High throughput evaluation of macrocyclization strategies for conformer stabilization

Dan Sindhikara & Ken Borrelli

While macrocyclization of a linear compound to stabilize a known bioactive conformation can be a useful strategy to increase binding potency, the difficulty of macrocycle synthesis can limit the throughput of such strategies. Thus computational techniques may offer the higher throughput required to screen large numbers of compounds. Here we introduce a method for evaluating the propensity of a macrocyclic compound to adopt a conformation similar that of a known active linear compound in the binding site. This method can be used as a fast screening tool for prioritizing macrocycles by leveraging the assumption that the propensity for the known bioactive substructural conformation relates to the affinity. While this method cannot identify new interactions not present in the known linear compound, it could quickly differentiate compounds where the three dimensional geometries imposed by the macrocyclization prevent adoption of conformations with the same contacts as the linear compound in their conserved region. Here we report the implementation of this method using an RMSD-based structural descriptor and a Boltzmann-weighted propensity calculation and apply it retrospectively to three macrocycle linker optimization design projects. We found the method performs well in terms of prioritizing more potent compounds.

Macrocycles are an important and pervasive class of molecule for drug design\(^1\). Though definitions vary, they are typically described as molecules with ring sizes of at least 8, 10, or 12 atoms\(^2,3\). They span a molecular weight range typically larger than small molecules and smaller than biological therapeutics\(^4\). Though some design projects begin with a macrocyclic native ligand, often a cyclization will be performed to introduce a conformational restriction to an otherwise linear molecule\(^5–7\). Regardless, the cyclization topology can drastically affect the conformational propensity of the molecule and alter the affinity of the molecule for its target receptor. Understanding this conformational effect on bioactivity is essential to macrocycle design.

\textit{In silico} sampling of macrocycles is particularly difficult due to the conformational restriction imposed by the cyclization. Several methods have been developed to overcome this sampling challenge including distance-geometry-based\(^8\), low-mode based\(^9\), normal-mode-based\(^10\), inverse kinematics-based\(^11\), and loop-sampling-based\(^12\) conformational searches.

Previous work has created workflows to screen macrocycles by combining these sampling methods with scoring functions such as molecular mechanics simulated annealing combined with quantum mechanical strain scoring\(^13\), inverse kinematics with ROSETTA\(^14\), implicit solvent/molecular mechanics scoring\(^15\), and normal-mode-based sampling with molecular mechanics scoring\(^16\). In this work, we implement an integrated, high-throughput method utilizing a loop-sampling-based macrocycle conformational sampling protocol along with a simple molecular mechanics strain-based scoring function.

The ability to access a given binding conformation is one of the many factors that determine the binding affinity of a specific ligand for a specific protein. When the binding mode is assumed to be both known and constant, this can be treated as a necessary, but not sufficient, condition for high affinity binding. While this applies to all molecules, it is especially important for macrocyclic ligands due to the restriction of conformational space caused by the cyclization. When comparing different potential cyclizations of a linear molecule where the linker is not interacting with the protein, this could even become the primary factor affecting the relative binding energy of a series of compounds. In these cases, determining the relative ability of a various cyclized versions of the same linear molecule to adopt a known binding model may be able to serve as a proxy for the relative binding affinity of those compounds. In this work we will show that the propensity of a macrocycle to adopt a specific binding...
mode, also termed the bioactive conformation of that macrocycle, can explain the differences in binding affinity for sets of congenic macrocyclic molecules when the only differences between those macrocycles are in the linker region and that linker region is not making any contacts with the protein.

Previous work has shown the ability of Prime Macrocycle Sampling to efficiently sample the conformational space of macrocycles \cite{McCoul2018} and here we combine it with the OPLS3 force field \cite{Jorgensen1996} to evaluate whether molecular mechanics approaches, which are significantly computationally cheaper than quantum approaches, are adequate to determine the relative strain energies of these compounds. Unlike the inverse kinematic approach previously reported, this method would be expected to be work with macrocycles that did not have additional crosslinks to reduce the conformational space to be sampled. The approach presented here is similar to that used by McCoul et al. to design macroyclic BCL6 inhibitors \cite{McCoul2016}, however here we demonstrate an integrated workflow and validate it on diverse test sets.

