field³⁵. Harmonic distance restraints to their initial positions are applied to atoms in the common core (identified by a maximum common structure search) using a stiff force constant (10^4 kcal/mol/Å²). In this way, we can enforce the conformations of the generated molecules to only vary from the core in the region of the added R-group. This region may additionally be extended by adding further atoms into the flexible substructure of the template. For convenience, we provide a minimal set of around 500 R-groups that are commonly used in medicinal chemistry optimisation³⁶. R-groups can be interactively selected from the library using the mols2grid package³⁷, or the user may prepare their own molecules for attachment (see Tutorial).

Geometry optimisation. The constrained conformer generation described above aims to enumerate all accessible, physically-reasonable conformers of the added R-group (and any other flexible regions) in vacuum. However, most of these

conformations will be incompatible with the protein binding site. Hence, a 3D filter and geometry optimiser aims to find the bioactive conformers of the designed ligands.

The protein is treated with PDBFixer³⁸ to add any missing atoms, residues, and hydrogen atoms. Water molecules (and other non-protein residues) are stripped by default, but can be optionally retained as part of the receptor, for example, if they are thought to form an important part of the hydrogen bonding network within the binding pocket (an example is shown later in Case Study I). A simple distance filter removes any ligand conformers that form a steric clash with the protein (any atom–atom distance less than 1 Å). Next, the remaining conformers are refined in the context of a rigid receptor via energy minimisation using OpenMM³⁸. All atoms of the protein, and any retained water molecules, are kept fixed during the optimisation in the positions provided by the user.

The energy minimisation uses the AMBER FF14SB³⁹ force field for the receptor and either GAFF2⁴⁰ (General AMBER force field) or the Open Force Field 1.0.0 (‘Parsley’)⁴¹ general force fields for the ligand, with the choice left to the user. Optionally the intramolecular interactions of the ligand can be modelled using the ANI-2x ML potential²⁸ in a hybrid ML/MM simulation. In this so-called mechanical embedding scheme, the total potential energy of the ML/MM system is composed of three terms⁴²:

$$E^{\text{tot}} = E^{\text{MM}}(R) + E^{\text{MM}}(RL) + E^{\text{ML}}(L), \quad (1)$$

where R, RL, and L correspond to receptor–receptor, receptor–ligand, and ligand intramolecular interactions, respectively. The second term acts as the coupling term between the ML and MM subsystems and consists of the standard Coulomb and Lennard-Jones 12-6 non-bonded interaction energies. Thus, a general force field (here, GAFF2 or Parsley) is still required for the ligand to model the non-bonded interactions with the receptor. The use of ANI helps to avoid known deficiencies in the potential energy surfaces predicted by force fields, while ensuring that the optimisations are significantly faster than could be achieved with full quantum mechanics. For example, it has been shown that the description of biaryl torsions, which are commonly found in drug-like molecules, is one area where ANI-2x performs better than contemporary general force fields⁴³. The hybrid ML/MM approach has also been shown to predict binding poses that overlap well with crystallographic electron density maps of bound ligands, even for those that contain

charged moieties that were not included in the training of the ANI potential⁴⁴. As such, in FEgrow, users may turn on the hybrid approach for binding pose refinement provided the molecule contains only elements covered by the model (H, C, N, O, F, S, and Cl), else the selected classical force field is used for the entire ligand.

Following the procedure recommended in BOMB, Lennard-Jones radii are scaled by a factor of 0.8 during optimisation. This is intended to mitigate to some extent the rigid protein approximation, by allowing extra space in the binding pocket to accommodate ligand growth. Furthermore, to account in an implicit manner for the neglected dielectric response of the protein and solvent molecules, the atomic charges are reduced by a factor of $\frac{1}{\sqrt{\epsilon}}$, where ϵ is the relative permittivity, in this case taken to be 4. Analysis of the effect of these scaling factors on structural and affinity predictions in Case Study I is provided in Supplementary Note 1, Table S1, and Fig. S1. The lowest energy optimised conformer, and all conformers within 5 kcal/mol, are output as PDB/SDF files for further analysis and scoring.

Binding pose scoring. Once the geometry optimisation is completed, the remaining (low energy) conformers are scored to predict their binding affinity. There are many choices available for scoring binding poses and their corresponding binding affinities, and these are usually classified as either force-field, empirical, or knowledge-based. In the latter case, input features (such as atom-atom pairwise contacts) are used to train models to reproduce data for known protein–ligand complexes. Recently, machine learning models have emerged, in which an arbitrary, nonlinear relationship between input and target prediction is learned. One such approach is the gnina convolutional neural network (CNN) model⁴⁵, which takes as its input features a 3D grid-based representation of the protein–ligand complex and the atom types. The model has been jointly trained for binding pose and affinity prediction on a cross-docked set containing examples of ligand poses generated by docking into multiple receptors⁴⁶. The resulting models are competitive with other grid-based CNN models, and outperform the traditional empirical Vina scoring function⁴⁶. They are available as part of the gnina docking software package⁴⁷, which is a fork of Smina⁴⁸ and AutoDock Vina⁴⁹. Here, we use gnina only for re-scoring the output ligand 3D structures, using the ‘score_only’ flag and the default ensemble of CNN scoring models. Gnina CNNAffinity scores (predicted pK) are output, and compared with experimental binding affinity (where available).

