Identification of Hot Spots within Druggable Binding Regions by Computational Solvent Mapping of Proteins

Melissa R. Landon,[†] David R. Lancia, Jr.,[‡] Jessamin Yu,[†] Spencer C. Thiel,[‡] and Sandor Vajda*[,§]

Bioinformatics Graduate Program, Boston University, 24 Cummington Street, Boston, Massachusetts, Department of Biomedical Engineering, Boston University, 44 Cummington Street, Boston, Massachusetts, SolMap Pharmaceuticals, Inc., 196 Broadway, 2nd Floor, Cambridge, Massachusetts

Received September 28, 2006

Here we apply the computational solvent mapping (CS-Map) algorithm toward the in silico identification of hot spots, that is, regions of protein binding sites that are major contributors to the binding energy and, hence, are prime targets in drug design. The CS-Map algorithm, developed for binding site characterization, moves small organic functional groups around the protein surface and determines their most energetically favorable binding positions. The utility of CS-Map algorithm toward the prediction of hot spot regions in druggable binding pockets is illustrated by three test systems: (1) renin aspartic protease, (2) a set of previously characterized druggable proteins, and (3) *E. coli* ketopantoate reductase. In each of the three studies, existing literature was used to verify our results. Based on our analyses, we conclude that the information provided by CS-Map can contribute substantially to the identification of hot spots, a necessary predecessor of fragment-based drug discovery efforts.

Introduction

Characterization of protein−ligand binding sites has been focused recently on the identification of hot spots, subsites of ligand binding regions of proteins that are major contributors to the binding energy.[1,2] Druggable binding pockets generally contain smaller regions that are crucial to the binding of functional groups and, hence, are the prime targets in drug design. As developed over the past decade and summarized in a recent paper,[3] NMR-based screening of a variety of protein targets with a large compound library demonstrates that the hot spot regions bind a large variety of small molecules. Although the binding of most compounds is weak, the authors also determined that a relatively high hit rate is predictive of target sites that are likely to bind inhibitors with high affinity. A more direct approach to the identification of hot spots within the NADPH binding region of ketopantoate reductase (KPR) was recently described Ciulli et al.[4] The strategy involved the breaking down of NADPH into smaller fragments and the subsequent characterization of each fragment's ligand efficiency and binding specificity using isothermal titration calorimetry, NMR spectroscopy, and inhibition studies. Despite the decreased affinity of smaller chemical fragments in a binding pocket as compared to larger molecules, the study by Ciulli et al.[4] made apparent that most of the binding free energy comes from select regions of the binding pocket and that the binding modality of NADPH fragments in a binding pocket can be predictive of these regions.

Despite tremendous progress in methodology, the identification of hot spot regions by NMR, X-ray crystallography, or biophysical methods is still expensive and time-consuming. Here we describe a computational alternative, namely, computational solvent mapping (CS-Map), and show that we can reliably reproduce the available experimental results,[5−7] hence providing an effective alternative to experimental hot spot identification. The CS-Map algorithm moves small organic functional groups around the protein surface to find their most favorable binding positions. The direct motivation for developing the method was to replicate in silico the results obtained by the multiple solvent crystal structures (MSCS) method introduced by Mattos and Ringe.[8−10] The MSCS method involves the soaking of a protein in a series of organic solvents, whereupon binding regions can be determined based on the crystallization of the protein in each solvent and subsequent superimposition of the X-ray structures to identify regions where multiple solvent molecules bind on the protein; the authors observed that these regions of aggregation typically correspond very well with known binding regions. While numerous computational techniques exist for the identification of binding sites on proteins,[11−16] it appears that CS-Map is the first algorithm shown to reliably reproduce the results of such MSCS experiments. Previously, the CS-Map algorithm was applied to a variety of well-characterized proteins[5−7] to identify the regions of their binding pockets that are most important for the binding of small molecules; these results were in good agreement with those determined by MSCS and other experimental approaches. The advantage of using CS-Map versus other methods rests primarily in the utilization of a library of solvent-like molecules as probes, making it analogous to the MSCS approach or to other biophysical-based analyses of the binding of small organic molecules to proteins.

