Scaffold Hopping Transformations Using Auxiliary Restraints for Calculating Accurate Relative Binding Free Energies

Junjie Zou[1,2,3,4*], Zhipeng Li[1], Shuai Liu[2], Chunwang Peng[1], Dong Fang[1], Xiao Wan[1], Zhixiong Lin[1], Tai-sung Lee[5], Daniel P. Raleigh[3,4]*, Mingjun Yang[1,]*, Carlos Simmerling[3,4*]

1. Shenzhen Jingtai Technology Co., Ltd. (XtalPi), 4F, No. 9 Hualian Industrial Zone, Dalang Street, Longhua District, Shenzhen 518000, China
2. Shenzhen Jingtai Technology Co., Ltd. (XtalPi), 245 Main St, 11th Floor, Cambridge, MA 02142, United States
3. Department of Chemistry, Stony Brook University, Stony Brook, New York 11794-3400, United States
4. Laufer Center for Physical and Quantitative Biology, Stony Brook University, Stony Brook, New York 11794-3400, United States
5. Laboratory for Biomolecular Simulation Research, Center for Integrative Proteomics Research, Rutgers University, Piscataway, New Jersey, 08854-8076, United States

Corresponding Authors

*Junjie Zou: junjie.zou@xtalpi.com

*Daniel Raleigh: daniel.raleigh@stonybrook.edu

*Mingjun Yang: mingjun.yang@xtalpi.com

*Carlos Simmerling: carlos.simmerling@stonybrook.edu

Abstract

In silico screening of drug target interactions is a key part of the drug discovery process. Changes in the drug scaffold via contraction or expansion of rings, the breaking of rings and the introduction of cyclic structures from acyclic structures are commonly applied by medicinal chemists to improve binding affinity and enhance favorable properties of candidate compounds. These processes, commonly referred to as scaffold hopping, are challenging to model computationally. Although relative binding free energy (RBFE) calculations have shown success in predicting binding affinity changes caused by perturbing R-groups attached to a common scaffold, applications of RBFE calculations to modeling scaffold hopping are relatively limited. Scaffold hopping inevitably involves breaking and forming bond interactions of quadratic functional forms, which is highly challenging. A novel method for handling ring opening/closure/contraction/expansion and linker contraction/expansion is presented here. To the best of our knowledge, RBFE calculations on linker contraction/expansion have not been previously reported. The method uses auxiliary restraints to hold the atoms at the ends of a bond in place during the breaking and forming of the bonds. The broad applicability of the method was demonstrated by examining perturbations involving small molecule macrocycles and mutations of proline in proteins. High accuracy was obtained using the method for most of the perturbations studied. Unlike other methods that rely on λ-dependent functional forms for bond interactions, the method presented here can be employed using modern MD software without modification of codes or force field functions.

Introduction

During the process of drug discovery and lead optimization, changes in the scaffold of a candidate compound are frequently performed by medicinal chemists in order to enhance the binding affinity and improve the drug like properties. These include ring opening and closure, as well as changes in ring and linker length. These changes are commonly referred to as scaffold-hopping. In addition to the improvement of pharmaceutical properties, scaffold hopping is used to expand patentable space[1]. Scaffold hopping is not limited to small molecules and can also be found in the building blocks of biomolecules; mutations involving proline are a notable and important example. The ring topology of the proline sidechain significantly restricts the backbone conformations of proteins and eliminates a backbone hydrogen bond donor, effects that make proline a disruptor of both α-helices and β-sheets. This makes proline substitutions a popular probe for studying the kinetics and thermodynamics of protein folding, binding and aggregation[2-7]. Mutations involving proline are also widely employed during protein design for shaping proteins into desired geometries and for modulating the thermodynamic/kinetic properties of the designed proteins[8-10].

In silico methods of varied accuracy and efficiency have been developed to aid the screening of molecules and reduce efforts in wet labs[11-15]. These are now a key part of the drug development process. These methods include both data-based machine learning or artificial intelligence models, and rule-based physical models. Among these, alchemical free energy calculations are believed to be capable of delivering highly accurate predictions of binding affinity[16-17]. One popular variant of the alchemical free energy method is the relative binding free energy (RBFE) calculation used for comparing the binding affinities between a pair of candidate compounds sharing some common chemical groups. This method can minimize the thermodynamic noise by limiting the perturbations to small portions of the two compounds[18-19]. In the past few years, RBFE calculations have shown high accuracy in benchmarking and validation studies[18, 20-23]. Moreover, more and more examples have been reported in which RBFE calculations have made a positive impact on real drug-discovery projects in an industry setting[24-26]. Broadening the impact of RBFE

calculations by improving their range of applicability is highly desirable. In particular, improved methodologies for handling scaffold hopping are desired.

RBFE calculations compute the free energy changes using alchemical Hamiltonians, in which the change between initial and final compounds is described as a function of a variable λ; λ=0 corresponds to the initial compound and λ =1.0 corresponds to the final compound. Conformational sampling at steps along λ is performed using molecular dynamics (MD) simulations; in principle, RBFE calculations can better evaluate two compounds with different degrees of flexibility in their scaffolds compared to other fast methods that rely solely on static structures. However, RBFE calculations that involve breaking and forming of covalent bonds were considered to be infeasible, thus the application of RBFE calculations has been limited to so-called R-group perturbations and heterocycle replacements, in which the forming and breaking of covalent bonds are not conducted. The problem arises from the quadratic form of the bond interactions in molecular mechanics (MM), whose energy increases drastically as the bond distance moves away from the equilibrium. Thus, if a bond breaking process takes place from λ=0 (bond present) to λ=1 (bond absent), the sampled conformations at λ=1 may have extremely high potential energy under Hamiltonians at other values of λ. This can cause slow convergence when estimating free energy changes using thermodynamic integration or the Bennet acceptance ratio, as well as rounding errors in computation. To overcome this issue, Wang et al. developed a λ-dependent bond interaction functional form, which is referred to as the soft-bond potential[27]. The method inherits the philosophy of the softcore potential for nonbonded interactions[28] and efficiently avoids high energy at extreme bond distances. The λ-dependent bond interaction approach has shown successes in perturbations involving ring opening/closure, ring contraction/expansion and macrocyclization of linear compounds[27, 29]. However, implementation of this method in MD simulation packages requires modification of the code handling the bond interactions, which is beyond the experience of many RBFE users and limits potential applicability.

This paper describes a novel method that utilizes auxiliary restraints to enable the breaking and forming of covalent bonds in RBFE calculations. The main purpose of the auxiliary restraints is to temporarily keep the atoms at the ends of the modified bond near their

equilibrium distance during the breaking and forming of the covalent bonds. This method can be employed using modern MD software without modification of codes because the auxiliary restraints use the most basic functional forms in molecular mechanics. For most of the perturbation pairs studied here, our method combined with the GAFF2 force field[30] achieved equivalent accuracy and reliability compared to the soft-bond method[27, 29] with OPLS3 force field[31]. The equivalence of the applicability of the new method and the soft-bond method is demonstrated by conducting multiple transformations involving ring opening/closure, ring and chain contraction/expansion for small molecules as well as transformation involving macrocycles and non-proline to proline mutations in proteins.

Methods

In this section, the details for ring opening and closure, ring or linker contraction and expansion are described first, followed by a description of the system setup and detailed simulation methods.

Notation used to identify compounds used as test cases. A number of examples have been taken from the literature to provide tests for the new method. We utilized the same numbering employed in the original publications which described these compounds for clarity. In some cases, the same number has been used in different publications for different compounds. In these cases, a Roman numeral is appended to the numerical identifier.

Ring opening/closure

For a ring opening scenario, after identifying the ring topology in ligand L0 that is absent in ligand L1, a bond in L0 is selected such that removal of the bond will result in a bonding topology most similar to the bonding topology of L1. The next step is to apply dihedral restraints on the atoms of the ring that will be opened. For an N-member ring, N-3 auxiliary dihedral restraints are applied on the ring atoms. For example, in order to break the bond between atoms a1 and a6 in a 6-membered ring (**Fig. 1**), three dihedral restraints are applied on the ring atoms a1-a2-a3-a4, a2-a3-a4-a5 and a3-a4-a5-a6. The strength of the dihedral restraints was chosen to be 10 kcal/mol and the reference angle of each dihedral restraint was obtained from the corresponding values in the equilibrated structure of L0. With these dihedral restraints, the selected bond can be broken as the auxiliary restraints keep the

distance between atom a1 and a6 near the reference distance of bond a1-a6. The dihedral restraints are released after the selected bond is fully broken. Applying the ring opening process in reverse order results in a ring closure process. During ring closure transformations, the dihedrals may need to be rotated under the auxiliary restraints so the atoms at the ends of the forming bonds adopt values near those with the bond present. Once the ring opening or ring closure process has been finished, additional chemical differences between L0 and L1 may still remain. Traditional R-group RBFE calculations can then be applied to fully transform L0 into L1. The thermodynamic cycle in **Fig. 2** illustrates the steps required to obtain the total free energy change of a ring opening transformation.

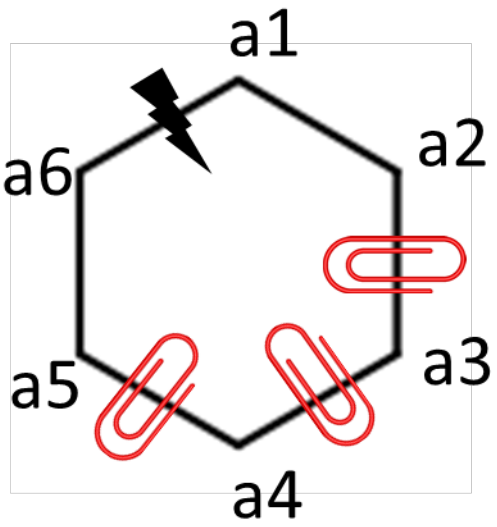

Figure 1. **An illustration of the process of ring breaking.** A 6-membered ring with a lightning bolt indicating the bond to be broken and red clips indicating the dihedrals to be restrained is shown.

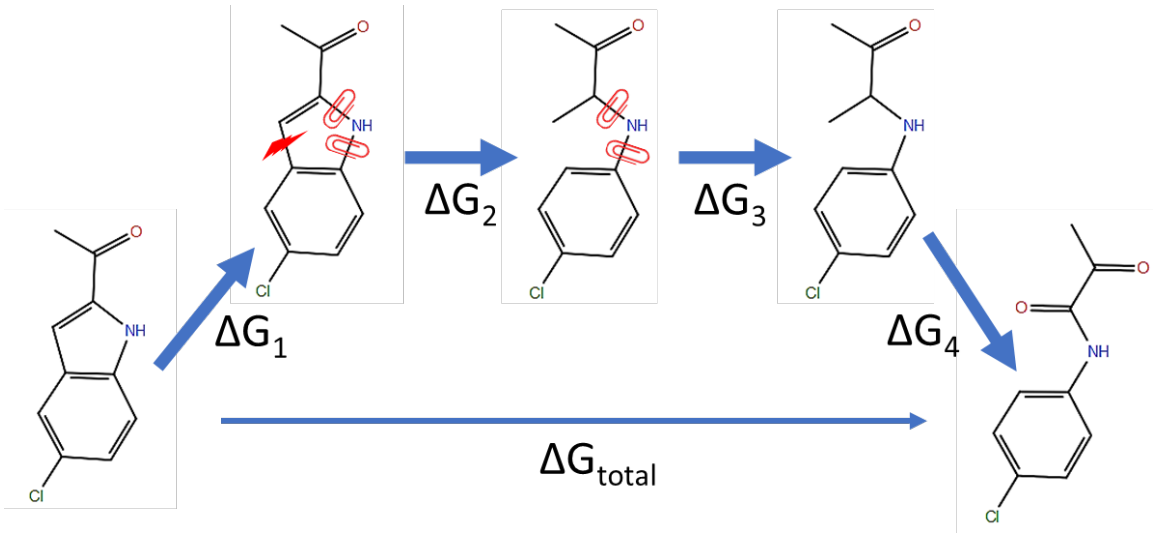

Figure 2. A thermodynamic cycle for a sample ring opening transformation. ΔG_1, ΔG_2, ΔG_3 and ΔG_4 are the free energy changes of applying the auxiliary restraints, breaking the covalent bond, releasing the auxiliary restraints and completing any remaining chemical differences, respectively. $\Delta G_{total} = \Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_4$.

