

RESEARCH ARTICLE

Hot-spot identification on a broad class of proteins and RNA suggest unifying principles of molecular recognition

John L. Kulp, III^{1,2}, Ian S. Cloudsdale¹, John L. Kulp, Jr¹, Frank Guarnieri^{3*}

1 Conifer Point Pharmaceuticals, Doylestown, Pennsylvania, United States of America, **2** Department of Chemistry, Baruch S. Blumberg Institute, Doylestown, Pennsylvania, United States of America, **3** PAKA Pulmonary Pharmaceuticals, Acton, Massachusetts, United States of America

* frankguarnieri@yahoo.com



OPEN ACCESS

Citation: Kulp JL, III, Cloudsdale IS, Kulp JL, Jr, Guarnieri F (2017) Hot-spot identification on a broad class of proteins and RNA suggest unifying principles of molecular recognition. PLoS ONE 12 (8): e0183327. <https://doi.org/10.1371/journal.pone.0183327>

Editor: Manuela Helmer-Citterich, Universita degli Studi di Roma Tor Vergata, ITALY

Received: January 24, 2017

Accepted: August 2, 2017

Published: August 24, 2017

Copyright: © 2017 Kulp et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Dr. Guarnieri is an employee of PAKA Pulmonary Pharmaceuticals and affiliated with Virginia Commonwealth University. Dr. Kulp III is an employee of Conifer Point Pharmaceuticals and an employee of the nonprofit, The Baruch S. Blumberg Institute. Dr. Kulp JR and Dr. Cloudsdale are employees of Conifer Point Pharmaceuticals. PAKA provided support in the form of salary for Dr.

Abstract

Chemically diverse fragments tend to collectively bind at localized sites on proteins, which is a cornerstone of fragment-based techniques. A central question is how general are these strategies for predicting a wide variety of molecular interactions such as small molecule-protein, protein-protein and protein-nucleic acid for both experimental and computational methods. To address this issue, we recently proposed three governing principles, (1) accurate prediction of fragment-macromolecule binding free energy, (2) accurate prediction of water-macromolecule binding free energy, and (3) locating sites on a macromolecule that have high affinity for a diversity of fragments and low affinity for water. To test the generality of these concepts we used the computational technique of Simulated Annealing of Chemical Potential to design one small fragment to break the RecA-RecA protein-protein interaction and three fragments that inhibit peptide-deformylase via water-mediated multi-body interactions. Experiments confirm the predictions that 6-hydroxydopamine potently inhibits RecA and that PDF inhibition quantitatively tracks the water-mediated binding predictions. Additionally, the principles correctly predict the essential bound waters in HIV Protease, the surprisingly extensive binding site of elastase, the pinpoint location of electron transfer in dihydrofolate reductase, the HIV TAT-TAR protein-RNA interactions, and the MDM2-MDM4 differential binding to p53. The experimental confirmations of highly non-obvious predictions combined with the precise characterization of a broad range of known phenomena lend strong support to the generality of fragment-based methods for characterizing molecular recognition.

Introduction

The nature of how small molecules bind to proteins continues to be the object of many experimental[1–6] and theoretical[7–9] studies. Early investigations into how diverse small organic compounds, which are commonly referred to as fragments, collectively bind to localized sites on proteins using X-ray crystallography[10, 11] or NMR[12] were originally developed as new

Guarnieri, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Conifer Point provided salary support for Dr. Kulp III and Dr. Kulp JR and these individuals had a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Specific roles are articulated in the 'authors contributions' section. Part of the research reported in this publication was supported by National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health under award number R43GM109549.

Competing interests: Dr. Guarnieri is an employee of PAKA Pulmonary Pharmaceuticals and affiliated with Virginia Commonwealth University. Dr. Guarnieri has received financial support from PAKA. Dr. Kulp III is an employee of Conifer Point Pharmaceuticals and an employee of the nonprofit, The Baruch S. Blumberg Institute. He receives salary support from both these organizations. He also receives financial support from Fox Chase Chemical Diversity Center and Small Molecule PPI Mimics. Dr. Kulp III is also a Regional Assistant Dean for the Geisinger Commonwealth School of Medicine where he teaches professional development, seminars, biochemistry, and pharmacology. Dr. Kulp JR is an employee of Conifer Point Pharmaceuticals and has received financial support from Conifer Point. Dr. Kulp JR is also the principal investigator on the NIH grant R43GM109549. Dr. Cloudsdale has received no financial support from Conifer Point. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

methods for generating lead compounds in drug design. The overlapping and adjacent binding of fragments suggests how these entities might be chemically linked into larger higher affinity molecules—a necessary process for lead generation since fragments usually bind to proteins with very low affinity,[13, 14] often in the millimolar range. Fragment-based approaches have produced drug candidates, validating the clinical relevance[15, 16] of these methods. They have also made fundamental contributions to the basic understanding of protein-protein[17–23] interactions (PPI). PPIs often have long extensive interfaces that enable a multitude of potentially stabilizing interactions spread across a vast surface. Studies of fragment-protein interactions, however, indicate that in many cases the association free energy responsible for PPIs occurs at highly localized positions within the interface: so called “hot spots”. This finding not only furthers our basic understanding of PPIs, but pragmatically indicates that drugs designed to break PPIs should be targeted to these sites. Thus, fragment-based approaches have now become important methods in both pharmaceutical and academic research.

Experimental fragment-based methods are extremely resource intensive and thus it would be highly desirable to perform complimentary computer simulations that could assist with focusing and minimizing such experiments. These considerations led to the development of the GRID[24] and MCSS[25] algorithms for studying protein-fragment interactions. The realization, however, that these methods incorrectly overestimate the number of hot spots on proteins motivated Vajda[26] to develop the “solvent mapping” technique, which has proven to be highly accurate in identifying protein hot spots in a range of important[27–32] studies. Solvent mapping is a heuristic algorithm that enhances the sampling of fragments over the whole protein surface by initially zeroing out the van der Waals term and part of the solvation energy. This has the effect of dramatically smoothing out the protein surface, which eliminates a large number of small local minima that a probe fragment could inadvertently get stuck in. After thousands of potential fragment-protein interaction sites are located with a multi-start simplex method, a second minimization with van der Waals interactions and all solvation terms turned on is performed, which now recreates a realistic representation of the protein surface, to determine the multitude of final positions for a single fragment. The last step is to identify and energy rank localized sites on the macromolecule predicted to have high affinity for a broad diversity of fragments, the so-called clustering step.

A compelling alternative to an *ad hoc* method like solvent mapping would be to generate a Boltzmann weighted distribution of fragment binding sites over the whole macromolecule with grand canonical Monte Carlo (GCMC), because this simulation produces the theoretically correct set of fragment-macromolecule binding free energies, not simply minimized enthalpies. The grand canonical ensemble probability density function— $\exp[-(\Delta E + \mu N)/RT]$ —is the Boltzmann distribution augmented with the chemical potential (μ) that accounts for changing the number of particles (N) at fixed temperature (T) with R being the universal gas constant. Using a standard Monte Carlo algorithm, a fragment is randomly inserted or deleted into the protein simulation cell, the ΔE is computed for this change, and this new trial configuration is accepted or rejected with the grand canonical probability density function. When this process is repeated for millions of steps the formally correct distribution of fragment binding states is produced in the (V, T, μ) ensemble. This notation means that volume, temperature, and chemical potential are initially set and kept constant throughout the simulation. At a given fixed chemical potential GCMC commonly produces dozens of sites on the macromolecule that have high affinity for the fragment. The term affinity here means binding free energy (not enthalpy), because chemical potential is a formally correct free energy. GCMC has nothing to say, however, about relative binding affinities—all sites populated with fragments are deemed equally probable and thus there is no way of distinguishing which one of these sites on the macromolecule has higher or lower affinity for the fragment. Quantitatively predicting the

rank order difference in binding free energy of these different and distinct populated sites can be done by running a sequence of GCMC simulations at successively decreasing values of the chemical potential—a technique known as Simulated Annealing of Chemical Potential[33] (SACP). As the chemical potential is gradually lowered, individual fragment-occupied sites are successively depopulated. The weaker binding sites are depopulated with a small lowering of the chemical potential while the stronger, higher affinity, binding sites require significantly more lowering of the chemical potential to depopulate. SACP gives the chemical potential when every particular site is vacated and thus a quantitative and formally correct free energy metric of the relative binding of a fragment to different sites on the macromolecule is obtained. Furthermore, SACP has proved to be especially valuable in understanding the hydration free energy on a macromolecular surface. Hydration free energy is especially difficult to predict, because it is often a multibody water-water-macromolecule phenomena. SACP is a uniquely powerful way of discovering multibody water-macromolecule interactions, which was first illustrated when the method was used to predict the differential hydration propensities between the major and minor grooves of DNA[33] in 1996.

The results of a SACP simulation using water as the probe fragment are illustrated in Fig 1. At high values of the chemical potential insertion of water into the protein cavities and bulk has a high probability and the system becomes densely hydrated (Fig 1a). The chemical potential is lowered by a small amount until equilibrium is re-established. This process is repeated, resulting in very little change in the system, until a chemical potential that is low enough to break the water-water solvent cohesion occurs. At this point, the chemical potential probabilistically favors water deletion over water insertion and the system goes through a phase transition (Fig 1b) resulting in evacuation of almost all waters from the protein simulation cell (Fig 1c). As the chemical potential is lowered further even the identified tightly bound water molecules can be removed from the protein—the additional increment of chemical potential required to finally eliminate these waters indicates their relative free energy of binding.

There are several key benefits of this phase transition method: (i) the high positive starting value for the chemical potential causes the simulation cell to become densely packed, forcing fragments into all interstices of the protein, which efficiently samples multi-fragment configurations—particularly critical for water modeling; (2) as the chemical potential is annealed to lower values, the system goes through a phase transition where the regions of the simulation cell away from the protein surface are voided—the chemical potential at which this happens distinguishes strongly bound fragments from transient weakly interacting fragments; and (3) as the chemical potential is lowered beyond the phase transition, fragment binding sites become isolated. In this latter regime, fragments have a negligible interaction with other fragments, and the chemical potential—the average free energy per molecule—can then be used to characterize the free energy of the fragment binding in each individual site.

While SACP is a powerful foundation for accurately predicting relative fragment-macromolecule binding free energies, it requires additional principles for investigating general molecular recognition phenomena. We previously proposed three governing principles—accurate prediction of 1) small organic fragment macromolecule binding free energy, 2) hydration free energy, and 3) macromolecule sites that have high affinity for a diversity of organic fragments and low affinity for water.[34, 35] These would apply without regard to any particular method used: either experimental or computational. If these three principles can be accomplished with any fragment-based technique, then a wide variety of seemingly disparate molecule recognition phenomena could be characterized and predicted. SACP is one particular method for assessing these principles, because it is easily generalized[36] beyond investigating water interactions—the binding free energy of any organic fragment to any macromolecule can be determined with these simulations. Specifically, a set of fragments

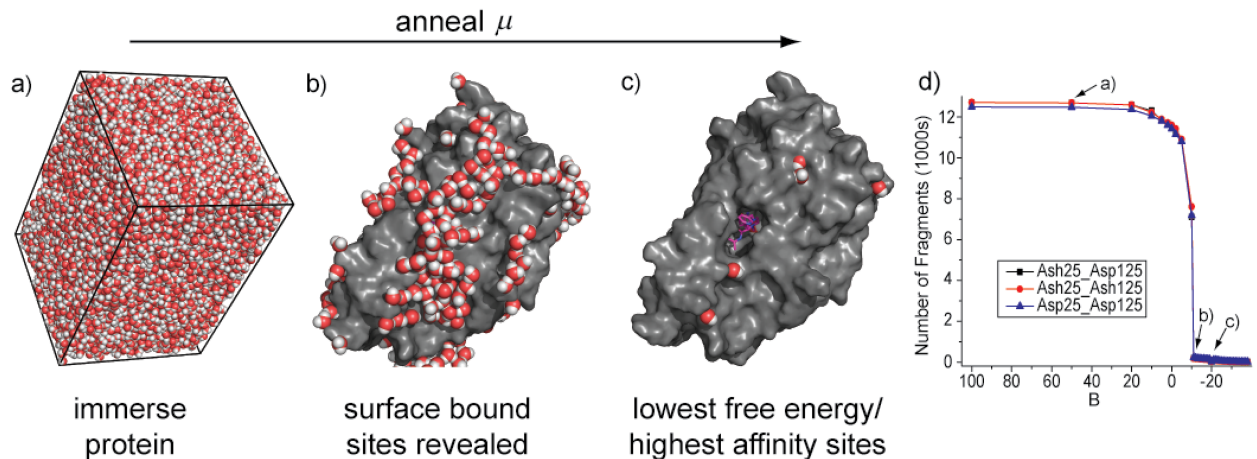


Fig 1. The chemical potential simulation. The protein is placed in a simulation box big enough to accommodate 3 layers of solvent—water in this case. The simulation is started with a high chemical potential and solvent molecules rapidly fill the box and occupy (a) all possible cavities at the surface and deep within the protein. As the chemical potential is lowered, not much happens until this free energy reaches a critical level that is capable of destroying the solvent cohesion and the system goes through a phase transition that evacuates the bulk solvent and most of the protein-bound solvent, exposing high affinity (b) binding sites. Continuing to lower the chemical potential further reveals the extremely high affinity (c) binding sites. A plot of the number of fragments in the simulation box versus the chemical potential shows the abrupt transition when the bulk solvent is evaporated from the cell. The charged state of the catalytic aspartates in HIV protease (Asp = charged, Ash = neutral) does not change the phase transition.

<https://doi.org/10.1371/journal.pone.0183327.g001>

representing a large variety of diverse organic functional groups and water are simulated with SACP on a macromolecule, a cluster analysis is performed to find the localized sites of the macromolecule that have a high affinity for a diversity of fragments, and a cluster site is eliminated if water is also predicted to bind there with high affinity (so-called water exclusion principle). The clustering set of fragments contains 13 entities (S2 Fig). The set contains all small relatively flat fragments. There is a mixture of hydrophobic and hydrophilic fragments as well as a variety of hydrogen bond done and accept types. These properties of the clustering set map out the chemical properties of the various biomolecule surface. SACP has been successfully applied (i) to show how mitochondrial aspartate amino transferase moonlights[37] as a lipid transporter, (ii) to the binding site predictions[35] of diverse wild-type and mutated hen egg white lysozymes, (iii) how pregnenolone binds to and modulates the cannabinoid[38] receptor, (iv) to the design of a very small molecule erythropoietin[39] agonist, (v) p38[40, 41] non-ATP site antagonists, and (vi) a cancer drug discovery[42] program.

To demonstrate the generality of this method, we ran the SACP system on a diverse set of examples: elastase, HIV protease, dihydrofolate reductase, HIV TAT-TAR, MDM2-MDM4-p53, RecA, and peptide deformylase (PDF). The first[43] co-crystal of elastase was reported in 1976 with subsequent structures[44–46] all producing a similar ligand binding position. In 1994, new co-crystals[47] with ligands almost identical to previous studies revealed that the binding site was dramatically larger. By contrast dihydrofolate reductase has an obviously very large[48] binding site, which holds the ligand and co-factor, but functional analysis reveals that the electron transfer process occurs at a pinpoint location within the binding site. In HIV Protease, there is a key bound water[49] molecule in the center of the binding site and thus this is an important protein for testing SACP. The binding of MDM2 and MDM4 to p53 provides a stringent test of a method's ability to distinguish between subtly different homologous protein-protein[50–55] interactions. Because p53 dysfunction occurs in more than[56] 50% of all cancers, the ability to dissect such interactions is vital for understanding carcinogenic transformation. TAR is an RNA domain[57, 58] present in all HIV transcripts. When

the HIV TAT protein binds to TAR, transcription is dramatically enhanced, thus this protein-RNA interaction[59] is an appealing target for antiviral drugs. The RecA protein[60] is responsible for repairing damaged[61, 62] bacterial DNA. Dozens of RecA proteins interlock[63] and bind to mutated DNA as a homo-oligomer that performs ATP driven repair operations in a directed[61] low fidelity manner that gives rise to resistant mutations. RecA is not only responsible for drug-resistant mutagenesis, but also for the sharing of the mutated genes resulting in the spreading of the developed[64] resistance. Peptide deformylase is a target for the development of antibacterial[65] and anticancer agents[66, 67]. These systems were chosen for this study, because collectively they represent a vast diversity of molecular recognition phenomena and they are all historically well characterized.

Experimental

Protein structures

Protein structures were downloaded from the Protein Data Bank (PDB).[68] For HIV protease, the 1SDT structure[69] was run under three different protonation states: (i) both catalytic aspartate residues deprotonated, the apo state[70] (Asp25_Asp125), (ii) one aspartate residue protonated and the other deprotonated, the ligand bound state,[70] where Asp25 has a proton on OD2[71] (Ash25_Asp125), and (iii) both aspartate residues protonated (Ash25_Ash125). The antiviral drug, indinavir, was removed prior to simulation. For dihydrofolate reductase, three PDB structures were simulated: 1J3I, 1J3J, and 1J3K, which represent the wild type, double mutant and quad mutant structures, respectively, in *Plasmodium falciparum*. [72] The ligand was removed for each structure and NADPH was retained. In an additional simulation on 1J3I, both the ligand and NADPH were removed. For the elastase enzyme, the 1ELA structure[47] was selected the peptide ligand removed. For the p53-MDM2 enzyme, the 1T4F structure[73] was simulated with the peptide ligand removed. For the p53-MDMX enzyme, the 3FDO structure[74] was used with the peptide ligand removed. For the RNA structure, the 2KX5 structure [59]—HIV TAR RNA with the peptide mimetic of Tat protein removed—was simulated. For the RecA protein, the 1xms structure[75] was simulated using a single monomeric unit with the ATP analog removed. Peptide deformylase, PDF, was modeled from the 1G27 pdb structure. All of the structures were checked, and corrected, for missing atoms or residues with the Profix program.[76] Hydrogen atoms were added with the Reduce program.[77] Protein protonation states were determined with PROPKA[78–81] and applied unless otherwise noted.