We used three test systems in this study to determine the effectiveness of CS-Map in the identification of hot spots for proteins that are both important pharmaceutical targets and whose binding pockets have been experimentally characterized. The test systems include renin aspartic protease, the training set of proteins used in the Abbott NMR-based screening study, and KPR. Renin and the angiotensin converting enzyme (ACE) are long-standing pharmaceutical targets for the treatment of hypertension.[17,18] The cleavage by renin of its peptide substrate, angiotensin, and subsequent formation of angiotensin I precedes the modification of angiotensin I by ACE. While both renin and ACE are viable drug targets for the treatment of hypertension, currently all hypertension therapeutics on the market are ACE inhibitors.[19] Inhibitors of renin consist primarily of

* To whom correspondence should be addressed. Phone: +1-617-353-4757. Fax: +1-617-353-7020. E-mail: vajda@bu.edu.
[†] Bioinformatics Graduate Program, Boston University.
[‡] SolMap Pharmaceuticals, Inc.
[§] Department of Biomedical Engineering, Boston University.

10.1021/jm061134b CCC: $37.00
Published on Web 02/17/2007

peptidomimetics, specifically those targeting the S2 and S3 subsites of the angiotensin binding pocket.[20,21] Despite their high affinity and specificity for the binding pocket, peptidomimetic inhibitors exhibit poor pharmacokinetics, largely attributed to their hydrophilicity and peptide character. For this reason, the development of drug-like inhibitors of renin has proven to be largely unsuccessful.

In 2000, Novartis Pharmaceuticals in collaboration with Speedel published a new class of renin inhibitors demonstrating oral availability.[22−24] These inhibitors lacked the peptide-like features of earlier classes of inhibitors, resulting in molecules that were more drug-like. Additionally, these inhibitors bound primarily in the S1 and S3 subsites of the binding pocket, demonstrating a marked departure from earlier efforts. The authors contended that crucial to the affinity of this class of inhibitors were interactions made in a previously uncharacterized region of the binding pocket, named S3SP due to its proximity to the S3 subsite. Further modification of these molecules has resulted in the development of aliskiren, a renin inhibitor now in clinical trials. In our analysis of renin, we sought to predict the preferential binding pose of aliskiren in terms of key subsite/residue contacts, as compared to that of the peptidomimetic inhibitors using CS-Map.

A training set of 23 proteins, comprising 28 binding sites, was used in the NMR-based study of druggability published by Fesik and co-workers at Abbott Pharmaceuticals.[3] The authors demonstrated a very high correlation between the frequency of screening hits for a particular binding pocket and the ability of that pocket to bind lead-like molecules with high affinity. For those binding regions that were deemed druggable from the training set, we performed a residue-based analysis of binding affinity in an effort to predict hot spots within each druggable pocket.

Last, we predicted hot spots for *E. coli* KPR and compared our results to those recently published by Ciulli et al.[4] KPR plays an important role in the biosynthesis of vitamin B5 by catalyzing the reduction of ketopantoate to pantoate. As vitamin B5 is an essential nutrient for bacteria, inhibition of enzymes involved in this pathway could result in novel antibacterial agents.[25,26] Using a fragment growth strategy to rebuild the structure of NADPH in its binding pocket on KPR, Abell and co-workers exposed via biophysical methods two regions that are primarily responsible for the binding efficiency of NADPH on opposite sides of the binding pocket, namely, the regions that bind the 2′-phosphate and the reduced nicotinamide groups. To confirm their findings, the authors performed mutational studies, where a residue located in each putative hot spot, Arg31 and Asn98, were mutated to alanines to demonstrate a loss of binding affinity upon mutation of residues in these regions. In the CS-Map based analysis of KPR, we sought to uncover these same regions of the binding pocket to directly compare our method to experimental fragment-based approaches for hot spot identification.