Ring and linker contraction/expansion

Ring contraction/expansion and linker contraction/expansion are treated similarly as they both involve the removal of atoms flanked by other atoms. **Fig. 3** shows an example for the contraction of a ring and a linker, in which the transformation from ligand L0 to L1 requires changing a topology of -a1-a2-a3- into a topology of -a1-a3-. This transformation requires the removal of atom a2 between atom a1 and a3 and the formation of bonded interactions between atom a1 and a3. The non-bonded interactions involving atom a2 and the hydrogen atoms attached to atom a2 are turned off first (ΔG_1). In the next step (ΔG_2), all bonded interactions involving atom a2 are removed and the bonded interactions involving both atom a1 and a3 are turned on. Simultaneously, all bond and angle restraints involving atom a2 and the hydrogen atoms attached to atom a2 are turned on and the associated free energy change calculated. No dihedral restraints are needed in this case. The participants, strength and reference distances/angles for the bond and angle restraints are listed in **Table 1**, which are the same for ring contraction and linker contraction. None

of the restraints in **Table 1** need to be turned off after the contraction as the restraints can be considered equivalent to the bonded interactions connecting dummy atoms in the end states of regular R-group RBFE calculations. These steps complete the contraction process and give a final state of the contracted ring or linker as illustrated on the right side of **Fig. 3**. Ring or linker expansion can be accomplished by applying the previous steps in reverse order.

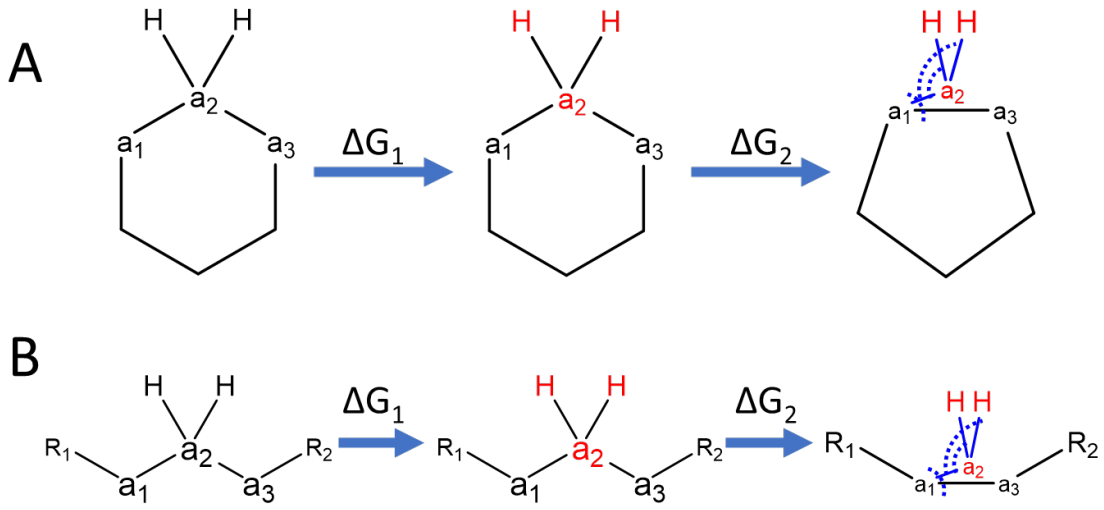

Figure 3. Illustration of the concept of A) ring and B) linker contraction transformations using auxiliary restraints. ΔG_1 represents the free energy change of removing the non-bonded interaction of atoms a2 and H. ΔG_2 represents the free energy change of applying the restraints listed in **Table 1**. Atoms shown in red are dummy atoms with no nonbonded interactions. The blue lines indicate the auxiliary bond restraints applied on a1-a2 and H-a2. The dashed arcs indicate the auxiliary angle restraints applied on a1-a2-a3 and H-a2-a1.

Table 1. The participants, strength and reference distances/angles for the bond and angle restraints used for ring and linker contraction transformations.

Bond/Angle	k	r_0/θ_0
a1- a2	100 kcal/mol/Å^2	half of the reference distance of bond a1-a3 in L1
H-a2	Same as the k of the native H-a2 bond	Same as the r_0 of the native H-a2 bond
a1 – a2 – a3	100 kcal /mol/rad^2	π rad (180˚)
H- a2 – a1	50 kcal/mol/rad^2	$\pi/2$ rad (90˚)

Protein and ligand setup for test cases

Small molecules were parameterized using GAFF2[30]. Partial charges were assigned using the AM1-BCC method[32]. The TIP3P water model[33] was used to solvate the molecules. The ff14SB force field[34] was used to parameterize the protein. For compounds listed in **Table 2 & 3**, the ligand poses were adopted from the published soft-bond method[27]. For compounds listed in **Table 4**, the ligand poses were manually constructed using the X-ray structures listed in **Table S1** as templates. The manually constructed ligands poses were briefly energy minimized by using RDKit[35] (v 2020.09.1) before parameterization. If present, ions were removed from the experimental structures. Missing hydrogens of the protein were added uingMolprobity[36]. The rotamer states of Asn/Gln/His were adjusted by following the suggestions of Molprobity. The protein/ligand complexes and free ligands were solvated in truncated octahedron water boxes with initial buffer sizes of 8 and 15 Å respectively. Systems were neutralized with minimal numbers of Na+ or Cl- ions.

We opted to use the publicly available GAFF2 force field[30] with AM1-BCC charges[32, 37] to parameterize the small molecule compounds. This will allow our method to be compared with other scaffold hopping methods under the same force field in the future. The force field parameters in Amber format can be found in SI. However, the method is not restricted to the use of this force field. Our in-house force field, XFF, was employed to recalculate transformations that showed large errors using the GAFF2 force field.

Full coordinate sets in PDB and SDF format for all the protein receptors and ligands studied here in their bound conformations are provided in the SI (Table S1). The 2-D structures of all studied molecules can be found in **Fig. S1**.

General simulation details

All simulations, including the equilibration and production runs, used a Langevin integrator with a 2 fs timestep and a friction coefficient of 2 ps^{-1}. Bonds to hydrogens were constrained via SHAKE[38], except when the bond connects a real atom and a softcore atom. Smooth particle mesh Ewald electrostatics with an 8 Å direct space cutoff was used[39]. The

same cutoff distance was used for the hard truncation of Lennard-Jones interactions, which has a long-range continuum correction on the dispersive term.

Equilibration

Initial structures of the complex and the free states of L0 and L1 were minimized using 100 steps of steepest descent plus 100 steps of conjugate gradient. During the minimization, Cartesian restraints were applied to all non-solvent heavy atoms with a strength of 10 kcal/(mol*Å^2). The systems were sequentially heated at fixed volume from 100 K to 298 K with an increase of 20 K every 10 ps. Cartesian restraints with a strength of 4 kcal/(mol*Å^2) were applied to all non-solvent heavy atoms. A 0.5 ns constant pressure simulation at 1 atm was carried out to equilibrate density and gradually release the Cartesian restraints by reducing the restraint force constant by 0.2 kcal/(mol*Å^2) every 0.025 ns. The pressure was regulated using the Monte Carlo barostat[40] with pressure coupling constant set to 0.2 ps^{-1}.

Topology and coordinate for the RBFE calculations were constructed by appending the appearing atoms of the equilibrated L1 to the structures of equilibrated L0 using Z-matrix by tLEaP. The appended structure was used for all the λ windows. Under each λ window the appended structures were briefly minimized using 100 steps of steepest descent. The systems were then heated to 298K over 50 ps under constant volume. Cartesian restraints with a strength of 5 kcal/mol*Å^{-2} were applied to all non-solvent heavy atoms during the minimization and heating. After heating the systems, constant pressure simulations at 1 atm and 298K with a length of 40 ps were used to further equilibrate the volume without any Cartesian restraints. The pressure was regulated by the Monte Carlo barostat[40] and the pressure coupling constant was set to be 0.2 ps^{-1}.

RBFE calculations

The softcore vdW potential was applied to all atoms that disappear or appear during the free energy calculations[41]. The value of α in the softcore vdW potential, which is the softcore radius, was set to be 0.5 Å. The auxiliary restraints were applied when calculating the ΔG for the protein/ligand complexes as well as the ΔG for the free ligands.

Hamiltonian replica exchange (HRE) between adjacent λ windows was used here to

facilitate the convergence of the calculations[42], but is not required for use of the new method. The number and spacing of λ windows were set to satisfy an exchange success rate of >15% for any two adjacent λ windows. For each λ window, the length of production runs was 5 ns and the number of exchange attempts was 4000. For each λ window at every exchange attempt, the potential energy of the coordinates under the Hamiltonian of the λ window were collected, as well as the potential energies of these coordinates under the Hamiltonian of all other λ windows. From the collected potential energies, the free energy change ΔG was computed by using multi-state Bennet acceptance ratio (MBAR)[43-44]. Three independent HRE simulations with different initial velocities were conducted to obtain the standard deviation of $\Delta\Delta G_{binding}$.

Software packages and example setup

The method was integrated into the XFEP platform[45] to carry out the free energy calculations presented in the main text. Use of other simulation packages should be straightforward, since customized bond interaction functional forms are not required. The auxiliary restraints can be applied as modified dihedral parameters. To illustrate this, a ring opening (CHK1 **20→17**), a ring contraction (ERα **3b→2d**) and a chain contraction (CatS **35→132**) FEP calculation were conducted using the standard Amber18[46] force field as examples. The calculated ΔΔG values using the Amber18 force field for these three example transformations are presented in **Table S2**. The standard Amber18 force field reproduced the ΔΔG values calculated using the XFEP platform for all three example transformations. More details as well as the input and output files are provided in the SI, in order to enable the method to be evaluated using publicly available software packages.

Results

We validated our new scaffold hopping method by calculating the relative binding free energy changes resulting from transformations involving ring opening/closure as well as ring and linker expansion. The validation systems include well documented small-molecule compounds that have been studied in previous publications using the soft-bond potential[27, 29]. Additionally, a set of protein-protein interactions involving proline

replacements were used to validate the application of the methodology to protein. In particular, the changes in binding free energy between the turkey ovomucoid third domain (OMTKY3) and its binding partners caused by non-proline to proline mutations were calculated using the new method. This is an excellent model system as high resolution structures of the free proteins and the bound state are available and careful thermodynamic measurements have been reported[47-50]. Firstly, we summarize the key features of the new method, then we describe the model systems and present results.

