BRIDGE an open platform for reproducible high throughput free energy simulations

Tharindu Senapathi[1], Miroslav Suruzhon[2], Christopher B. Barnett[1], Jonathan Essex[2] and Kevin J. Naidoo[1,3]*

[1]Scientific Computing Research Unit and Department of Chemistry, University of Cape Town, Rondebosch, 7701, South Africa; [2]School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom, [3]Institute of Infectious Disease and Molecular Medicine, Faculty of Health Science, University of Cape Town, Rondebosch, 7701, South Africa

KEYWORDS: ProtoCaller, Galaxy, Free Energy of Binding, Solvation Free Energy, Workflows, ST3Gal-I, BRIDGE

ABSTRACT

Biomolecular Reaction and Interaction Dynamics Global Environment (BRIDGE) is an open-source web platform developed with the aim to provide an environment for the design of reliable methods for, and to conduct reproducible, biomolecular simulations. It is built on the well-known Galaxy bioinformatics platform. Through this BRIDGE hosts computational chemistry tools on public web servers for internet use as well as provide machine and operating system independent portability using the Docker container platform for local use. This construction improves the accessibility, shareability, and reproducibility of computational methods for molecular simulations. Here we present integrated free energy tools (or apps) to calculate absolute binding free energies (ABFE) and relative binding free energies (RBFE) illustrated through use cases. We present Free Energy Perturbation (FEP) methods contained in various software packages such as GROMACS and YANK that are independent of hardware configuration, software libraries or operating systems when ported in the Galaxy-BRIDGE Docker container platform. By performing Cyclin-dependent kinase 2 (CDK2) FEP calculations on geographically dispersed webservers (in Africa and Europe) we illustrate that large scale computations can be performed using the exact same codes and methodology by collaborating groups through publicly shared protocols and workflows. The ease of public sharing and independent reproduction of simulations via BRIDGE makes possible an open review of methods and complete simulation protocols. This makes the discovery of inhibitors for drug targets accessible to the non-experts and the computer experiments that are used to arrive at leads verifiable by experts and reviewers. We illustrate this on beta-galactoside alpha-2,3-sialyltransferase I (ST3Gal-I), a breast cancer drug target, where a combination of RBFE and ABFE methods are used to compute the binding free energies of three inhibitors.

INTRODUCTION

The challenges facing novice molecular modelers, looking to compute protein-ligand interactions accurately using vetted protocols and large-scale drug design collaboratives, requiring common methods and models to perform accurate high throughput intermolecular computations, are common. Both require reproducibility and shareability. The obstacles to use are often that various codes and associated protocols are not easily accessible to those that are not part of development groups or are privy to in-house expert laboratory experiences and protocols. Transferability and web access of code is a largely solved problem as much of e-commerce notably giant corporations such as *amazon* or web-based mail services such as *Gmail* rely on software container functionality to allow ease of access no matter where the user is geographically located or what hardware is being used for the application. Recently we reported an open-source Biomolecular Reaction and Interaction Dynamics Global Environment (BRIDGE)[1] that provides, on a single platform, computational chemistry simulation data and tools alongside the inherent Galaxy[2, 3] features where biological data and bioinformatics tools are seamlessly accessed. A feature of Galaxy is its prepackaging using mostly Docker container technology, allowing the transferable use of scientific code that is independent of the hardware platform or operating system. We chose Galaxy as a development platform as its genetics and proteomics databases, and schemes can be linked to molecular dynamics (MD) simulation and analysis methods through BRIDGE. Here we report the BRIDGE molecular *Interaction* functionality that provides a robust framework to predict ligand affinities, which includes alchemical free energy methods such as free energy perturbation (FEP) and thermodynamic integration (TI) methods. [4-6]