Simulated Annealing of Chemical Potential calculations (SACP)

Each protein was placed in a simulation box with minimal dimensions to accommodate at least three layers of fragments. Next, a grand canonical Monte Carlo method was used to calculate fragment distributions.[82] A chemical potential is imposed between the simulation box and an ideal gas reservoir. The fragments are repeatedly inserted and deleted until the system equilibrates at this chemical potential as indicated by the average number of fragments becoming stable.[15, 83] During the simulation, the chemical potential is annealed in a succession of steps.[33] Annealing is carried out by gradually lowering the parameter $B = \mu_{ex}/kT + \ln \langle N \rangle$ where μ_{ex} is the excess chemical potential related to protein interactions and $\langle N \rangle$ is the average amount of fragment molecules in the simulation. The series of steps starts at B equal to +100.0 and there are no fragments in the simulation cell. A defining characteristic of SACP is the clear occurrence of a phase transition. At high chemical potential the entire simulation cell is completely packed with fragments bound in every crevice of the protein and forming three tight packed solvation layers. As the chemical potential is gradually lowered very little change occurs in the system and suddenly within a very small ΔB , the simulation cell becomes

evacuated leaving only tightly bound protein fragments. Further lowering of the chemical potential reveals the relative free energy of binding between these bound fragments as they successively depopulate the protein. We use energy cutoffs instead of the traditional distance cutoffs by pre-computing an optimal cutoff distance for each fragment type based on the maximum interaction energy between any pair of fragment types or residues. 0.1 kT was the cutoff energy contribution applied. Various cutoff distances are applied for atom-atom models and less costly multipole models for residues or fragments. This method has been described in further detail elsewhere.[35, 84]

A simplified but representative SACP protocol is to generate a random number (RanX) between 0 and 1 and to insert a fragment into the protein simulation cell if $\text{RanX} < 0.5$ or delete a fragment if $\text{RanX} > 0.5$. In either case this is a trial move with attempted insertion assigned a probability of $P = \exp[-[E(n+1)-E(n)]/(RT)] * \exp[B]/(N+1)$ and attempted fragment deletion a probability of $P = N * \exp[-[E(n-1)-E(n)]/(RT)] * \exp[-B]$ where $E(n)$ is the energy of the system before an insert or delete, $E(n+1)$ is the energy of the system after a fragment is inserted, and $E(n-1)$ is the energy of the system after a fragment has been deleted., R is the gas constant and T is the temperature in Kelvin. B is the chemical potential of the simulation cell containing the protein and the only adjustable parameter. Operationally, a random number (RanX) is generated between 0 and 1 and the inserted or deleted fragment is accepted as the new configuration if $P < \text{RanX}$. When B is set high, the probability of inserting a fragment into the protein simulation cell is dramatically enhanced resulting in saturation with fragments. As the chemical potential is gradually lowered by decreasing the value of B , fragments persist in the simulation cell until a low enough B -value occurs that causes a dramatic evacuation of the simulation cell with almost all fragments exiting. Generally about 5 million simulation steps are done at each B -value and the simulation is run for about 30 different B -values. Note, when using somewhat more sophisticated procedures such as Mezei's cavity-bias[85] to determine where to insert fragments more efficiently, the probability equation has to be adjusted to maintain detailed balance, but this is a technical matter that does not change the conceptual process. Details of the compute time and hardware configurations can be found in the Supporting Information.

Fragment clustering and water exclusion

Hot-spot prediction was done using a simple clustering algorithm. The algorithm for fragment clustering and water exclusion for hot-spot identification contains four parameters: (i) the number of different fragment types used, typically 10–15 probe fragments including some amino acid side chains (Nature's fragments) and a few other organic molecules such as pyridine and furan (ii) a chemical potential energy past the phase transition for each fragment type; (iii) the chemical potential energy of waters, which is used to exclude fragments, and typically occurs at the phase transition; and (iv) a radius for water exclusion, so that any fragment within 1 Å of the oxygen atom of a bound water will be excluded. After applying the simple clustering algorithm, there will often be multiple clusters or hot spots. Post-processing the data, two additional parameters are varied: (i) chemical diversity, the number of fragment types in a fragment cluster, and (ii) cluster radius, a radius for fragment cluster assignment—the center of mass of one fragment to the center of mass to another fragment. Chemical diversity and cluster radius allow the identification of tightly clustered hot spots or hot spots that have a larger area, such as binding sites with multiple subpockets.

Compounds and assays

6-hydroxydopamine was purchased from Sigma-Aldrich. RecA inhibition assays were run as previously described.⁷⁹ Indole derivatives were synthesized as previously described⁸⁷ and

NMR plus MS characterization can be found in the supporting information. Peptide deformylase assays described in supporting information.

Results

The small fragment 6-hydroxydopamine potently inhibits the RecA-RecA protein-protein interaction

RecA forms complicated protein-protein, protein-nucleic acid and protein-ATP interactions. Fig 2 shows the results of SACP simulations on RecA with clustering and water exclusion. The first important thing to observe is that three binding sites are predicted (Fig 2a), which correspond to the three known RecA interaction sites. Detailed analysis of the predicted RecA-RecA Protein-Protein interaction site shows that the highest affinity fragment interactions occur at the universally conserved amino acids K216 and F217 (Fig 2b). All mutations of F217 dramatically reduce or eliminate the RecA-RecA interactions except for the F217Y mutant[86] that increases the protein-protein interaction by over two orders of magnitude indicating the importance of a single hydroxyl substitution at this position. The SACP simulations predict that phenol binds with higher affinity at this site in accord with the known experiments. SACP also predicts the new finding that benzene rings with multiple hydroxyl substituents bind in the position of F217 (Fig 2b) with even higher affinity than the mono-hydroxylated fragment and alkylamines bind in the position of K216 (Fig 2b) in a way that creates 6-hydroxydopamine. This prediction was so compelling that we decided to purchase 6-hydroxydopamine (6-HD) and perform the inhibition[87, 88] assay. As shown in Fig 3, 6-HD inhibits RecA with an IC_{50} of 23 μ M. This is experimental validation of a notable and very surprising prediction, because 6-HD only weighs 166 Da. An interesting aspect of these results is that they do not rule out that 6-HD might be binding at the ATP-binding site of RecA instead of the predicted RecA-RecA interaction site. This possibility is unlikely, however, because many studies of kinases have demonstrated that 6-HD does not bind at the ATP site. A study of the ATP-binding purenergic receptor[89] P2X7, for example, indicates that 6-HD induces Parkinson's like symptoms in an animal model in a P2X7 independent manner, which leads the investigators to conclude that 6-HD does not bind at the ATP site. What is also especially compelling about this finding is that the Singleton lab tested a variety of well-known ATP mimetics to inhibit RecA and all of them were ineffective, which is further explained in the Discussion Section.

Hydroxamic acid fragments bind to and inhibit peptide deformylase through water mediated interactions

N-formylmethionine is the universal amino acid used by bacteria to initiate[90] protein synthesis. Because the mature protein requires removal of the formyl group by the enzyme peptide-deformylase (PDF), small molecule inhibitors[65] designed to block this process are potentially a new class of antibiotics. The discovery that PDF is also active in mitochondria and that its inhibition can cause cells to enter apoptosis makes PDF an anticancer[66, 67] target as well. The binding of small molecules such as heterocyclic hydroxamic acids[91] to PDF are apparently complicated by water-mediated interactions[92] at the active site. Therefore, predicting and characterizing how fragments interact with PDF is especially challenging, due to the multi-body effects of bound waters. SACP simulations with clustering and water exclusion were applied to PDF with the results summarized in Fig 4. The PDF binding site (Fig 4a) is formed by the deeply buried Fe atom (orange) and Glu33 and the surface exposed residues Glu42 and Arg97. The simulations predict that a triplet of high affinity waters (Fig 4b) exists between Glu42 and Arg97, which are just outside of the binding site and effectively extend and

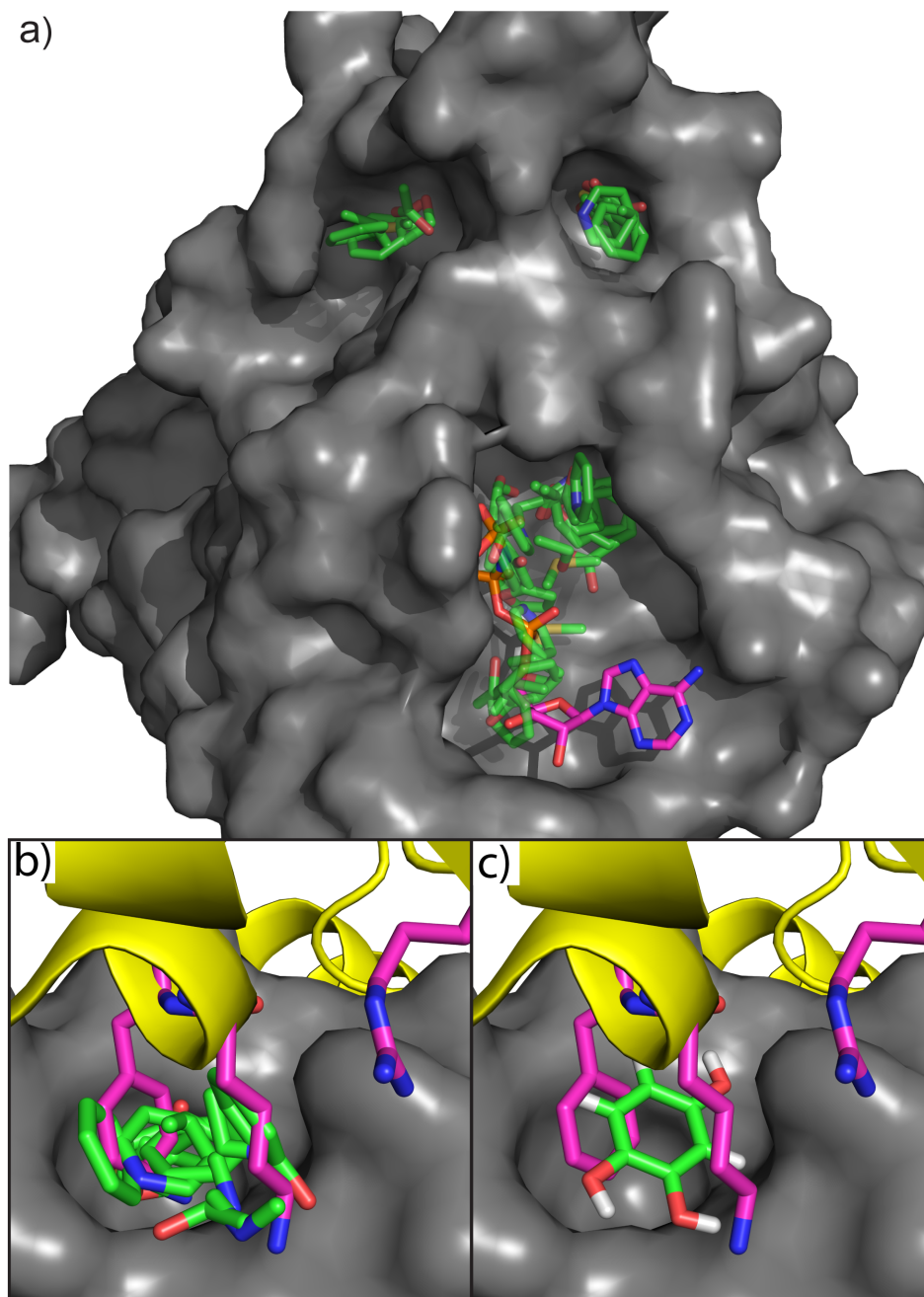


Fig 2. Hot spots on RecA. a) Sites where a diversity of fragments simultaneously binds with high affinity and where no high affinity waters appear. The sites correspond to the ATP binding site (ATP is shown in magenta), an interaction site between RecA monomers (top right site), and the DNA binding site (upper left site). Three universally conserved residues (K216, F217, and R222) from the neighboring subunit (yellow ribbons) are shown with magenta carbons and the interactions at the RecA-RecA site for these residues are shown in (b). Mutations of these three residues will result in the loss of RecA function, although F217Y results in a 250-fold increase in the interaction between RecA subunits. SACP predicts that alkylamines bind in the RecA pocket mimicking K216 interactions (b) and that phenol mimics the F217 interaction. SACP also predicts that adding extra hydroxyl groups to the benzene ring results in a higher affinity (c). When the fragment patterns are combined, SACP creates 6-Hydroxydopamine in the RecA-RecA interaction site.

<https://doi.org/10.1371/journal.pone.0183327.g002>

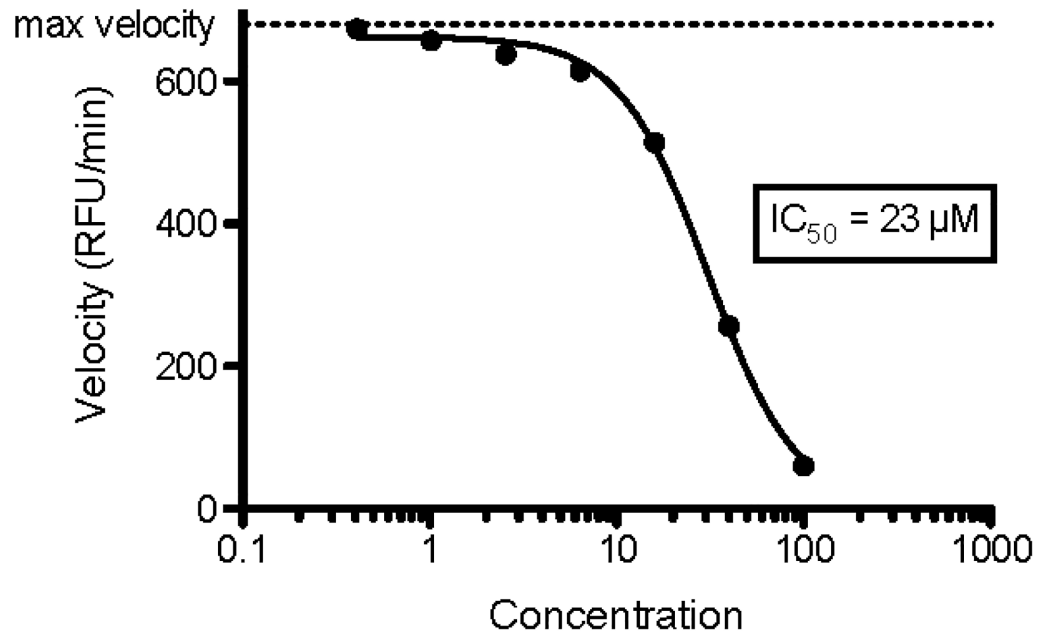


Fig 3. RecA data. RecA hydrolyzes ATP to ADP that can be enzymatically converted to H₂O₂, which oxidizes amplex red to the fluorescent molecule resorufin. The steady state rate of creation of the fluorescence is plotted on the y-axis in relative fluorescence units (RFU). The addition of a RecA inhibitor decreases the rate of production of RFUs in a concentration dependent manner. 6-Hydroxydopamine inhibits RecA with an IC₅₀ of 23 μM. To our knowledge, this is an almost singular example of potently breaking a protein-protein interaction with such a small fragment and quite a remarkable prediction given how difficult it is to inhibit RecA (see Discussion section). The experiments were done in triplicate and all gave virtually the same binding affinity.