Methods

Conformer generation. Although generation of conformational ensembles of macrocycles is difficult, many methods have been developed to overcome the particular difficulties of sampling introduced by macrocyclization. Here we use the Prime Macrocycle Conformational Sampling algorithm (Prime-MCS) which can quickly produce accurate, diverse macrocycle conformations \cite{Jorgensen1996}. Default calculation parameters were used including unbiased PrimeMCS sampling and OPLS3.0 force field parameters \cite{Jorgensen1996} in vacuum. This assumes that unbiased default sampling is sufficient to cover representative conformational space and that the OPLS3 force field can sufficiently represent the compound. It is possible that, depending on the design context, especially the macrocycle topological complexity, that these assumptions will be broken and the sampling protocol should be modified. Some such modifications are discussed in the Discussion section.

Ensemble scoring. Conformer structural metric. To properly score the conformational ensemble, a metric for the similarity to the bioactive substructure is required. For many projects, the conserved substructure is obvious since the linker modifications leave the original acyclic molecule largely unchanged. Here, we apply the maximum common substructure (MCS) algorithm, as implemented in Canvas software \cite{Schrodinger2017}, to the designed macrocycles and the bioactive reference structure to determine the conserved region. Though this approach may not be optimal for use in prospective projects, the MCS is useful under the assumption that the topological differences between the macrocyclic designs and the reference are purely in the cyclization linker, and thus the bioactive conformation of the remaining substructure is intended to be conserved. We take the RMSD of the heavy atoms in the conserved region, \(\text{RMSD}_{\text{cons}}\), as a simple metric for similarity to the bioactive substructure under the assumption that, for the sets of macrocycles we will be considering, the ability for these atoms to adopt a conserved binding pose is the primary determinant of binding affinity. Alternate structural metrics are discussed below in the Discussion section. When measuring the \(\text{RMSD}_{\text{cons}}\), we account for symmetry by generating a SMARTS pattern from the MCS atoms in the reference ligand and choosing the minimum RMSD for all matches to the SMARTS patterns in the test compounds. Here, the MCS atoms were calculated using Schrodinger’s canvas_MCS utility (\texttt{SCHRODINGER/utilities/canvas_MCS}) using default settings in Schrodinger release 2017-3. This method technique distinguishes atoms by atomic number, bond order, and aromaticity.

Ensemble weighting and the proxy metric for affinity. If the conformational ensemble were complete and representative, ensemble weighting would not be necessary. However for biased or clustered methods, reweighting is necessary to recover the correct ensemble. The conformationer generation algorithm here, Prime-MCS, uses structure-based cluster output. Such clustered structures may be weighted to the canonical ensemble using simple Boltzmann-weighting (assuming these clusters are evenly distributed in conformational state space). That is, that each conformer is weighted according to its energy \((P_i = \exp(-E_i/kT)/Z)\). This approach has been used, for example, for finite-reservoir sampling methods \cite{Geyer1992, Lesfari2008}.

After weighting the ensemble, we can calculate the probability distribution of observables. We assume that the affinity is a function of this conformational distribution. Here, we simply use the expectation value of the metric, \(\langle \text{RMSD}_{\text{cons}} \rangle = \sum_i P_i \text{RMSD}_{\text{cons},i} \) as the proxy for affinity. The expectation value is useful if the affinity gets better the closer the compound gets to the bioactive structure. Alternate proxy metrics are discussed later in the Discussion section.

Dataset preparation. For this study we use systems from macrocycle design projects taken from the literature. In a related paper submitted recently, seven macrocyclization projects were curated to evaluate free energy perturbation (FEP) technology on affinity calculations for macrocycles \cite{McCoul2018}. For that article, papers were chosen describing projects which contained macrocyclizations (conversion of linear compounds to macrocyclic), modification of macrocycle linker sizes, or both. Projects were only used if there were no apparent issues with the data that would complicate an evaluation by atom-scale biophysical modeling such as significant missing structural data and insufficient affinity data. Here we select from those seven projects the only three where there were at least two macrocycles within the set. Those systems were Chk1 (PDBID 2E9P), Bace-1 (PDBID 2Q15) \cite{Xu2008}, and Hsp90 (PDBID 3RKZ) \cite{Nakamura2009}, see Table 1 and Figure 1. Though protein structures are available, they were not used for any part of these calculations except to confirm that the various modifications to the linker are on the solvent exposed side of the ligand and therefore less likely to be involved in specific protein-ligand interactions that could affect the binding affinities of these molecules.