Molecular property filters. Having assembled the 2D and 3D structures of the core and user-defined R-groups, we include some simple tools for assessing the drug-likeness and synthetic tractability of the designed compounds. Several sets of rules exist to investigate the likelihood of a molecule displaying drug-like behaviour. While there are many examples of approved drugs which violate these considerations, they still provide a useful indication of whether a molecule is worth testing (that is, if it disobeys all of the conditions discussed below, it is most likely a poor candidate). FEgrow reports Lipinski’s rule of five (Ro5) counts,⁵⁰ the synthetic accessibility score (SAScore)⁵¹ and flags describing whether the proposed molecule is Ro5 compliant and if it contains undesirable features based on the PAINS,⁵² NIH^{53,54} and unwanted substructure⁵⁵ filters. Our implementation is adapted from the TeachOpenCADD⁵⁶ Talktorials 2 and 3, using functionality from the Descriptors and FilterCatalog modules of RDKit.³² Further details are provided in Supplementary Note 2.

Case study I: Protein–ligand benchmarks. The protein–ligand benchmark of Hahn et al.^{57,58} is an open, curated set of high quality structural (e.g., high similarity between crystallised and simulated ligands and no missing atoms) and bioactivity (e.g., taken from a single data source with adequate dynamic range) data, which has been collected with the goal of assessing the accuracy of free energy methods. For each target, modelled structures of the protein in complex with a congeneric series of ligands are provided as starting points for free energy calculations, but the methods used to position the R-groups are, to our knowledge, not necessarily consistent or documented.

Here, we apply the FEgrow workflow to ten targets from the protein–ligand benchmark set. Starting from the crystal structure of each target, we truncate the bound ligand to a common core, which is shared across the congeneric series to be modelled. A summary of the targets, the crystal structures used, the number of R-groups grown, and their common core and net charge is provided in Tables S2 and S3. We use the methods outlined in the previous section to re-grow the congeneric series of ligands in the binding pockets, including enumeration and optimisation of possible R-group conformers, and scoring of final poses. Figure 2 shows overlays of the modelled and crystal structures (where there is an exact match between the crystallised ligand and one of the modelled R-groups), as well as the measured root-mean-square deviation (RMSD) between the predicted and experimental coordinates of the heavy atoms of the functional groups.

For the targets, TYK2 (Fig. 2a) and Thrombin (Fig. 2b), we obtain a good overlap between the grown R-groups and the

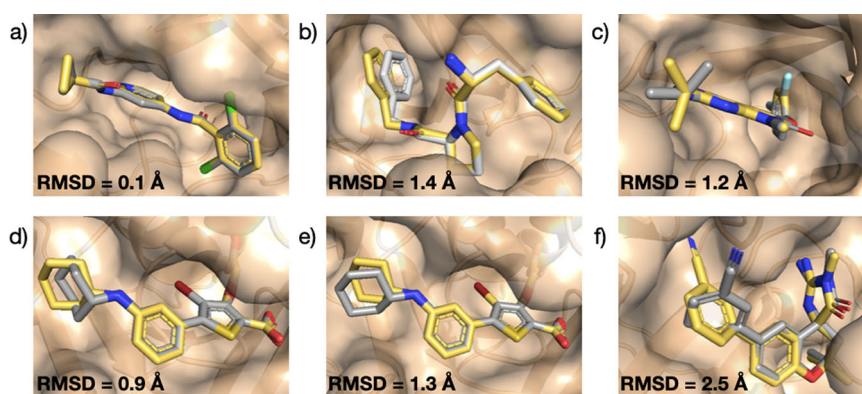


Fig. 2 Overlay of experimental and predicted protein–ligand benchmark dataset structures. Crystal structures are shown in yellow and grown compounds in grey. **a** TYK2 (PDB: 4GIH⁵⁹), **b** Thrombin (PDB: 2ZFF⁶⁰), **c** P38 (PDB: 3FLY⁶¹), **d** PTP1B with force field optimisation (PDB: 2QBS⁶²), **e** PTP1B using ML/MM optimisation, and **f** BACE(Hunt) (PDB: 4JPC⁶³). Root-mean-square distances (RMSD) between predicted and experimental coordinates of atoms in the built R-groups were calculated using RDKit³².

crystal structures. In the former case, the dihedral angle formed between the grown cyclopropyl C2 and C3 carbons and the amide carbonyl oxygen core (30° and -37°) are in agreement with those reported experimentally⁵⁹. For Thrombin, although the added R-group here is a rigid phenyl moiety, we make use of the option to add atoms from the core to the flexible region and allow the linking $-\text{CH}_2-$ group to freely rotate during structural optimisation. This added flexibility leads to a rotation of the phenyl group of less than 10° , compared to the corresponding crystal structure⁶⁰.

