Modeling and System preparation

System preparation, modeling and initial docking calculations were performed using the Schrödinger molecular modeling suite (version 2015-4), unless otherwise noted. The protein structure was obtained from the Protein Data Bank (PDB)1 and prepared using the Protein Preparation Wizard2. In this step, force field atom types and bond orders are assigned, missing atoms are added, tautomer/ionization states are assigned, water orientations are sampled, Asn, Gln, and His residues are flipped to optimize the hydrogen bond network, and a constrained energy minimization is performed. All resolved crystal water molecules were retained during the preparation.

Potential binding sites were explored and characterized using the SiteMap3,4 tool. Ligands that have shown activity in experiments together with known inactives were docked in the putative binding sites using Glide SP5,6 in order to test enrichment of known actives. Reasonable scores for the ‘Ro’ series was shown for the ‘Gossypol’ binding site described by Lan *et al*.7

Since receptor structure may not be in the optimal conformation to bind small molecule inhibitors, induced fit docking8 of ligand Ro 08-2750 was performed to this binding pocket. Induced fit docking results were validated with the Metadynamics protocol described by Clark *et al*9. The pose ranked second using the Induced Fit Docking score came out best. This receptor configuration was furthermore validated towards a virtual screening using a Glide SP docking of known actives and inactives. Furthermore, a WaterMap10,11 calculation was done for this receptor.

The virtual screening was then preformed with this receptor conformation using Glide SP by docking the March 2016 collection of the eMolecules dataset. All ligand structures were prepared with LigPrep including a minimization with the OPLS3 force field12. One low energy ring conformation per compound was generated. Ionization states and tautomer forms were enumerated at pH 7.0 ± 2.0 with Epik13,14.

The top 5000 hits from virtual screening were filtered by applying filters according to Lipinski’s rule of five15, flagging REOS16 and PAINS

17.

The hitlist was ranked in addition to the Glide SP DockingScore also by a Pareto ranking of DockingScore and number of WaterMap hydration sites with Δ*G* > 2 kcal/mol which overlap with the ligand pose. The top 200 ranked hits from both lists were combined. Finally, a leader-follower clustering using dendridic fingerprints was performed using Canvas18,19 resulting in 243 unique cluster hits.

Induced Fit Docking of Ro-A6 and Ro-OH compounds

Induced Fit Docking was performed against the receptor pose from the selected Ro 08-2750 pose, using Schödinger molecular modeling suite (version 2017-4). Poses for Ro-A6 and Ro-OH, the top and second scored poses respectively, were selected to most closely match the Ro 08-2750 pose.

Alchemical Free Energy Calculations

*System Preparation* *and modeling*. The protein and ligand poses generated by induced fit docking were selected for input files. Because the proteins and ligands were already prepared, they were simply run through the pdbfixer command line tool with add-atoms and add-residues set to None. This was done to convert residue and atom names to be tleap compatible.

*Parameterization.* tleap (ambermini 16.16.0) was used to solvate the complex in a cubic box with a 12Å buffer of TIP3P water molecules around the protein. The system was parameterized using AMBER’s forcefield ff14sb and GAFF 1.8. Missing ligand parameters were determined using antechamber. The ligand was assigned charges using the AM1-BCC implementation in OpenEye (OEtoolkit 2017.6.1 through openmoltools 0.8.1).

*Minimization.* Minimization was performed using the implementation of the L-BFGS algorithm in OpenMM 7.1.120 with a tolerance of 1kJ/(M\*nm).

*Production Simulation.* Production simulation was run using YANK 0.19.4 using OpenMMTools 0.13.4. The ligand was confined to the binding site using a Harmonic restraint (K =0.33 kcal/mol\*Å2) centered on the following residues in the receptor: 2, 4, 46, 76, 78, and 80. The calculation was performed using an explicit PME solvent, with a nonbonded forces cutoff using a 9Å cutoff and four neutralizing Cl-. The calculation was carried out using a Langevin integrator (VRORV splitting) set at 300K with a 2fs timestep, and a Monte Carlo barostat was used to maintain 1 atm pressure . Ro 08-2750 and Ro A6 were run for 10000 iterations with 2500 timesteps per iteration, while Ro-OH was run for 15000 iterations with 2500 timesteps per iteration. A Hamiltonian Replica Exchange step was performed at each iteration using the Gibbs sampling scheme described previously21. The alchemical pathway was automatically determined for each compound using the YANK autoprotocol feature.