The main challenges and the theoretical background involved in the breaking and forming bonds in free energy calculations are well documented[27]. In short, the problem originates from the functional form of the bond interaction in MM, where the energy increases drastically as the bond distance deviates from the reference value. When a covalent bond is broken as λ goes from 0 to 1, the strength of the bond will gradually be turned off. At λ=1, with zero force constant, the two atoms in the bond can move far apart, which has extremely high bond energy under Hamiltonians (H) at other λ values with non-zero force constants. For thermodynamic integration, $\partial H/\partial\lambda$ values (equal to $H_{\lambda=1}$ - $H_{\lambda=0}$ with linear mixing of $H_{\lambda=1}$ and $H_{\lambda=0}$) are collected at each λ to compute the integration. Thus, $\partial H/\partial\lambda$ values can exceed the optimal numerical precision range and exhibit high fluctuations at λ=1. For MBAR calculations, this leads to poor phase space overlap as coordinates sampled at λ=1 have extreme energy at other λ values. The soft-bond method utilizes a λ-dependent functional form to prevent extreme values of the bond energy when evaluating coordinates from λ=1 using Hamiltonians at other λ values, including λ=0. This is similar to the softcore potential for turning on and off the van der Waals and electrostatic interactions during free energy calculations[28].

In contrast to the soft-bond method, our method does not require modification of the functional form of bonded interactions. For ring opening transformations several dihedrals restraints are applied on the ring topology to hold the two atoms at the ends of the bond in place and prevent the bond length deviating from the reference distance when breaking the bonded interaction. This prevents extreme values of bond energies that can arise evaluating coordinates sampled at λ=1 using Hamiltonians at λ=0 during the breaking of a bond interaction. After the bond interaction has been removed, the dihedral restraints are also

removed and the associated free energy contribution is accumulated in the overall transformation. The dihedral interactions in MM have a cosine functional form ($k*\cos(n*\theta-\phi)$), which is a periodic function with a maximal difference of 2k between the energy minima and maxima. Thus, the removal of the dihedral restraints in free energy calculations is a much milder transformation than the removal of bond interactions with a quadratic functional form. Ring closure transformations can be accomplished by reversing the steps of ring opening transformations.

Our method is conceptually similar to the attach-pull-release method used for calculating absolute binding free energies between receptors and ligands[51]. Both methods utilize angle and dihedral restraints to minimize the fluctuation of free energy changes during the more challenging parts of the transformation.

As an example of the application of auxiliary restraints to ring opening, we show the effect of the restraints on bond removal using the transformation of compound **20** to **17** (**Fig. 4A)**. A bond connecting atoms C and O was selected to be broken in the ring opening transformation. The bond interaction between the C and O atoms is linearly scaled by λ. The bond interaction is intact when $\lambda = 0.00$ and is removed when $\lambda = 1.00$. We compared the distance distribution between the atoms C and O of **20** at λ=0.00, 0.80, 0.95, 0.99 and 1.00 with dihedral restraints of k = 10 kcal/mol (**Fig. 4B**) and without the dihedral restraints (**Fig. 4C**). The distance distribution overlaps are similar among λ=0.00, 0.80, 0.95, 0.99 with the dihedral restraints and without the dihedral restraints. However, under no dihedral restraints (**Fig. 4C**), a significantly poorer overlap between $\lambda = 0.99$ and 1.00 was obtained compared to the overlap obtained with the dihedral restraints (**Fig. 4B**).

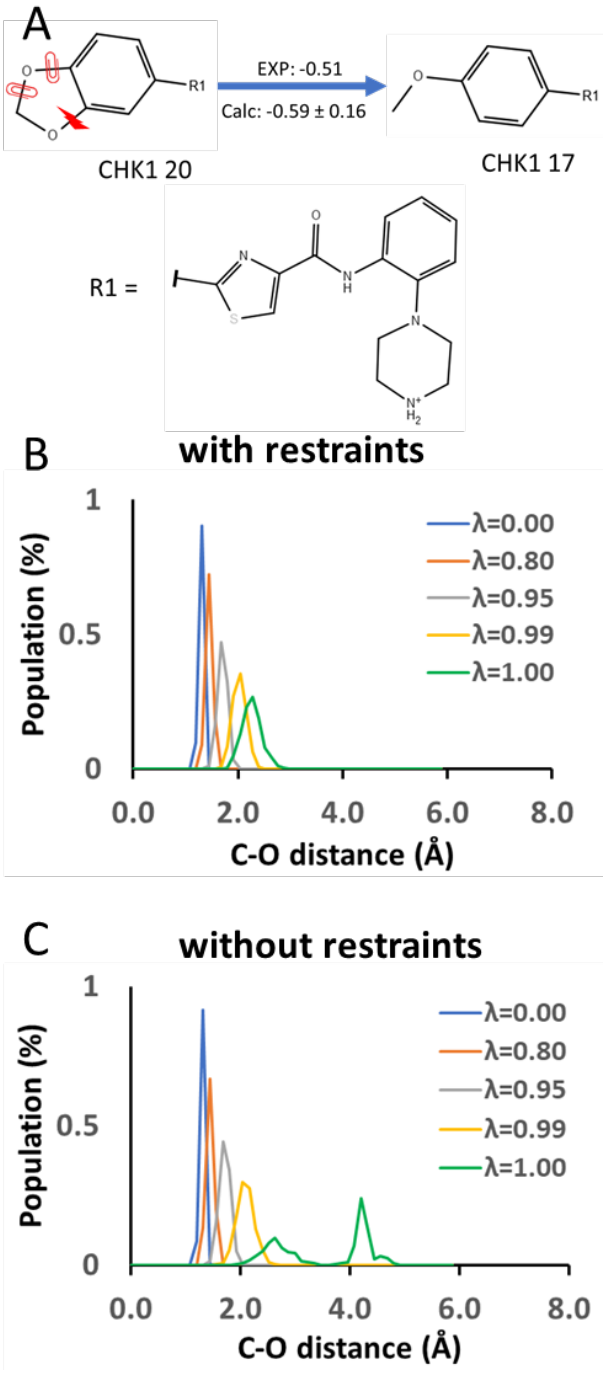

Figure 4. A) An example of the ring opening transformation between ligands **20** and **17** of Chk1. The red lightning bolt indicates the bond being broken during the ring opening transformation, and the red clips indicate the auxiliary dihedral restraints. B) Distance distribution between atoms C and O at the ends of the breaking bond under dihedral restraints of k= 10 kcal/mol at different λ values. C) Distance distribution between atoms

C and O at the ends of the breaking bond under no dihedral restraint at different λ values. The bond interaction is scaled linearly by λ.

We tested our method on the perturbation pairs studied previously using the soft-bond method[27]. Indexing of the compounds was adopted from the citations associated with each protein target. Here, we provide a brief summary of these model systems and compounds used for the RBFE calculations.

Applications to ring opening/closure

Checkpoint kinase 1 (CHK1) is a key participant in the S and G2/M checkpoints and is critical for the survival of cancer cells with p53 mutations[52]. A series of compounds, including compounds **17**, **19**, **20** and **21,** were designed to improve the affinity to CHK1 and the selectivity against CDK2(cyclin-dependent kinase 2) of the initial hit compound (**1-i**)[53].

Factor Xa (FacX), which converts prothrombin to thrombin in the coagulation cascade, is a target enzyme for treating thromboembolic disease. Edoxaban and the two other candidates proposed during drug development, **4c** and **4d**, all bind to the S1 and the S4 pockets of FacX[54]. Edoxaban exhibits both higher affinity to FacX and anticoagulant activity than **4c** and **4d**[54-56].

β-tryptase (TPSB2) is reported to be strongly correlated with inflammatory and allergic disorders[57]. Although compound **2** has a lower affinity to TPSB2 compared to compound **1-ii**, compound **2** has significantly less off-target binding, which was believed due to the higher rigidity of the tropanylamide scaffold of compound **2** than the piperidinylamide scaffold of compound **1-ii**[58].

Beta-secretase 1 (BACE-1) is a promising target for treating Alzheimer's disease[59]. Knockout of BACE-1 completely abolishes the generation of amyloid β-peptides and reverses the cognitive decline in a mouse model[60-61]. Compound **7** is a BACE-1 inhibitor with a bridged ring scaffold which is intentionally designed to be more conformationally

constrained than the lead compound **6** which has a bicyclic ring scaffold[62]. Compound **31** was designed to achieve a higher *in vivo* brain penetration.

RBFE calculations of these compounds can be accomplished by ring opening/closure transformations, which serve as good examples to validate our method. The RBFE values calculated for these systems, using the auxiliary restraints method with the GAFF2 force field, generally show good correlation with the experimentally measured ΔΔG values (**Table 2)**. The mean unsigned error (MUE), R^2 and p-value are 0.63 kcal/mol, 0.66 and <0.01 which are comparable to the results obtained using the soft-bond method with the OPLS3 force field.

Table 2. Comparison of the ΔΔG values for the ring opening/closure transformations calculated using the auxiliary restraints method, calculated using the soft-bond method, and measured by experiments. Units are kcal/mol. "Edo" signifies the compound edoxaban.

Pairs	Auxiliary restraints/GAFF2	Soft-bond/OPLS3[27]*	Experimental[53-54, 56, 58]	Auxiliary restraints/ XFF
CHK1 21 →19	-0.23±0.27	0.03	0.59	
CHK1 21→17	-1.10±0.32	0.22	-0.57	
CHK1 1→19	0.35±0.53	0.95	1.15	
CHK1 20→17	-0.59±0.16	-0.03	-0.51	
CHK1 1-i→17	-0.39±0.12	0.70	-0.02	
FacX edo→4d	0.88±0.06	1.69	0.87	
FacX edo→4c	0.93±0.46	1.48	0.8	
TPSB2 2→1-ii	-2.39±0.21	-0.16	-0.62	-0.28±0.07
BACE-1 7→6	-1.48±0.35	-0.67	-0.12	
BACE-1 7→31	-1.90±0.53	-1.24	-0.64	
MUE/R^2/p-value	0.71/0.75/p<0.01	0.54/0.62/p< 0.01		

*: the reported ΔΔG with cycle closure corrections were shown. The reported uncertainties for the soft-bond/OPLS3 method are not shown, as those reported uncertainties were calculated as BAR errors and cycle closure errors rather than independent runs as calculated for our data.

Some transformation pairs resulted in larger errors. Significant deviations (>1 kcal/mol) between the ΔΔG values calculated using our method and measured by experiments were found for pairs TPSB2 **2→1-ii**, BACE-1 **7→6** and BACE-1 **7→31**.

For BACE-1 **7→6** and BACE-1 **7→31**, the sidechain rotamer state of Y120 depends on whether the core contains the bridged ring of ligand **7** or the bicyclic rings of ligand **6** and **31**, which can be observed in PDB 4ZSQ and 4ZSP (**Fig. S2**). Besides the sidechain of Y120, the backbone traces of the β-sheets harboring Y120 do not align well between the bridged ring complex (PDB 4ZSQ) and the bicyclic ring complex (PDB 4ZSP). Thus, accurate modeling of the perturbations of BACE-1 **7→6** and BACE-1 **7→31** requires sufficient sampling of the protein conformational changes of Y120 and its neighboring residues. Even if a sufficient sampling of the conformational changes can be achieved using enhanced sampling methods, the accuracy of the calculated ΔΔG values for the perturbations will still depend on whether the protein force fields can accurately describe the energetics of the two conformations observed in 4ZSP and 4ZSQ. For these reasons, we suggest avoiding BACE-1 **7→6** and BACE-1 **7→31** in future benchmarking of scaffold hopping methods, as the complications prevent a fair evaluation of the scaffold hopping methods themselves.

For TPSB2 **2→1-ii**, the ΔΔG value was recalculated using ligands parameterized by the XFF force field instead of GAFF2 (**Table 2**). Using the XFF force field significantly decreases the error of the ΔΔG between TPSB2 **2** and **1-ii**, which indicates that the GAFF2 force field may be inadequate for modeling the binding free energy difference between TPSB2 **2** and **1-ii**. By visually examining the conformational ensemble of TPSB2 **1-ii & 2** in the unbound state, we found that the conformations of TPSB2 **1-ii & 2** given by the GAFF2 and XFF force fields mainly differ at the planarity of the thiophene-2-carboxamide. The GAFF2 force field prefers a more planar conformation of the thiophene-2-carboxamide, while the XFF force field prefers a more perpendicular conformation. The more perpendicular conformation given by the XFF force field is consistent with the conformation of **2** observed in the X-ray structure of **2** co-crystalized with TPSB2[58]. TPSB2 **2** may experience more internal steric clashes in the perpendicular conformation, which explains why TPSB2 **2** is overly favored in the unbound state when using the GAFF2 force field.