Computational Methods

CS-Map Algorithm. CS-Map entails a five-step process by which the energetically favorable binding positions of a series of small, solvent-like probes are determined for a protein.[5−7] The regions in which the highest number of different probe clusters overlap are then regarded as the hot spots predicted by the method. Input for the algorithm is an X-ray structure of the protein of interest in either apo or liganded form. All ligands are removed from bound complexes prior to initiation of the algorithm. A brief description of the four algorithmic steps utilized in this study is provided below; earlier publications should be consulted for a more detailed explanation. An additional subclustering step, described in previous mapping studies, was not used in this analysis.

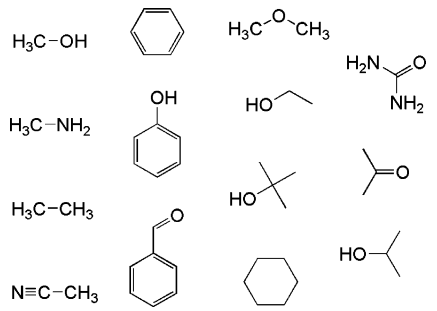

Figure 1. Set of 14 probes used for binding site identification by the CS-Map algorithm.

Step 1: Rigid Body Search. A series of fourteen organic probes, shown in Figure 1, are used to define binding pockets using CS-Map. Each probe is run separately to avoid clashes. A set of 222 initial probe positions is determined based on a placement algorithm, where the protein is placed on a dense grid, followed by the identification of buried and surface points. Surface points in the vicinity of buried points are considered to be in a pocket-like area and are used as initial probe positions to ensure the sampling of more buried regions. Nonpocket-like surface points are then clustered such that the number of cluster centers obtained is equivalent to the remaining number of available probe positions; a probe is then placed at each of these cluster centers.

A multistart simplex method[27] is used to move the probes to energetically optimal positions, where a free energy score of the probe−protein complex is determined by the equation $\Delta G_s = \Delta E_{elec} + \Delta G_{des} + V_{exc}$. Here ΔE_{elec} denotes the coulombic component of the electrostatic energy, ΔG_{des} describes the desolvation free energy, and V_{exc} is an excluded volume penalty term that is set to zero if the probe does not overlap with the protein. The electrostatic term is calculated as the summation of the products of the charge of each probe atom and the solvated protein's electrostatic field at that position; atom charges are assigned by the partial charges employed in the Quanta program (http://www.accelyrs.com), while the electric field term is calculated according to the finite difference Poisson−Boltzmann method[28] implemented in the software package CONGEN (http://www.congenomics.com). The desolvation term, ΔG_{des}, is determined using the atomic contact potential model.[29]

Step 2: Free Energy Refinement and Final Docking. The rigid body search described above yields over 6000 conformations of each probe type in complex with the protein. A modified free energy term is used to further minimize this complex, where now $\Delta G = \Delta E_{elec} + \Delta E_{vdw} + \Delta G^*_{des}$. This free energy equation includes a van der Waals energy term for the protein−ligand complexes, ΔE_{vdw}, and a revised desolvation term, ΔG^*_{des}, that includes the change in solute−solvent van der Waals interaction energy. The ACE model[30] implemented in Charmm version 27[31] is used to compute the electrostatic and desolvation contributions to the free energy. Minimization of each protein−probe complex is performed using the Newton−Raphson method also found in Charmm, where the protein is held fixed while the probe atoms move freely. At most, one thousand minimization steps are allowed, although most complexes require far fewer steps to achieve convergence.

Step 3: Clustering, Scoring, and Ranking. Minimized probe positions are clustered in an interaction-based fashion where a contact vector is used to represent the interactions between the probes and the protein. Each index in u represents a distinct residue i, and $\bar{u}(i)$ is set to 1 if a probe atom is in contact with residue i; otherwise $\bar{u}(i) = 0$. Contact is defined as a probe atom being within 6 Å of atoms of the residue under consideration. Distances are calculated in a pairwise fashion between contact vectors of different probe molecules, denoted $\bar{u}$ and $\bar{v}$, using a normalized distance $D = 1 - \bar{u}\cdot\bar{v}/(|\bar{u}||\bar{v}|)$. The distance is zero if and only if all interactions are the same and is equal to one if vectors $\bar{u}$ and $\bar{v}$ are orthogonal. Contact vectors are then clustered based on their D-values to all