The P38 benchmark set includes a series of alkyl amino substitutions originally investigated as part of a structure-activity relationship study into kinase inhibitors⁶¹. Here, the added amino group is correctly positioned to form a hydrogen bond with the protein backbone, though the *i*-Pr group is rotated by around 60° compared to the crystal structure (Fig. 2c). In PTP1B, the grown cyclohexyl substituent is able to rotate quite freely, with many conformations predicted to lie within 5 kcal/mol of the minimum. The minimum energy structure shows good overlap with the crystal structure⁶² with low RMSD, but connects to the core at the axial position of the cyclohexyl group (Fig. 2d). In this case, the core structure contains a Br atom, so we are unable to optimise with the ANI-2x potential (the workflow defaults to the Parsley force field for the ligand). Interestingly, if we remove the Br atom, and re-run the workflow using hybrid ML/MM optimisation, we recover the equatorial connection as the lowest energy conformer, in agreement with the crystal structure (Fig. 2e). This demonstrates the potential advantages of employing hybrid ML/MM structure prediction methods in binding mode determination.

Finally, the BACE(Hunt) target, includes a series of substituted phenyl additions to a spirocyclic core. Here, the grown cyanophenyl group is rotated by approximately 90° , relative to the crystal structure⁶³, which shows the *meta*-CN group accommodated in the binding pocket (Fig. 2f). An exact match with the crystal structure is also output, but it is predicted to be around 3 kcal/mol higher in energy. Closer examination of the experimental structure reveals a crystal water molecule, close to the binding pocket, that is capable of forming a hydrogen bond with the $-\text{CN}$ group, and a further network of water molecules that would be displaced by the conformation shown in Fig. 2f. Figure S2 investigates the effects of including the hydrogen-bonding water molecule in the rigid receptor structure, and changing the force fields used, but no input settings recover the crystal structure.

As discussed, we include with the FEGrow workflow the option to score the output poses of the designed ligands with a scoring function. In particular, we use the gnina convolutional neural network score, which has been trained on both binding pose and affinity prediction⁴⁶. While accurate recovery of experimental binding affinity is not necessarily expected for current scoring functions, it is useful to evaluate to what extent they can be used to provide guidance in early stage design, ahead of more rigorous physics-based scoring methods. The root-mean-square error between gnina CNN affinities (converted to free energies) and experiment is quite acceptable (Table S4), ranging from 0.9 kcal/mol (BACE(P2)) to 1.7 kcal/mol (Jnk1), which indicates that the CNN scoring function is able to predict the affinity range of most of these series. In fact, the errors may be lower than typically expected⁴⁶, because we are using here additional information from the experiment (the binding pose of the core) and not relying on the scoring function to determine the bioactive conformation.

The R^2 correlation coefficients between the predicted and experimental affinities are more variable (Table S4), however, ranging from close to zero (the BACE targets) to 0.68 (Thrombin). The full set of CNN-predicted binding affinity data is plotted in Fig. S3, and reveals that most of the predictions lie in quite a narrow range, compared to the experimental data. We

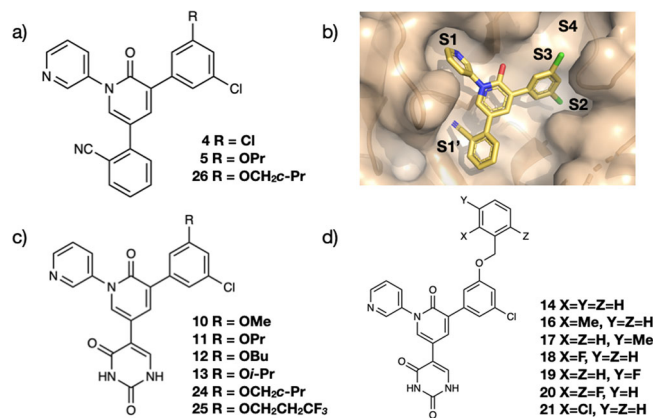


Fig. 3 Structures of the series of cyanophenyl- and uracil-based compounds SARS-CoV-2 main protease (M^{Pro}) inhibitors investigated here. **a** Cyanophenyl-based M^{Pro} inhibitors. **b** X-ray crystal structure of **4** in complex with the protease, with discussed binding pockets labelled. **c, d** Uracil-based M^{Pro} inhibitors.

note that this is quite a challenging test for the scoring function, since the modifications made to the core are relatively small and cover a smaller dynamic range in affinity than most test sets. Nevertheless, it seems that current scoring functions have some utility in guiding design, but that more accurate physics-based scoring is required to accurately discriminate between structural changes in the hit-to-lead stage.

Case study II: SARS-CoV-2 main protease. The main protease (M^{Pro}) of SARS-CoV-2, the virus responsible for the COVID-19 pandemic, is an attractive target for the development of antiviral agents⁶⁴. The Jorgensen lab has focused on the development of drug-like, non-covalent inhibitors of the protease through lead optimisation of virtual screening hits²⁴. In particular, starting from the anti-epileptic drug, perampanel, researchers combined model building with the BOMB software, with free energy calculations, to rapidly yield potent antiviral compounds. Figure 3 shows the structures of the two main series of cyanophenyl- and uracil-based compounds investigated. A high-resolution X-ray crystal structure of **4** with M^{Pro} confirmed binding to the S1, S1' and S2 pockets, with space to grow into the S3–S4 region²⁴.