Expansion/contraction of ring and linker

Contraction of rings and linkers involves the removal of atom a2 and formation of a bond between atom a1 and a3 in a bond topology of -a1-a2-a3- (**Figure 3**). We believe that the choice of location to restrain the atom a2 after the contraction is critical for ring and linker contraction transformations in free energy calculations. To prevent extreme bond distances between a1 and a2 as well as between a2 and a3, a2 must be kept within the proximity of both a1 and a3 after the contraction. We chose to keep a2 at the exact center between a1 and a3 after the contraction. We believe that placing a2 at the center between a1 and a3 increases the symmetry of a2 and reduces the available conformational space of a2, which also may benefit the convergence of the free energy calculations involving chain and linker contractions. The expansion of rings and linkers can be accomplished by reversing the contraction transformations of rings and linkers.

We tested our ring contraction/expansion method on perturbation pairs studied previously by using the soft-bond method[27], involving the estrogen receptor subtype alpha (ERα). ERα is expressed in many cells and tissues with key roles in physiological function, which makes it a popular target for treating various diseases[63]. A series of compounds, including compounds **3b**, **2e** and **2d**, were designed to improve the selectivity against ERβ of the lead compound SERBA-1[64]. The strategy for a representative ring contraction using auxiliary restraints is shown in **Figure 5**.

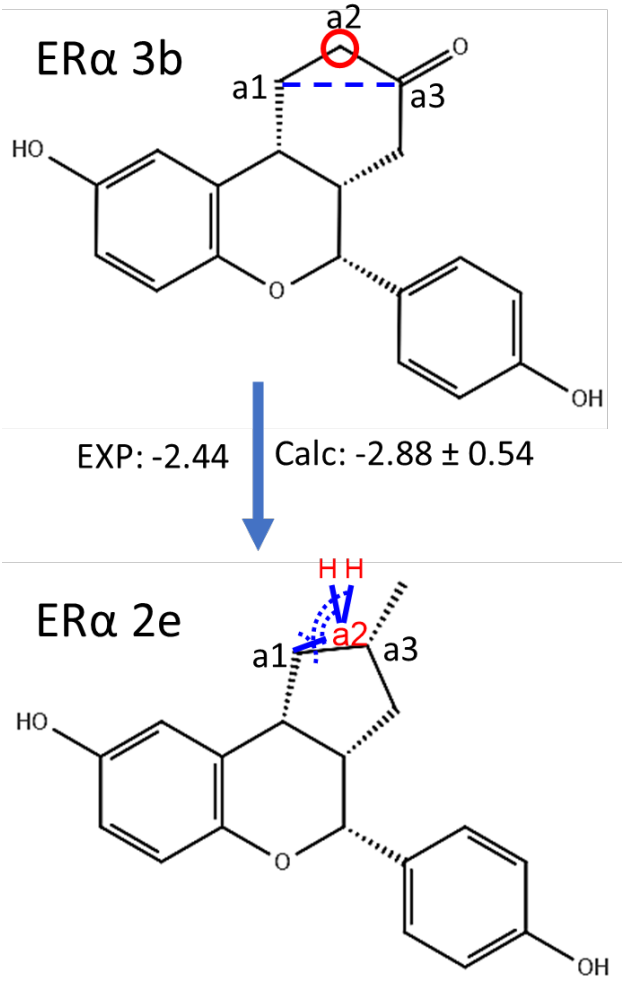

Figure 5. An example illustrating the ring contraction transformation from Erα ligands **3b** to **2d** studied using the auxiliary restraints method. The atom to be removed from the ring (a2) is indicated by a red circle, and the bond to be formed (a1-a3) is indicated by a blue dashed line. The blue lines indicate the auxiliary bond restraints applied on a1-a2 and H-a2. The dashed arcs indicate the auxiliary angle restraints applied on a1-a2-a3 and H-a2-a1. Chirality is indicated using the wedge-dash notation. Further changes are required in addition to the ring contraction.

Cathepsin S (CatS) mediates the cleavage of major histocompatibility class II (MHC-II) associated invariant chain (Ii), which is crucial in the initiation of MHC-II related immune response to an antigen. Inhibition of CatS can treat various autoimmune disorders and

inflammatory diseases[65]. A perturbation between a pair of Cathepsin S (CatS) ligands (compound **35** and **132**) was also included here, which involve contraction of the linker between the core and the thiophene group (**Fig. 6**). To our knowledge, examples for handling expansion and contraction of linkers in free energy calculations have not been published previously.

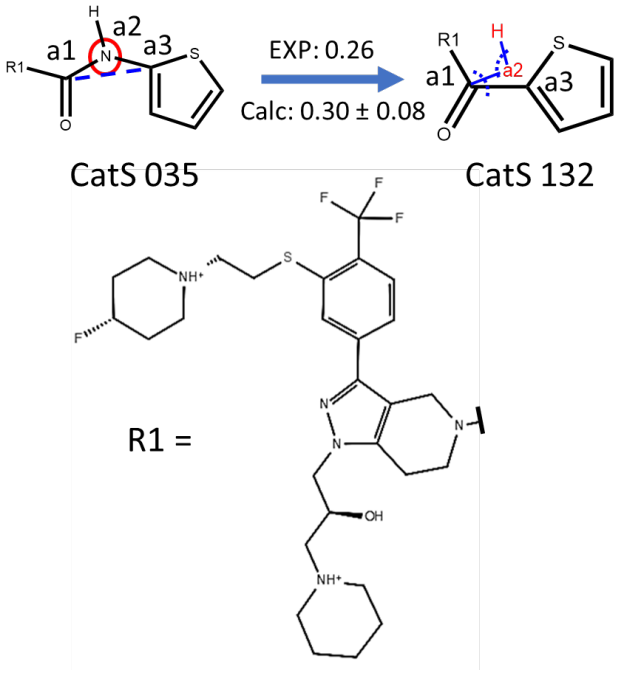

Figure 6. An example illustrating the linker contraction transformation from CatS ligands **35** to **132** studied using the auxiliary restraints method. Atom (a2) to be removed from the linker is indicated by the red circle and the bond (a1-a3) to be formed is indicated by the blue dashed line. The blue lines indicate the auxiliary bond restraints applied on a1-a2 and H-a2. The dashed arcs indicate the auxiliary angle restraints applied on a1-a2-a3 and H-a2-a1. Chirality is indicated using the wedge-dash notation. Units: kcal/mol.

Calculated and experimental RBFE values for the ring and linker contraction/expansion transformations are provided in **Table 3**. For all of these transformations, the auxiliary restraints method and the GAFF2 force field show errors under 0.5 kcal/mol, which is comparable to the soft-bond method with the OPLS3 force field.

Table 3. Comparison of the ΔΔG values for the ring and linker contraction/expansion transformations calculated using the auxiliary restraints method, calculated using the soft-bond method and measured by experiments. (Units: kcal/mol)

Pairs	Auxiliary restraints/GAFF2	Soft-bond/OPLS3[27]*	Experimental[64, 66-67]
CHK1 21→20	-0.04±0.15	-0.19	-0.07
ERα 3b→2d	-1.34±0.40	-1.45	-1.78
ERα 3b→2e	-2.88±0.54	-2.80	-2.44
CatS 35→132	0.30±0.08		0.26
MUE/R^2/p-value	0.24/0.94/0.03	0.27/0.92/0.17	

*: The reported ΔΔG with cycle closure corrections are shown. The reported uncertainties for the soft-bond/OPLS3 method are not shown here as the uncertainties, as those reported uncertainties were calculated as BAR errors and cycle closure errors rather than independent runs as calculated for our data.

Macrocyclization, ring-size changing and rigidification of macrocycles. Macrocyclization of drug compounds is frequently used to improve binding affinities of acyclic compounds and can be considered as a special case of scaffold hopping[68-71]. The binding affinity and drug likeness of macrocyclized compounds often can be further improved by rigidifying the cyclic scaffolds and changing the ring sizes of the macrocycles[69-74]. The ability to model these changes is an important part of any *in silico* method. Consequently, the performance of the auxiliary restraints method was also tested on transformations that involve macrocyclization of linear compounds, changing ring-size of macrocycles and rigidification of macrocycles by adding sub-rings onto the primary rings. All the transformations were accomplished using the method described above for ring opening/closure and expansion/contraction.

We chose several model systems which have been previously studied using the soft-bond method[29]. A brief description of the test compounds follows: The down-regulation of casein kinase 2 (CK2), a serine/threonine kinase, has been shown to decrease proliferation and increased apoptosis of cancer cells[75]. The macrocyclic compound **2** has a significantly lower binding affinity to CK2 compared to the acyclic compound **1**[71], which is contradictory to expectation. However, compound **2** shows significantly enhanced cellular activity due to its higher membrane permeability. The macrocyclization of BACE-1 compound **3** results in compound **4**, which shows a higher binding affinity[74]. Compound

19 binds to BACE-1 with a hairpin shape. A series of macrocycles were designed based on this structure, including compounds **20**-**23**, by linking the two ends of compound **19**[73].

MTH1 is a member of the Nudix phosphohydrolase superfamily of enzymes, which hydrolyses oxidized purines and prevents their incorporation into DNA[76]. The macrocyclized compound **7** shows a significantly improved binding affinity to MTH1 compared to the two acyclic compounds **5** and **6**[69]. However, inhibition of MTH1 did not display any significant suppression of cancer cells[69].

Compound **8** to **11** are macrocyclic CHK1 inhibitors which mainly differ at the ring size. Synthesis and study of compounds **8** to **11** were meant to find the optimal ring size of the macrocycles[70].

Cancer cells can adapt to solo target inhibition by up-regulating alternative pathways. Inhibition of chaperon proteins, such as heat shock protein 90 (HSP90), disrupts the function of a wide range of client proteins, which can eliminate alternative pathways for cancer cell survival[77]. A series of macrocyclic compounds (**12-18**) were designed to improve the binding affinity to HSP90 of an initial lead compound with satisfactory pharmacokinetics[72].

Compounds **24** to **32** are from another series of macrocyclic HSP90 inhibitors which have significantly improved binding affinity compared to the acyclic compound **33**[68].

Macrocycles require more dihedral auxiliary restraints during the removal of bond interactions. It is possible that the length of the bond may still be too long to be broken due to the accumulation of small fluctuations of restrained dihedrals far along the chain from the bond being removed. To test this, the distance distribution between C-C for CK2 **1→2** was collected during the removal of the bond interaction under dihedral auxiliary restraints with k = 10 kcal/mol (**Fig. 7A**) at λ=0.00, 0.80, 0.95, 0.99 and 1.00. For the macrocyclic compound CK2 **2**, a wider distance distribution at λ = 1.00 (**Fig. 7B**) was obtained than the distance distribution (**Fig. 3B**) for CHK1 **20**, which has a small ring, under dihedral auxiliary restraints with the same strength. However, the auxiliary restraints still provide sufficient overlap of the distance distribution between C-C of CK2 **2** over the same set of

λ values, especially between λ = 0.99 and 1.00. In contrast, there is almost no overlap between the distance distribution for λ =0.99 and 1.00, when no dihedral restraints are applied (**Fig. 7C**).