*Free Energy Estimates.* ΔG of binding for each compound was estimated using MBAR22

*Clustering analysis.* The fully interacting trajectory from YANK was extracted to a pdb file, discarding initial iterations prior to equilibration23: 1500 for Ro 08-2750, 1600 for RoOH, and 1600 for RoA6. These trajectories were aligned in MDTraj24 using only protein backbone atoms. The small molecules were then sliced out and clustered on Cartesian coordinates using the MSMBuilder25 implementation of RegularSpatial clustering, using a 1Å RMSD cutoff. For the most populated clusters for Ro 08-2750 and RoOH, cluster centers were selected and shown with 10 randomly sampled cluster members. RoA6 produced a large number of lowly populated clusters with highly heterogeneous binding poses, and were therefore not shown.

*Conformational Heterogeneity analysis.* To investigate the conformational heterogeneity in the presence or absence of the ligand, the fully interacting state and fully non-interacting states for all three ligands were extracted using a 4-frame skip, discarding the initial frames as above.

Author affiliations:

Daniel Cappel: Schrödinger GmbH, Dynamostraße 13, 68165 Mannheim, Germany

Steven K. Albanese: 1) Gerstner Sloan Kettering Graduate School, Memorial Sloan Kettering Cancer Center, New York, NY 10065 2) Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York, USA,

Andrea Rizzi: 1) Tri-Institutional Training Program in Computational Biology and Medicine, New York, NY, USA 2) Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Levi Naden: Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York, USA,

1. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28,** 235–242 (2000).

2. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R. & Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **27,** 221–234 (2013).

3. Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model* **49,** 377–389 (2009).

4. Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem Biol Drug Des* **69,** 146–148 (2007).

5. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47,** 1739–1749 (2004).

6. Halgren, T. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47,** 1750–1759 (2004).

7. Lan, L. *et al.* Natural product (-)-gossypol inhibits colon cancer cell growth by targeting RNA-binding protein Musashi-1. *Mol Oncol* **9,** 1406–1420 (2015).

8. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A. & Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **49,** 534–553 (2006).

9. Clark, A. J. *et al.* Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J. Chem. Theory Comput.* **12,** 2990–2998 (2016).

10. Abel, R., Young, T., Farid, R., Berne, B. J. & Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.* **130,** 2817–2831 (2008).

11. Young, T., Abel, R., Kim, B., Berne, B. J. & Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein-ligand binding. *PNAS* **104,** 808–813 (2007).

12. 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).

13. Shelley, J. C. *et al.* Epik: a software program for pK( a ) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **21,** 681–691 (2007).

14. Greenwood, J. R., Calkins, D., Sullivan, A. P. & Shelley, J. C. Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **24,** 591–604 (2010).

15. Lipinski, C. A., Lombardo, F., Dominy, B. W. & Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **46,** 3–26 (2001).

16. Walters, W. P., Stahl, M. T. & Murcko, M. A. Virtual screening—an overview. *Drug Discov. Today* **3,** 160–178 (1998).

17. Baell, J. B., chemistry, G. H. J. O. M.2010. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *ACS Publications*

18. Duan, J., Dixon, S. L., Lowrie, J. F. & Sherman, W. Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J. Mol. Graph. Model.* **29,** 157–170 (2010).

19. Sastry, M., Lowrie, J. F., Dixon, S. L. & Sherman, W. Large-Scale Systematic Analysis of 2D Fingerprint Methods and Parameters to Improve Virtual Screening Enrichments. *J Chem Inf Model* **50,** 771–784 (2010).

20. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* **13,** e1005659 (2017).

21. Chodera, J. D. & Shirts, M. R. Replica exchange and expanded ensemble simulations as Gibbs sampling: simple improvements for enhanced mixing. *J Chem Phys* **135,** 194110 (2011).

22. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* **129,** 124105 (2008).

23. Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **12,** 1799–1805 (2016).

24. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109,** 1528–1532 (2015).

25. Beauchamp, K. A. *et al.* MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J. Chem. Theory Comput.* **7,** 3412–3419 (2011).