For the calculations here, the experimental structural data for the macrocyclic structures were removed by converting the structures to SMILES then back to 3D. These 3D structures were then inspected to ensure correct stereoisomer states as well as trans amide bonds. These, as well as the linear reference structure, were prepared using LigPrep \cite{Schrodinger2017}.

\[ \text{RMSD}_{\text{cons}} = \sum_i P_i \text{RMSD}_{\text{cons},i} \]
Table 1. Markush representations of enumerated macrocycles with associated substitutions and experimental affinities ($\Delta G_{\text{bind}}$ in kcal/mol).

| Chk1 | Cpd | R | n | Exp. dG$_{\text{bind}}$ |
|------|-----|---|---|-----------------------|
| 1    | H   | 2 |   | −10.88                |
| 2    | CN  | 1 |   | −11.18                |
| 3    | CN  | 2 |   | −11.09                |
| 4    | CN  | 3 |   | −10.27                |

| Hsp90 | Cpd | R$_1$ | R$_2$ | R$_3$ | R$_4$ | n1 | n2 | Exp. dG$_{\text{bind}}$ |
|-------|-----|-------|-------|-------|-------|----|----|-----------------------|
| 5     |     |       |       |       |       | 0  | 1  | −6.67                 |
| 6     |     | CH$_3$| CH$_3$|       |       | 0  | 2  | −9.32                 |
| 7     |     |       |       |       |       | 1  | 2  | −8.89                 |
| 8     |     | CH$_3$|       | CH$_3$|       | 0  | 2  | −9.28                 |
| 9     |     | −CH$_2$−CH$_2$−| −CH$_2$−CH$_2$−|       |       | 0  | 2  | −8.45                 |
| 10    |     | CH$_3$| CH$_3$|       |       | 0  | 2  | −13.14                |
| 11    |     | CH$_3$| CH$_3$|       |       | 0  | 2  | −12.74                |

| Bace1 | Cpd | R1  | R2   | n1 | n2 | Exp. dG$_{\text{bind}}$ |
|-------|-----|-----|------|----|----|-----------------------|
| 12    |     | H   | Cyclohexyl| 1  | 2  | −9.82                 |
| 13    |     | Cyclohexyl| Cyclohexyl| 2  | 1  | −11.29                |
| 14    |     | Cyclohexyl| Cyclohexyl| 2  | 2  | −10.21                |
| 15    |     | Cyclohexyl|       | 2  | 1  | −10.57                |

Figure 1. (a–c) Markush representations of macrocycle candidate compounds.
**Results**

**Determining conserved bioactive substructure.** Figure 2a–c shows the macrocyclic designs restrained to and superimposed onto their crystal bioactive references with conserved atoms highlighted. As is shown in this figure, the conserved atoms comprise the majority of the heavy atoms. While the MCS approach to determining the conserved atoms may be a reasonable naive guess, the MCS atoms do not necessarily represent the most important atoms in terms of bioactive structure. For example in Chk1, the conserved includes atoms span across the aryl ethers. It could be conjectured that these ethers serve only as structural atoms and thus the representative bioactive substructure should be curtailed at the aromatic rings. Additionally, the MCS algorithm is unaware of conformation, so for Chk1, there is an off-linker carbon included in the conserved region. Since the \( \text{RMSD}_{\text{cons}} \) calculations use the minimum RMSD across all degenerate matches, this is partially accounted for. But the calculation could still incorrectly favor macrocycles which prefer the wrong conformation of this MCS atom. Such issues would not occur if conserved atoms were manually selected by the macrocycle designer.

**RMSD vs energy plots.** Figure 3 shows the relative energy of each output conformation (strain) vs the \( \text{RMSD}_{\text{MCS}} \). The nature of the conformation generation algorithm generates an ensemble of conformers spanning both \( \text{RMSD}_{\text{cons}} \) and strain space. Different compounds will have different conformational landscapes, and, importantly, may or may not be able to adopt low strain, low \( \text{RMSD}_{\text{cons}} \) conformations (bottom left corner of the graphs). As can be seen, the compounds with better affinity (lower \( \Delta \text{G}_{\text{bind}} \)) tend to better occupy this space. The only clear exception is Chk1 ligand 3 which despite having binding affinity only 0.09 kcal/mol greater (within experimental error) of compound 5, does not contain any conformations in this space. This is likely due to the conserved atom selection (discussed in the Discussion section).