<https://doi.org/10.1371/journal.pone.0183327.g003>

close off the binding site. While we generally run about 100 fragments against any protein target, we focused the simulations and analysis on indoles carrying a hydroxamic acid functionality. We did this because Boularot[93] and coworkers, using NMR fragment-based screening, discovered that some of these derivatives are PDF inhibitors. Specifically, they found that indole substituted at the 3-position with acetyl-hydroxamic acid blocked PDF function with micromolar potency. We therefore used this molecule as a fragment in SACP simulations of PDF and we also created the 2-substituted molecule even though there are no reports to our knowledge on this compound. We usually create several different variations of each fragment, because in our experience it often results in surprising outcomes. In fact, this is exactly what happened; the previously unstudied 2-substituted indole is predicted by SACP to bind with higher affinity to PDF than the known 3-substituted indole. Higher affinity means that the 2-substituted fragment continues to bind at the site as the chemical potential is lowered, but the 3-substituted fragment has already disappeared. So as the simulation proceeds to lower chemical potentials there is a point where the 3-substituted molecule no longer binds to PDF but the 2-substituted compound still does and thus the 2-substituted molecule has a predicted higher binding affinity. We synthesized the 2-substituted molecule, and it does in fact have a higher potency than the 3-substituted molecule confirming the prediction. To discover the mechanistic reasons why this new 2-substituted compound has higher affinity than the reported 3-substituted molecule, we analyzed the SACP fragment data. Surprisingly, the 2-substituted molecule binds deeply (Fig 4c) into the PDF pocket making a hydrogen bond with Glu33. By contrast, the 3-substituted molecule (Fig 4d) cannot make this hydrogen bond—it is rotated 180 degrees and it binds near the protein surface and thus the interaction is somewhat weaker. SACP also makes another surprising and interesting prediction,

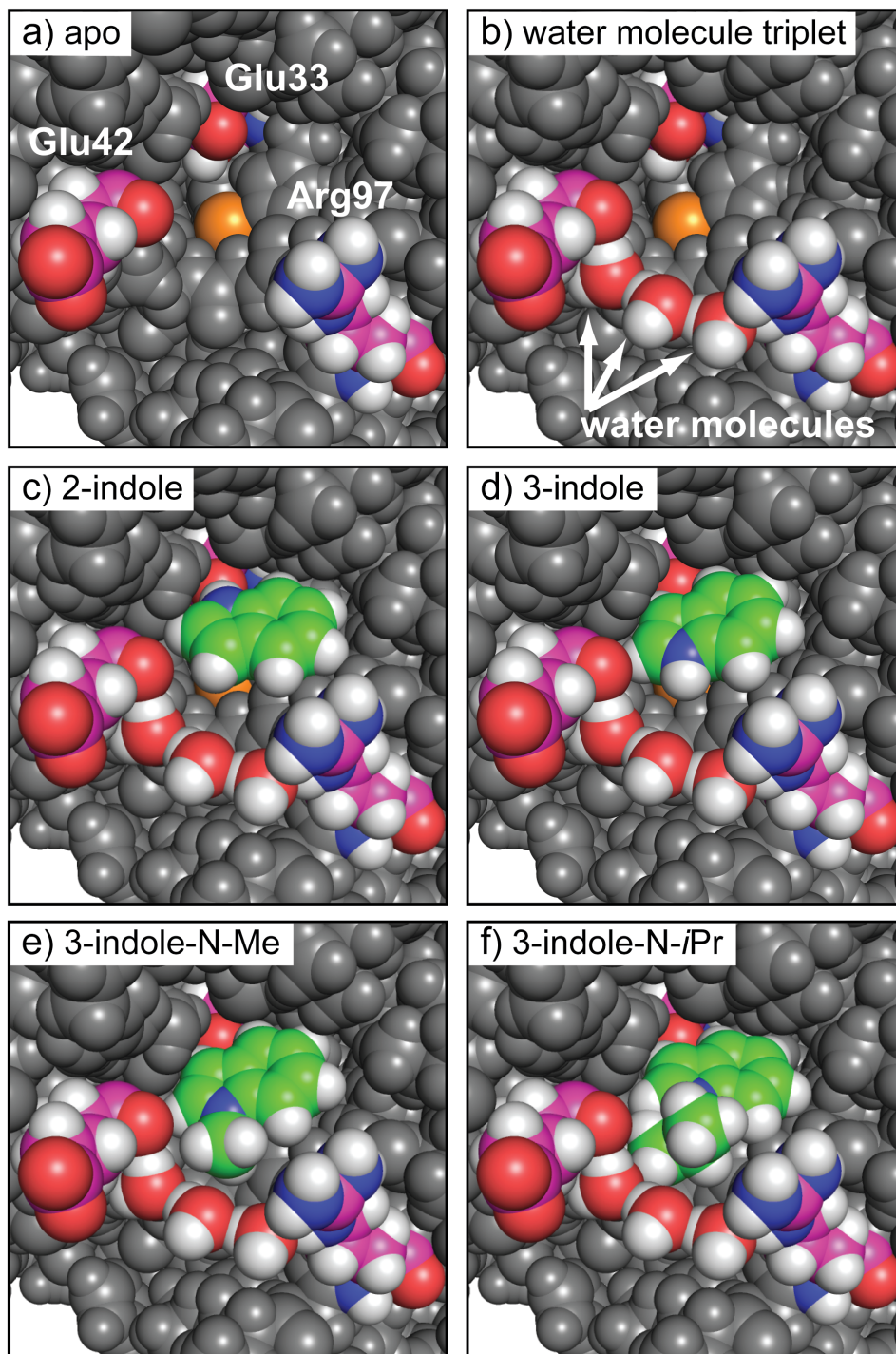


Fig 4. Binding poses on PDF. a) The PDF binding site consists of the buried Fe (orange) and Glu133 and the exposed Glu42 and Arg97. SACP predicts a water triplet (b) that bridges Glu42 and Arg97. SACP predicts that the 2-hydroxamic indole (c) binds deeply in the pocket H-bonding to Glu133 and has a higher affinity than the 3-substituted molecule, which is predicted to rotate out (d) of the pocket. SACP predicts that N-methylating the 3-substituted compound (e) is destabilizing, because the methyl group is floating in a vacuum. The N-isopropyl compound fills the pocket (f) created by the waters and thus is predicted to have higher affinity.

<https://doi.org/10.1371/journal.pone.0183327.g004>

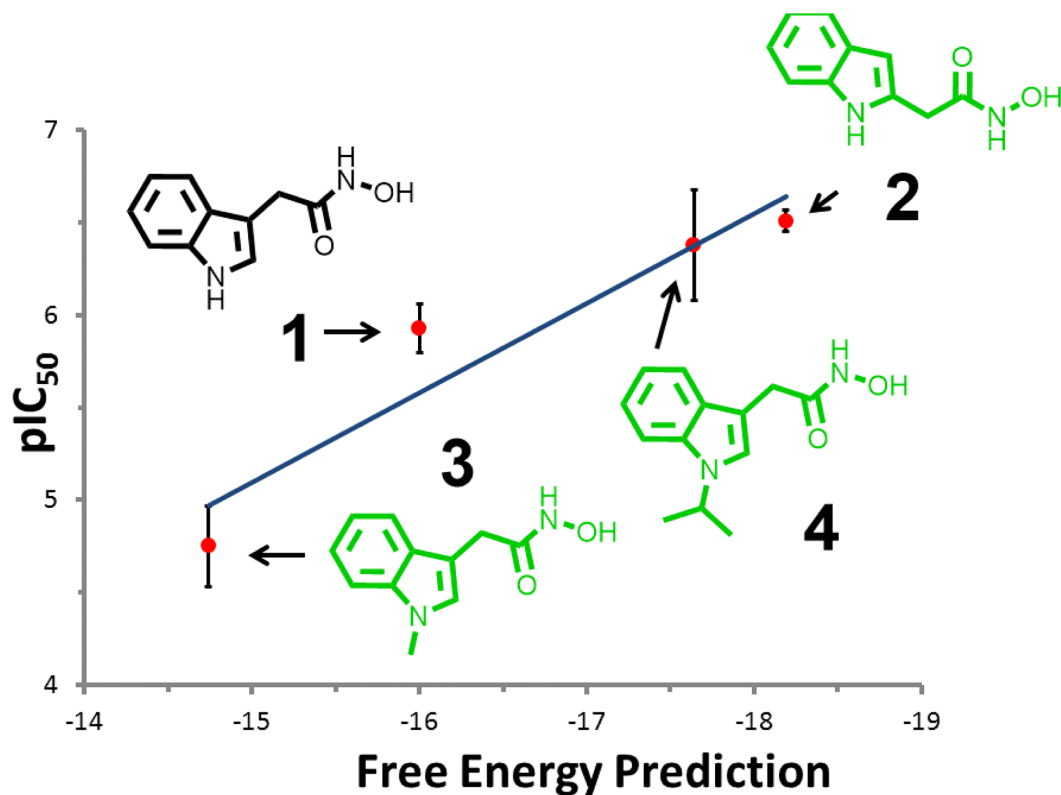


Fig 5. PDF data. The Boularot compound (1) is predicted to bind with fairly high affinity by SACP. The simulations predict that simply moving the acetyl-hydroxamic acid from the 3-position on the indole to the 2-position (2) will result in a more potent compound, which is confirmed by the experiments. SACP predicts that N-methylating this compound (3) will reduce the affinity, but that the N-isopropyl (4) derivative will increase the affinity, which is confirmed by the experiments.

<https://doi.org/10.1371/journal.pone.0183327.g005>

N-methylating the 3-substituted molecule will destabilize the binding while the N-isopropyl derivative will enhance the affinity, which is seemingly very perplexing but reminiscent of hydrophobic[94] hydration. So we synthesized both molecules and in fact the assay data experimentally confirms the simulation predictions (Fig 5). The mechanistic explanation for these unexpected findings is that the methyl group does not fill the pocket created by the bound waters (Fig 4e) and hence it is partially floating in a vacuum, which is unstable. By contrast, the isopropyl moiety precisely fills the pocket (Fig 4f) formed by the waters and thus is predicted to be more stable, which is confirmed by the binding experiments.

To summarize the results of the SACP simulations, the 3-hydroxamic indole points the nitrogen of the indole towards the predicted bound waters. Another surprising result is the prediction that a propane fragment binds with a tight van der Waals interaction with these bound waters (Fig 4f) in a way that alkylates the nitrogen of the indole. Specifically, simply N-methylating this nitrogen is predicted to decrease the affinity of the 3-hydroxamic acid indole because it does not fill the cavity—the methyl group essentially sits in a vacuum and is thus destabilizing, but the n-isopropyl derivative is predicted to enhance the affinity because it does fill the cavity with the right van der Waals contacts. These molecules were synthesized and assayed, which confirm the computational predictions of the relative rank ordering of these four compounds as shown in the graph displayed in Fig 4. Trial and error synthetic expansion on the nitrogen might possibly have found these derivatives, but SACP predictions narrow

down the variations and focuses on the most likely improvements. This set of results shows how complicated fragment-protein interactions can be when water mediated interactions must also be accounted for. Thus, using SACP simulations to first locate the tightly bound waters and then using these waters in subsequent SACP simulations to discover their influence on how organic fragments bind is a powerful way of interrogating the multibody protein-fragment-water architecture.

Accurate water mapping is essential for characterizing the binding site of HIV protease

We have found that in almost all cases protein binding sites and sites of high affinity protein-protein interaction have very low affinity for water and thus the water exclusion principle is an essential part of our hypothesis on molecular recognition. HIV protease is an important exception, because this enzyme has one essential molecule of water at the center of its binding site. The co-crystal structure of HIV protease[49] with one of the earliest drug-like inhibitors[95] crivivan, immediately established the central role of water as a key mediator of ligand binding in the active site. Numerous recent studies[96–99] have demonstrated that resistance mutations to protease inhibitors is critically dependent on the hydration state of the active site, and thus water binding must be taken into account in the development of any new inhibitors. Early efforts to understand the mechanism of HIV protease led to novel protein constructs that enabled high resolution crystal structures[100, 101] of the Apo protein. All of these efforts revealed a key structural water molecule that bound to both catalytic aspartates of the HIV protease dimer. Thus, a principled way of understanding and predicting the characteristic hydration of HIV protease is essential for understanding its fundamental molecular recognition processes.

An important factor in the study of HIV protease is the protonation states[102] of the catalytic aspartates, so SACP of water was run on the structure (1SDT) with both aspartates deprotonated, with one protonated and the other deprotonated, and with both protonated. In all cases SACP finds one isolated water molecule at the center of the binding site at very low chemical potentials indicating that it has very high affinity for the protein at this site. What is very interesting is that HIV protease, with this one singular exception, otherwise obeys the water exclusion principle. Fig 6a shows the SACP predicted water binding to both catalytic aspartates in the exact position found in the crystal structures with this water present even at a chemical potential of -29 in the singly deprotonated state and -25 in the doubly deprotonated state. Paradoxically, it might be expected that have both aspartic acid residues (ASPs) negatively charged would have a higher affinity for water, but SACP in accord with both experiments[103] and computations[102] indicates that the pKa for this site means that one ASP is protonated and one is deprotonated. As the chemical potential is increased by approximately 20 units to -8 for the singly protonated state and -5 for the doubly protonated state low affinity water molecules fill in the boundaries of the binding site but no new additional water molecules are found in the binding site itself (Fig 6b and 6c). Even in this case over a large range of chemical potential values SACP shows that the water exclusion principle is obeyed by HIV protease with one functional exception. Finally, applying the principle of high affinity fragment clustering, the HIV protease binding site is readily mapped out as shown in Fig 7.

To summarize these results, Fig 6a shows that the singular water molecule at the HIV active site identified by X-ray crystallography is identical to the water molecule predicted by SACP simulations. Additionally SACP also predicts that the lower affinity bound waters define the boundaries of the binding site, which obeys the water exclusion principle that binding sites are generally devoid of very high affinity waters. Thus, the key water at the center of this binding

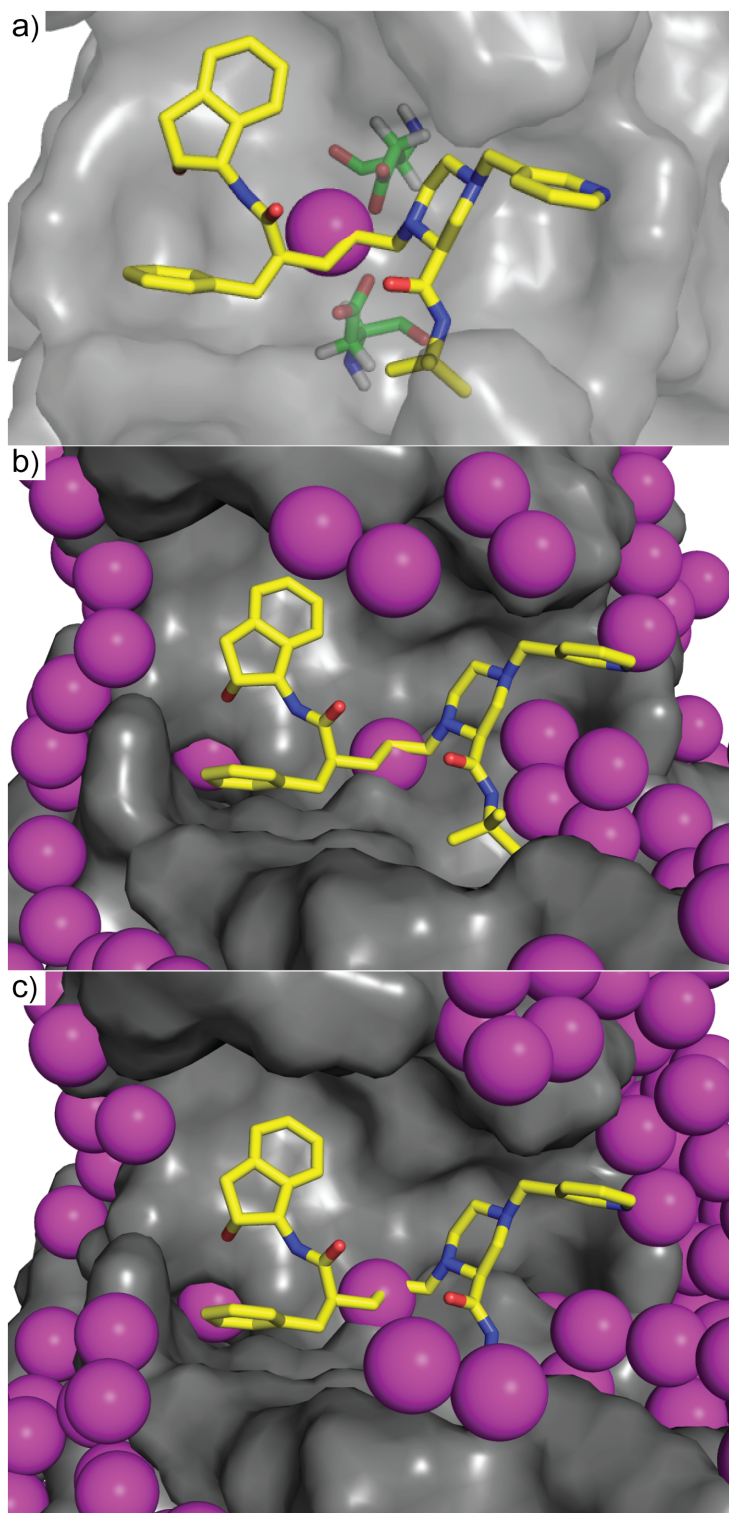


Fig 6. Water data on HIV protease. The SACP water molecules predicted to bind to HIV protease. The simulation produces that key water molecule (a) experimentally known to be at the center of the binding site. All of the other water molecules from SACP obey the water exclusion principle and bind in a ring just outside of the site (b-c). Two key aspartates are at the center of the binding site (Asp25 and Asp125 colored with green carbons in panel a). The ligand indinavir is shown with yellow carbons in all three panels. Part of the protein surface is hidden to reveal the binding site. Three different protonation states were simulated, a) ASP25 and ASP125, b) ASH25 and ASP125, c) ASH25 and ASH125 with all giving comparable results.