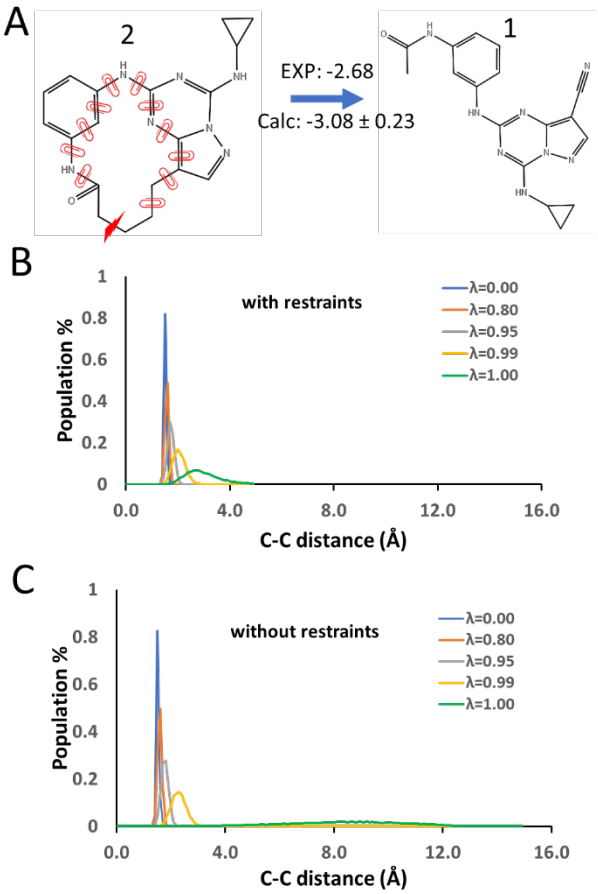

Figure 7. A) The ring opening transformation between macrocycles CK2 **2** and **1**. The red lightning bolt indicates the bond broken during ring opening transformation. B) Distance distribution between atoms C and C at the ends of the breaking bond under dihedral restraints of k= 10 kcal/mol. C) Distance distribution between atoms C and C at the ends of the breaking bond under no dihedral restraint. Bond interaction is scaled linearly by λ.

The resulting ΔΔG values are compared to those reported from calculations using the soft-bond method[29] in **Table 4**. For macrocyclization transformations, the auxiliary restraints method successfully predicted **1→2** to be the only pair with significantly weakened binding affinity after macrocyclization. Macrocyclization was predicted to have mild effect on the binding affinity of compound **19** and strong enhancement on the binding affinity of

compounds **3, 5, 6,** and **33**, which are consistent with the experimentally measured changes in affinity. Our method achieved an R^2 and p-value of 0.91 and $<10^{-3}$, which is the same as the reported results calculated using the soft-bond method. The MUE of the calculated ΔΔG values using our method is 0.88 kcal/mol, which is slightly higher than the MUE of the reported results calculated using the soft-bond method.

Table 4. Comparison of the ΔΔG values for macrocycles, calculated using the auxiliary restraints method, calculated using the soft-bond method and measured by experiments. (Units: kcal/mol)

Pairs	Auxiliary restraints/GAFF2	Soft-bond/OPLS3[29]*	Experimental[68-74]
Macrocyclization			
1→2	3.08±0.23	2.83	2.68
3→4	-3.11±1.03	-2.13	-2.09
5→7	-4.44±0.18	-6.55	-5.12
6→7	-5.69±1.52	-4.22	-4.14
19→22	-0.94±0.64	-0.64	0.61
19→23	-0.66±0.54	-0.43	0.25
33→30	-5.21±1.43	-6.04	-5.7
33→31	-5.46±0.76	-4.17	-5.89
MUE/R^2/p-value	0.88/0.91/$p<10^{-3}$	0.71/0.91/$p<10^{-3}$	
Addition of sub-ring			
16→13	-1.16±0.28	-0.81	0.87
16→15	-0.50±0.74	0.17	0.83
27→30	-0.06±0.36	-2.4	0.54
29→32	2.31±0.47	-0.51	1.91
MUE/R^2/p-value	1.09/0.74/p<1	1.93/0.19/p<1	
Contraction of ring size			
10→9	0.15±0.22	0.11	-0.09
11→10	0.09±0.37	-0.11	-0.82
11→9	-1.47±0.51	-0.01	-0.91
13→12	3.06±0.41	3.66	2.65
14→12	3.40±0.67	3.75	2.22
14→13	0.22±0.73	0.1	-0.43
22→20	1.39±1.06	1.02	0.39
22→21	-0.03±0.80	-1.72	-1.08
22→23	0.64±0.28	0.2	-0.36
25→24	-0.34±0.76	0.08	0.65
26→24	0.26±0.85	-0.28	0.21
26→25	-0.67±0.30	-0.35	-0.44
26→27	-1.70±0.97	-2.29	-2.49
28→27	0.71±0.77	-1.54	-1.04
29→28	-0.45±0.46	-0.43	0.11
31→30	-0.90±0.41	1.87	0.19
32→31	-0.49±0.59	1.78	0.19
8→9	-1.65±0.91	-1.31	-0.3
MUE/R^2/p-value	0.81/0.58/$p<10^{-3}$	0.74/0.78/$p<10^{-5}$	
Overall			
MUE/R^2/p-value	0.86/0.82/$p<10^{-11}$	0.89/0.78/$p<10^{-9}$	

***:** Data were taken from the 5ns simulations with cycle closure corrections. The reported uncertainties for the soft-bond/OPLS3 method are not shown, as those reported

uncertainties were calculated as BAR errors and cycle closure errors rather than independent runs as calculated for our data.

Our method achieved MUE and R^2 of 1.09 kcal/mol and 0.74, respectively, for the transformation involving addition of sub-rings, which is better than the reported results calculated using the soft-bond method. However, the ΔΔG value of **16→13** calculated using auxiliary restraints has an error of > 2.0 kcal/mol .

For transformation involving ring contractions, our method combined with GAFF2 achieved MUE, R^2 and p-value of 0.81, 0.58 and $< 10^{-3}$. Our method successfully predicted that ring contractions, like **13→12**, **14→12** and **26→27**, have dramatic effects on the binding affinities. However, our method appears less accurate than the soft-bond method in predicting the consequences of ring contractions involving smaller changes in binding affinities.

Though the overall correlation and MUE of our results are slightly better than the results calculated using the soft-bond method, it is important to note that macrocycles usually have high internal friction due to their cyclic nature and large sizes, which significantly dampens the sampling of their conformations in MD simulations. This can lead to issues in the calculation of ΔΔG values. For example, in previous studies of the same set of macrocycles using the soft-bond method[29], MD simulation lengths of 5 ns and 25 ns led to differences in the calculated ΔΔG of up to 2.1 kcal/mol. Thus, having macrocycles with initial conformations in high energy local minima affects both the accuracy and precision of the calculated ΔΔG, apart from issues with the quality of force fields and the method used for handling scaffold hopping [78]. Methods have been developed to aid the search of bioactive conformations of macrocycles, but they emphasize the search of conformations instead of giving thermodynamically equilibrated conformational ensembles[79-80]. For these reasons, we opted not to use the XFF force field to verify the source of errors for macrocycles whose calculated ΔΔG values significantly deviated from the experimental values. However, the results clearly showed that our method and the soft-bond method are of comparable applicability for studying various scaffold hopping transformations involving macrocycles[29].

Overall correlation for scaffold hopping in drug-like compounds

The correlation between the calculated (using auxiliary restraints) and experimental ΔΔG for all scaffold hopping transformations of drug compounds studied here is shown in **Fig. 8**. The correlation is excellent. For all the ring opening/closure, ring and linker contraction/expansion transformations of non-macrocyclic and macrocyclic compounds, the calculated and experimental ΔΔG values were fit to a trendline of y = 0.95*x - 0.23 with a p-value of $<10^{-15}$. The MUE, R^2 and Kendall's τ were 0.77 kcal/mol, 0.79 and 0.54 respectively. A similar comparison cannot be tested for the softcore bond approach due to the lack of data for many of the transformations calculated here.

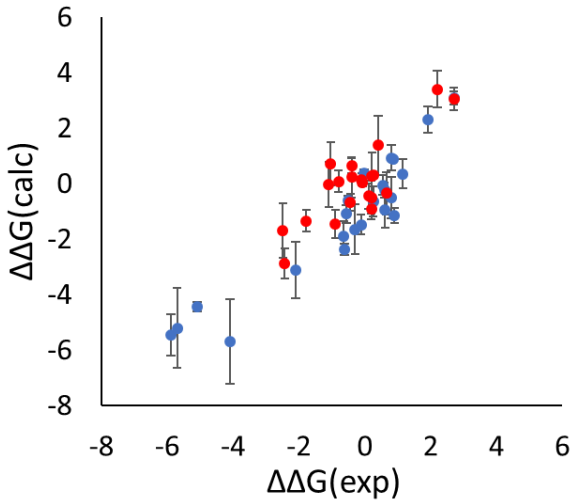

Figure 8. Comparison of the calculated and experimental ΔΔG for the ring opening/closure transformations (blue dots), the linker and chain contraction/expansion transformations (red dots). A total of 44 transformations were studied. Calculated values used the auxiliary restraints method. Units: kcal/mol.

Proline mutations

Correct handling of the ring topology of the amino acid proline is critical for modeling mutations involving proline in free energy calculations. The gain and loss of the backbone conformations will not be accounted for in free energy calculations if the ring topology of

proline remains intact or is absent during the perturbation. Here, proline to non-proline mutations and non-proline to proline mutations were treated as ring opening and closure transformations respectively using the auxiliary restraints.

We opted to validate our method by calculating the protein-protein binding free energy changes caused by proline mutations rather than calculating the thermostability change induced by proline mutations. The reason is that thermostability changes involve modeling both the folded and unfolded state of proteins. It is still infeasible to rigorously model the unfolded state of proteins in explicit water simulations as the unfolded state is highly expanded and dynamic. Though modeling the unfolded state using a short peptide model has shown successes in some cases[81-83], the short peptide models will be inadequate if the unfolded state is highly structured and has significant long-range interactions, which is common for many proteins[84-88].

The effect of Leu18-to-Pro mutations on the binding of OMTKY3 to four of its receptors, bovine chymotrypsin Aa[47] (CHYM), *Streptomyces griseus* proteinase B[48] (SGPB), human leukocyte elastase[49] (HLE) and subtilisin Carlsberg[50] (CARL), were calculated using the auxiliary restraints method. The OMTKY3 binding complexes are popular testing systems used for validating the performance of free energy calculations on predicting ΔΔG caused by mutations[23, 89]. Leu18-to-Pro mutations have a deleterious effect on the binding between OMTKY3 and its receptor as Pro18 has to adopt a backbone conformation with φ/ψ=-89.5˚/41.6˚, which is energetically unfavorable, in the bound complex[90] (**Fig. 9**).

The comparison between the calculated and experimental ΔΔG values are presented in **Table 5**. The MUE and R^2 are 0.63 kcal/mol and 0.92 respectively (**Fig. 10**), which is comparable to the MUE and R^2 of 0.57 kcal/mol and 0.82 for the non-proline to non-proline mutations studied in our previous publication[23]. This indicates that the auxiliary restraints method extends the range of RBFE calculations to those involving ring transformations, while maintaining the same accuracy that we previously obtained for these systems for transformations that did not involve ring opening.