Site-directed mutagenesis and comparative protein-ligand binding are primary tools to biochemically map a disease and with which to design drugs from an understanding of binding mechanisms. Computational design and informatics play an increasingly significant role in the

discovery of molecular mechanisms of small molecules that alter disease pathways.[7] The alchemical free energy methods incorporated into BRIDGE sample a large number of conformations using MD or Monte Carlo simulations and so improving on account of entropic binding contributions.[8] FEP based methods can be used to calculate the free energy (FE) of a small molecule bound to a receptor (absolute binding FE) or the relative affinity of a series of related ligands to a receptor (relative binding FE). Relative binding FE (RBFE) calculations are most commonly used in lead optimization structure-based drug designs because of the comparative speed and accuracy advantages over absolute binding FE (ABFE) calculations. ABFE calculations are most useful when the reference structures are absent, or the perturbations between structures are computationally inaccessible. Extensive reviews of FE methods are available[9, 10]. With attempts at ligand-protein FE computation now more than 50 years in the making[11], there is consensus on methodology[9] and protocols[10] needed for accuracy and reliability of these calculations. Robust FE methods accompanied by accurately parameterized biomolecular and small molecule force fields are embedded in packages such as CHARMM,[12] AMBER[13], and GROMOS[14]. The BRIDGE open-access platform presented here necessitates open-source and well-supported molecular modeling software; for this reason, the development was based on YANK[15] and GROMACS[16].

THEORY and METHODS

The first BRIDGE development contained the *Dynamics Global Environment* modules. Here the development of the *Interaction* module that enables seamless setup, computation, and protocol sharing of ligand-protein FE is reported. The existing BRIDGE tools and the Docker image were extended. The FE tools were wrapped using the Galaxy Tool XML language and are available on GitHub for open access. Software dependencies for the tools are resolved using the Anaconda

Python distribution system. Most public Galaxy servers have quite strict channel specifications; thus, tools requiring custom channels are available in the Docker image and at https://galaxy-compchem.ilifu.ac.za.

Absolute Binding Free Energy tools

The absolute binding of a ligand (L) to a protein (P) is estimated as:

$$\Delta G_{bind} = \left[(G(L)_{water}) - \left(G(L)_{gas}\right)\right] - \left[(G(PL)_{water}) - \left(G(P)_{water} + G(L)_{gas}\right)\right] \quad (1)$$

The enabling tools developed for ABFE computations are Absolute Solvation Free Energy, Absolute Binding Free Energy, and YANK Analysis. These tools are based on YANK[15], which expands on OpenMM toolkit[17] and supports the AMBER force field[18] for proteins and the General Amber Force Field (GAFF)[19] for small molecules. YANK is well suited for parallel and GPU computing using a Hamiltonian Replica Exchange MD[20] to efficiently sample ligands, which may be hindered in the protein binding site.

In the design of novel ligands such as glycomimetic inhibitors[21], a reference structure may not be present to perform RBFE. In these cases, ABFE is useful to arrive at an estimate of the FE, which requires the protein structure as a PDB file and the ligand structure as a MOL2 file. The ligand need not be docked in the active site, as the ABFE tool can be used to find the binding pocket as well as the free energy of binding depending on the restraints used between the protein and the ligand (for example Harmonic, Flat-Bottom, Boresch, or RMSD restraints). ABFE simulations are computationally intensive; they can be run on the CPU like the other tools but are best suited to run on GPUs. If no GPU is available locally, the computation can be set up and executed on remote GPU resources. The Absolute Solvation Free Energy (ASFE) tool was also developed and tested on a use case described in Shivakumar et al.[22] (supporting information).

Relative Binding Free Energy tools

RBFE simulations measure the simulated binding affinity of one ligand compared with another through the 'mutation' of the reference ligand (A) to the final ligand (B). The transition is made through intermediate steps on the λ-coordinate. The tools developed for RBFE are Alchemical Setup, Alchemical Run, and Alchemical Analysis. These tools are based on ProtoCaller[23] and GROMACS[16], and Alchemical Analysis[24]. To initiate a RBFE simulation or series of simulations, a protein structure (PDB ID or PDB file) and ligand structure (SMILES, InChI, or SDF file) need to be uploaded to the BRIDGE platform (Figure 1A). We developed the Alchemical Setup wrapper that includes ProtoCaller to link several specialized tools to perform protein setup and parametrization. ProtoCaller supports standard AMBER force fields and includes an enhancement to the maximum common substructure algorithm to improve ligand-ligand mapping for stereoisomers. This ligand mapping is crucial as for RBFE the ligand must be placed in the active site. ProtoCaller handles the full setup, including mapping the ligands, charge parameterization, and adding dummy atoms that are needed for the RBFE calculations. The RBFE simulations are carried out using the Alchemical Run tool developed based on GROMACS. The slow-growth method as implemented in GROMACS along with the topologies created by the Alchemical Setup tool were used to run FEP simulations. The Alchemical Run tool has an inbuilt workflow that carries out energy minimization and equilibration (to avoid any possible Hamiltonian lagging) followed by production runs for each free energy window. The Alchemical Analysis tool can be used to carry out extensive analysis of the RBFE simulations using various free energy estimators such as Bennett Acceptance Ratio (BAR), Multiscale Bennett Acceptance Ratio (MBAR), and Thermodynamic Integration (TI). From these estimators graphical and textual output can be produced to compute the free energy differences and evaluate the quality of the simulations.