<https://doi.org/10.1371/journal.pone.0183327.g006>

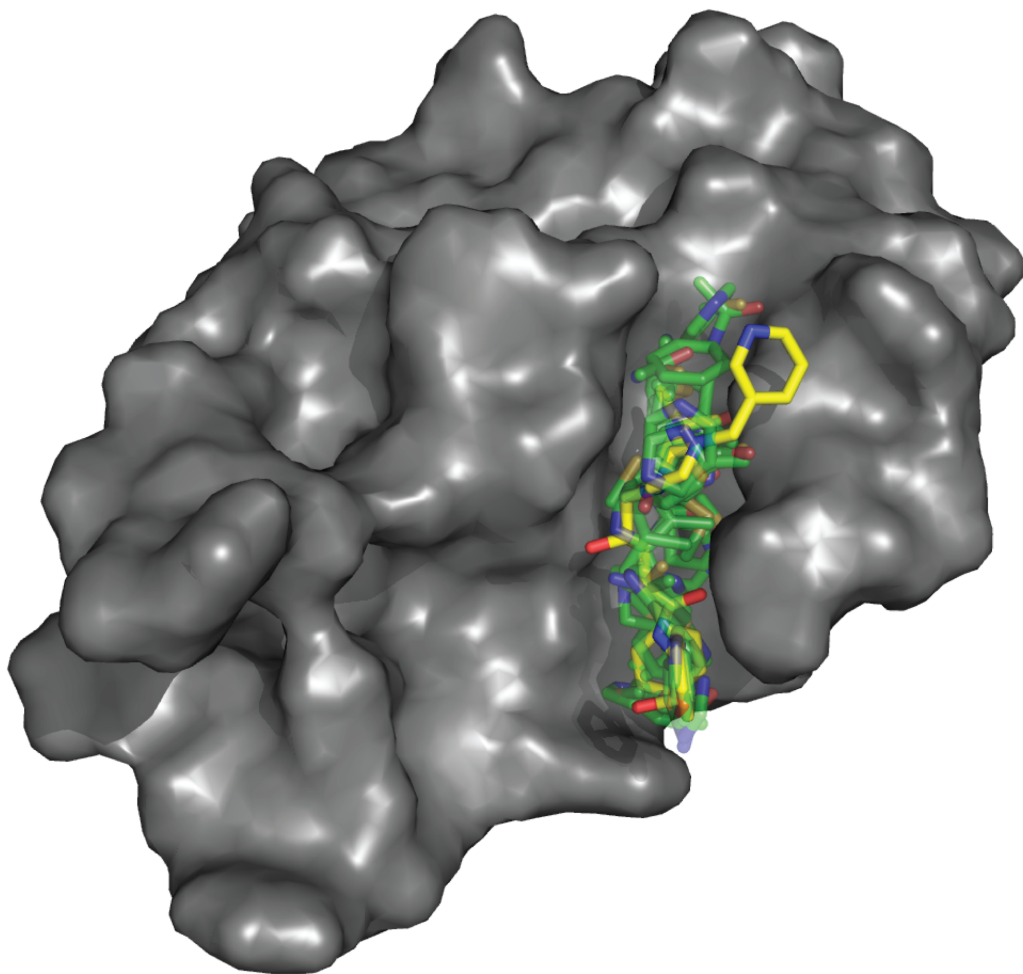


Fig 7. Hot spot on HIV protease. High affinity diverse fragment clustering completely maps out the indinavir binding site on HIV protease.

<https://doi.org/10.1371/journal.pone.0183327.g007>

site is predicted by SACP to be an exception to the water exclusion principle. Finally, in Fig 7 we show clustered organic fragment binding data superimposed[95] over indinavir, illustrating that SACP with no free parameter adjustments actually recreates known ligand binding patterns.

Locating the point of electron transfer in the large extensive binding site of dihydrofolate reductase

Dihydrofolate reductase is an enzyme that catalyzes the reduction of folic acid by carrying out an electron transfer from the cofactor NADP. A simple binding site analysis of wild-type *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase (PfDHFR-TS) complexed with WR99210—displaying the residues that are $< 4 \text{ \AA}$ from any ligand atom shown in red in Fig 8a—clearly shows that the two ligands contact a large and extended binding site of the protein. A more detailed analysis shown in Fig 8b reveals that one aromatic ring from each ligand, although interacting noncovalently, are only separated by 1.7 \AA . These nonbonded rings are placed within almost covalent distance, because this is the region where electron transfer occurs. SACP simulations strikingly predicts the highest affinity cluster with water exclusion

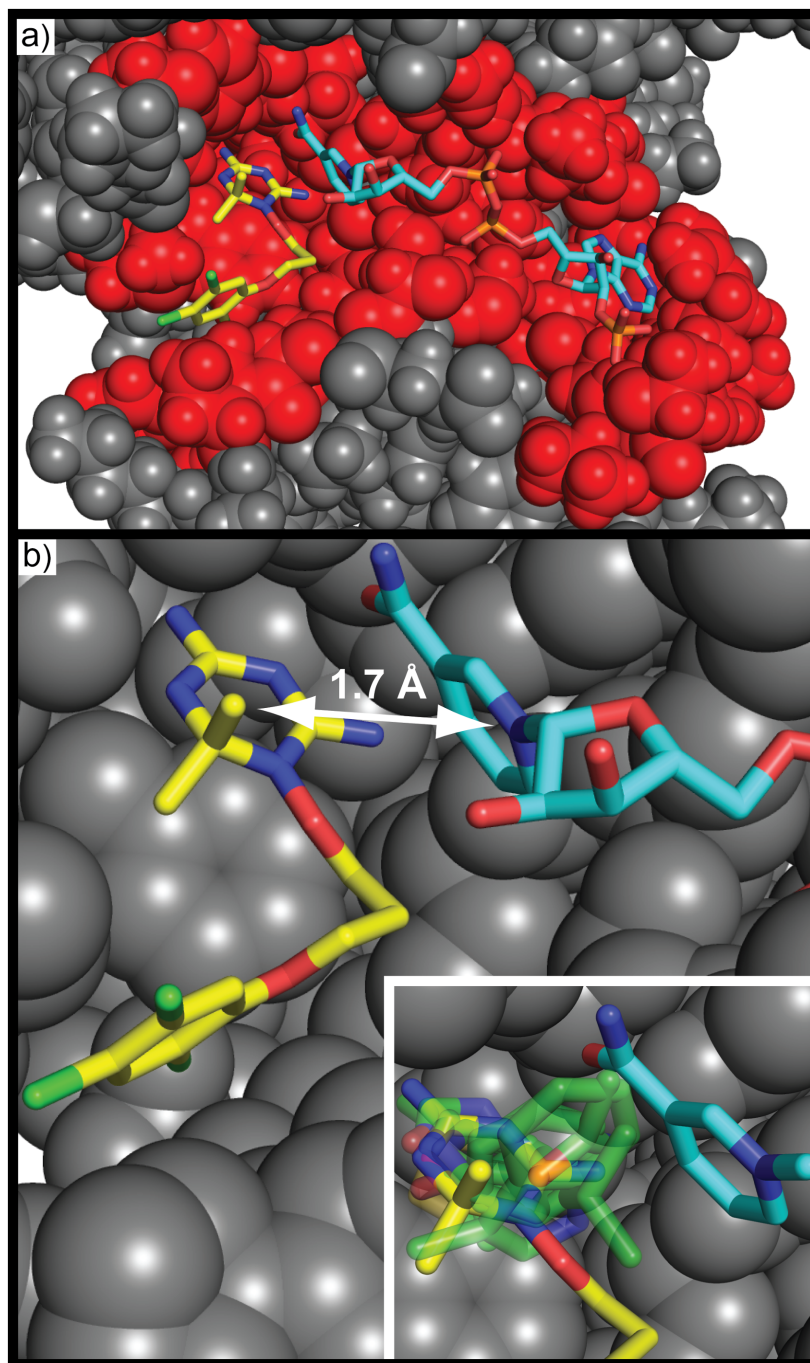


Fig 8. Hot spot on DHFR. DHFR complexed with WR99210 (yellow carbon atoms) and NADPH (cyan carbon atoms) is shown. Some residues are hidden to reveal the binding site. DHFR has an extensive (a) binding site; all residues within 4 Å of ligand and cofactor displayed with red spheres. Two aromatic rings (one on WR99210 and another on NADPH) are separated (b) by ~1.7 Å—this is an impossibly high energetic configuration under normal circumstances, which the protein sets up. SACP (inset: fragment cluster & water exclusion presented with partially transparent green carbon atoms) predicts a high affinity pinpoint site between these rings showing that the extensive binding site is scaffolding that creates this functional site.

<https://doi.org/10.1371/journal.pone.0183327.g008>

lies exactly between these two rings as shown in the inset of Fig 8. To summarize these results, although DHFR apparently has a very large extensive binding site, the electron transfer function occurs at a point between two aromatic rings separated by only 1.7 Å—an impossibly high energy state that can only occur within this type of binding site. SACP with clustering and water exclusion easily locates this singular functional hot spot as the one key site of this protein.

Experimental and computational fragment-based studies locate the binding subsites in elastase

Elastase has been a historically important protein[104, 105] since 1969, with the discovery that its hyperactivity is the central mechanistic cause[106, 107] of emphysema. The first co-crystal structure of elastase bound with a substrate[43] was published in 1976 by Alber, Petsko, and Tsenoglou, but the first publicly available structure deposited in the PDB in 1986 was solved in 1982 by Hughes[44] and coworkers. The elastase co-crystals solved by Meyer[45] and colleagues and the subsequent inhibitor-elastase high resolution structure solved by Navia[46] and co-workers showed the ligands in the same configuration as the Hughes structure confirming comparable binding modes. Petsko, 18 years after solving the first co-crystal, demonstrated with his colleague Ringe that in fact almost identical substrates could bind in radically different configurations, thus discovering new binding sub-sites in elastase[47] almost 2 decades after the first structure was solved.

The long progression of co-crystal structures required to uncover hidden binding subsites in elastase demonstrates some of the challenges and mysteries of molecular recognition, which makes it an excellent test case for fragment-based molecular recognition studies. The results of SACP simulations with clustering and water exclusion are summarized in Fig 9. Three peptide co-crystals with the ligands superimposed (9a) show experimentally that there is a tightly focused hot-spot where all three ligands overlap. Typically, we perform a preliminary initial set of SACP calculations on 30 organic fragments and water. Recall, that these 31 simulations are performed simultaneously and independently on different CPU processors to discover the free energy of binding to the protein in all cases separately, so different fragments have no knowledge of each other. Also recall, that for a given simulation the fragment-fragment self-energy is part of the simulation, because the protein is completely solvated by this organic fragment and annealing of the chemical potential goes through a phase transition that evacuates the “solvent” leaving behind tightly bound fragments. This is identical to the fragment soaking experiments done with experimental X-ray crystallographic methods. In our previous communication on hen egg white lysozyme, we present SACP results using acetonitrile as a fragment and the post phase transition of the dozen fragment binding positions matches the X-ray studies done in the Ringe Lab.

Typically, after all of the individual fragment simulations are complete the clustering algorithm processes the data and discovers where 10 diverse fragments bind to the protein within 2Å or less of each other. In the case of HIV protease as shown in Fig 7, the fragments cover the extensive binding site exactly where crivivan is located. This happens, because each fragment is slightly offset from a different fragment and thus a chain across the whole binding site is formed. In elastase, by contrast, the fragments all pile up in a tight location (Fig 8b) right at the place where the 3 different ligands intersect (Fig 8a). This localized region is discovered by SACP when only the highest affinity (lowest chemical potential) binding fragments are used for the clustering algorithm and as always water exclusion is applied so that we know that water does not bind tightly at this site. Water exclusion using explicit water binding with SACP is a rigorous way of knowing whether or not organic fragments can effectively bind, because water does not bind tightly at this location.

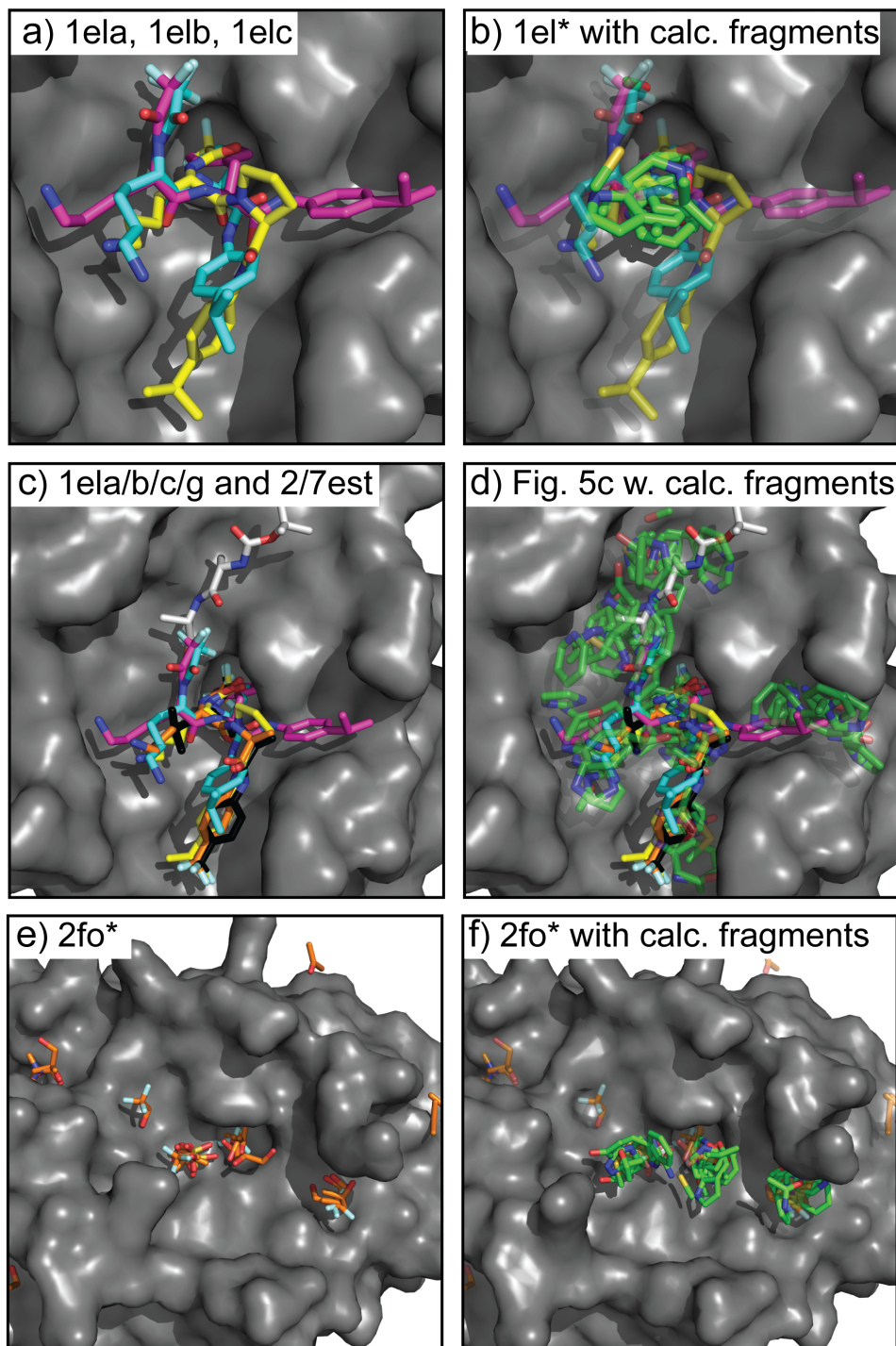


Fig 9. Hot spots on elastase. Overlay of 3 different (a) elastase co-crystal structures 1ela (yellow), 1elb (cyan), and 1elc (magenta). Beginning in 1976 and continuing until 1994 all of the elastase co-crystals had the ligand binding in the same mode as 1ela. The 1elb and 1elc structures solved in 1994 revealed new binding subsites. The highest affinity (b) SACP fragment clusters (green carbons) with water exclusion are at the site where the 3 ligands overlap. This result is the same for simulations using the 1ela, 1elb, and 1elc structures. Nine co-crystal (c) structures (2fo9, 2foa, 2fob, 2foc, 2fod, 2foe, 2fof, 2fog, 2foh) are needed to map out the entire binding site in elastase. SACP fragment binding predictions using the top three highest interactions (d) cover the subsites from all 9 co-crystals, indicating that these sites were long undiscovered because that have lower affinity for the ligands. Importantly, experimental X-ray fragment-based soaking experiments (e) done in

the Ringe Lab did reveal the lower affinity hard-to-find subsites. Using this crystal structure as input into SACP, the simulations (f) also find all of the subsites.

<https://doi.org/10.1371/journal.pone.0183327.g009>

Fig 9c shows a superimposition of 6 different co-crystal structures—the same three from Fig 9a and three more solved by Petsko and Ringe. The important thing to note is that one of the new co-crystals binds in the same mode as the older 3 structures so it provides no new information, but the two others bind in totally different modes defining previously unidentified sub-sites. The method for fragment clustering involves several parameters: (i) the number of fragment types used (chemical diversity), (ii) a radius for fragment cluster assignment, and (iii) the predicted affinity of the fragments. To fully map out the binding sites of the elastase peptides, we included the second highest affinity binding fragments, increased the clustering radius from 2 to 3 Å, and lowered the chemical diversity from 10 fragments down to 9. By doing this, all the new subsites from the 6 agglomerated co-crystals were identified (Fig 9d).