**Expected RMSD vs affinity.** Though the \( \text{RMSD}_{\text{cons}} \) vs energy plots can be useful for analyzing the conformational space of these candidate compounds, for quantitative comparison, a single metric as a function of these data is necessary. Figure 4a–c shows the Boltzmann-weighted expectation value of the \( \text{RMSD}_{\text{cons}} \) among the output ensemble. Here, we expect the lower expected RMSD to have the lower \( \Delta \text{G}_{\text{bind}} \). For Hsp90 and Bace-1, the more potent compounds are clearly separated from the rest. For Chk1 despite correctly ranking the most potent compound (ligand 2) as having the lowest expected RMSD, the intermediate compounds do not correlate well (especially ligand 3 as noted in the previous section). The Pearson correlation coefficients were 0.21, 0.90, and 0.90 for Chk1, Hsp90, and Bace-1 respectively, giving a weighted average \( R^2 \) of 0.60. But for Chk1 and Bace1, which had only four data points each and dynamic ranges less than 1.5 k cal/mol, these correlations were not significant (p-values were 0.78 and 0.10 respectively). The Hsp90 set rather, containing seven data points across 7 kcal/mol, had a p-value of 0.005.

---

**Figure 2.** (a–c) Macro cyclic ligands restraint and superimposed to their bioactive linear references. Reference ligands shown with green carbons. MCS atoms shown as balls and sticks. For Chk1, Bace-1 and HSP90, the MCS SMARTS patterns were, \( c1c(Cl)ccc(OC)c1NC(=O)Nc2cncc(n2)OCC, CCCN(C1)C(N)=[N+]c(c12)ccc(c2)Oc3cccc3, \) and \( c1cc(C(N)=O)ccc1n(c(c2C)CC)c(c23)CC(C)(C)CC3=O \) respectively.
Calculation time. The vast majority of the calculation time for this algorithm is in the sampling since the RMSD<sub>cons</sub> and <RMSD<sub>cons</sub> algorithm takes only a few seconds per compound on a CPU. The PrimeMCS sampling algorithm is parallelized over 10 threads per compound. For all the sets here, the median parallel calculation time on commodity CPU workstations was 88 seconds per compound, or about 12 minutes per compound on a single CPU. Thus, each CPU could profile over 120 compounds in a single day, giving designers with even modest resources the ability to screen thousands of compounds per day for similarly-sized macrocycles. As noted in the PrimeMCS reference<sup>12</sup>, speed will vary depending on the conformational complexity of the macrocycle and additional sampling or appropriate restraints may increase the accuracy of the method.

Discussion

Limitations of the hypothesis. As mentioned previously, there are strong assumptions in the hypothesis underlying this method. It assumes that higher propensity for the purported bioactive substructural conformation leads to higher affinity. First, this assumption excludes possibility of interactions of the nonconserved regions with the protein and any binding solvation effects. However, since many linker designs are in solvent exposed regions, such effects may be small depending on the choice and size of linker. Secondly, the assumption excludes the possibility that the known active structure may not adopt the ideal geometry, that alternate conformations may also imbue activity, or even that some ligand flexibility is required to maintain interactions in a flexible active site. The ability to dismiss these possibilities must be determined on a per-project basis. Due to these limitations, the method may not be applicable to all macrocycle design projects or may only be applicable in portions of the project when the assumptions hold true.

Limitations of the metric. Here, we applied the <RMSD<sub>cons</sub> as our proxy for affinity. With respect to RMSD<sub>cons</sub> for example, the linker may contribute, only a small subset of the conserved region could contribute, or other structural features may be more important, such as r-group torsional propensity or relative polar group vectors. Though RMSD<sub>cons</sub> may offer a good naive metric, we recommend that designers use judgment and testing to choose the appropriate structural metric. Although the expectation value is a simple non-parametric measure of the distribution of the values, it may not be the ideal way to combine results from multiple conformers. An alternative metric could be, for example, the probability of the compound meeting an RMSD<sub>cons</sub> threshold. This

Figure 3. (a–c) Relative energy of conformation (strain) vs RMSD in the MCS region relative to the reference for Chk1, Hsp90 and Bace1. Each conformation represents one point. Only the 12 lowest strain conformations for each compound are shown. In parentheses the legends indicate the experimental binding affinity.
would make sense if the compound only binds if it meets that RMSD threshold (but binds no better or worse
the otherwise larger or smaller the RMSD cons
gets). For example, if a 1.0 Å structure were necessary to bind, the
function would be

$$\text{P}(\text{RMSD cons} \leq 1.0 \text{ Å})$$

A similar metric was applied by McCoull et al. in a BCL6 macrocycle
design project.16