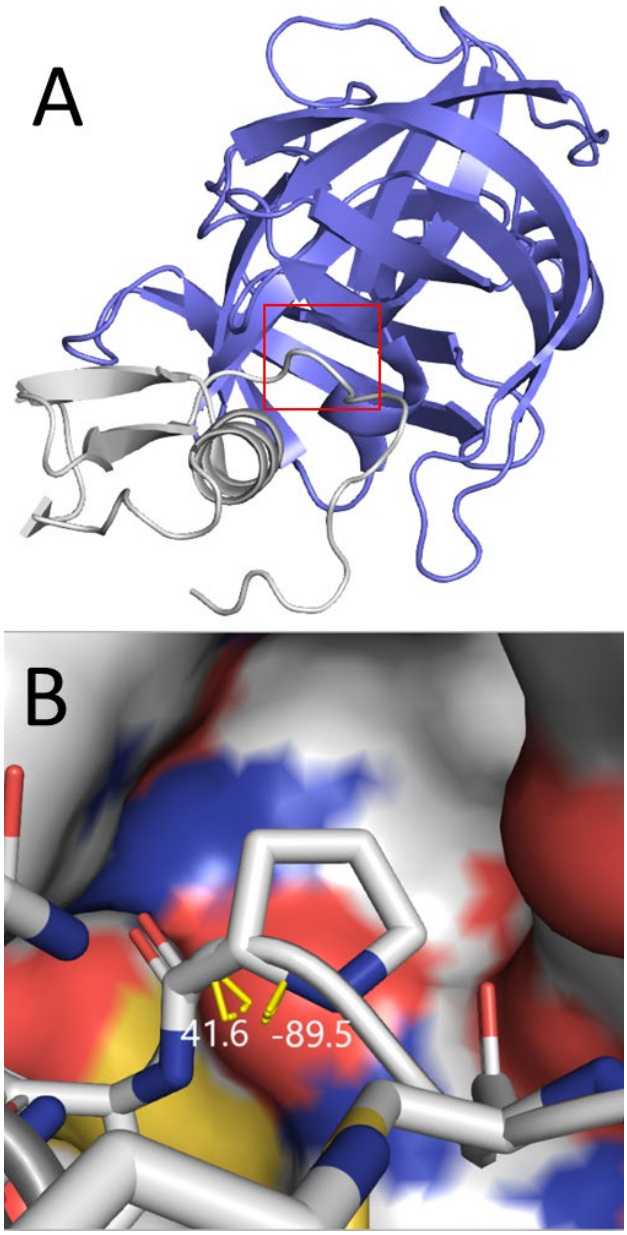

Figure 9. A) Cartoon representation of the SGPB/OMTKY3-L18P complex (PDB code 2sgp). The red box indicates the enlarged area shown in panel B. B) The binding complex of OMTKY3-L18P and SGPB. SGPB is shown as a surface. OMTKY3 is shown as licorice, with P18 in the center. Carbon, Nitrogen, Oxygen and Sulfur atoms are shown in white, blue, red and yellow respectively. The backbone φ/ψ dihedrals are indicated by yellow bars and their values are shown in white. Hydrogen atoms and water are omitted.

Table 5. Comparison between the calculated (using auxiliary restraints) and experimental ΔΔG values of Leu18-to-Pro mutations in OMTKY3 complexes. (Units: kcal/mol).

	Calculated ΔΔG	Experimental ΔΔG[90]
CHYM/OMTKY3-L18P	7.64 ± 0.18	8.82
SGPB /OMTKY3-L18P	8.02 ± 0.24	8.46
HLE /OMTKY3-L18P	5.73 ± 0.13	6.16
CARL/OMTKY3-L18P	7.22 ± 0.62	7.70
MUE/R^2/p-value	0.63/0.92/p=0.04	

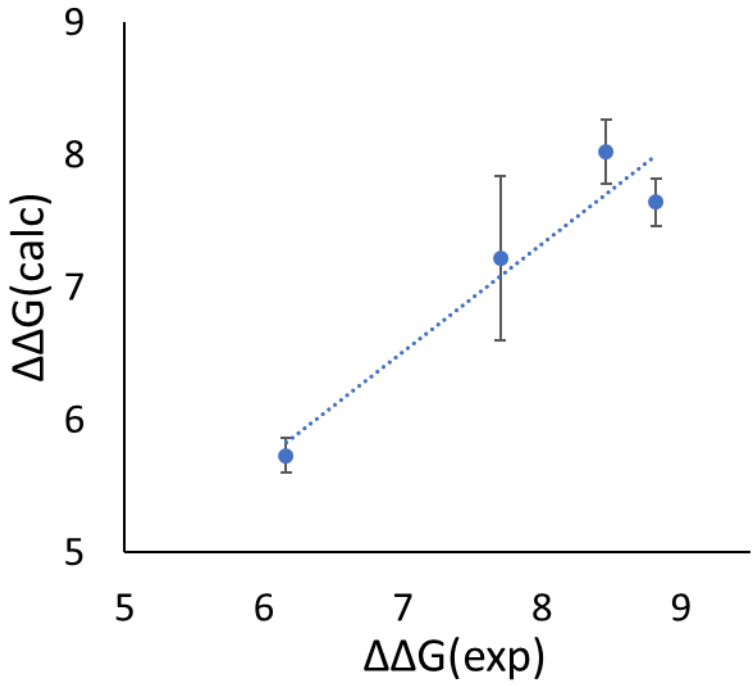

Figure 10. Comparison of the calculated and experimental ΔΔG for the Leu18-to-Pro mutations in OMTKY3 complexes. Units: kcal/mol.

Force constants and reference values of auxiliary restraints

For all the above perturbations, the strength and reference value (length or angle) of the auxiliary restraints were set to the values given in Methods. These appear to be robust enough to cover all the transformations presented here, but it may be possible to further optimize these values to improve phase space overlap between intermediate states. In general, overly-weak restraints are inadequate to hold the bond at the desired distance during the breaking of bond interactions, while too-strong restraints are more difficult to released. Optimal auxiliary restraint parameters may be system dependent, just as the

number and spacing of λ windows used in a perturbation is. However, we believe that the auxiliary restraint parameters are likely to be less important for the calculated ΔΔG than other factors, such as docking poses and force fields, since they serve to constrain sampling, and the energetic impact of these restraints is expected to cancel across the free energy cycle.

Conclusions

In this study, we have presented a novel method for handling scaffold hopping transformations in RBFE calculations. The method relies on auxiliary restraints to prevent sampling of extreme bond distances during the breaking and forming of bonds. Free energy calculations involving ring opening/closure transformations, and ring contraction/expansion transformations were performed using the new method for datasets previously studied using the soft-bond method, augmented with additional transformation pairs. In total 44 transformations of drug-like compounds were tested. In addition, a linker contraction transformation was applied using the new method, which, to our knowledge, has not been reported before. Our method in combination with the GAFF2 force field shows satisfactory accuracy for most of the compounds. It is noteworthy that accuracy is comparable to the accuracy of transformations that did not involve these challenging topology changes. The auxiliary restraints method presented here achieved the same applicability as the soft-bond method, as demonstrated by the transformations involving macrocycles and the proline mutations. However, precautions should be taken when comparing the quantitative ΔΔG values calculated using the auxiliary restraints and the ΔΔG values calculated using the soft-bond approach, because different small molecule force fields were used in the two studies. An important feature of the auxiliary restraints method is that it is not tied to any specific force field, and can be used without modification as small-molecule force fields are improved[91-94]. We also showed that the method can be accomplished without modifying the code of simulation packages using examples calculated by off-the-shelf programs such as Amber.

To apply the auxiliary restraints method on a large scale, platforms for analyzing molecular topologies, setting up multiple perturbation steps for applying and releasing the auxiliary restraints and preparing input files with the corresponding restraints for each compound

will be useful. While a description of their development is beyond the scope of this work we note that the XFEP platform[45] offers a highly automated workflow for handling the non-trivial preparations necessary for carrying out the auxiliary restraint method.

In summary, we have presented a new approach to calculate free energy changes arise from scaffold hopping transformations which can be utilized with existing MD code without modification. The accuracy and precision of the methodology were validated using a wide range of compounds including examples of macrocycles and protein-protein complexes. The method is expected to broaden the impact and applicability of RBFE calculations during drug discovery and protein characterization.

Funding

This study was partially supported by NIH grant GM078114 to D.P.R and NIH grant GM107104 to C.S. JZ was partially supported by the Laufer Center for Physical and Quantitative Biology at Stony Brook University.

Author Contributions. JZ designed the project, built the workflow in XFEP, conducted calculation, analyzed data and wrote the manuscript. ZL(i) built the workflow in XFEP and conducted calculation. SL conducted calculations. CP reviewed code of the workflow in XFEP. DF conducted calculation. XW trained the XFF force field for TPSB2 compounds. ZL(in) analyzed data. TL modified code of the XFEP simulation engine. MY directed and guided research. DPR directed and guided research and wrote the manuscript, CS directed and guided research and wrote the manuscript.

Conflict of interest

JZ, ZL(i), SL, CP, DF, XW, ZL(in) and MY are employees of Xtalpi Inc. TL, CS and DPR declare no conflicts.

Supporting information

Additional tables and figures referenced in the main text; files for the complex structures used in this study in pdb and sdf format, complete description and simulation inputs for the example transformations conducted using standard Amber18, XFF force field parameters for the transformations reported here.

References

1. Sun, H.; Tawa, G.; Wallqvist, A., Classification of scaffold-hopping approaches. *Drug Discov. Today* **2012,** *17* (7-8), 310-324.
2. Osváth, S.; Gruebele, M., Proline can have opposite effects on fast and slow protein folding phases. *Biophys. J.* **2003,** *85* (2), 1215-1222.
3. Abedini, A.; Raleigh, D. P., Destabilization of human IAPP amyloid fibrils by proline mutations outside of the putative amyloidogenic domain: Is there a critical amyloidogenic domain in human IAPP? *J. Mol. Biol.* **2006,** *355* (2), 274-281.
4. Buchanan, L. E.; Dunkelberger, E. B.; Tran, H. Q.; Cheng, P. N.; Chiu, C. C.; Cao, P.; Raleigh, D. P.; de Pablo, J. J.; Nowick, J. S.; Zanni, M. T., Mechanism of IAPP amyloid fibril formation involves an intermediate with a transient β-sheet. *Proc. Natl. Acad. Sci. U. S. A.* **2013,** *110* (48), 19285-19290.
5. Hardy, J. A.; Nelson, H. C. M., Proline in α-helical kink is required for folding kinetics but not for kinked structure, function, or stability of heat shock transcription factor. *Protein Sci.* **2000,** *9* (11), 2128-2141.
6. Creighton, T. E., Possible implications of many proline residues for the kinetics of protein unfolding and refolding. *J. Mol. Biol.* **1978,** *125* (3), 401-406.
7. Williams, A. D.; Portelius, E.; Kheterpal, I.; Guo, J.-t.; Cook, K. D.; Xu, Y.; Wetzel, R., Mapping Aβ amyloid fibril secondary structure using scanning proline mutagenesis. *Journal of Molecular Biology* **2004,** *335* (3), 833-842.
8. Choi, E. J.; Mayo, S. L., Generation and analysis of proline mutants in protein G. *Protein Eng., Des. Sel.* **2006,** *19* (6), 285-289.
9. Melnikov, S.; Mailliot, J.; Rigger, L.; Neuner, S.; Shin, B. S.; Yusupova, G.; Dever, T. E.; Micura, R.; Yusupov, M., Molecular insights into protein synthesis with proline residues. *EMBO Rep.* **2016,** *17* (12), 1776-1784.
10. Remeeva, A.; Nazarenko, V. V.; Goncharov, I. M.; Yudenko, A.; Smolentseva, A.; Semenov, O.; Kovalev, K.; Gülbahar, C.; Schwaneberg, U.; Davari, M. D.; Gordeliy, V.; Gushchin, I., Effects of proline substitutions on the thermostable LOV domain from Chloroflexus aggregans. *Crystals* **2020,** *10* (4).
11. Ain, Q. U.; Aleksandrova, A.; Roessler, F. D.; Ballester, P. J., Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015,** *5* (6), 405-424.
12. Wang, C.; Greene, D. A.; Xiao, L.; Qi, R.; Luo, R., Recent developments and applications of the MMPBSA method. *Front. Mol. Biosci.* **2018,** *4*.
13. Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004,** *3* (11), 935-949.
14. Jiménez, J.; Škalič, M.; Martínez-Rosell, G.; De Fabritiis, G., KDEEP: Protein–ligand absolute binding affinity prediction via 3D-convolutional neural networks. *J. Chem. Inf. Model.* **2018,** *58* (2), 287-296.
15. Mobley, D. L.; Klimovich, P. V., Perspective: Alchemical free energy calculations for drug discovery. *J. Chem. Phys.* **2012,** *137* (23).
16. Zwanzig, R. W., High‐temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics* **1954,** *22* (8), 1420-1426.