A different way of doing a comparable analysis of elastase binding subsites is to crystallize the protein in the presence of different small organic molecules, a process known as multiple solvent crystal structures (MSCS). The protein is crystallized in the presence of one fragment acting as the solvent. This is done again with a single different fragment and then repeated with many fragments individually. The protein is not crystallized in the presence of multiple fragments simultaneously. High resolution X-ray structures are obtained in each case, which always reveals that each fragment has a unique spectrum of binding sites spread out over the entire surface of the protein. Locating the sites on the protein that have a high affinity for a diversity of fragments invariably reveals the binding site. Mattos and coworkers mapped the binding surface of elastase with this technique to find all of the local hot spots (Fig 9e). [108] MSCS is a form of fragment-based X-ray crystallography that successfully mapped out all of the binding sub-sites of elastase. The ease of finding all of the elastase binding sub-sites with theoretical or experimental fragment-based methods such as SACP or MSCS, respectively, may be contrasted with the difficulty of identifying them with experimental co-crystallography, because not only did it take almost two decades, but Petsko and Ringe had to solve fifteen different co-crystals before all of the hot-spots were experimentally located and characterized. SACP identified these sites (Fig 9f) using only one crystal structure. The definition of the hot spots, using chemical diversity and chemical radius, with clustering and water exclusion in order to find all hot spots demonstrates these principles can be used to identify and characterize the distribution of multiple binding modes in elastase and the singular functional site in DHFR.

MDM2-MDMX-p53—Identifying the subtle differences in homologous protein-protein interactions

Protein-protein interactions (PPI) typically have large surface area and appear [109] featureless, which hinders the development of small molecule inhibitors. A great challenge in the PPI of cancer is to identify the subtle differences between how MDM2 and MDM4 bind to p53. From previous computational [110] and experimental [52] studies it has been discovered that the three most important residues in this PPI are F19, W23, and L26 (Fig 10a and 10b). SACP Hot-spot identification using clustering with a 2Å fragment radius and water exclusion on MDM2 and MDM4 identified the binding site at W23 indicating that this is the highest affinity interaction point. Increasing the cluster radius to 3 Å with lower fragment diversity returns a second cluster that binds at F19 on both MDM2 (Fig 10a) and MDM4 (Fig 10b). Interestingly, these two clusters are tighter and better focused in MDM2 than in MDM4. If we continue to lower the restraints by making the cluster radius 4 Å and asking for less diversity, there is a

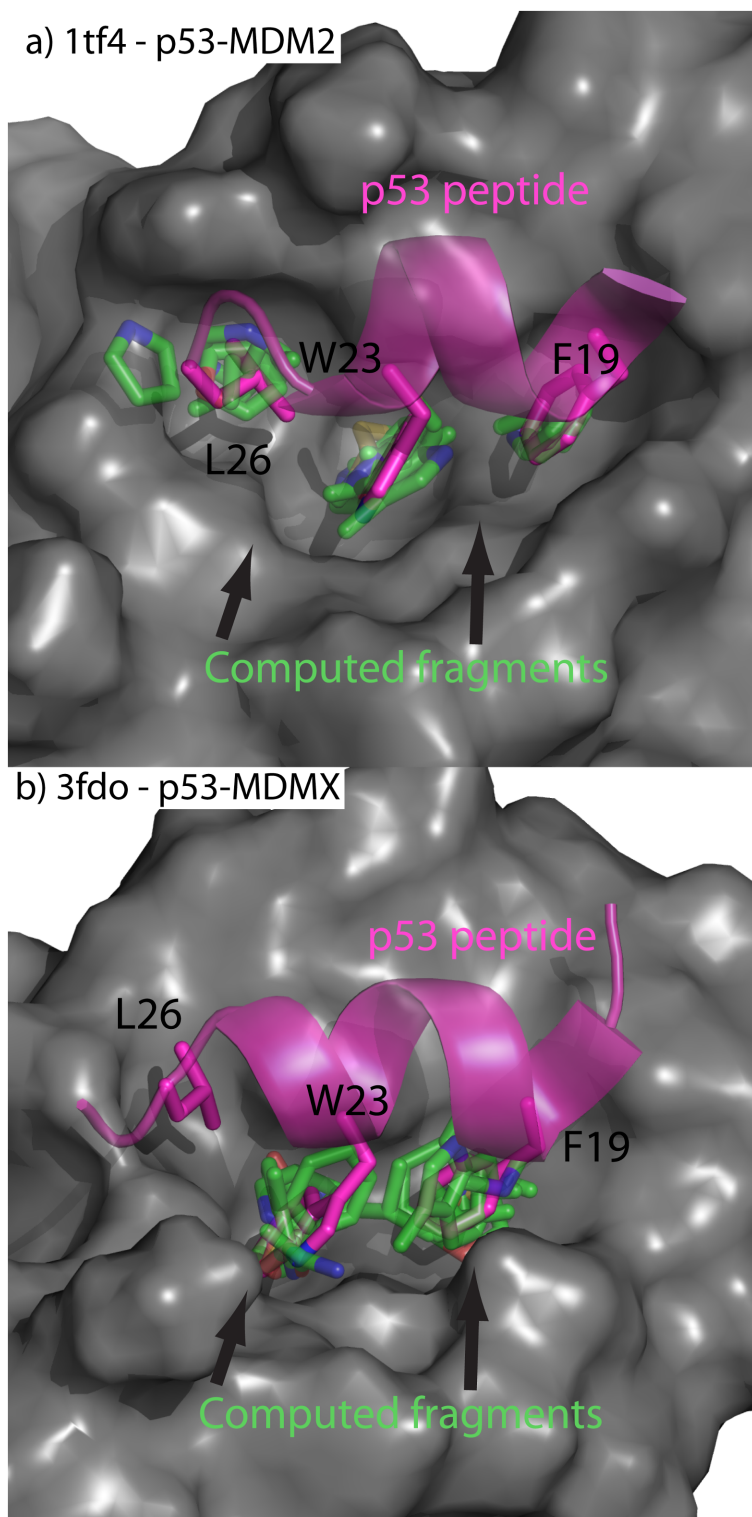


Fig 10. Hot spots on MDM2/X. a) p53-MDM2 protein-protein interaction displayed with overlay of computed fragments. b) p53-MDMX protein-protein interaction displayed with overlay of computed fragments. The MDM surfaces are presented in gray. The p53 peptide is colored magenta. The computed fragments are shown in green. Strikingly, the fragments predict a high affinity site at L26 for MDM2 but not MDMX, which is in accord with the experimental finding on the differential binding.

<https://doi.org/10.1371/journal.pone.0183327.g010>

binding hot-spot at L26 on MDM2 (Fig 10a) that does not exist for MDM4. To discover the reason for this subtle difference, we analyzed the binding clusters further and discovered that MDM4 has a methionine at position 54 that impinges on the pocket around L26, whereas by comparison[54] MDM2 has a corresponding leucine in a comparable spatial position at residue 53 that does not impinge on the L26 pocket. Another important difference between the structures of MDM2 and MDM4 is the presence of a second hydrophobic binding site, adjacent to the L26 site, in MDM4—formed by L33, V52, and L106 and separated from the P53 binding site by M53.

Further reducing the restraints of the hot-spot algorithm, the second hydrophobic site is revealed on MDM4 but no site is identified on MDM2. This site could be utilized to discover novel, and selective, MDM4 inhibitors. SACP hot-spot analysis not only found clusters over the three most critical residues, but also ranked these clusters, which is in agreement with a computational alanine screen[110] and an experimental alanine screen.[52] By rank-ordering the fragments by affinity, using a diverse fragment set, and excluding sites that have waters, we locate the sites of high interaction energy and can rank those sites and thus uncover the differential binding to p53 of these 2 highly analogous proteins.

Understanding the HIV TAT-TAR protein-nucleic acid interactions requires locating and characterizing the binding sites on the viral RNA

To further demonstrate the generality of the SACP system for identifying and characterizing key sites of molecular recognition, the simulations were run on the HIV TAR RNA stem loop. RNA offers a unique system for hot-spot mapping because RNA can fold into unique 3-dimensional structures, as proteins do, but RNA is only made up of four different bases, which can appear chemically uniform. TAR RNA is found at the end of all viral transcripts, and cyclic peptides that inhibit this protein-RNA interaction have been shown to slow viral replication in all clades[59] tested. The cyclic peptides clearly indicate a small set of neighboring highly localized protein-RNA molecular recognition sites. I12, R7, and K8 are the three critical residues required for peptides mimicking the native HIV binding protein, TAT. Cluster analysis with water exclusion predicts two neighboring sites of molecular recognition; one cluster overlapping K7 and part of R8 with another cluster over I12 and part of K7 (Fig 11). Even though RNA is chemically uniform, hot-spot mapping with SACP was able to reveal the binding sites on TAR RNA that are associated with the key residues on the Tat mimetic in the 2kx5 conformation. The only input to the simulation is the RNA structure and the fragments including water. Post phase transition cluster analysis and water exclusion immediately produce the site of RNA-protein interaction where known peptide inhibitors bind.

Discussion

Four crucial aspects of these studies are, a) the sampling technique that will be illustrated by the multibody water-water-protein interactions, b) generating a phase transition to determine locations on the protein that have high affinity for a fragment, c) interpreting the data in order to predict new experiments, and d) the need for complete and accurate water mapping over the whole protein surface.

Predicting the well-known experimental multi-body water triplet in BPTI illustrates SACP sampling

The bovine pancreatic trypsin inhibitor (BPTI) has been studied extensively with X-ray[111–115] crystallography, NMR[116], and neutron[117] diffraction. The most striking motif of

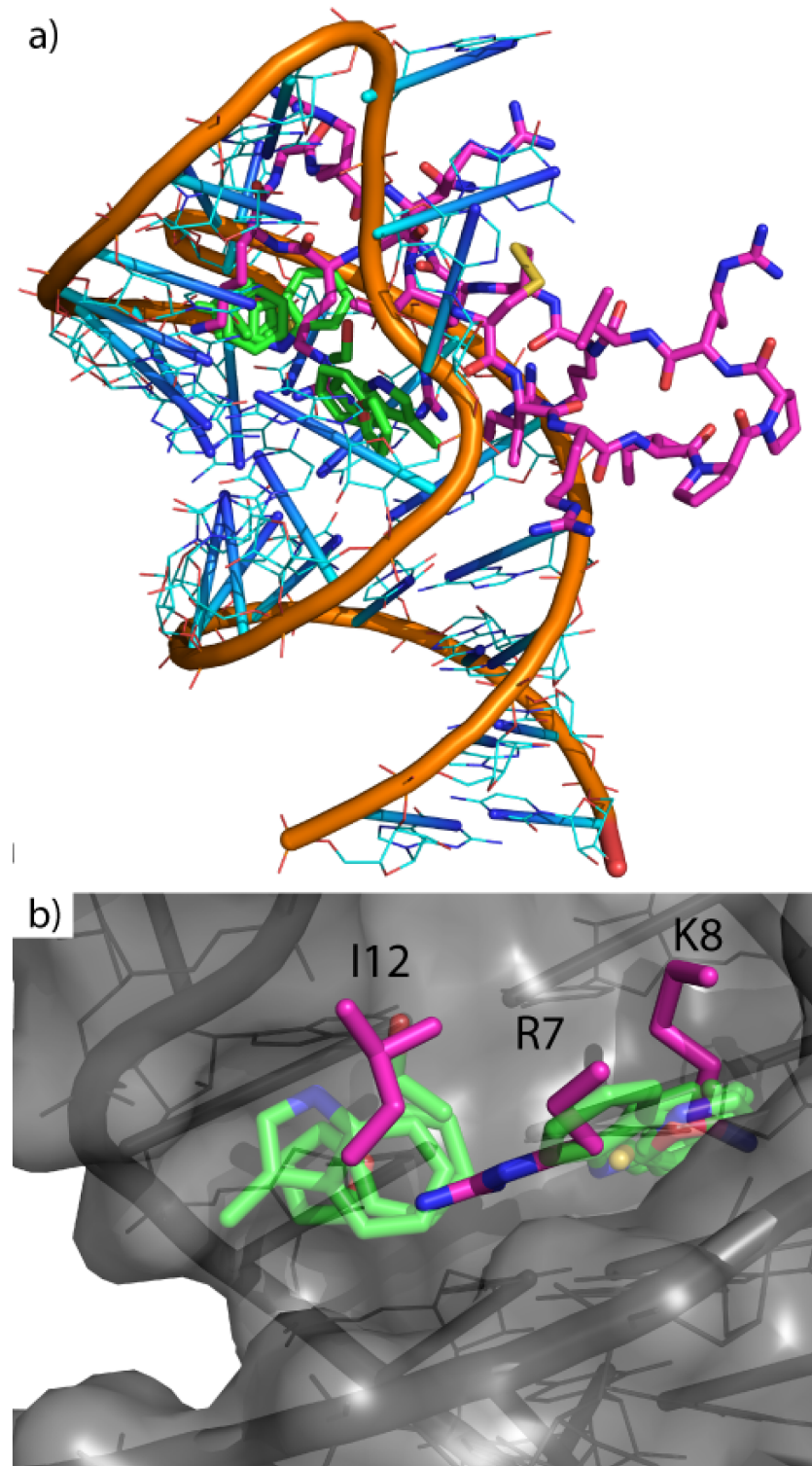


Fig 11. Hot spots on HIV TAR. a) HIV RNA (orange backbone with cyan nucleosides) and TAT peptide (magenta carbons) with calculated clusters (transparent green carbons). b) Zoomed in and displays overlap of two clusters on residues I12, R7, and K8, which are the experimentally known key interactions.

<https://doi.org/10.1371/journal.pone.0183327.g011>

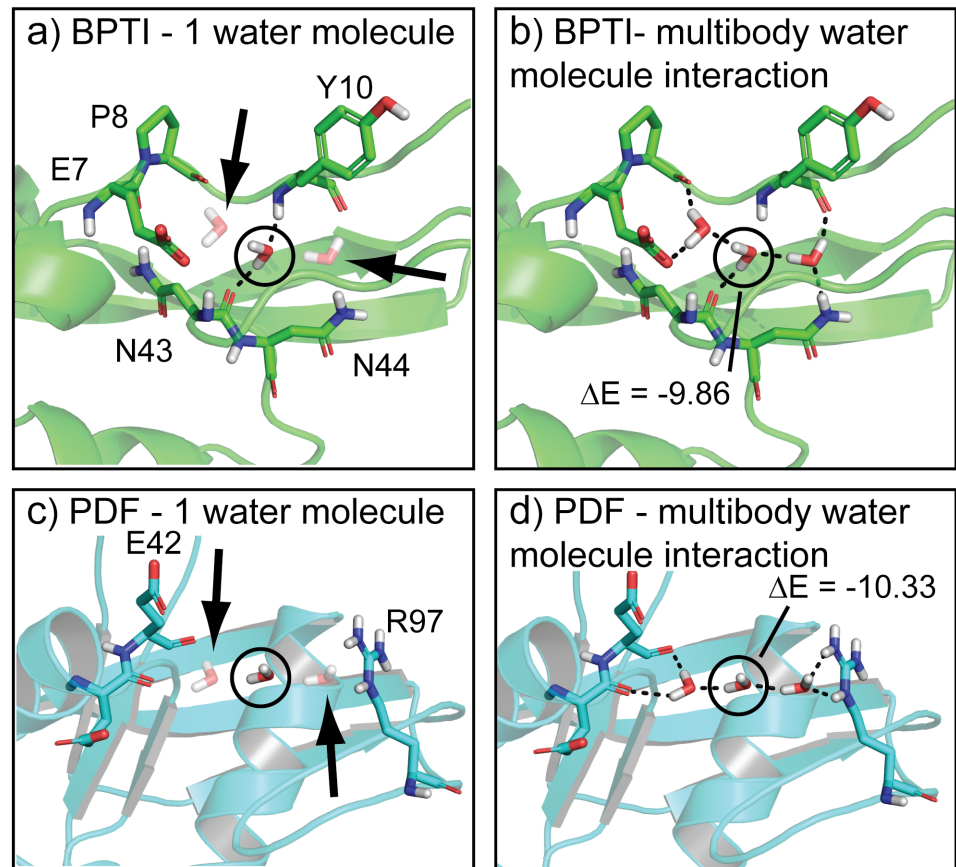


Fig 12. Multi-body water interactions. a) The water triplet of BPTI is shown with the central water circled and its neighbors faded (arrows) to indicate that we performed an energy calculation on the middle water with the others not considered. When all the waters are included (b) the energy of the central water drops by approximately 10 kcal/mol. Annealing the chemical potential with SACP simulations correctly predicts the BPTI water triplet, because the unstable water is acceptable at high chemical potentials. Continued sampling at high chemical potential enables discovery of the other water molecules and creation of the highly stable multi-body triplet that survives driving the system through the phase transition. SACP predicts a similar water triplet in PDF. Like BPTI, the central water has a high energy when the other two waters are not considered (c), but when all three waters are taken into account (d) the energy of the central water drops by more than 10 kcal/mol.