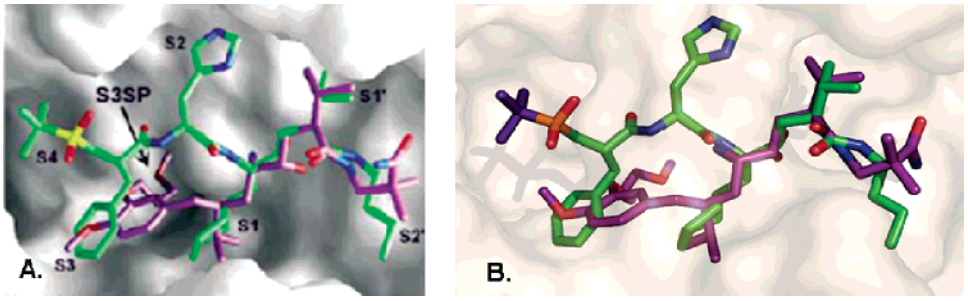

Figure 2. (A) Published figure from Novartis depicting the conformation of aliskiren with respect to a peptidomimetic. Reprinted from *Biochemical and Biophysical Research Communications* (Volume 308; J. M. Wood et al.; Structure-Based Design of Aliskiren, a Novel Orally Effective Renin Inhibitor; pages 698−705), ref 22. Copyright 2003, with permission from Elsevier. Structural data on the renin−aliskiren complex is currently not available from the Protein Data Bank. (B) Resulting docked conformation of aliskiren, shown in purple, in the peptide-binding pocket of renin. The relative conformation of the same peptidomimetic from (A) is shown in green. The docked conformation of aliskiren is qualitatively consistent with the published figure. Key structural features, such as the occupation of the S3SP pocket, conjectured to be responsible for the affinity of aliskiren, are preserved in the docked conformation.

other vectors. The clustering algorithm works by seeding a cluster with the unclustered probe with the lowest free energy. Remaining unclustered probes are searched to find the probe with the lowest distance score D to the seed probe. If this score is below 0.35, the probe is added to the cluster. The maximum allowable score for addition to a cluster, 0.35, was chosen based on a qualitative analysis. Once more than two probes populate the cluster, additional probes are added to the cluster by checking that the average distance score between the new probe and all existing members is below 0.35. If the new probe has an average score above 0.35, the probe is rejected and a new search begins. After all probes are checked for admittance to a cluster and no additional probes can be added, a new cluster begins by repeating the process above. Clusters are optimized after initial creation by reclustering the probes such that if a probe could be moved from one cluster to another so that the average D score would be lowered in both clusters, then the probe is moved. Small clusters consisting of less than fifteen members are excluded from consideration. For each of the remaining clusters, we calculate the probability $p_i = Q_i/Q$, where the partition function Q is the sum of the Boltzmann factors over all conformations, $Q = \sum_j \exp(-\Delta G_j/RT)$, and Q_i is obtained by summing the Boltzmann factors over the conformations in the ith cluster only. The cluster average of a property x for the ith cluster is calculated by $\langle x \rangle_i = \sum_j p_{ij} x_j$, where $p_{ij} = \exp(-\Delta G_j/RT)/Q_i$, and the sum is taken over the members of the ith cluster. Average free energy terms $\langle \Delta G \rangle$, $\langle \Delta E_{elec} \rangle$, $\langle \Delta E_{vdw} \rangle$, and $\langle \Delta G^*_{des} \rangle$ are calculated for each cluster; the average free energy term, $\langle \Delta G \rangle$, is used to rank each cluster.