In what follows, we employ FEGrow to retrospectively build and score the listed analogues (Fig. 3) to demonstrate the potential utility of the workflow in guiding future design efforts. Starting from the crystal structure of **4** (PDBID: 7L10), we begin by replacing one of the *meta* chlorine atoms by propoxy to form **5**. The modelled structure agrees well with the corresponding high-resolution crystal structure (Fig. 4a). In particular, the propoxy OCCO dihedral angle in the lowest energy structure (53°) matches the experimental *gauche* conformation (47°), which allows hydrophobic contact with Met165 and Leu167. Similarly, good agreement is obtained for the cyclopropyl analogue **26** with the corresponding experimental crystal structure (Fig. 4b).

Turning attention to the uracil series, the core molecule was again built from the crystal structure of **4**, by removing the cyanophenyl group. The added uracil group has three low energy conformations, and in this case, we retained the second lowest energy structure, which forms key hydrogen bonding interactions with the backbone of Thr26 and the catalytic Cys145. In agreement with the original modelling, performed using the BOMB software²⁴, we find that again a range of substituents are permitted in the S3/S4 pocket, including substituted benzyloxy side chains (Fig. 3). Figure 4c shows that the modelled uracil group in the S1' pocket is in good agreement with the

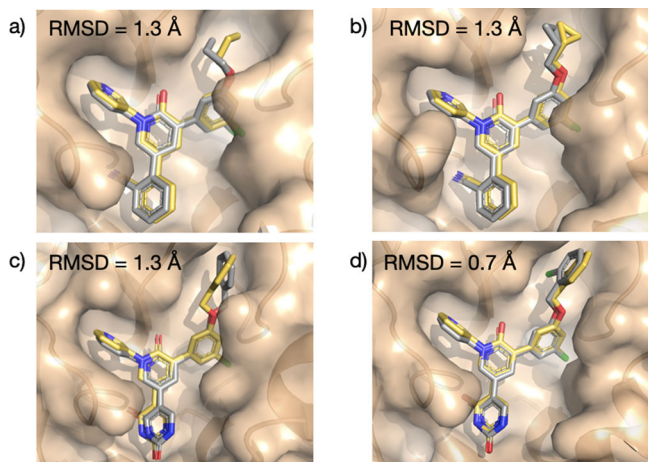


Fig. 4 Comparison between experimental and predicted structures of SARS-CoV-2 main protease (MP^{pro}) inhibitors. Overlay of **a** **5** and PDBID: 7L11, **b** **26** and 7L14, **c** **14** and 7L12, **d** **21** and 7L13. Crystal structures are coloured in yellow, and modelled binding poses in grey. Root-mean-square distances (RMSD) between predicted and experimental coordinates of atoms in the built R-groups were calculated using RDKit³².

corresponding crystal structure (7L12). However, the predicted conformation of the unsubstituted benzyloxy side chain is at odds with the crystal structure (7L12). The correct conformer is output as an alternative low energy conformer and, interestingly, the majority of the modelled larger, substituted benzyloxy groups adopt the crystal conformation. This is exemplified by **21** in Fig. 4d, which also correctly orients the *ortho*-Cl down into the S4 pocket.

The uracil series comprises a set of 13 analogues, spanning around 2.5 kcal/mol in binding free energy, and as such provides a useful benchmark for demonstrating the next stage of the workflow. Although the gnina CNN affinities for these compounds are reasonably well correlated with experimental IC₅₀ measurements in a kinetic assay (Fig. S5)²⁴, it is desirable to investigate whether more rigorous free energy methods can be used to improve accuracy. Hence, relative binding free energies were computed using the SOMD software³⁰, starting from the structures output by the FEgrow workflow in complex with the receptor (see “Computational Methods”). Note that we have used the lowest energy structures as input to the free energy calculations (using instead the structure of e.g., **14** that corresponds most closely to the crystal structure can introduce differences of up to 0.4 kcal/mol in free energies in our tests, but this information would not be available for prospective studies). Figure 5 shows the agreement between experiment and simulations (MUE = 0.45 kcal/mol, $R^2 = 0.53$), and the raw data is provided in Table S5. Here, we can see that even though we have only used information from a single crystal structure of **4** bound to the protease, the combination of structure building and optimisation with the FEgrow workflow and free energy calculations with SOMD allows the (retrospective) prioritisation of compounds, such as compounds **20** and **21** for synthesis and testing.