<https://doi.org/10.1371/journal.pone.0183327.g012>

BPTI is a deeply buried water triplet that is completely enclosed by the protein. There is no way to discover this water-water-protein multi-body triplet with molecular dynamics or canonical Monte Carlo, because both techniques require that water diffuse through the protein—one method driven by Newtonian forces and the other with a random walk. In principle, grand canonical Monte Carlo can discover these deeply buried waters, because the process can insert and delete new water molecules into cavities, but it does not (we ran many such simulations on BPTI and it never produced the water triplet), because it maintains a constant chemical potential. SACP, however, does find the water triplet and is the only method in our hands capable of doing so. The reasons for this are illustrated in Fig 12. During a grand canonical simulation, any particular water inserted into this cavity has a very high unstable energy, because it is at least partially flanked by a vacuum and is therefore rejected as a probable configuration. At the elevated chemical potentials of the early phase of SACP simulations that overcome high energy barriers, however, one water molecule has a dramatically enhanced

probability of being an acceptable configuration. Subsequently, when additional trial waters are inserted in the presence of this first isolated water, the collective water-water-protein interactions are so stable with such a dramatic drop in energy the state is immediately accepted. Because this is a very low free energy state, a high fraction of the randomly generated replicates subject to deletion steps fail to break up these collective interactions and thus this water triplet remains intact even as the SACP simulations go through the phase transition and proceed to very low values of the chemical potential. Obtaining the free energy of protein hydration and dehydration is essential for accurate molecular recognition predictions and SACP is the only method to our knowledge capable of doing this.

In almost all cases that we have studied, binding sites do not have tightly bound water molecules—going through the phase transition depopulates water from these regions of the protein. As pointed out in the results section, HIV protease is an exceptional case with one tightly bound water molecule at the center of the binding site. PDF turned out to be another unusual case—SACP produced a water triplet (Fig 3c and 3d) at the very edge of the binding site that survives driving the simulation through the phase transition. This is reminiscent of the triplet in BPTI—the multibody interactions are essential for the stability of this configuration. These very low free energy interactions combined with the compelling way that the edge of the binding site is closed and completed, motivated us to keep these waters in the analysis of other fragment simulations and use the predictions to carry out new binding experiments. These results, which could in principle be observed with experimental fragment-based techniques, can only be obtained computationally with SACP to the best of our knowledge, because it is the only method in our hands capable of doing the extensive sampling required.

The phase transition

All multi-body fragment-fragment interactions are taken into account, which forms a collective network of solvent-solvent interactions around the protein. A phase transition is generated by successively lowering the chemical potential until a free energy capable of breaking apart the solvent-solvent (fragment-fragment) interactions is discovered and suddenly fragments are rapidly deleted from the simulation cell. At this point, fragment-protein interactions are effectively in competition with fragment-fragment interactions. Going through the phase transition, a fragment will only continue to bind at a specific location on the protein if this fragment-protein interaction has a higher affinity than the fragment-fragment collective interactions.

To be clear, at high chemical potentials, the chemical potential free energy dominates the simulations and fragment insertions have a high probability, which fills the entire simulation cell and every cavity in the protein large enough to accommodate a fragment. As the chemical potential is lowered, very little happens until this chemical potential approaches the free energy of the multibody bulk fragment-fragment interactions. When the chemical potential is lowered beyond this point, these bulk interactions are broken essentially destroying the global solvent cohesion and fragment deletion becomes highly probable and the simulation cell becomes rapidly depopulated of fragments. Fragments that have an interaction free energy with the protein that is much greater than the fragment-fragment interactions survive driving through the phase transition. This process discovers the sites on a protein that have higher affinity for a fragment than a fragment has for itself. Experimental fragment-based studies bathe the protein in a solution of a particular fragment and then crystallize this protein to see where fragments stick or use NMR to determine where they bind. The SACP process mimics these fragment soaking experiments in a rigorous way.

Designing experiments

It is highly nontrivial to construct new experiments to test simulation predictions, because there is a great deal of data produced by SACP that needs to be analyzed. One convenient starting point is to examine how well the simulations reproduce what is already known. With regard to RecA, the protein-protein interaction site contains several amino acids that are universally conserved in all bacteria, with the most important ones being a phenylalanine and a lysine. The SACP simulations not only reproduced these key interactions, but also predicted that tyrosine has a higher affinity than phenylalanine, which RecA investigators have experimentally demonstrated with mutagenesis experiments. Totally new findings from the simulations were that multiply hydroxylated benzene rings are predicted to have even higher affinity, and that alkylamines could overlay the lysine side chain with the correct geometric bond lengths and angles for binding to the benzene ring. This analysis resulted in the prediction that 6-hydroxydopamine would have high affinity for the RecA-RecA protein-protein interaction site. A review of the RecA literature [87, 88, 118–121] indicates that all screening efforts have failed to produce any molecule that inhibits RecA including a very wide variety of historically important ATP mimetic inhibitors (Fig 13) and thus finding anything to block RecA activity would be an extremely difficult and unprecedented challenge. Nevertheless, the SACP results were so compelling that we decided to have 6-HD assayed, and as described in the Results section it is very potent for such a low molecular weight entity.

Knowledge of alternative binding modes provides new and different avenues for optimization in drug discovery. When SACP uncovered that different substitution patterns of hydroxamic acids can bind in different orientations to PDF, this compelled us to synthesize a new derivative and perform binding experiments. When the experiments confirmed the hard-to-explain SACP predictions, we dug a little deeper into a possible mechanistic reason for the findings. This led to an even more surprising result that bound waters play a crucial role for some derivatives. Before synthesizing any more compounds, we checked on the validity of the bound triplet water predictions for PDF by simulating a very well-known protein, BPTI. When the simulations accurately reproduced the known water triplet deeply buried in BPTI, we had enough confidence to trust the new predictions on PDF and to synthesize a couple of compounds to test the water-mediated binding predictions. The binding experiments confirmed the computational predictions. It is a difficult decision to commit experimental resources using only the results of a simulation, so we will often run closely related test cases whenever possible. Success on the test cases provides a measure of confidence that motivates doing the experiments.

The crucial nature of water binding

Water plays a special and complicated role in molecular recognition. It is absolutely essential to determine the sites on a macromolecule that have high affinity for water. Dehydrating a protein or nucleic acid by driving it through a phase transition quantitatively determines the details of water binding at every site of the macromolecule. When the system approaches a chemical potential that enters the phase transition, it is at a free energy that begins to break the water-water-water solvent cohesion. Thus, an additional very small lowering of the chemical potential completely destroys the multi-body solvent cohesion and all bulk water molecules are evacuated from the system. Many of the waters bound to the macromolecule are also removed during the phase transition, but a few finite sites on the macromolecular continue to have bound waters often in a multi-body configuration after the phase transition. What does this mean physically? The sites on the macromolecule that are unable to continue holding on to the water molecules through the phase transition indicate that these sites have lower affinity

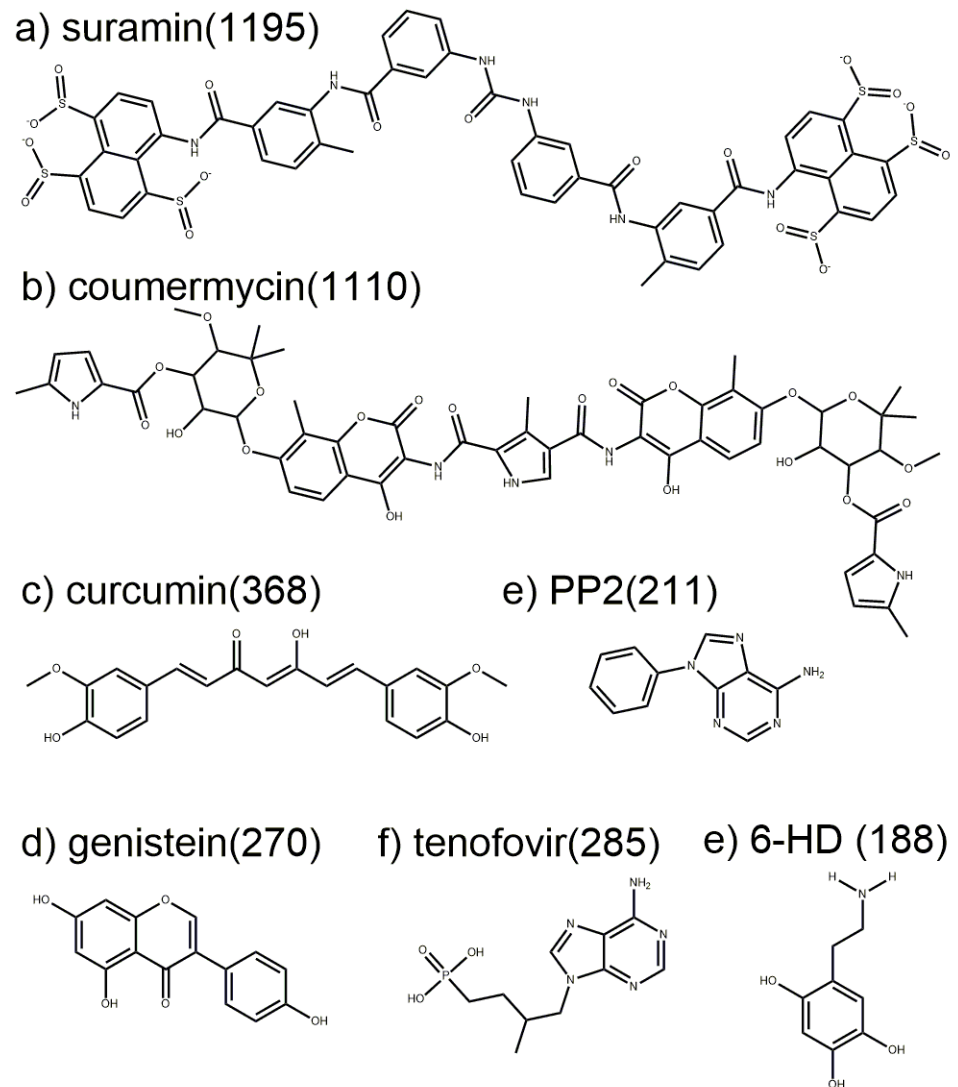


Fig 13. RecA ATPase inhibitors. Many different classes of ATPase inhibitors were tested on RecA and shown to have no effect. a) Suramin was developed in the early 1900s to treat parasitic diseases like African sleeping sickness and is on the World Health Organization's list of essential medicines. b) Coumermycin is a natural product produced by *Streptomyces*, which potently inhibits staphylococcus gyrase. c) Curcumin and d) genistein are natural products that have been widely studied as antibiotics and anticancer agents. e) PP2 inhibits tyrosine kinases and has been widely studied as an anticancer agent. f) Tenofovir is an FDA-approved treatment for human immunodeficiency virus 1 (HIV-1) and hepatitis B virus infections. g) 6-HD discovered with SACP is the only selective inhibitor of RecA within this group. Numbers in parentheses are molecular weight for each compound.

<https://doi.org/10.1371/journal.pone.0183327.g013>

for water than the bulk water-water-water interactions. By extension, this means that at these sites the bound waters rapidly and spontaneously exchange with the bulk. The sites on the macromolecule that retain water binding through the phase transition have a higher affinity for water than the bulk water-water-water interactions, and thus these are ordered bound waters.

Protein binding sites in almost all cases do not have any bound waters when driven through the phase transition. This can be understood from the perspective that Nature needs to have ready access to binding sites to perform functional operations and that bound waters cannot

be in the way, because the operations have to be reversible and repeatedly carried out. Cycles of binding and removing high affinity water molecules would be too costly energetically to be part of this function. What is very interesting, in some of our simulations on proteins that do use water in catalysis, we generally find the water bound at the functional site in the region of the inflection point of the phase transition. Thus, this water has high enough affinity to participate in functional operations, but binds with only modest affinity and thus can be readily removed.

High affinity binding of diverse fragment clusters with water exclusion is a hypothesis about the general nature of a protein interaction site. Physically, this hypothesis states that a large diversity at a localized site must exist to have enough interactions to generate specificity, and this set of interactions must readily out-compete water. Given all the ligand-macromolecule and macromolecule-macromolecule interactions that occur in nature, a combinatorial diversity of many different interactions must occur in order to attain specificity. Since water is omnipresent, we hypothesize that the interaction sites must be configured, in general, to not bind water with high affinity, because displacing high affinity waters would be too costly.

We always first use SACP to determine the sites on a macromolecule that have high affinity for water—again this is quantitatively and automatically determined by generating a dehydrating phase transition. We do the same for every fragment to find high affinity protein binding sites. By locating sites on the protein that have high affinity for a fragment cluster and only considering the sites that also do not bind water, we are explicitly determining a cluster of interactions that can easily out-compete water binding.

For all cases in this study, the fragment cluster analysis was done only at sites where water was predicted to not bind except for HIV protease, PDF, and BPTI. The reason for this is because these are water molecules that continue to bind to the protein after the phase transition and thus are the exception to the water exclusion principle. The HIV case was explicitly chosen to see if SACP correctly predicted this well-known water molecule at the center of the binding site. This was an important test, because we have run SACP on over 100 proteins and in all cases the binding sites do not retain water after the phase transition. So the fact that SACP creates a ring of waters around the outside of the HIV protease binding site in accord with the water exclusion principle, but also predicts the singular high affinity water at the center of the binding site is extremely important. HIV protease inhibitor design often incorporates protein interactions mediated by this water, so it is important that a truly predictive technique finds this water since it provides key bridging interactions with the inhibitor. We include the BPTI simulations to illustrate the need for multi-body sampling that overcomes high energy barriers and also to illustrate the nature of how water binds as a collective. Finally, the only fragment simulations that included bound waters were for PDF. In all other cases the bound waters were far away from the binding site, so they were immaterial. For PDF, however, the triplet of waters that completed the outer edge of the binding site remained after the phase transition. These waters mediate the predicted ligand binding that was experimentally confirmed.

Conclusions

The collection of systems chosen for this study were selected to test a computational fragment-based method in a broad diversity of molecular recognition events that include electron transfer (DHFR), multiple ligand binding co-crystals that dramatically expand our knowledge of the protein binding site over time (Elastase), water-mediated recognition (HIV protease), protein-protein interactions with subtle differences in molecular recognition at the same site (p53-MDM2 versus p53-MDMX), and protein-RNA interactions (TAT-TAR of HIV). The

surprising and unexpected results on two additional systems, RecA and PDF, were so intriguing that we decided to conduct fragment binding experiments to test these predictions. In all cases using SACP to uncover the high affinity locations on a macromolecule for an organic fragment or water by driving the chemical potential through a phase transition coupled with clustering and water exclusion identified the key locations responsible for molecular recognition. Taken together, these studies provide strong evidence that fragment-based techniques coupled with our three-part hypothesis is a powerful general way of investigating molecular recognition phenomena.

Supporting information

S1 File. Supporting information. This file contains the details of the biological assays, MS and NMR characterization of the small molecules described in this study. The files also contains additional information about the details of the simulations.

(PDF)

S2 File. PDB data files. PDB data files containing computed fragments.

(ZIP)

S1 Fig. An illustration of the thermodynamics of ligand-protein binding. The protein binding site must be dehydrated with a DG(P-H₂O) as shown in the top line in order to bind a ligand. The ligand must be dehydrated with a DG(L-H₂O) as shown in the bottom line in order to interact with the protein. These two lines converge in the middle with the ligand coming together with the protein DG(P-L). Fragment binding and protein hydration-dehydration are rigorously computed with Simulated Annealing of Chemical Potential (SACP). Ligand dehydration was neglected for the reasons described in the text.

(TIF)

S2 Fig. Basis fragment set for hot-spot simulations. The basis fragment set used to probe the various targets described in the manuscript. This is similar to the dataset in our previous publication.

(TIF)

Acknowledgments

Part of the research reported in this publication was supported by National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health under award number R43GM109549.

Author Contributions

Conceptualization: John L. Kulp, III, Ian S. Cloudsdale, John L. Kulp, Jr, Frank Guarnieri.

Data curation: John L. Kulp, III, Ian S. Cloudsdale.

Formal analysis: John L. Kulp, III, Ian S. Cloudsdale, John L. Kulp, Jr, Frank Guarnieri.

Funding acquisition: John L. Kulp, Jr.

Investigation: John L. Kulp, III, Ian S. Cloudsdale, John L. Kulp, Jr, Frank Guarnieri.

Methodology: John L. Kulp, III, Ian S. Cloudsdale, John L. Kulp, Jr, Frank Guarnieri.

Project administration: John L. Kulp, Jr, Frank Guarnieri.

Resources: John L. Kulp, III, John L. Kulp, Jr, Frank Guarnieri.

Software: John L. Kulp, III, John L. Kulp, Jr.

Supervision: John L. Kulp, Jr, Frank Guarnieri.

Validation: John L. Kulp, III, Ian S. Cloudsdale, Frank Guarnieri.

Visualization: John L. Kulp, III, Frank Guarnieri.

Writing – original draft: John L. Kulp, III, Frank Guarnieri.

Writing – review & editing: John L. Kulp, III.