17. Kirkwood, J. G., Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics* **1935,** *3* (5), 300-313.
18. Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M., Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* **2020,** *60* (11), 5595-5623.
19. Cournia, Z.; Allen, B.; Sherman, W., Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. *J. Chem. Inf. Model.* **2017,** *57* (12), 2911-2937.
20. He, X.; Liu, S.; Lee, T.-S.; Ji, B.; Man, V. H.; York, D. M.; Wang, J., Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein–Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega* **2020,** *5* (9), 4611-4619.
21. Raman, E. P.; Paul, T. J.; Hayes, R. L.; Brooks, C. L., Automated, accurate, and scalable relative protein–ligand binding free-energy calculations using lambda dynamics. *J. Chem. Theory. Comput.* **2020**.
22. Zou, J.; Tian, C.; Simmerling, C., Blinded prediction of protein-ligand binding affinity using Amber thermodynamic integration for the 2018 D3R grand challenge 4. *J. Comput.-Aided Mol. Des.* **2019,** *33* (12), 1021-1029.
23. Zou, J.; Simmerling, C.; Raleigh, D. P., Dissecting the energetics of intrinsically disordered proteins via a hybrid experimental and computational approach. *J. Phys. Chem. B* **2019,** *123* (49), 10394-10402.
24. Schindler, C. E. M.; Baumann, H.; Blum, A.; Böse, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; Eguida, M. K. I.; Follows, B.; Fuchß, T.; Grädler, U.; Gunera, J.; Johnson, T.; Jorand Lebrun, C.; Karra, S.; Klein, M.; Knehans, T.; Koetzner, L.; Krier, M.; Leiendecker, M.; Leuthner, B.; Li, L.; Mochalkin, I.; Musil, D.; Neagu, C.; Rippmann, F.; Schiemann, K.; Schulz, R.; Steinbrecher, T.; Tanzer, E.-M.; Unzue Lopez, A.; Viacava Follis, A.; Wegener, A.; Kuhn, D., Large-scale assessment of binding free energy calculations in active drug discovery projects. *J. Chem. Inf. Model.* **2020,** *60* (11), 5457-5474.
25. Kuhn, B.; Tichý, M.; Wang, L.; Robinson, S.; Martin, R. E.; Kuglstatter, A.; Benz, J.; Giroud, M.; Schirmeister, T.; Abel, R.; Diederich, F.; Hert, J., Prospective evaluation of free energy calculations for the prioritization of cathepsin L inhibitors. *J. Med. Chem.* **2017,** *60* (6), 2485-2497.
26. Lenselink, E. B.; Louvel, J.; Forti, A. F.; van Veldhoven, J. P. D.; de Vries, H.; Mulder-Krieger, T.; McRobb, F. M.; Negri, A.; Goose, J.; Abel, R.; van Vlijmen, H. W. T.; Wang, L.; Harder, E.; Sherman, W.; Ijzerman, A. P.; Beuming, T., Predicting binding affinities for GPCR ligands using free-energy perturbation. *ACS Omega* **2016,** *1* (2), 293-304.
27. Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R., Accurate modeling of scaffold hopping transformations in drug discovery. *J. Chem. Theory. Comput.* **2016,** *13* (1), 42-54.
28. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F., Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chem. Phys. Lett* **1994,** *222* (6), 529-539.
29. Yu, H. S.; Deng, Y.; Wu, Y.; Sindhikara, D.; Rask, A. R.; Kimura, T.; Abel, R.; Wang, L., Accurate and reliable prediction of the binding affinities of macrocycles to their protein targets. *J. Chem. Theory. Comput.* **2017,** *13* (12), 6290-6300.
30. 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* (9), 1157-1174.

31. Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A., OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theory. Comput.* **2015,** *12* (1), 281-296.
32. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002,** *23* (16), 1623-41.
33. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.
34. 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* (8), 3696-713.
35. Lee, J. C.; Gutell, R. R., Helix capping in RNA structure. *PloS one* **2014,** *9* (4), e93664.
36. Williams, C. J.; Headd, J. J.; Moriarty, N. W.; Prisant, M. G.; Videau, L. L.; Deis, L. N.; Verma, V.; Keedy, D. A.; Hintze, B. J.; Chen, V. B.; Jain, S.; Lewis, S. M.; Arendall, W. B., 3rd; Snoeyink, J.; Adams, P. D.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C., MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **2018,** *27* (1), 293-315.
37. Jakalian, A.; Jack, D. B.; Bayly, C. I., Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* **2002,** *23* (16), 1623-41.
38. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of N-alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327-341.
39. Darden, T.; York, D.; Pedersen, L., Particle mesh ewald - an N.Log(N) method for ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.
40. Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O., Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chem. Phys. Lett* **2004,** *384* (4-6), 288-294.
41. Steinbrecher, T.; Joung, I.; Case, D. A., Soft-Core potentials in thermodynamic integration: Comparing one- and two-step transformations. *J. Comput. Chem.* **2011,** *32* (15), 3253-63.
42. Meng, Y.; Dashti, D. S.; Roitberg, A. E., Computing alchemical free energy differences with Hamiltonian replica exchange molecular dynamics (H-REMD) simulations. *J. Chem. Theory. Comput.* **2011,** *7* (9), 2721-2727.
43. Bennett, C. H., Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976,** *22* (2), 245-268.
44. Shirts, M. R.; Chodera, J. D., Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008,** *129* (12).
45. Zhixiong, L.; Junjie, Z.; Chunwang, P.; Shuai, L.; Zhipeng, L.; Xiao, W.; Dong, F.; Jian, Y.; Gianpaolo, G.; Yongpan, C.; Jian, M.; Shuhao, W.; Peiyu, Z.; Mingjun, Y., A cloud computing platform for scalable relative and absolute binding free energy prediction: New opportunities and challenges for drug discovery. ChemRxiv, 2020.
46. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; Cruzeiro, I., V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H., *AMBER 2018*. University of California, San Francisco, 2018.

47. Fujinaga, M.; Sielecki, A. R.; Read, R. J.; Ardelt, W.; Laskowski, M.; James, M. N. G., Crystal and molecular structures of the complex of α-chymotrypsin with its inhibitor Turkey ovomucoid third domain at 1.8 Å resolution. *J. Mol. Biol.* **1987,** *195* (2), 397-418.
48. Huang, K.; James, M. N. G.; Lu, W.; Laskowski, M.; Anderson, S., Water molecules participate in proteinase-inhibitor interactions: Crystal structures of Leu18, Ala18, and Gly18variants of turkey ovomucoid inhibitor third domain complexed withStreptomyces griseusproteinase B. *Protein Sci.* **1995,** *4* (10), 1985-1997.
49. Bode, W.; Wei, A. Z.; Huber, R.; Meyer, E.; Travis, J.; Neumann, S., X-ray crystal structure of the complex of human leukocyte elastase (PMN elastase) and the third domain of the turkey ovomucoid inhibitor. *EMBO J.* **1986,** *5* (10), 2453-2458.
50. Horn, J. R.; Ramaswamy, S.; Murphy, K. P., Structure and energetics of protein–protein interactions: The role of conformational heterogeneity in OMTKY3 binding to serine proteases. *J. Mol. Biol.* **2003,** *331* (2), 497-508.
51. Heinzelmann, G.; Henriksen, N. M.; Gilson, M. K., Attach-pull-release calculations of ligand binding and conformational changes on the first BRD4 bromodomain. *Journal of Chemical Theory and Computation* **2017,** *13* (7), 3260-3275.
52. Xiao, Z.; Chen, Z.; Gunasekera, A. H.; Sowin, T. J.; Rosenberg, S. H.; Fesik, S.; Zhang, H., Chk1 mediates S and G2 arrests through Cdc25A degradation in response to DNA-damaging agents. *J. Biol. Chem.* **2003,** *278* (24), 21767-21773.
53. Huang, X.; Cheng, C. C.; Fischmann, T. O.; Duca, J. S.; Yang, X.; Richards, M.; Shipps, G. W., Discovery of a novel series of CHK1 kinase inhibitors with a distinctive hinge binding mode. *ACS Med. Chem. Lett.* **2012,** *3* (2), 123-128.
54. Nagata, T.; Yoshino, T.; Haginoya, N.; Yoshikawa, K.; Nagamochi, M.; Kobayashi, S.; Komoriya, S.; Yokomizo, A.; Muto, R.; Yamaguchi, M.; Osanai, K.; Suzuki, M.; Kanno, H., Discovery of N-[(1R,2S,5S)-2-{[(5-chloroindol-2-yl)carbonyl]amino}-5-(dimethylcarbamoyl)cyclohexyl]-5-methyl-4,5,6,7-tetrahydrothiazolo[5,4-c]pyridine-2-carboxamide hydrochloride: A novel, potent and orally active direct inhibitor of factor Xa. *Bioorg. Med. Chem.* **2009,** *17* (3), 1193-1206.
55. Pinto, D. J. P.; Smallheer, J. M.; Cheney, D. L.; Knabb, R. M.; Wexler, R. R., Factor Xa inhibitors: Next-generation antithrombotic agents. *J. Med. Chem.* **2010,** *53* (17), 6243-6274.
56. Furugohri, T.; Isobe, K.; Honda, Y.; Kamisato-Matsumoto, C.; Sugiyama, N.; Nagahara, T.; Morishima, Y.; Shibano, T., DU-176b, a potent and orally active factor Xa inhibitor:in vitroandin vivopharmacological profiles. *J. Thromb. Haemostasis* **2008**.
57. Galli, S. J.; Costa, J. J., Mast-cell-leukocyte cytokine cascades in allergic inflammation. *Allergy* **1995,** *50* (11), 851-862.
58. Liang, G.; Choi-Sledeski, Y. M.; Shum, P.; Chen, X.; Poli, G. B.; Kumar, V.; Minnich, A.; Wang, Q.; Tsay, J.; Sides, K.; Kang, J.; Zhang, Y., A β-tryptase inhibitor with a tropanylamide scaffold to improve in vitro stability and to lower hERG channel binding affinity. *Bioorg. Med. Chem. Lett.* **2012,** *22* (4), 1606-1610.
59. Cai, H.; Wang, Y.; McCarthy, D.; Wen, H.; Borchelt, D. R.; Price, D. L.; Wong, P. C., BACE1 is the major β-secretase for generation of Aβ peptides by neurons. *Nat. Neurosci.* **2001,** *4* (3), 233-234.
60. Hu, X.; Das, B.; Hou, H.; He, W.; Yan, R., BACE1 deletion in the adult mouse reverses preformed amyloid deposition and improves cognitive functions. *J. Exp. Med.* **2018,** *215* (3), 927-940.
61. Luo, Y.; Bolon, B.; Kahn, S.; Bennett, B. D.; Babu-Khan, S.; Denis, P.; Fan, W.; Kha, H.; Zhang, J.; Gong, Y.; Martin, L.; Louis, J.-C.; Yan, Q.; Richards, W. G.; Citron, M.; Vassar, R., Mice