Discussion

We have introduced here FEgrow, an open-source molecular builder and free energy preparation workflow. Taking as input a receptor and ligand core structure, FEgrow aims to build a user-defined library of chemical functional groups of the sort that would typically be used to explore structure-activity relationships with free energy calculations. Inspired by the BOMB approach to molecular design¹³, we grow from a fixed ligand core in order to

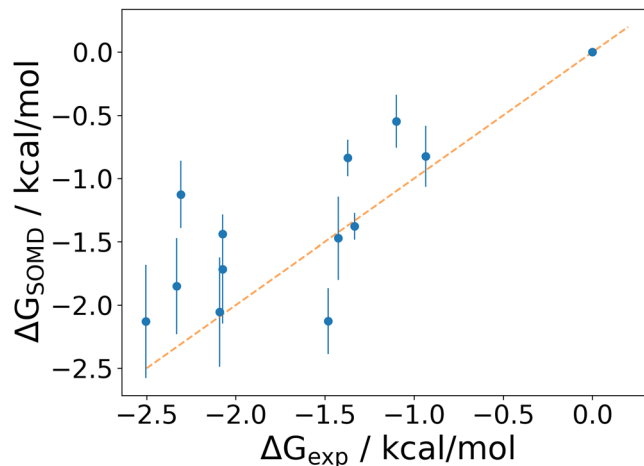


Fig. 5 Comparison between free energy calculations and experiment. Binding free energies of 13 analogues of the uracil-based MP^{pro} inhibitors, relative to compound **10**. The error bars indicate one standard error based on least square fitting⁷⁵.

maximise the use of binding mode information from structural biology sources, and rely on the user’s medicinal chemistry expertise to suggest functional groups that improve binding affinity whilst remaining synthetically tractable. Alternative, generative methods for fragment growth^{11,12} could be incorporated in future, but testing of expert medicinal chemist designs still remains popular today and FEgrow aims to automate this process.

The modular workflow of FEgrow allows us to experiment with functionalities, such as new optimisation or scoring methods. With the use of hybrid ML/MM structural optimisation, in particular, we aim to obtain reliable coordinates for the added R-groups. In this respect, the ANI neural network potential (within the ML/MM approximation) has already been shown to be capable of predicting protein–ligand binding poses in agreement with electron density distributions determined by X-ray crystallography⁴⁴, and should be significantly more reliable than the general purpose force fields (such as UFF) that are typically used for structure refinement in de novo design packages. Updated machine learning potentials or semi-empirical methods⁶⁵ can readily be included in future versions of FEgrow.

Ligand designs are evaluated for simple molecular properties, and their binding affinity predicted using the gnina CNN scoring function. Despite the challenge of discriminating between relatively small functional group modifications, the scoring function performs quite well and is useful in providing initial guidance for a number of targets from the protein–ligand benchmark set used here. Nevertheless, we envisage the primary use of FEgrow being as a source of input structures for more rigorous free energy-based affinity predictions. We demonstrate this functionality here, using SOMD to calculate the relative binding free energies of 13 uracil-based inhibitors of the SARS-CoV-2 main protease. Using only a single crystal structure as input (PDB: 7L10) and the FEgrow workflow to build the remaining structures, we obtain excellent agreement with experimental binding affinities (MUE = 0.45 kcal/mol, $R^2 = 0.53$).

We envisage future improvements including the use of a flexible receptor for the growth phase, and future use cases including seeding free energy calculations with multiple low energy conformers. The BACE(Hunt) target in Case Study I highlighted the difficulty of accurately including the energetics and effects on the binding affinity of displacing water networks in hydrated binding pockets. There does not currently appear to be a

satisfactory means to include water networks into the optimisation or scoring phases of FEgrow, but output structures could be passed to molecular dynamics or Monte Carlo-based simulations to assess optimal hydration sites for predicted poses^{66–68}. FEgrow is available for download from <https://github.com/cole-group/FEgrow>, and we welcome suggestions from the community for added functionality.

Computational methods

Free energy calculations. Structures of 13 inhibitors of the main protease (M^{Pro}) of SARS-CoV-2 were built using the FEgrow workflow and taken through to free energy calculations for accurate physics-based scoring. The PDB structure, 7L10, was used for the receptor. Missing residues (E47 and D48) were added using MODELLER⁶⁹, which uses optimisation of a pseudo energy function for loop modelling, and hydrogen atoms were added using Chimera²⁰, which includes options for optimisation of the hydrogen bond network. The BioSimSpace package⁷⁰ was used for free energy setup, along with a relative binding free energy protocol used previously⁷¹. The lowest energy conformer for each ligand was parameterised with the GAFF2 force field, using the AM1-BCC charge model. The AMBER FF14SB³⁹ force field was used for the protein, along with the TIP3P water model. Each ligand was then solvated in a 35 Å cube, or 90 Å cube in the presence of the protein. The bound and unbound structures then underwent a short equilibration using the default procedure in BioSimSpace⁷⁰. Namely, the structure was minimised, then heated to 300 K in the NVT ensemble over a period of 10 ps. It was then equilibrated for a further 10 ps in the NpT ensemble at 300 K and 1 bar, using the Langevin thermostat and Berendsen barostat. Atoms in the protein backbone were restrained to their initial positions throughout, and a 8 Å nonbonded cutoff was applied.