References

1. Edink E, Rucktooa P, Retra K, Akdemir A, Nahar T, Zuiderveld O, et al. Fragment growing induces conformational changes in acetylcholine-binding protein: a structural and thermodynamic analysis. *J Am Chem Soc.* 2011; 133(14):5363–71. <https://doi.org/10.1021/ja110571r> PMID: 21322593
2. Egger J, Weckerle C, Cutting B, Schwaradt O, Rabbani S, Lemme K, et al. Nanomolar E-selectin antagonists with prolonged half-lives by a fragment-based approach. *J Am Chem Soc.* 2013; 135(26):9820–8. <https://doi.org/10.1021/ja4029582> PMID: 23742188
3. Guan JY, Keizers PH, Liu WM, Lohr F, Skinner SP, Heeneman EA, et al. Small-molecule binding sites on proteins established by paramagnetic NMR spectroscopy. *J Am Chem Soc.* 2013; 135(15):5859–68. <https://doi.org/10.1021/ja401323m> PMID: 23509882
4. Hay DA, Fedorov O, Martin S, Singleton DC, Tallant C, Wells C, et al. Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomains. *J Am Chem Soc.* 2014; 136(26):9308–19. <https://doi.org/10.1021/ja412434f> PMID: 24946055
5. Miller RM, Paavilainen VO, Krishnan S, Serafimova IM, Taunton J. Electrophilic fragment-based design of reversible covalent kinase inhibitors. *J Am Chem Soc.* 2013; 135(14):5298–301. <https://doi.org/10.1021/ja401221b> PMID: 23540679
6. Wilcken R, Liu X, Zimmermann MO, Rutherford TJ, Fersht AR, Joerger AC, et al. Halogen-enriched fragment libraries as leads for drug rescue of mutant p53. *J Am Chem Soc.* 2012; 134(15):6810–8. <https://doi.org/10.1021/ja301056a> PMID: 22439615
7. Hao GF, Wang F, Li H, Zhu XL, Yang WC, Huang LS, et al. Computational discovery of picomolar Q (o) site inhibitors of cytochrome bc1 complex. *J Am Chem Soc.* 2012; 134(27):11168–76. <https://doi.org/10.1021/ja3001908> PMID: 22690928
8. Raman EP, MacKerell AD Jr. Spatial analysis and quantification of the thermodynamic driving forces in protein-ligand binding: binding site variability. *J Am Chem Soc.* 2015; 137(7):2608–21. <https://doi.org/10.1021/ja512054f> PMID: 25625202
9. Raman EP, Yu W, Lakkaraju SK, MacKerell AD Jr. Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J Chem Inf Model.* 2013; 53(12):3384–98. <https://doi.org/10.1021/ci4005628> PMID: 24245913
10. Allen KN, Bellamacina CR, Ding X, Jeffery CJ, Mattos C, Petsko GA, et al. An Experimental Approach to Mapping the Binding Surfaces of Crystalline Proteins. *J Phys Chem.* 1996; 100(7):2605–11.
11. Mattos C, Ringe D. Locating and characterizing binding sites on proteins. *Nat Biotechnol.* 1996; 14(5):595–9. <https://doi.org/10.1038/nbt0596-595> PMID: 9630949
12. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW. Discovering high-affinity ligands for proteins: SAR by NMR. *Science.* 1996; 274(5292):1531–4. PMID: 8929414
13. Mashalidis EH, Sledz P, Lang S, Abell C. A three-stage biophysical screening cascade for fragment-based drug discovery. *Nat Protoc.* 2013; 8(11):2309–24. <https://doi.org/10.1038/nprot.2013.130> PMID: 24157549
14. Silvestre HL, Blundell TL, Abell C, Ciulli A. Integrated biophysical approach to fragment screening and validation for fragment-based lead discovery. *Proc Natl Acad Sci U S A.* 2013; 110(32):12984–9. <https://doi.org/10.1073/pnas.1304045110> PMID: 23872845
15. Chessari G, Woodhead AJ. From fragment to clinical candidate—a historical perspective. *Drug Discov Today.* 2009; 14(13–14):668–75. <https://doi.org/10.1016/j.drudis.2009.04.007> PMID: 19427404
16. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H. Twenty years on: the impact of fragments on drug discovery. *Nat Rev Drug Discov.* 2016; 15(9):605–19. <https://doi.org/10.1038/nrd.2016.109> PMID: 27417849

17. Davies TG, Wixted WE, Coyle JE, Griffiths-Jones C, Hearn K, McMenamin R, et al. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1: NRF2) Protein-Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *J Med Chem.* 2016; 59(8):3991–4006. <https://doi.org/10.1021/acs.jmedchem.6b00228> PMID: 27031670
18. Gable JE, Lee GM, Acker TM, Hulce KR, Gonzalez ER, Schweigler P, et al. Fragment-Based Protein-Protein Interaction Antagonists of a Viral Dimeric Protease. *ChemMedChem.* 2016; 11(8):862–9. <https://doi.org/10.1002/cmdc.201500526> PMID: 26822284
19. Liu LK, Finzel BC. Fragment-based identification of an inducible binding site on cell surface receptor CD44 for the design of protein-carbohydrate interaction inhibitors. *J Med Chem.* 2014; 57(6):2714–25. <https://doi.org/10.1021/jm5000276> PMID: 24606063
20. Mikuni J, Kato M, Taruya S, Tsuganezawa K, Mori M, Ogawa N, et al. A fluorescence correlation spectroscopy-based assay for fragment screening of slowly inhibiting protein-peptide interaction inhibitors. *Anal Biochem.* 2010; 402(1):26–31. <https://doi.org/10.1016/j.ab.2010.03.019> PMID: 20298671
21. Rouhana J, Hoh F, Estaran S, Henriquet C, Boublik Y, Kerkour A, et al. Fragment-based identification of a locus in the Sec7 domain of Arno for the design of protein-protein interaction inhibitors. *J Med Chem.* 2013; 56(21):8497–511. <https://doi.org/10.1021/jm4009357> PMID: 24112024
22. Waaijers S, Koorman T, Kerver J, Boxem M. Identification of human protein interaction domains using an ORFeome-based yeast two-hybrid fragment library. *J Proteome Res.* 2013; 12(7):3181–92. <https://doi.org/10.1021/pr400047p> PMID: 23718855
23. Winter A, Sigurdardottir AG, DiCara D, Valenti G, Blundell TL, Gherardi E. Developing Antagonists for the Met-HGF/SF Protein-Protein Interaction Using a Fragment-Based Approach. *Mol Cancer Ther.* 2016; 15(1):3–14. <https://doi.org/10.1158/1535-7163.MCT-15-0446> PMID: 26712116
24. Goodford PJ. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J Med Chem.* 1985; 28(7):849–57. PMID: 3892003
25. Miranker A, Karplus M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins.* 1991; 11(1):29–34. <https://doi.org/10.1002/prot.340110104> PMID: 1961699
26. Dennis S, Kortvelyesi T, Vajda S. Computational mapping identifies the binding sites of organic solvents on proteins. *Proc Natl Acad Sci U S A.* 2002; 99(7):4290–5. <https://doi.org/10.1073/pnas.062398499> PMID: 11904374
27. Bohnuud T, Beglov D, Ngan CH, Zerbe B, Hall DR, Brenke R, et al. Computational mapping reveals dramatic effect of Hoogsteen breathing on duplex DNA reactivity with formaldehyde. *Nucleic Acids Res.* 2012; 40(16):7644–52. <https://doi.org/10.1093/nar/gks519> PMID: 22705795
28. Buhrman G, O'Connor C, Zerbe B, Kearney BM, Napoleon R, Kovrigina EA, et al. Analysis of binding site hot spots on the surface of Ras GTPase. *J Mol Biol.* 2011; 413(4):773–89. <https://doi.org/10.1016/j.jmb.2011.09.011> PMID: 21945529
29. Chuang GY, Kozakov D, Brenke R, Beglov D, Guarnieri F, Vajda S. Binding hot spots and amantadine orientation in the influenza A virus M2 proton channel. *Biophys J.* 2009; 97(10):2846–53. <https://doi.org/10.1016/j.bpj.2009.09.004> PMID: 19917240
30. Golden MS, Cote SM, Sayeg M, Zerbe BS, Villar EA, Beglov D, et al. Comprehensive experimental and computational analysis of binding energy hot spots at the NF-kappaB essential modulator/IKK-beta protein-protein interface. *J Am Chem Soc.* 2013; 135(16):6242–56. <https://doi.org/10.1021/ja400914z> PMID: 23506214
31. Hall DH, Grove LE, Yueh C, Ngan CH, Kozakov D, Vajda S. Robust identification of binding hot spots using continuum electrostatics: application to hen egg-white lysozyme. *J Am Chem Soc.* 2011; 133(51):20668–71. <https://doi.org/10.1021/ja207914y> PMID: 22092261
32. Kozakov D, Hall DR, Chuang GY, Cencic R, Brenke R, Grove LE, et al. Structural conservation of druggable hot spots in protein-protein interfaces. *Proc Natl Acad Sci U S A.* 2011; 108(33):13528–33. <https://doi.org/10.1073/pnas.1101835108> PMID: 21808046
33. Guarnieri F, Mezei M. Simulated Annealing of Chemical Potential: A General Procedure for Locating Bound Waters. Application to the Study of the Differential Hydration Propensities of the Major and Minor Grooves of DNA. *J Am Chem Soc.* 1996; 118(35):8493–94.
34. Kulp JL 3rd, Blumenthal SN, Wang Q, Bryan RL, Guarnieri F. A fragment-based approach to the SAMPL3 Challenge. *J Comput Aided Mol Des.* 2012; 26(5):583–94. <https://doi.org/10.1007/s10822-012-9546-1> PMID: 22290624
35. Kulp JL 3rd, Kulp JL Jr., Pompliano DL, Guarnieri F. Diverse fragment clustering and water exclusion identify protein hot spots. *J Am Chem Soc.* 2011; 133(28):10740–3. <https://doi.org/10.1021/ja203929x> PMID: 21682273

36. Guarnieri F. Computational Protein Probing to Identify Binding Sites. Computational Protein Probing to Identify Binding Sites. 2004;U.S. Patent 6735530.
37. Bradbury MW, Stump D, Guarnieri F, Berk PD. Molecular modeling and functional confirmation of a predicted fatty acid binding site of mitochondrial aspartate aminotransferase. *J Mol Biol.* 2011; 412(3):412–22. <https://doi.org/10.1016/j.jmb.2011.07.034> PMID: 21803047
38. Vallee M, Vitiello S, Bellocchio L, Hebert-Chatelain E, Monlezun S, Martin-Garcia E, et al. Pregnenolone can protect the brain from cannabis intoxication. *Science.* 2014; 343(6166):94–8. <https://doi.org/10.1126/science.1243985> PMID: 24385629
39. Guarnieri F. Designing a small molecule erythropoietin mimetic. *Fragment-Based Methods in Drug Discovery*; Klon Anthony, Editor. 2015; 1289:185–210.
40. Guarnieri F. Designing an Orally Available Nontoxic p38 Inhibitor with a Fragment-Based Strategy. *Fragment-Based Methods in Drug Discovery*; Klon Anthony, Editor. 2015; 1289:211–26.
41. Moffett K, Konteatis Z, Nguyen D, Shetty R, Ludington J, Fujimoto T, et al. Discovery of a novel class of non-ATP site DFG-out state p38 inhibitors utilizing computationally assisted virtual fragment-based drug design (vFBDD). *Bioorg Med Chem Lett.* 2011; 21(23):7155–65. <https://doi.org/10.1016/j.bmcl.2011.09.078> PMID: 22014550
42. Moore WR Jr. Maximizing discovery efficiency with a computationally driven fragment approach. *Curr Opin Drug Discov Devel.* 2005; 8(3):355–64. PMID: 15892251
43. Alber T, Petsko GA, Tsernoglou D. Crystal structure of elastase-substrate complex at—55 degrees C. *Nature.* 1976; 263(5575):297–300. PMID: 958484
44. Hughes DL, Sieker LC, Bieth J, Dimicoli JL. Crystallographic study of the binding of a trifluoroacetyl dipeptide anilide inhibitor with elastase. *J Mol Biol.* 1982; 162(3):645–58. PMID: 6926029
45. Meyer EF Jr., Radhakrishnan R, Cole GM, Presta LG. Structure of the product complex of acetyl-Ala-Pro-Ala with porcine pancreatic elastase at 1.65 Å resolution. *J Mol Biol.* 1986; 189(3):533–9. PMID: 3640831
46. Navia MA, McKeever BM, Springer JP, Lin TY, Williams HR, Fluder EM, et al. Structure of human neutrophil elastase in complex with a peptide chloromethyl ketone inhibitor at 1.84-Å resolution. *Proc Natl Acad Sci U S A.* 1989; 86(1):7–11. PMID: 2911584
47. Mattos C, Rasmussen B, Ding X, Petsko GA, Ringe D. Analogous inhibitors of elastase do not always bind analogously. *Nat Struct Biol.* 1994; 1(1):55–8. PMID: 7656008
48. Sawaya MR, Kraut J. Loop and subdomain movements in the mechanism of Escherichia coli dihydrofolate reductase: crystallographic evidence. *Biochemistry.* 1997; 36(3):586–603. <https://doi.org/10.1021/bi962337c> PMID: 9012674
49. Chen Z, Li Y, Chen E, Hall DL, Darke PL, Culberson C, et al. Crystal structure at 1.9-Å resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *J Biol Chem.* 1994; 269(42):26344–8. PMID: 7929352
50. Hu B, Gilkes DM, Farooqi B, Sebt SM, Chen J. MDMX overexpression prevents p53 activation by the MDM2 inhibitor Nutlin. *J Biol Chem.* 2006; 281(44):33030–5. <https://doi.org/10.1074/jbc.C600147200> PMID: 16905541
51. Joseph TL, Madhumalar A, Brown CJ, Lane DP, Verma CS. Differential binding of p53 and nutlin to MDM2 and MDMX: computational studies. *Cell Cycle.* 2010; 9(6):1167–81. <https://doi.org/10.4161/cc.9.6.11067> PMID: 20190571
52. Li C, Pazgier M, Li C, Yuan W, Liu M, Wei G, et al. Systematic Mutational Analysis of Peptide Inhibition of the p53–MDM2/MDMX Interactions. *Journal of Molecular Biology.* 2010; 398(2):200–13. <https://doi.org/10.1016/j.jmb.2010.03.005> PMID: 20226197
53. Murray K, Gellman S.H. Targeting protein–protein interactions: Lessons from p53/MDM2. *Peptide Science.* 2007; 88(5):657–86. <https://doi.org/10.1002/bip.20741> PMID: 17427181
54. Popowicz GM, Dömling A, Holak TA. The Structure-Based Design of Mdm2/Mdmx—p53 Inhibitors Gets Serious. *Angewandte Chemie, International Edition.* 2011; 50(12):2680–8.
55. Qin L, Yang F, Zhou C, Chen Y, Zhang H, Su Z. Efficient reactivation of p53 in cancer cells by a dual MdmX/Mdm2 inhibitor. *J Am Chem Soc.* 2014; 136(52):18023–33. <https://doi.org/10.1021/ja509223m> PMID: 25453499
56. Feki A, Irminger-Finger I. Mutational spectrum of p53 mutations in primary breast and ovarian tumors. *Critical reviews in oncology/hematology.* 2004; 52(2):103–16. <https://doi.org/10.1016/j.critrevonc.2004.07.002> PMID: 15501075
57. Massari S, Sabatini S, Tabarrini O. Blocking HIV-1 replication by targeting the Tat-hijacked transcriptional machinery. *Curr Pharm Des.* 2013; 19(10):1860–79. PMID: 23092279