deficient in BACE1, the Alzheimer's β-secretase, have normal phenotype and abolished β-amyloid generation. *Nat. Neurosci.* **2001,** *4* (3), 231-232.
62. Winneroski, L. L.; Schiffler, M. A.; Erickson, J. A.; May, P. C.; Monk, S. A.; Timm, D. E.; Audia, J. E.; Beck, J. P.; Boggs, L. N.; Borders, A. R.; Boyer, R. D.; Brier, R. A.; Hudziak, K. J.; Klimkowski, V. J.; Garcia Losada, P.; Mathes, B. M.; Stout, S. L.; Watson, B. M.; Mergott, D. J., Preparation and biological evaluation of conformationally constrained BACE1 inhibitors. *Bioorg. Med. Chem.* **2015,** *23* (13), 3260-3268.
63. Paterni, I.; Granchi, C.; Katzenellenbogen, J. A.; Minutolo, F., Estrogen receptors alpha (ERα) and beta (ERβ): Subtype-selective ligands and clinical potential. *Steroids* **2014,** *90*, 13-29.
64. Richardson, T. I.; Dodge, J. A.; Durst, G. L.; Pfeifer, L. A.; Shah, J.; Wang, Y.; Durbin, J. D.; Krishnan, V.; Norman, B. H., Benzopyrans as selective estrogen receptor β agonists (SERBAs). Part 3: Synthesis of cyclopentanone and cyclohexanone intermediates for C-ring modification. *Bioorg. Med. Chem. Lett.* **2007,** *17* (17), 4824-4828.
65. Gupta, S.; Singh, R. K.; Dastidar, S.; Ray, A., Cysteine cathepsin S as an immunomodulatory target: present and future trends. *Expert Opin. Ther. Targets* **2008,** *12* (3), 291-299.
66. Parks, C. D.; Gaieb, Z.; Chiu, M.; Yang, H.; Shao, C.; Walters, W. P.; Jansen, J. M.; McGaughey, G.; Lewis, R. A.; Bembenek, S. D.; Ameriks, M. K.; Mirzadegan, T.; Burley, S. K.; Amaro, R. E.; Gilson, M. K., D3R grand challenge 4: blind prediction of protein-ligand poses, affinity rankings, and relative binding free energies. *J. Comput.-Aided Mol. Des.* **2020,** *34* (2), 99-119.
67. Wiener, J. J. M.; Wickboldt, A. T.; Wiener, D. K.; Lee-Dutra, A.; Edwards, J. P.; Karlsson, L.; Nguyen, S.; Sun, S.; Jones, T. K.; Grice, C. A., Discovery and SAR of novel pyrazole-based thioethers as cathepsin S inhibitors. Part 2: Modification of P3, P4, and P5 regions. *Bioorg. Med. Chem. Lett.* **2010,** *20* (7), 2375-2378.
68. Suda, A.; Koyano, H.; Hayase, T.; Hada, K.; Kawasaki, K.-i.; Komiyama, S.; Hasegawa, K.; Fukami, T. A.; Sato, S.; Miura, T.; Ono, N.; Yamazaki, T.; Saitoh, R.; Shimma, N.; Shiratori, Y.; Tsukuda, T., Design and synthesis of novel macrocyclic 2-amino-6-arylpyrimidine Hsp90 inhibitors. *Bioorg. Med. Chem. Lett.* **2012,** *22* (2), 1136-1141.
69. Kettle, J. G.; Alwan, H.; Bista, M.; Breed, J.; Davies, N. L.; Eckersley, K.; Fillery, S.; Foote, K. M.; Goodwin, L.; Jones, D. R.; Käck, H.; Lau, A.; Nissink, J. W. M.; Read, J.; Scott, J. S.; Taylor, B.; Walker, G.; Wissler, L.; Wylot, M., Potent and selective inhibitors of MTH1 probe its role in cancer cell survival. *J. Med. Chem.* **2016,** *59* (6), 2346-2361.
70. Tao, Z.-F.; Wang, L.; Stewart, K. D.; Chen, Z.; Gu, W.; Bui, M.-H.; Merta, P.; Zhang, H.; Kovar, P.; Johnson, E.; Park, C.; Judge, R.; Rosenberg, S.; Sowin, T.; Lin, N.-H., Structure-based design, synthesis, and biological evaluation of potent and selective macrocyclic checkpoint kinase 1 inhibitors. *J. Med. Chem.* **2007,** *50* (7), 1514-1527.
71. Nie, Z.; Perretta, C.; Erickson, P.; Margosiak, S.; Lu, J.; Averill, A.; Almassy, R.; Chu, S., Structure-based design and synthesis of novel macrocyclic pyrazolo[1,5-a] [1,3,5]triazine compounds as potent inhibitors of protein kinase CK2 and their anticancer activities. *Bioorg. Med. Chem. Lett.* **2008,** *18* (2), 619-623.
72. Zapf, C. W.; Bloom, J. D.; Li, Z.; Dushin, R. G.; Nittoli, T.; Otteng, M.; Nikitenko, A.; Golas, J. M.; Liu, H.; Lucas, J.; Boschelli, F.; Vogan, E.; Olland, A.; Johnson, M.; Levin, J. I., Discovery of a stable macrocyclic o-aminobenzamide Hsp90 inhibitor which significantly decreases tumor volume in a mouse xenograft model. *Bioorg. Med. Chem. Lett.* **2011,** *21* (15), 4602-4607.
73. Huang, Y.; Strobel, E. D.; Ho, C. Y.; Reynolds, C. H.; Conway, K. A.; Piesvaux, J. A.; Brenneman, D. E.; Yohrling, G. J.; Moore Arnold, H.; Rosenthal, D.; Alexander, R. S.; Tounge, B. A.; Mercken, M.; Vandermeeren, M.; Parker, M. H.; Reitz, A. B.; Baxter, E. W., Macrocyclic BACE

inhibitors: Optimization of a micromolar hit to nanomolar leads. *Bioorg. Med. Chem. Lett.* **2010,** *20* (10), 3158-3160.
74. Stachel, S. J.; Coburn, C. A.; Sankaranarayanan, S.; Price, E. A.; Pietrak, B. L.; Huang, Q.; Lineberger, J.; Espeseth, A. S.; Jin, L.; Ellis, J.; Holloway, M. K.; Munshi, S.; Allison, T.; Hazuda, D.; Simon, A. J.; Graham, S. L.; Vacca, J. P., Macrocyclic inhibitors of β-secretase: Functional activity in an animal model. *J. Med. Chem.* **2006,** *49* (21), 6147-6150.
75. Wang, G.; Unger, G.; Ahmad, K. A.; Slaton, J. W.; Ahmed, K., Downregulation of CK2 induces apoptosis in cancer cells – A potential approach to cancer therapy. *Mol. Cell. Biochem.* **2005,** *274* (1-2), 77-84.
76. Freudenthal, B. D.; Beard, W. A.; Perera, L.; Shock, D. D.; Kim, T.; Schlick, T.; Wilson, S. H., Uncovering the polymerase-induced cytotoxicity of an oxidized nucleotide. *Nature* **2014,** *517* (7536), 635-639.
77. Blagg, B. S. J.; Kerr, T. D., Hsp90 inhibitors: Small molecules that transform the Hsp90 protein folding machinery into a catalyst for protein degradation. *Med. Res. Rev.* **2006,** *26* (3), 310-338.
78. Paulsen, J. L.; Yu, H. S.; Sindhikara, D.; Wang, L.; Appleby, T.; Villaseñor, A. G.; Schmitz, U.; Shivakumar, D., Evaluation of free energy calculations for the prioritization of macrocycle synthesis. *J. Chem. Inf. Model.* **2020,** *60* (7), 3489-3498.
79. Ugur, I.; Schroft, M.; Marion, A.; Glaser, M.; Antes, I., Predicting the bioactive conformations of macrocycles: a molecular dynamics-based docking procedure with DynaDock. *J. Mol. Model.* **2019,** *25* (7).
80. Watts, K. S.; Dalal, P.; Tebben, A. J.; Cheney, D. L.; Shelley, J. C., Macrocycle conformational sampling with MacroModel. *J. Chem. Inf. Model.* **2014,** *54* (10), 2680-2696.
81. Zou, J.; Song, B.; Simmerling, C.; Raleigh, D., Experimental and computational analysis of protein stabilization by Gly-to-d-Ala substitution: A convolution of native state and unfolded state effects. *J. Am. Chem. Soc.* **2016,** *138* (48), 15682-15689.
82. Duan, J.; Lupyan, D.; Wang, L., Improving the accuracy of protein thermostability predictions for single point mutations. *Biophys. J.* **2020,** *119* (1), 115-127.
83. Seeliger, D.; de Groot, B. L., Protein thermostability calculations using alchemical free energy simulations. *Biophys. J.* **2010,** *98* (10), 2309-2316.
84. Voelz, V. A.; Singh, V. R.; Wedemeyer, W. J.; Lapidus, L. J.; Pande, V. S., Unfolded-state dynamics and structure of protein L characterized by simulation and experiment. *J. Am. Chem. Soc.* **2010,** *132* (13), 4702-4709.
85. Bowler, B. E., Residual structure in unfolded proteins. *Curr. Opin. Struct. Biol.* **2012,** *22* (1), 4-13.
86. Anil, B.; Li, Y.; Cho, J.-H.; Raleigh, D. P., The unfolded state of NTL9 is compact in the absence of denaturant. *Biochemistry* **2006,** *45* (33), 10110-10116.
87. Meng, W.; Shan, B.; Tang, Y.; Raleigh, D. P., Native like structure in the unfolded state of the villin headpiece helical subdomain, an ultrafast folding protein. *Protein Sci.* **2009,** *18* (8), 1692-1701.
88. Fitzkee, N. C.; García-Moreno E, B., Electrostatic effects in unfolded staphylococcal nuclease. *Protein Sci.* **2008,** *17* (2), 216-227.
89. Clark, A. J.; Negron, C.; Hauser, K.; Sun, M.; Wang, L.; Abel, R.; Friesner, R. A., Relative binding affinity prediction of charge-changing sequence mutations with FEP in protein–protein interfaces. *J. Mol. Biol.* **2019,** *431* (7), 1481-1493.
90. Lu, W.; Apostol, I.; Qasim, M. A.; Warne, N.; Wynn, R.; Zhang, W. L.; Anderson, S.; Chiang, Y. W.; Ogin, E.; Rothberg, I.; Ryan, K.; Laskowski, M., Binding of amino acid side-chains to S 1 cavities of serine proteinases. *J. Mol. Biol.* **1997,** *266* (2), 441-461.

91. Wang, L.-P.; Mobley, D.; Bayly, C. I.; Chodera, J.; Gilson, M.; Shirts, M. R.; Tjanaka, B.; Lucas, X.; Rizzi, A.; Stern, C.; Lim, V. T.; Gokey, T.; Bannan, C. C.; Wagner, J.; Jang, H.; Boothroyd, S.; Smith, D.; Qiu, Y., Development and benchmarking of Open Force Field v1.0.0, the Parsley small molecule force field. **2020**.
92. Vanommeslaeghe, K.; MacKerell, A. D., Automation of the CHARMM general force field (CGenFF) I: Bond perception and atom typing. *J. Chem. Inf. Model.* **2012,** *52* (12), 3144-3154.
93. Chatterjee, P.; Heid, E.; Schröder, C.; MacKerell, A. D., Polarizable general force field for drug-like molecules: Drude general force field (DGenFF). *Biophys. J.* **2019,** *116* (3).
94. Wu, J. C.; Chattree, G.; Ren, P., Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012,** *131* (3).

TOC figure