The network of alchemical transformations was built manually to include cycle closures for error analysis, and is shown in Fig. S4. Table S6 shows that the absolute cycle closure errors are typically less than 0.5 kcal/mol, and less than 1 kcal/mol for all cycles. The overlap for each perturbation was determined using a maximum common substructure search to determine the atoms to be morphed. Each transformation leg was simulated using the SOMD software package³⁰ for 4 ns, and the first 400 ps were discarded as equilibration. Eleven equally-spaced λ windows were employed between 0 and 1, along with the default soft core. The time step was set to 2 fs, with constraints applied to unperturbed hydrogen bonds. Simulations were performed in the NpT ensemble, using an Andersen thermostat with collision frequency of 10.0 ps⁻¹ and a Monte Carlo barostat with a frequency of 25 time steps. Periodic boundary conditions and a tapered nonbonded cutoff distance of 10 Å were applied. Electrostatic interactions were calculated using the reaction-field method with a dielectric constant of 78.3 outside the nonbonded cutoff⁷². All transformations reported here were run in both forward and backward directions, and in duplicate. Free energy changes and their errors were calculated from the output with MBAR using the asymptotic covariance method⁷³. Final free energies and their associated error bars (Fig. 5) were calculated from the network with the freenrgworkflows package⁷⁴, using the method of Yang et al.⁷⁵. All protocols used and raw data are provided in the accompanying Supporting Data (<https://doi.org/10.5281/zenodo.7112943>).

Data availability

Analysis of scaling factors used during geometry optimisation, background information on molecular property filters, full details of protein–ligand benchmark targets studied and CNN scoring function results, free energy networks and raw data for M^{Pro} free

energy calculations (Supplementary Information). Data accompanying this paper are freely available at <https://doi.org/10.5281/zenodo.7112943>.

Code availability

FEgrow, and an accompanying tutorial, are freely available at <https://github.com/cole-group/FEgrow>. The version 1.0.2 of FEgrow used in this study is available for download from <https://doi.org/10.5281/zenodo.7105647>.

Received: 20 May 2022; Accepted: 11 October 2022;

Published online: 27 October 2022

References

1. Bender, B. J. et al. A practical guide to large-scale docking. *Nat. Protoc.* **16**, 4799–4832 (2021).
2. Schneider, G. & Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **4**, 649–663 (2005).
3. Chéron, N., Jasty, N. & Shakhnovich, E. I. OpenGrowth: An automated and rational algorithm for finding new protein ligands. *J. Med. Chem.* **59**, 4171–4188 (2016).
4. Durrant, J. D., Amaro, R. E. & McCammon, J. A. AutoGrow: A novel algorithm for protein inhibitor design. *Chem. Biol. Drug Des.* **73**, 168–178 (2009).
5. Yuan, Y., Pei, J. & Lai, L. Ligbuilder 2: A practical de novo drug design approach. *J. Chem. Inf. Model.* **51**, 1083–1091 (2011).
6. Sousa, T., Correia, J., Pereira, V. & Rocha, M. Generative deep learning for targeted compound design. *J. Chem. Inf. Model.* **61**, 5343–5361 (2021).
7. Schneider, G. & Clark, D. E. Automated de novo drug design: Are we nearly there yet? *Angew. Chem. Int. Ed.* **58**, 10792–10803 (2019).
8. Pearce, B. C., Langley, D. R., Kang, J., Huang, H. & Kulkarni, A. E-Novo: An automated workflow for efficient structure-based lead optimization. *J. Chem. Inf. Model.* **49**, 1797–1809 (2009).
9. Cross, S. & Cruciani, G. FragExplorer: GRID-based fragment growing and replacement. *J. Chem. Inf. Model.* **62**, 1224–1235 (2022).
10. Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **28**, 849–857 (1985).
11. Green, H., Koes, D. R. & Durrant, J. D. DeepFrag: A deep convolutional neural network for fragment-based lead optimization. *Chem. Sci.* **12**, 8036–8047 (2021).
12. Imrie, F., Hadfield, T. E., Bradley, A. R. & Deane, C. M. Deep generative design with 3d pharmacophoric constraints. *Chem. Sci.* **12**, 14577–14589 (2021).
13. Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **42**, 724–733 (2009).
14. Cournia, Z., Allen, B. & Sherman, W. Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. *J. Chem. Inf. Model.* **57**, 2911–2937 (2017).
15. Cournia, Z. et al. Rigorous free energy simulations in virtual screening. *J. Chem. Inf. Model.* **60**, 4153–4169 (2020).
16. Mey, A. S. J. S. et al. Best practices for alchemical free energy calculations [article v1.0]. *LiveCoMS* **2**, 18378 (2020).
17. Mobley, D. L. & Gilson, M. K. Predicting binding free energies: Frontiers and benchmarks. *Annu. Rev. Biophys.* **46**, 531–558 (2017).
18. Gapsys, V. et al. Pre-exascale computing of protein–ligand binding free energies with open source software for drug design. *J. Chem. Inf. Model.* **62**, 1172–1177 (2022).
19. Citations | Schrödinger. <https://www.schrodinger.com/citations#Maestro>. Accessed 4 March 2022 (2022).
20. Pettersen, E. F. et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
21. Jorgensen, W. L. et al. Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **16**, 663–667 (2006).
22. Lee, W.-G. et al. Picomolar inhibitors of HIV reverse transcriptase featuring bicyclic replacement of a cyanovinylphenyl group. *J. Am. Chem. Soc.* **135**, 16705–16713 (2013).
23. Dziedzic, P. et al. Design, synthesis, and protein crystallography of biaryltriazoles as potent tautomerase inhibitors of macrophage migration inhibitory factor. *J. Am. Chem. Soc.* **137**, 2996–3003 (2015).
24. Zhang, C.-H. et al. Potent noncovalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug peramppanel guided by free energy perturbation calculations. *ACS Cent. Sci.* **7**, 467–475 (2021).