User Interface, workflows, use cases

The BRIDGE interface (Figure 1A) is straightforward and follows the Galaxy design. Tools are in the left panel; a history of progress is in the right panel, and information about the current tool or dataset selected is in the central panel.

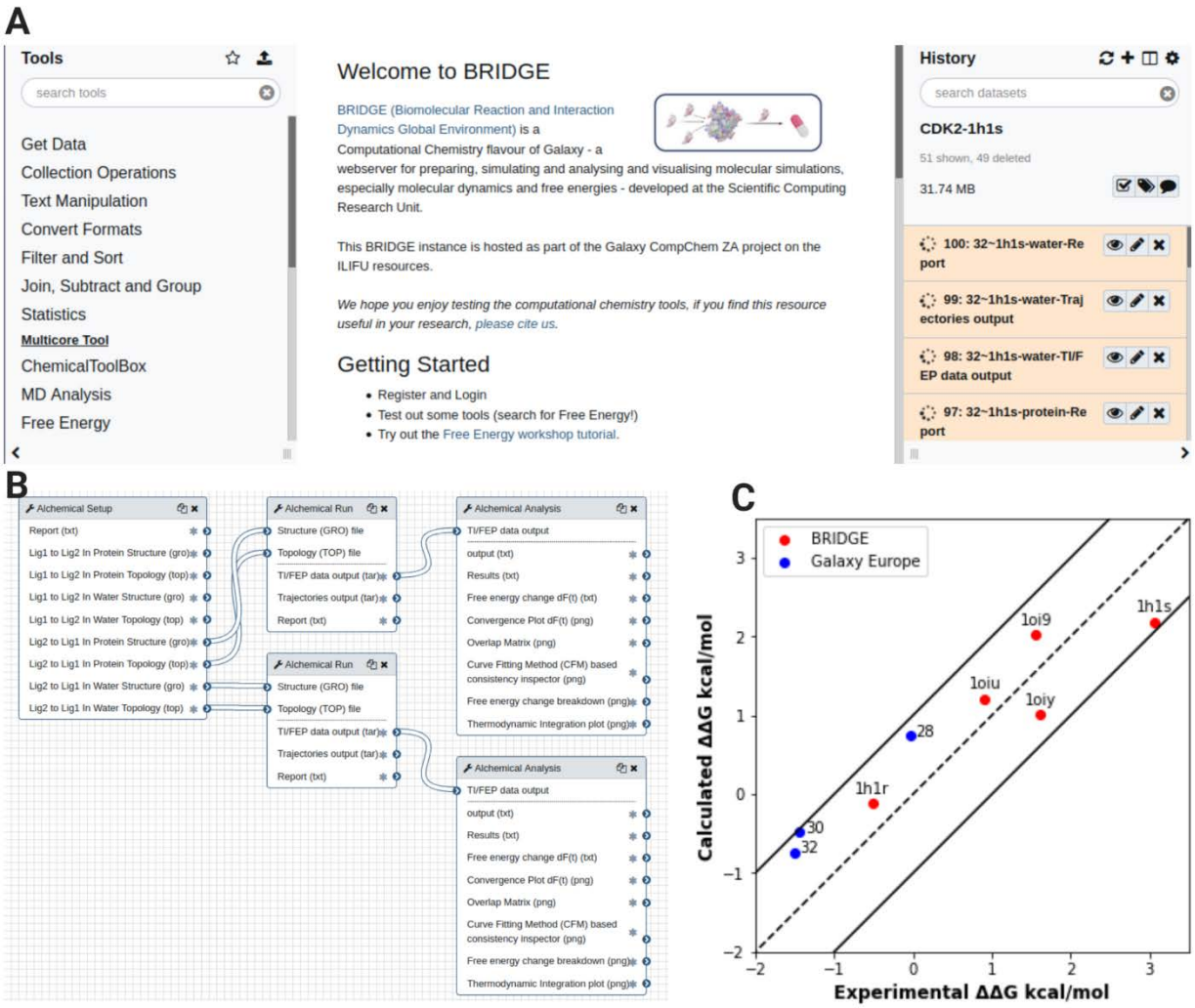


Figure 1: (A) The Galaxy/BRIDGE interface (B) workflow illustrating RBFE calculations, (C) Data can be plotted directly in Galaxy. Aggregated experimental and simulation data were plotted for the RBFEs for ligands in CDK2, which were simulated on geographically distant Galaxy servers.

Data can be uploaded using the upload icon at the top of the Tools menu (left panel), while the top menu bar includes share resources, user account, and other information. By way of illustration, an example workflow for an RBFE calculation is given (Figure 1B). This includes the setup, instructions for the FEP simulations using the GROMACS tool, where the structure and topology file outputs are selected from the alchemical setup tool and the desired simulation parameters defined. The four steps, (i) energy minimization, (ii) equilibration in an *NVT* ensemble, (iii) equilibration in an *NpT* ensemble, and (iv) production FEP simulation in an *NpT* ensemble is fully automated (Figure 1B) and executed for each thermodynamic window.

On completion of the RBFE computations, an Alchemical Analysis[24] tool can be deployed to interrogate the output from the Alchemical Run tool and verify the convergence of the simulations. The RBFE values are obtained by subtracting the FE change in water from the change in the active site, and the quality of the FEP simulation and selection of windows is validated with these tools (note the windows can be changed at runtime manually or with the following tool https://github.com/scientificomputing/rbfe_sequence_generator). Once a protocol has been established, a workflow connecting all the setup, run, and analysis steps can be shared. Both ligand in water and ligand in the receptor active site simulations can be run in the same workflow. Workflows provide a means to train novices, ensure reproducibility of the calculations, and build pipelines for high throughput computations evaluating multiple ligand-protein binding experiments. Workflows allow for the independent validation of simulations by project supervisors, collaborators, or article referees.

To illustrate how BRIDGE applies to collaborative drug discovery, we include the well-known CDK2[25, 26] use cases (workflows and data are available[27]). Use cases detailing absolute solvation, absolute binding, and relative binding free energy calculations are in the supporting information.