Step 4: Creation and Ranking of Consensus Sites. The five to ten lowest average free energy clusters of each probe type are used to create consensus sites, for example, regions of the protein where clusters of multiple probe types are located. Superimposition of these low-energy probe clusters on the protein structure allows for identification of these consensus sites. Ranking of consensus sites is based on both the total number of probe clusters that comprise the site as well as the number of different probe types represented by the clusters. For example, a consensus site containing 13 of the 14 different probe molecules is ranked higher than a site with only 7 of the 14 probe molecules if the total number of clusters is the same. If two sites share the same number of probe types, then duplicate types within the consensus site are also considered in the count.

Docking of Aliskiren. We used the GOLD (Genetic Optimization for Ligand Docking) program[32,33] to generate conformations of aliskiren in the peptide binding of renin. The input structure was PDB code 1RNE, where the binding pocket was defined as residues within 10 Å of the α carbon of Asp32. Rings, amides, and nitrogens bound to sp2 carbons were allowed to rotate during the optimizations. Ten conformations of aliskiren were generated using GOLD and then compared to the published conformation.[22]

Detection of Hot Spots Using CS-Map and HBPLUS. The HBPLUS program[34] created by Thornton and co-workers was utilized to generate both nonbonded and hydrogen-bonded interactions on an atomic level between the residues comprising the binding pockets of the proteins used in our studies and the low energy conformations of probe molecules resulting from the mapping of each protein. For each protein, we defined its binding pocket as those residues within 6 Å of a bound ligand (see the Results and Discussion section for a description of each ligand). Solvent probes that were assigned to the top ten lowest free energy probe clusters (see Computational Methods section) and were located in the binding region were used in our analysis. Subsequent to the calculation of nonbonded and hydrogen-bonded interactions, the number of atom interactions for each residue was tallied and normalized by the total number of residue interactions to determine the percentage of atom interactions for each residue relative to all residues in the binding pocket. We considered residues accounting for greater than four percent of the atom interactions in a pocket to be potentially located in a hot spot. Separate interaction analyses were performed between each protein and its bound ligand in a similar manner for comparison.

Results and Discussion

Conformational Analysis of Aliskiren in the Binding Pocket of Renin. Structural information for the renin−aliskiren complex is not available currently from the PDB; however, a figure from the original Novartis publication,[22] reproduced here as Figure 2A, illustrates the position of aliskiren, shown in purple, relative to a peptidomimetic inhibitor, shown in green. It is apparent from this view that, while aliskiren occupies in part the same region of the binding pocket as the peptidomimetic inhibitor does, namely, the S1 and S1′ subsites, significant deviations in the conformations of the two molecules exist in the S3 and S2′ subsites and both molecules occupy subsites uniquely. Specifically, the S2 and S4 subsites are occupied solely by an imidazole functional group and sulfonyl group, respectively, of the peptidomimetic inhibitor, while a hydrophobic functional group of aliskiren resides in the S3SP subsite, a region of the binding site previously unoccupied by inhibitors and considered to be largely responsible for the affinity of aliskiren.

Docking studies of aliskiren in the renin binding pocket using PDB code 1RNE were performed using the GOLD docking program (see Computational Methods) to determine a conformation of aliskiren that exhibited high similarity to the published conformation. Details regarding key contacts and hydrogen-bonding positions from the literature were used in conjunction with qualitative information, specifically Figure 2A, to determine the most appropriate docked conformation for use in our analyses. Figure 2B represents the most similar docked conformation of aliskiren, colored in purple, in the binding pocket of renin. The conformation of the same peptidomimetic inhibitor shown in Figure 2A, for which structural information is