25. The COVID Moonshot Consortium. COVID Moonshot: Open Science Discovery of SARS-CoV-2 Main Protease Inhibitors by Combining Crowdsourcing, High-Throughput Experiments, Computational Simulations, and Machine Learning. Accessed 4 March 2022 <https://doi.org/10.26434/chemrxiv.13158218.v1> (2020).
26. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer generation with omega: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **50**, 572–584 (2010).
27. choderalab. perses. <https://github.com/choderalab/perses>. Accessed 4 March 2022 (2022).
28. Devereux, C. et al. Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens. *J. Chem. Theory Comput.* **16**, 4192–4202 (2020).
29. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
30. Woods, C., Hedges, L., Michel, J. Sire Molecular Simulation Framework. <http://siremol.org> (2021).
31. Rego, N. & Koes, D. 3Dmol.js: Molecular visualization with WebGL. *Bioinformatics* **31**, 1322–1324 (2014).
32. Landrum, G. Rdkit: Open-source cheminformatics. <http://www.rdkit.org/> (2022).
33. O'Boyle, N. M. et al. Open babel: An open chemical toolbox. *J. Cheminformatics* **3**, 33 (2011).
34. Riniker, S. & Landrum, G. A. Better informed distance geometry: Using what we know to improve conformation generation. *J. Chem. Inf. Model.* **55**, 2562–2574 (2015).
35. Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *J. Am. Chem. Soc.* **114**, 10024–10035 (1992).
36. Takeuchi, K., Kunimoto, R. & Bajorath, J. R-group replacement database for medicinal chemistry. *Future Sci. OA* **7**, 8 (2021).
37. Bouysset, C. mols2grid - Interactive molecule viewer for 2D structures. <https://github.com/cbouy/mols2grid> (2022).
38. Eastman, P. et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **13**, 7 (2017).
39. Maier, J. A. et al. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
40. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
41. Qiu, Y. et al. Development and benchmarking of open force field v1.0.0—the Parsley small-molecule force field. *J. Chem. Theory Comput.* **17**, 6262–6280 (2021).
42. Cole, D. J., Mones, L. & Csányi, G. A machine learning based intramolecular potential for a flexible organic molecule. *Faraday Discuss.* **224**, 247–264 (2020).
43. Lahey, S.-L. J., Thien Phuc, T. N. & Rowley, C. N. Benchmarking force field and the ani neural network potentials for the torsional potential energy surface of biaryl drug fragments. *J. Chem. Inf. Model.* **60**, 6258–6268 (2020).
44. Lahey, S.-L. J. & Rowley, C. N. Simulating protein-ligand binding with neural network potentials. *Chem. Sci.* **11**, 2362–2368 (2020).
45. Ragoza, M., Hochuli, J., Idrobo, E., Sunseri, J. & Koes, D. R. Protein-ligand scoring with convolutional neural networks. *J. Chem. Inf. Model.* **57**, 942–957 (2017).
46. Francoeur, P. G. et al. Three-dimensional convolutional neural networks and a cross-docked data set for structure-based drug design. *J. Chem. Inf. Model.* **60**, 4200–4215 (2020).
47. McNutt, A. T. et al. GNINA 1.0: Molecular docking with deep learning. *J. Cheminf.* **13**, 1–20 (2021).
48. Koes, D. R., Baumgartner, M. P. & Camacho, C. J. Lessons learned in empirical scoring with smina from the csar 2011 benchmarking exercise. *J. Chem. Inf. Model.* **53**, 1893–1904 (2013).
49. Trott, O. & Olson, A. J. Autodock vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
50. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**, 3–25 (1997).
51. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminformatics* **1**, 8 (2009).
52. Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (pains) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740 (2010).
53. Jadhav, A. et al. Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease. *J. Med. Chem.* **53**, 37–51 (2010).
54. Doveston, R. G. et al. A unified lead-oriented synthesis of over fifty molecular scaffolds. *Org. Biomol. Chem.* **13**, 859–865 (2014).
55. Brenk, R. et al. Lessons learnt from assembling screening libraries for drug discovery for neglected diseases. *ChemMedChem* **3**, 435–444 (2008).
56. Sydow, D., Morger, A., Driller, M. & Volkamer, A. Teachopencadd: A teaching platform for computer-aided drug design using open source packages and data. *J. Cheminformatics* **11**, 29 (2019).
57. Hahn, D. F. et al. Best practices for constructing, preparing, and evaluating protein-ligand binding affinity benchmarks [article v1.0]. LiveCoMS 4 <https://doi.org/10.33011/livecoms.4.1.1497> (2022).
58. Hahn, D. F. & Wagner, J. Protein-ligand-benchmark. <https://doi.org/10.5281/zenodo.4813735>. Accessed October 2021 (2021).
59. Liang, J. et al. Lead identification of novel and selective tyk2 inhibitors. *Eur. J. Med. Chem.* **67**, 175–187 (2013).
60. Baum, B. et al. More than a simple lipophilic contact: A detailed thermodynamic analysis of nonbasic residues in the s1 pocket of thrombin. *J. Mol. Biol.* **390**, 56–69 (2009).
61. Goldstein, D. M. et al. Discovery of 6-(2,4-difluorophenoxy)-2-[3-hydroxy-1-(2-hydroxyethyl)propylamino]-8-methyl-8h-pyrido[2,3-d]pyrimidin-7-one (pamapimod) and 6-(2,4-difluorophenoxy)-8-methyl-2-(tetrahydro-2h-pyran-4-ylamino)pyrido[2,3-d]pyrimidin-7(8h)-one (r1487) as orally bioavailable and highly selective inhibitors of p38 α mitogen-activated protein kinase. *J. Med. Chem.* **54**, 2255–2265 (2011).
62. Wilson, D. P. et al. Structure-based optimization of protein tyrosine phosphatase 1b inhibitors: From the active site to the second phosphotyrosine binding site. *J. Med. Chem.* **50**, 4681–4698 (2007).
63. Hunt, K. W. et al. Spirocyclic β -site amyloid precursor protein cleaving enzyme 1 (bace1) inhibitors: From hit to lowering of cerebrospinal fluid (csf) amyloid β in a higher species. *J. Med. Chem.* **56**, 3379–3403 (2013).
64. Zhang, L. et al. Crystal structure of sars-cov-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* **368**, 409–412 (2020).
65. Bannwarth, C., Ehlert, S. & Grimme, S. Gfn2-xtb-an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).
66. Samways, M. L., Bruce Macdonald, H. E. & Essex, J. W. grand: A python module for grand canonical water sampling in openmm. *J. Chem. Inf. Model.* **60**, 4436–4441 (2020).
67. Abel, R., Young, T., Farid, R., Berne, B. J. & Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor xa ligand binding. *J. Am. Chem. Soc.* **130**, 2817–2831 (2008).
68. Ge, Y. et al. Enhancing sampling of water rehydration on ligand binding: A comparison of techniques. *J. Chem. Theory Comput.* **18**, 1359–1381 (2022).
69. Webb, B. & Sali, A. Comparative protein structure modeling using modeller. *Curr. Protoc. Bioinform.* **54**, 5.6.1–5.6.37 (2016).
70. Hedges, L. et al. Biosimspace: An interoperable python framework for biomolecular simulation. *J. Open Source Softw.* **4**, 1831 (2019).
71. Nelson, L. et al. Implementation of the QUBE force field in SOMD for high-throughput alchemical free-energy calculations. *J. Chem. Inf. Model.* **61**, 2124–2130 (2021).
72. Kuhn, M. et al. Assessment of binding affinity via alchemical free-energy calculations. *J. Chem. Inf. Model.* **60**, 3120–3130 (2020).
73. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **129**, 124105 (2008).
74. Mey, A. S., Jiménez, J. J. & Michel, J. Impact of domain knowledge on blinded predictions of binding energies by alchemical free energy calculations. *J. Comput. Aided Mol. Des.* **32**, 199–210 (2018).
75. Yang, Q. et al. Optimal designs for pairwise calculation: An application to free energy perturbation in minimizing prediction variability. *J. Comput. Chem.* **41**, 247–257 (2020).