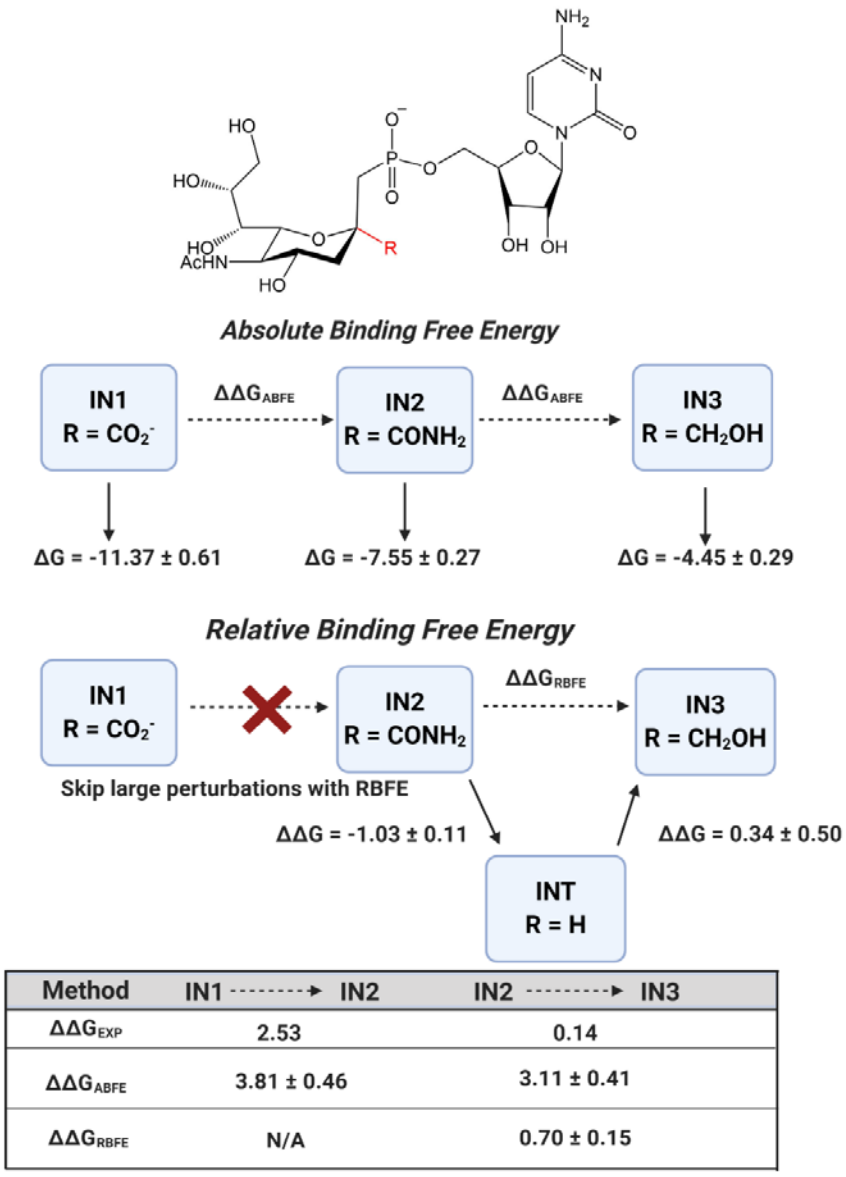
For CDK2 ligands, the RBFEs were calculated using the same workflow but calculated on geographically distant resources (the University of Cape Town affiliated ilifu server and Galaxy Europe at the University of Freiburg). The results were combined, and a scatter plot (Figure 1C, supporting information) reproduces the benchmarked results where ligand RBFEs are within a 1 kcal/mol tolerance and compare well with experimental data.

RESULTS AND DISCUSSION

To illustrate the advantage of drug development using diverse FE methodologies that are co-located on a single platform sharing common data, we computed the binding of possible inhibitors to beta-galactoside alpha-2,3-sialyltransferase I (ST3Gal-I) requiring a combination of absolute and relative free energies. Overexpression of ST3Gal-I in breast cancer affects not only the onset of carcinogenesis but may influence early tumor development.[28] Inhibiting the ST3Gal-I enzyme may prevent the formation of tumors, trigger apoptosis, and arrest metastasis. However, MD simulations and FE computations of ligand glycosyltransferase binding are more challenging than most protein-ligand cases. This presents a scenario where a protocol developed using expertise in glycosyltransferase modeling can be easily shared.

Previously, we computed RBFE for a set of ligands with marginal success given the lack of high throughput screening and the limitations of RBFE.[29] We now reconsider binding free energies of three previously studied inhibitors for ST3Gal-I.[30] We compute $\Delta\Delta G$ and compare these with the experimental $\Delta\Delta G$ (calculated from IC_{50}) using the RTln (IC50 ratio). The three ligands have a common frame and differ in only one functional group (Figure 2). While charge perturbation calculation from IN1 to IN2 would be preferred, it has the computational challenge of slow convergence in GROMACS. Nonetheless a workflow for RBFE with charge perturbation can be developed in BRIDGE for open sharing. Similarly, the platform lends itself to the inclusion of

state-of-the-art methods developed by a community of users. However, a ready solution to computing the IN1 to IN2 free energy differences is ABFE to calculate the binding free energies of the individual ligands separately.