Diagnosing the method. Due to the many assumptions of the method, it may not be clear a priori whether
this method is applicable to the macrocycle design project at hand. It is recommended that the designers confirm
that the strain-rmsd profiles (Fig. 3a–c) behave as expected, that the low energy conformations of any known
macrocyclic binders can access the bioactive substructure. Limitations of the metric (as mentioned above), and
sampling protocol may need to be adjusted. For example, if we were to use the Chk1 results as a retrospective
to propose prospective linkers, the aberrant behavior of compounds 3 would draw further critique. This could
possibly be explained by the large MCS region (visible in Fig. 2a) which extends through most of the linker. One
could argue that though this region is “conserved” (at least by the MCS method), it is not a good representation
of the bioactive substructure. Performing the calculation with an abbreviated SMARTS pattern to determine the
conserved region,

$$n1c(O)cncc1NC(=O)Nc2c(O)ccc(c2)Cl$$

which removes now omits most of the linker other than
the first atom on each end, suggests improved results ($R^2$ goes from from 0.05 to 0.27, though the p-value, 0.48,
does not suggest this is statistically significant), see Fig. 5.

The sampling itself can potentially be an area where modification to the protocol is necessary. For example,
for an extremely large macrocycle (>40 ring atoms), increasing sampling settings or constraints to focus the
sampling may be necessary. Additionally, since the hypothesis relies on the impact of the cycle topology on the
conformation, interactions of extended side chains, for example, may introduce noise into the strain calculation.

Additional concerns about proxy metric, force field, and solvent parameters may also be investigated by the
modeller in a project-dependent manner. The default solvent and force field parameters were used here. We did
also redo these calculations in VSGB2.1 solvent model16 to see if that would significantly affect the results. The
results were approximately unchanged, yielding a weighted $R^2$ of 0.61 (from 0.60). This change is insignificant and
we do not believe the improvement is worth rationalization.

Use as a screening tool. The assumptions of this method can be managed when evaluating a large number
of cyclizations or otherwise enumerated macrocycle topologies for purpose of structural stabilization. For such
cases, thousands of compounds can be quickly triaged down to those few compounds likely to maintain the
bioactive conformation. This tool could potentially be coupled with automated design tools such as linker enu-
meration or genetic-algorithm-type linker evolution methods. However, beyond this, more accurate screening
methods, such as free energy perturbation, will be necessary to further probe interactions, solvent effects, or more
precisely determine entropic factors contributing to the free energy of binding.
Conclusions

We have described a high-throughput method for profiling macrocycles propensity to adopt the bioactive conformation of their linear progenitor and shown that this propensity is a useful proxy for the binding affinity in cases where the binding mode is both constant and known and when the linker is not forming interactions with the protein. The systems here were limited to publically available data. But despite this limitation and the simplicity of the macrocycle stabilization algorithm, good performance was observed ($R^2$ weighted average 0.60). We expect that in a design context, the datasets will be larger and the $<\text{RMSD}_{\text{con}}>$ metric will likely have to be modified to properly test the macrocyclization strategy based on the particular structural motivation. For example, as demonstrated in the discussion section, a slight, rational modification significantly improved the correlation. We have also proposed options and rationales for varying the metric for such cases. We note that, though care must be taken to ensure that the results here are not over-extrapolated (e.g. since the metric cannot, in any way, predict new interactions or differential solvation effects), this method could be extremely useful in rapidly triaging large numbers macrocycle candidate compounds.