Acknowledgements

D.J.C., M.K.B., and J.H. acknowledge support from a UKRI Future Leaders Fellowship (grant MR/T019654/1). B.C. and R.P. are grateful for support from the EPSRC Centre for Doctoral Training in Molecular Sciences for Medicine (grant EP/S022791/1) and an EPSRC Doctoral Training Partnership studentship (grant EP/R51309X/1). This work made use of the facilities of the N8 Centre of Excellence in Computationally Intensive Research (N8 CIR) provided and funded by the N8 research partnership and EPSRC (grant EP/T022167/1). We are grateful to Dr Leela Dodda (ORCID: 0000-0002-3584-1729), Dr Antonia Mey (ORCID: 0000-0001-7512-5252), and Dr Julien Michel (ORCID: 0000-0003-0360-1760) for helpful discussions.

Author contributions

M.B.: Conceptualisation, data curation, formal analysis, investigation, methodology, software, validation, visualisation, and writing - original draft. B.C.: Conceptualisation, data curation, formal analysis, investigation, methodology, software, validation,

visualisation, and writing - original draft. R.P.: Conceptualisation, software, validation, and writing - original draft. J.H.: Conceptualisation, investigation, methodology, software, validation, and writing - original draft. N.T.: Conceptualisation, funding acquisition, project administration, supervision, validation, and writing - review & editing. D.C.: Conceptualisation, funding acquisition, methodology, project administration, resources, supervision, writing - original draft, and writing - review & editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42004-022-00754-9>.

Correspondence and requests for materials should be addressed to Daniel J. Cole.

Peer review information *Communications Chemistry* thanks Olexandr Isayev and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022