Method	IN1 ⇢ IN2	IN2 ⇢ IN3
$\Delta\Delta G_{EXP}$	2.53	0.14
$\Delta\Delta G_{ABFE}$	3.81 ± 0.46	3.11 ± 0.41
$\Delta\Delta G_{RBFE}$	N/A	0.70 ± 0.15

Figure 2: The binding FE of S_N1 type donor analog inhibitors for ST3Gal-I enzyme computed via ABFE (IN1⇢IN2, IN2⇢IN3) and RBFE (IN2⇢IN3) compared with experimental values reported in kcal/mol, errors calculated with the MBAR estimator. ABFE and RBFE performance is system dependent and may vary greatly. [31]

ABFEs were calculated for *IN1* and *IN2*. Following this, an RBFE workflow was designed to compute $\Delta\Delta G_{INS2\leftrightarrow INS3}$. The relatively large perturbation of the amide group to primary alcohol of the *IN2* to *IN3* perturbation was done in two steps. The amide group was perturbed to a hydrogen atom and then the hydrogen atom perturbed to the alcohol. The combined ABFE and RBFE results suggest that the binding FE increases in the order *IN3* < *IN2* < IN1 (also the case for ABFE only), which correlates with the experimental findings. Since the only difference between the three molecules is this R group, we reason that the carboxylic acid at the anomeric carbon is essential for binding.

The ST3Gal-I binding site is flexible and presents a ready environment for multiple O-glycan catalysis; this creates a challenge for specific ST3Gal-I inhibitor design. We chose a set of inhibitors that are highly flexible and have a charged phosphate. Including charged groups results in slow convergence for both YANK and GROMACS. For flexible ligands, such as IN3, ABFE requires significant computational resources and may not converge as well as RBFE.

Li et al. carried out a pairwise comparison of RBFE and ABFE for potential inhibitors of CDK2, TYK2, Thrombin, and JNK1. [31] While there is sometimes comparable accuracy (best R^2 is 0.79), the performance depends on the system and errors for individual ligands can vary greatly. Note, inconsistent results from two methodological approaches, such as the quantitative differences for the IN2 to IN3 perturbation from ABFE and RBFE (Figure 2), can be probed on a common platform. In this case, the poor ABFE convergence, especially for IN3, despite 16 days of computing time on an NVIDIA Tesla V100 (run at the CHPC, Cape Town, South Africa), is the likely reason.

The combination of charged, flexible ligands, and a flexible protein binding site (ST3Gal-I) shown here demonstrate the value of developing repeatable protocols for complex simulations that

can be shared and used for large ligand libraries. The comparative study of three inhibitors with varied functional group character is illustrative of rational drug design using reproducible simulations that can be shared between collaborating groups. The same computation and protocols are used across the collaboration but without requiring cumbersome installation and software tailoring.

We illustrate drug lead optimization using ST3Gal-I. This is accomplished through a pipeline that starts with the *Interaction* workflows to identify leads followed by the *Dynamics* module *and* analytics workflows that rationalize the emergence of the lead molecule. Using the *Interaction* module workflows, relative FE of binding for the three ligands that differ in charge and functionality is made possible using a combination of ABFE and RBFE methods. On discovering that IN1 has the best binding, amongst this small illustrative set, we investigated the reasons for this binding (supporting information); this involves an analysis of intermolecular interaction (in this case, ligand-protein hydrogen bonding) and analytics (in this case, PCA) performed on the protein. These essential tools and pre-designed workflows are the basis for lead optimization.

CONCLUSIONS

The *Interaction* module of the BRIDGE is reported, and the platform provides an open-access development space for FE modeling and best practice protocol sharing. BRIDGE makes protein-ligand binding FE calculations accessible, repeatable and shareable for novice and advanced users and collaborating users that are geographically dispersed. The open-source FE codes GROMACS and YANK and a setup tool ProtoCaller were wrapped for use in Galaxy, and the automation of FE high throughput screening methods can be accomplished by using workflows.

The process, results, and analysis of the FE calculations are automatically combined into a history, which can be used to share methods, protocols, and data that make FE simulations

reproducible. We illustrate this by performing RBFEs for ligands in CDK2 in two distant locations and recombining them through data sharing to produce a scatter plot for ligand evaluation and analysis. We showed that when demanding computations require hardware located at a distant site, the protocol can be shared on that site to undertake the computations. In a more advanced usage of the BRIDGE platform, we compute the relative FEPs (on ilifu) and absolute FEPs (CHPC GPU cluster) of a series of donor-based inhibitors binding to ST3Gal-I as an illustration of using BRIDGE to undertake collaborative drug discovery strategies such as lead discovery and optimization. Here we illustrate the benefits of an open platform to share protocols for challenging perturbations. However, as methods improve and protocols advance, they can be incorporated into BRIDGE and publicaly shared.