References

1. Villar, E. A. et al. How proteins bind macrocycles. Nature Chemical Biology 10, 1–10 (2014).
2. Still, W. C. & Galyaner, I. Chemical consequences of conformation in macrocyclic compounds. Tetrahedron 37, 3981–3996 (1981).
3. Dunietz, J. D. & Ibers, J. A. Perspectives in Structural Chemistry (1968).
4. Giordanetto, F. & Kihlberg, J. Macrocyclic Drugs and Clinical Candidates: What Can Medicinal Chemists Learn from Their Properties? J. Med. Chem. 57, 278–295 (2014).
5. Vendeville, S. & Cummings, M. D. Synthetic Macrocycles in Small-Molecule Drug Discovery. Annual Reports in Medicinal Chemistry 48, 371–386 (Copyright © 2013 Elsevier Inc. All rights reserved., 2013).
6. Mallinson, J. & Collins, I. Macrocycles in new drug discovery. Future Medicinal Chemistry 4, 1409–1438 (2012).
7. Marsault, E. & Peterson, M. L. Macrocycles Are Great Cycles: Applications, Opportunities, and Challenges of Synthetic Macrocycles in Drug Discovery. J. Med. Chem. 54, 1961–2004 (2011).
8. Weininger, D. Daylight Information Systems: Rubicon. (2011). Available at: http://www.daylight.com/dayhtml/doc/rubicon/#RTFToC32. (Accessed: 2017)
9. Labute, P. LowModeMD—Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. J Chem Inf Model 50, 792–800 (2010).
10. Watts, K. S., Dalal, P., Tebben, A. J., Cheney, D. L. & Shelley, J. C. Macrocycle Conformational Sampling with MacroModel. J Chem Inf Model 54, 2680–2696 (2014).
11. Couttias, A. E., Lexa, K. W., Wester, M. J., Pollock, S. N. & Jacobson, M. P. Exhaustive Conformational Sampling of Complex Fused Ring Macrocycles Using Inverse Kinematics. J. Chem. Theory Comput. acs.jctc.6b00250–63, https://doi.org/10.1021/acs.jctc.6b00250 (2016).
12. Sindhikara, D. et al. Improving Accuracy, Diversity, and Speed with Prime Macrocycle Conformational Sampling. J Chem Inf Model acs.jcim.7b00052–14, https://doi.org/10.1021/acs.jcim.7b00052 (2017).
13. Priestley, E. S. et al. Structure-Based Design of Macrocyclic Coagulation Factor VIIa Inhibitors. J. Med. Chem. 58, 6225–6236 (2015).
14. Bhardwaj, G. et al. Accurate de novo design of hyperstable constrained peptides. Nature 1–21, https://doi.org/10.1038/nature19791 (2016).
15. Tran, H. L. et al. Structure–Activity Relationship and Molecular Mechanics Reveal the Importance of Ring Entropy in the Biosynthesis and Activity of a Natural Product. J. Am. Chem. Soc. 139, 2541–2544 (2017).
16. McCoull, W. et al. Discovery of Pyrazolo[1,5- a]pyrimidine B-Cell Lymphoma 6 (BCL6) Binders and Optimization to High Affinity Macrocyclic Inhibitors. J. Med. Chem. 60, 4386–4402 (2017).
17. Harder, E. et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. J. Chem. Theory Comput. 12, 281–296 (2016).
18. Schrödinger, L. L. Schrödinger Release 2017-3: Canvas. Available at: www.schrodinger.com/canvas (Accessed: 2017).
19. Li, H., Li, G., Berg, B. A. & Yang, W. Finite reservoir replica exchange to enhance canonical sampling in rugged energy surfaces. J. Chem. Phys. 125, 144902 (2006).
20. Ruscio, J. Z., Fawzi, N. L. & Head Gordon, T. How hot? Systematic convergence of the replica exchange method through multiple reservoirs. J Comput Chem 31, 620–627 (2010).
21. Okur, A., Miller, B. T., Joo, K., Lee, J. & Brooks, B. R. Generating Reservoir Conformations for Replica Exchange through the Use of the Conformational Space Annealing Method. J. Chem. Theory Comput. 9, 1115–1124 (2013).
22. Yu, H. S., Deng, Y., Wu, Y. & chemical, D S. J. O. Accurate and Reliable Prediction of the Binding Affinities of Macromolecules to Their Protein Targets. ACS Publications, https://doi.org/10.1021/acs.jctc.7b00885 (2017).

Figure 5. Proxy affinity metric, $<\text{RMSD}_{\text{con}}>$ vs experimentally measured affinity for Chk1 with a smaller conserved region.
23. Baxter, E. W. et al. 2-Amino-3,4-dihydroquinazolines as Inhibitors of BACE-1 (β-Site APP Cleaving Enzyme): Use of Structure Based Design to Convert a Micromolar Hit into a Nanomolar Lead. J. Med. Chem. 50, 4261–4264 (2007).
24. Zapf, C. W. et al. Discovery of a stable macrocyclic o-aminobenzamide Hsp90 inhibitor which significantly decreases tumor volume in a mouse xenograft model. Bioorganic & Medicinal Chemistry Letters 21, 4602–4607.
25. Schrödinger Release 2017-3: LigPrep. Available at: www.schrodinger.com/ligprep (Accessed: 2017)
26. Li, J. et al. The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. Proteins 79, 2794–2812 (2011).

Acknowledgements
We would like to thank Tyler Day for helpful discussions.

Author Contributions
D.S. Conceived of the method, wrote the code, and ran analyses. K.B. advised research. All authors reviewed the manuscript.

Additional Information
Competing Interests: D.S. and K.B are employees of Schrodinger, LLC. Distributor of Schrodinger software.

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

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

© The Author(s) 2018