58. Mousseau G, Valente S. Strategies to Block HIV Transcription: Focus on Small Molecule Tat Inhibitors. *Biology (Basel)*. 2012; 1(3):668–97.
59. Davidson A, Patora-Komisarska K, Robinson JA, Varani G. Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. *Nucleic Acids Res*. 2011; 39(1):248–56. <https://doi.org/10.1093/nar/gkq713> PMID: 20724442
60. Roca AI, Cox MM. RecA protein: structure, function, and role in recombinational DNA repair. *Prog Nucleic Acid Res Mol Biol*. 1997; 56:129–223. PMID: 9187054
61. Indiani C, Patel M, Goodman MF, O'Donnell ME. RecA acts as a switch to regulate polymerase occupancy in a moving replication fork. *Proc Natl Acad Sci U S A*. 2013; 110(14):5410–5. <https://doi.org/10.1073/pnas.1303301110> PMID: 23509251
62. Kohanski MA, Dwyer DJ, Hayete B, Lawrence CA, Collins JJ. A common mechanism of cellular death induced by bactericidal antibiotics. *Cell*. 2007; 130(5):797–810. <https://doi.org/10.1016/j.cell.2007.06.049> PMID: 17803904
63. Bell JC, Plank JL, Dombrowski CC, Kowalczykowski SC. Direct imaging of RecA nucleation and growth on single molecules of SSB-coated ssDNA. *Nature*. 2012; 491(7423):274–8. <https://doi.org/10.1038/nature11598> PMID: 23103864
64. Steffens LS, Nicholson S, Paul LV, Nord CE, Patrick S, Abratt VR. *Bacteroides fragilis* RecA protein overexpression causes resistance to metronidazole. *Res Microbiol*. 2010; 161(5):346–54. <https://doi.org/10.1016/j.resmic.2010.04.003> PMID: 20435137
65. Corey R, Naderer OJ, O'Riordan WD, Dumont E, Jones LS, Kurtinecz M, et al. Safety, tolerability, and efficacy of GSK1322322 in the treatment of acute bacterial skin and skin structure infections. *Antimicrob Agents Chemother*. 2014; 58(11):6518–27. <https://doi.org/10.1128/AAC.03360-14> PMID: 25136015
66. Randhawa H, Chikara S, Gehring D, Yildirim T, Menon J, Reindl KM. Overexpression of peptide deformylase in breast, colon, and lung cancers. *BMC Cancer*. 2013; 13:321. <https://doi.org/10.1186/1471-2407-13-321> PMID: 23815882
67. Sheth A, Escobar-Alvarez S, Gardner J, Ran L, Heaney ML, Scheinberg DA. Inhibition of human mitochondrial peptide deformylase causes apoptosis in c-myc-overexpressing hematopoietic cancers. *Cell Death Dis*. 2014; 5:e1152. <https://doi.org/10.1038/cddis.2014.112> PMID: 24675470
68. Katada T. The inhibitory G protein G(i) identified as pertussis toxin-catalyzed ADP-ribosylation. *Biol Pharm Bull*. 2012; 35(12):2103–11. PMID: 23207763
69. Wernick NL, Chinnapen DJ, Cho JA, Lencer WI. Cholera toxin: an intracellular journey into the cytosol by way of the endoplasmic reticulum. *Toxins (Basel)*. 2010; 2(3):310–25.
70. Lax A. The *Pasteurella multocida* toxin: a new paradigm for the link between bacterial infection and cancer. *Curr Top Microbiol Immunol*. 2012; 361:131–44. https://doi.org/10.1007/82_2012_236 PMID: 22695919
71. Siegert P, Schmidt G, Papatheodorou P, Wieland T, Aktories K, Orth JH. *Pasteurella Multocida* Toxin Prevents Osteoblast Differentiation by Transactivation of the MAP-Kinase Cascade via the Galphaq/11—p63RhoGEF—RhoA Axis. *PLoS Pathog*. 2013; 9(5):e1003385. <https://doi.org/10.1371/journal.ppat.1003385> PMID: 23696743
72. Aktories K. Bacterial protein toxins that modify host regulatory GTPases. *Nat Rev Microbiol*. 2011; 9(7):487–98. <https://doi.org/10.1038/nrmicro2592> PMID: 21677684
73. Grasberger BL, Lu T, Schubert C, Parks DJ, Carver TE, Koblisch HK, et al. Discovery and cocrystal structure of benzodiazepinedione HDM2 antagonists that activate p53 in cells. *Journal of Medicinal Chemistry*. 2005; 48(4):909–12. <https://doi.org/10.1021/jm049137g> PMID: 15715460
74. Czarna A, Popowicz GM, Pecak A, Wolf S, Dubin G, Holak TA. High affinity interaction of the p53 peptide-analogue with human Mdm2 and Mdmx. *Cell Cycle*. 2009; 8(8):1176–84. <https://doi.org/10.4161/cc.8.8.8185> PMID: 19305137
75. Scott DE, Ehebauer MT, Pukala T, Marsh M, Blundell TL, Venkiteshwaran AR, et al. Using a fragment-based approach to target protein-protein interactions. *Chembiochem*. 2013; 14(3):332–42. <https://doi.org/10.1002/cbic.201200521> PMID: 23344974
76. Wellendorph P, Johansen LD, Brauner-Osborne H. Molecular pharmacology of promiscuous seven transmembrane receptors sensing organic nutrients. *Mol Pharmacol*. 2009; 76(3):453–65. <https://doi.org/10.1124/mol.109.055244> PMID: 19487246
77. Zahringer U, Lindner B, Inamura S, Heine H, Alexander C. TLR2—promiscuous or specific? A critical re-evaluation of a receptor expressing apparent broad specificity. *Immunobiology*. 2008; 213(3–4):205–24. <https://doi.org/10.1016/j.imbio.2008.02.005> PMID: 18406368
78. Eakin AE, Green O, Hales N, Walkup GK, Bist S, Singh A, et al. Pyrrolamide DNA gyrase inhibitors: fragment-based nuclear magnetic resonance screening to identify antibacterial agents.

- Antimicrob Agents Chemother. 2012; 56(3):1240–6. <https://doi.org/10.1128/AAC.05485-11> PMID: 22183167
79. Chilingaryan Z, Yin Z, Oakley AJ. Fragment-based screening by protein crystallography: successes and pitfalls. *Int J Mol Sci.* 2012; 13(10):12857–79. <https://doi.org/10.3390/ijms131012857> PMID: 23202926
 80. Rojas A, Delgado-Lopez F, Gonzalez I, Perez-Castro R, Romero J, Rojas I. The receptor for advanced glycation end-products: a complex signaling scenario for a promiscuous receptor. *Cell Signal.* 2013; 25(3):609–14. <https://doi.org/10.1016/j.cellsig.2012.11.022> PMID: 23200851
 81. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pKa values. *Proteins: Structure, Function, and Bioinformatics.* 2005; 61(4):704–21.
 82. Adams DJ. Grand canonical ensemble Monte Carlo for a Lennard-Jones fluid. *Mol Phys* 1975; 29(1):307–11.
 83. Winter A, Higuero AP, Marsh M, Sigurdardottir A, Pitt WR, Blundell TL. Biophysical and computational fragment-based approaches to targeting protein-protein interactions: applications in structure-guided drug discovery. *Q Rev Biophys.* 2012; 45(4):383–426. <https://doi.org/10.1017/S0033583512000108> PMID: 22971516
 84. Van Molle I, Thomann A, Buckley DL, So EC, Lang S, Crews CM, et al. Dissecting fragment-based lead discovery at the von Hippel-Lindau protein:hypoxia inducible factor 1alpha protein-protein interface. *Chem Biol.* 2012; 19(10):1300–12. <https://doi.org/10.1016/j.chembiol.2012.08.015> PMID: 23102223
 85. Mezei M. Grand-canonical ensemble Monte Carlo study of dense liquid Lennard-Jones, soft spheres and water. *Mol Phys.* 1987; 61:565–82.
 86. Kelley De Zutter J, Forget AL, Logan KM, Knight KL. Phe217 regulates the transfer of allosteric information across the subunit interface of the RecA protein filament. *Structure.* 2001; 9(1):47–55. PMID: 11342134
 87. Lee AM, Wigle TJ, Singleton SF. A complementary pair of rapid molecular screening assays for RecA activities. *Anal Biochem.* 2007; 367(2):247–58. <https://doi.org/10.1016/j.ab.2007.04.021> PMID: 17601483
 88. Peterson EJ, Janzen WP, Kireev D, Singleton SF. High-throughput screening for RecA inhibitors using a transcriber adenosine 5'-O-diphosphate assay. *Assay Drug Dev Technol.* 2012; 10(3):260–8. <https://doi.org/10.1089/adt.2011.0409> PMID: 22192312
 89. Marcellino D, Suarez-Boomgaard D, Sanchez-Reina MD, Aguirre JA, Yoshitake T, Yoshitake S, et al. On the role of P2X(7) receptors in dopamine nerve cell degeneration in a rat model of Parkinson's disease: studies with the P2X(7) receptor antagonist A-438079. *J Neural Transm.* 2010; 117(6):681–7. <https://doi.org/10.1007/s00702-010-0400-0> PMID: 20387084
 90. Blaha G, Stanley RE, Steitz TA. Formation of the first peptide bond: the structure of EF-P bound to the 70S ribosome. *Science.* 2009; 325(5943):966–70. <https://doi.org/10.1126/science.1175800> PMID: 19696344
 91. Huguet F, Melet A, Alves de Sousa R, Lieutaud A, Chevalier J, Maigre L, et al. Hydroxamic acids as potent inhibitors of Fe(II) and Mn(II) E. coli methionine aminopeptidase: biological activities and X-ray structures of oxazole hydroxamate-EcMetAP-Mn complexes. *ChemMedChem.* 2012; 7(6):1020–30. <https://doi.org/10.1002/cmdc.201200076> PMID: 22489069
 92. Cui P, Wang Y, Chu W, Guo X, Yang F, Yu M, et al. How water molecules affect the catalytic activity of hydrolases—A XANES study of the local structures of peptide deformylase. *Sci Rep.* 2014; 4:7453. <https://doi.org/10.1038/srep07453> PMID: 25503313
 93. Boularot A, Giglione C, Petit S, Duroc Y, Alves de Sousa R, Larue V, et al. Discovery and refinement of a new structural class of potent peptide deformylase inhibitors. *J Med Chem.* 2007; 50(1):10–20. <https://doi.org/10.1021/jm060910c> PMID: 17201406
 94. Cheng YK, Rosky PJ. Surface topography dependence of biomolecular hydrophobic hydration. *Nature.* 1998; 392(6677):696–9. <https://doi.org/10.1038/33653> PMID: 9565030
 95. Dorsey BD, Levin RB, McDaniel SL, Vacca JP, Guare JP, Darke PL, et al. L-735,524: the design of a potent and orally bioavailable HIV protease inhibitor. *J Med Chem.* 1994; 37(21):3443–51. PMID: 7932573
 96. Amano M, Tojo Y, Salcedo-Gomez PM, Campbell JR, Das D, Aoki M, et al. GRL-0519, a novel oxatri-cyclic ligand-containing nonpeptidic HIV-1 protease inhibitor (PI), potently suppresses replication of a wide spectrum of multi-PI-resistant HIV-1 variants in vitro. *Antimicrob Agents Chemother.* 2013; 57(5):2036–46. <https://doi.org/10.1128/AAC.02189-12> PMID: 23403426
 97. Liu Z, Wang Y, Yedidi RS, Dewdney TG, Reiter SJ, Brunzelle JS, et al. Conserved hydrogen bonds and water molecules in MDR HIV-1 protease substrate complexes. *Biochem Biophys Res Commun.* 2013; 430(3):1022–7. <https://doi.org/10.1016/j.bbrc.2012.12.045> PMID: 23261453

98. Prashar V, Bihani SC, Das A, Rao DR, Hosur MV. Insights into the mechanism of drug resistance: X-ray structure analysis of G48V/C95F tethered HIV-1 protease dimer/saquinavir complex. *Biochem Biophys Res Commun.* 2010; 396(4):1018–23. <https://doi.org/10.1016/j.bbrc.2010.05.049> PMID: [20471372](https://pubmed.ncbi.nlm.nih.gov/20471372/)
99. Yedidi RS, Liu Z, Wang Y, Brunzelle JS, Kovari IA, Woster PM, et al. Crystal structures of multidrug-resistant HIV-1 protease in complex with two potent anti-malarial compounds. *Biochem Biophys Res Commun.* 2012; 421(3):413–7. <https://doi.org/10.1016/j.bbrc.2012.03.096> PMID: [22469467](https://pubmed.ncbi.nlm.nih.gov/22469467/)
100. Kumar M, Kannan KK, Hosur MV, Bhavesh NS, Chatterjee A, Mittal R, et al. Effects of remote mutation on the autolysis of HIV-1 PR: X-ray and NMR investigations. *Biochem Biophys Res Commun.* 2002; 294(2):395–401. [https://doi.org/10.1016/S0006-291X\(02\)00482-5](https://doi.org/10.1016/S0006-291X(02)00482-5) PMID: [12051725](https://pubmed.ncbi.nlm.nih.gov/12051725/)
101. Liu F, Kovalevsky AY, Louis JM, Boross PI, Wang YF, Harrison RW, et al. Mechanism of drug resistance revealed by the crystal structure of the unliganded HIV-1 protease with F53L mutation. *J Mol Biol.* 2006; 358(5):1191–9. <https://doi.org/10.1016/j.jmb.2006.02.076> PMID: [16569415](https://pubmed.ncbi.nlm.nih.gov/16569415/)
102. Mehler EL, Guarnieri F. A self-consistent, microenvironment modulated screened coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys J.* 1999; 77(1):3–22. [https://doi.org/10.1016/S0006-3495\(99\)76868-2](https://doi.org/10.1016/S0006-3495(99)76868-2) PMID: [10388736](https://pubmed.ncbi.nlm.nih.gov/10388736/)
103. Wang Z, Zhu G, Huang Q, Qian M, Shao M, Jia Y, et al. X-ray studies on cross-linked lysozyme crystals in acetonitrile-water mixture. *Biochimica et Biophysica Acta.* 1998; 1384(2):335–44. PMID: [9659395](https://pubmed.ncbi.nlm.nih.gov/9659395/)
104. Pierce JA, Eisen AZ, Dhingra HK. Relationship of antitrypsin deficiency to the pathogenesis of emphysema. *Trans Assoc Am Physicians.* 1969; 82:87–97. PMID: [5310288](https://pubmed.ncbi.nlm.nih.gov/5310288/)
105. Turino GM, Seniorrm, Garg BD, Keller S, Levi MM, Mandl I. Serum elastase inhibitor deficiency and alpha 1-antitrypsin deficiency in patients with obstructive emphysema. *Science.* 1969; 165(894):709–11.
106. Guarnieri F. *The Dichotomy Between Understanding and Treating Emphysema.* Emphysema. 2012; Publisher: Intech; Editor Mahadeva R.(Chapter 4).
107. Guarnieri F, Spencer JL, Lucey EC, Nugent MA, Stone PJ. A human surfactant peptide-elastase inhibitor construct as a treatment for emphysema. *Proc Natl Acad Sci U S A.* 2010; 107(23):10661–6. <https://doi.org/10.1073/pnas.1001349107> PMID: [20534582](https://pubmed.ncbi.nlm.nih.gov/20534582/)
108. Mattos C, Bellamacina CR, Peisach E, Pereira A, Vitkup D, Petsko GA, et al. Multiple solvent crystal structures: probing binding sites, plasticity and hydration. *J Mol Biol.* 2006; 357(5):1471–82. <https://doi.org/10.1016/j.jmb.2006.01.039> PMID: [16488429](https://pubmed.ncbi.nlm.nih.gov/16488429/)
109. Mullard A. Protein–protein interaction inhibitors get into the groove. *Nat Rev Drug Discov.* 2012; 11(3):173–5. <https://doi.org/10.1038/nrd3680> PMID: [22378255](https://pubmed.ncbi.nlm.nih.gov/22378255/)
110. Massova I, Kollman PA. Computational Alanine Scanning To Probe Protein–Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *Journal of the American Chemical Society.* 1999; 121(36):8133–43.
111. Czapinska H, Otlewski J, Krzywda S, Sheldrick GM, Jaskolski M. High-resolution structure of bovine pancreatic trypsin inhibitor with altered binding loop sequence. *J Mol Biol.* 2000; 295(5):1237–49. <https://doi.org/10.1006/jmbi.1999.3445> PMID: [10653700](https://pubmed.ncbi.nlm.nih.gov/10653700/)
112. Danishefsky AT, Housset D, Kim KS, Tao F, Fuchs J, Woodward C, et al. Crevice-forming mutants in the rigid core of bovine pancreatic trypsin inhibitor: crystal structures of F22A, Y23A, N43G, and F45A. *Protein Sci.* 1993; 2(4):577–87. <https://doi.org/10.1002/pro.5560020409> PMID: [8518731](https://pubmed.ncbi.nlm.nih.gov/8518731/)
113. Housset D, Kim KS, Fuchs J, Woodward C, Wlodawer A. Crystal structure of a Y35G mutant of bovine pancreatic trypsin inhibitor. *J Mol Biol.* 1991; 220(3):757–70. PMID: [1714504](https://pubmed.ncbi.nlm.nih.gov/1714504/)
114. Wlodawer A, Deisenhofer J, Huber R. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J Mol Biol.* 1987; 193(1):145–56. PMID: [2438420](https://pubmed.ncbi.nlm.nih.gov/2438420/)
115. Wlodawer A, Nachman J, Gilliland GL, Gallagher W, Woodward C. Structure of form III crystals of bovine pancreatic trypsin inhibitor. *J Mol Biol.* 1987; 198(3):469–80. PMID: [2448484](https://pubmed.ncbi.nlm.nih.gov/2448484/)
116. Berndt KD, Guntert P, Orbons LP, Wuthrich K. Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J Mol Biol.* 1992; 227(3):757–75. PMID: [1383552](https://pubmed.ncbi.nlm.nih.gov/1383552/)
117. Wlodawer A, Walter J, Huber R, Sjolín L. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. *J Mol Biol.* 1984; 180(2):301–29. PMID: [6210373](https://pubmed.ncbi.nlm.nih.gov/6210373/)
118. Sexton JZ, Wigle TJ, He Q, Hughes MA, Smith GR, Singleton SF, et al. Novel Inhibitors of E. coli RecA ATPase Activity. *Curr Chem Genomics.* 2010; 4:34–42. <https://doi.org/10.2174/1875397301004010034> PMID: [20648224](https://pubmed.ncbi.nlm.nih.gov/20648224/)

119. Singleton SF, Roca AI, Lee AM, Xiao J. Probing the structure of RecA-DNA filaments. Advantages of a fluorescent guanine analog. *Tetrahedron*. 2007; 63(17):3553–66. <https://doi.org/10.1016/j.tet.2006.10.092> PMID: 17955055
120. Wigle TJ, Sexton JZ, Gromova AV, Hadimani MB, Hughes MA, Smith GR, et al. Inhibitors of RecA activity discovered by high-throughput screening: cell-permeable small molecules attenuate the SOS response in *Escherichia coli*. *J Biomol Screen*. 2009; 14(9):1092–101. <https://doi.org/10.1177/1087057109342126> PMID: 19675313
121. Wigle TJ, Singleton SF. Directed molecular screening for RecA ATPase inhibitors. *Bioorg Med Chem Lett*. 2007; 17(12):3249–53. <https://doi.org/10.1016/j.bmcl.2007.04.013> PMID: 17499507