Data and Materials

BRIDGE is available through GitHub and DockerHub (https://github.com/scientificomputing/BRIDGE; https://github.com/galaxycomputationalchemistry/galaxy-tools-compchem), the ilifu hosted South African instance (https://galaxy-compchem.ilifu.ac.za) and the European Galaxy server (https://cheminformatics.usegalaxy.eu). The South African National Integrated Cyber Infrastructure System (NICIS) hosts a BRIDGE deployment that can interface with the Centre for High-Performance Computing (CHPC) service.

Supporting Information Available: Supporting data, figures, tables, and details of methods and test cases.

AUTHOR INFORMATION

Corresponding Author

*Kevin J. Naidoo E-mail: kevin.naidoo@uct.ac.za

Author Contributions

All authors have given approval to the final version of the manuscript.

Funding Sources

This work is based on research supported by the South African Research Chairs Initiative (SARChI) of the Department of Science and Technology (DST) and National Research Foundation (NRF) grant 449130 and supported by the South African Medical Research Council under a Self-Initiated Research Grant and the Medical Research Council grant (KJN). Even though the work is supported by the MRC, the views and opinions expressed are not those of the MRC but of the authors of the material produced or publicized. TS thanks SARChI for a doctoral award. We thank the Center for High Performance Computing for the use of their clusters. CB thanks the Univeristy of Cape Town for support. JE and MS thank AstraZeneca, GSK, and Syngenta for funding in this project. MS is grateful for the support from the EPSRC Centre for Doctoral Training, Theory and Modelling in Chemical Sciences, under Grant EP/L015722/1.

ACKNOWLEDGMENT

With thanks to Björn Grüning, Simon Bray, the usegalaxy.eu team and the Galaxy Project community.

ABBREVIATIONS

ABFE, ASFE, BRIDGE, FEP, MD, RBFE, FE

REFERENCES

1. Senapathi, T.; Bray, S.; Barnett, C. B.; Gruning, B.; Naidoo, K. J., Biomolecular Reaction and Interaction Dynamics Global Environment (BRIDGE). *Bioinformatics* **2019**, 35, 3508-3509.
2. Goecks, J.; Nekrutenko, A.; Taylor, J., Galaxy: A Comprehensive Approach for Supporting Accessible, Reproducible, and Transparent Computational Research in the Life Sciences. *Genome Biol.* **2010**, 11, R86.
3. Afgan, E.; Baker, D.; Batut, B.; van den Beek, M.; Bouvier, D.; Cech, M.; Chilton, J.; Clements, D.; Coraor, N.; Gruning, B. A.; Guerler, A.; Hillman-Jackson, J.; Hiltemann, S.; Jalili,

V.; Rasche, H.; Soranzo, N.; Goecks, J.; Taylor, J.; Nekrutenko, A.; Blankenberg, D., The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses: 2018 Update. *Nucleic Acids Res.* **2018**, 46, W537-W544.
4. Christ, C. D.; Mark, A. E.; Van Gunsteren, W. F., Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, 31, 1569-1582.
5. Michel, J.; Essex, J. W., Prediction of Protein–Ligand Binding Affinity by Free Energy Simulations: Assumptions, Pitfalls and Expectations. *J. Comput. Aided Mol. Des.* **2010**, 24, 639-658.
6. Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S., Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, 21, 150-160.
7. Lounnas, V.; Ritschel, T.; Kelder, J.; McGuire, R.; Bywater, R.; Foloppe, N., Current Progress in Structure-Based Rational Drug Design Marks a New Mindset in Drug Discovery. *Comput. Struct. Biotechnol. J.* **2013**, 5.
8. Barnett, C. B.; Wilkinson, K. A.; Naidoo, K. J., Pyranose Ring Transition State Is Derived from Cellobiohydrolase I Induced Conformational Stability and Glycosidic Bond Polarization. *J. Am. Chem. Soc.* **2010**, 132, 12800-12803.
9. Pohorille, A.; Jarzynski, C.; Chipot, C., Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, 114, 10235-10253.
10. Hansen, N.; van Gunsteren, W. F., Practical Aspects of Free-Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, 10, 2632-2647.
11. Platzer, K. E.; Momany, F. A.; Scheraga, H. A., Conformational Energy Calculations of Enzyme-Substrate Interactions. I. Computation of Preferred Conformations of Some Substrates of α-Chymotrypsin. *Int. J. Pept. Protein Res.* **1972**, 4, 187-200.
12. Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, 30, 1545-1614.
13. Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz, K. M., Using AMBER18 for Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2019**, 59, 3128-3135.
14. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F., The GROMOS Biomolecular Simulation Program Package. *J. Phys. Chem. A* **1999**, 103, 3596-3607.
15. Wang, K.; Chodera, J. D.; Yang, Y.; Shirts, M. R., Identifying Ligand Binding Sites and Poses Using GPU-Accelerated Hamiltonian Replica Exchange Molecular Dynamics. *J. Comput. Aided Mol. Des.* **2013**, 27, 989-1007.
16. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism From Laptops to Supercomputers. *SoftwareX* **2015**, 1-2, 19-25.
17. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S., OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comp. Biol.* **2017**, 13, e1005659.

18. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters From ff99SB. *J. Chem. Theory Comput.* **2015**, 11, 3696-3713.
19. 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.* **2004**, 25, 1157-1174.
20. Fukunishi, H.; Watanabe, O.; Takada, S., On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, 116, 9058-9067.
21. Ernst, B.; Magnani, J. L., From Carbohydrate Leads to Glycomimetic Drugs. *Nat. Rev. Drug Discov.* **2009**, 8, 661.
22. Shivakumar, D.; Williams, J.; Wu, Y.; Damm, W.; Shelley, J.; Sherman, W., Prediction of Absolute Solvation Free Energies Using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J. Chem. Theory Comput.* **2010**, 6, 1509-1519.
23. Suruzhon, M.; Senapathi, T.; Bodnarchuk, M. S.; Viner, R.; Wall, I. D.; Barnett, C. B.; Naidoo, K. J.; Essex, J. W., ProtoCaller: Robust Automation of Binding Free Energy Calculations. *J. Chem. Inf. Model.* **2020**.
24. Klimovich, P. V.; Shirts, M. R.; Mobley, D. L., Guidelines for the Analysis of Free Energy Calculations. *J. Comput. Aided Mol. Des.* **2015**, 29, 397-411.
25. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R., Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, 137, 2695-2703.
26. Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L., Large Scale Relative Protein Ligand Binding Affinities Using Non-Equilibrium Alchemy. *Chem. Sci.* **2020**, 11, 1140-1152.
27. scientificomputing/bridge-interaction-paper-2020-sm: Data and workflows - BRIDGE an open platform for reproducible high throughput free energy simulations. https://doi.org/10.5281/zenodo.3818324
28. Picco, G.; Julien, S.; Brockhausen, I.; Beatson, R.; Antonopoulos, A.; Haslam, S.; Mandel, U.; Dell, A.; Pinder, S.; Taylor-Papadimitriou, J.; Burchell, J., Over-Expression of ST3Gal-I Promotes Mammary Tumorigenesis. *Glycobiology* **2010**, 20, 1241-1250.
29. Crous, W.; Naidoo, K. J., Conformational and Electrostatic Analysis of SN1 Donor Analogue Glycomimetic Inhibitors of ST3Gal-I Mammalian Sialyltransferase. *Biorg. Med. Chem.* **2016**, 24, 4998–5005.
30. Izumi, M.; Wada, K.; Yuasa, H.; Hashimoto, H., Synthesis of Bisubstrate and Donor Analogues of Sialyltransferase and Their Inhibitory Activities. *J. Org. Chem.* **2005**, 70, 8817-8824.
31. Li, Z.; Huang, Y.; Wu, Y.; Chen, J.; Wu, D.; Zhan, C.-G.; Luo, H.-B., Absolute Binding Free Energy Calculation and Design of a Subnanomolar Inhibitor of Phosphodiesterase-10. *J. of Med. Chem.* **2019**, 62, 2099-2111.