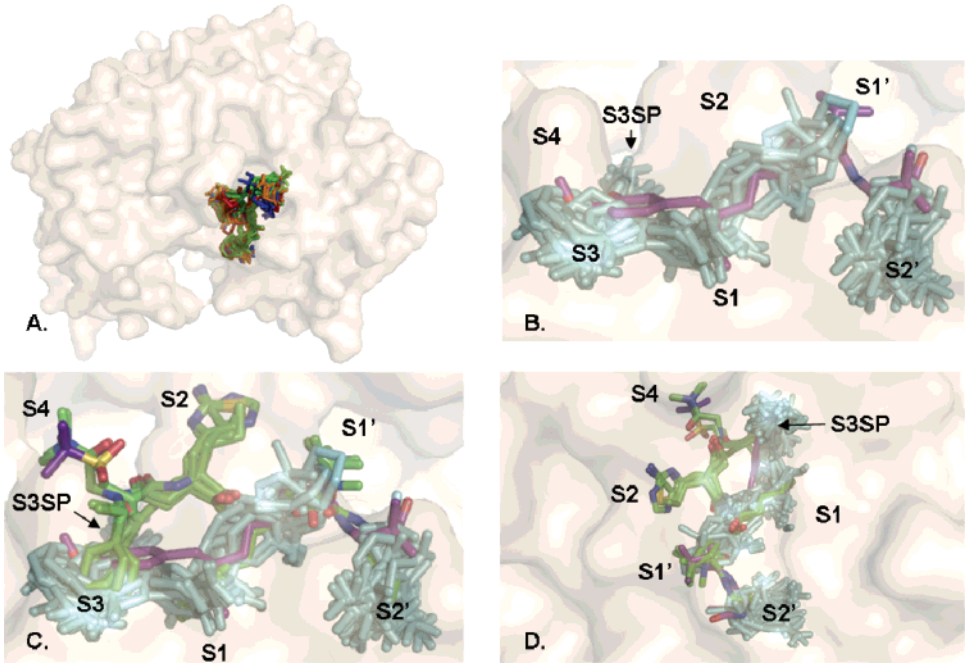

Figure 3. Mapping results for five structures of renin. (A) The first and second ranked consensus sites resulting from the mapping of the different structures are superimposed in the peptide binding pocket of renin, demonstrating the reproducibility of the results. Each color represents the results of a distinctive protein. (B) Closer examination of the consensus sites depicted in (A), now all colored light blue and shown in relation to the docked conformation of aliskiren, supports the importance of the S1, S2, S3, and S3SP subsites for ligand affinity. (C,D) The preferred binding mode of aliskiren as compared to the peptidomimetics, shown in green, is confirmed by the mapping results. The S2 and S4 subsites are bound preferentially by the peptidomimetics, but not by aliskiren or the CS-Map probes.

available, is shown in green to allow for comparison of the docked and experimental conformations. Only minor deviations are apparent between the docked conformation of aliskiren and the actual conformation. In particular, the regions of the binding pocket that make significant contact with aliskiren, namely, the S1, S3, and S2′ regions, are preserved in the docked structure shown in Figure 2A. The conformation of the chain of aliskiren that occupies the S3SP pocket is also preserved. The resulting contact data between aliskiren and renin were used to compare our mapping results to residue interactions made by aliskiren and other peptidomimetic inhibitors of renin.

Identification of High Affinity Subsites in the Binding Pocket of Renin. We mapped five renin structures to predict favorable binding locations within the peptide binding pocket. As shown in Figure 3A, all consensus sites that were found to be located in the binding region are shown superimposed onto the structure of PDB code 1RNE. Consensus sites are colored in Figure 3A according to the structure from which they were derived; structures utilized from the PDB for the study were 1BIL, 1BIM, 1HRN, 1RNE, and 2REN, where the latter structure represents the unliganded form of renin, while the four previous structures are bound by peptidomimetic inhibitors. Given the confinement of all consensus sites to the same region of the binding pocket, we concluded that small changes in the conformation of residues in the binding region do not affect the mapping results significantly.

Figure 3B is a close-up view of the consensus sites located in the binding pocket. In this depiction, all consensus sites resulting from the five structures are colored uniformly in light blue. For clarity, only the probe cluster representatives that comprise the consensus sites are displayed. The docked conformation of aliskiren is shown in purple in Figure 3B to allow for comparison to the mapping results. It is readily apparent from this view that the consensus sites in the binding region overlap significantly with the region occupied by aliskiren, making contacts primarily in the S1, S3, and S2′ subsites. The absence of probe clusters in the S2 and S4 subsites is illustrated in Figures 3C and 3D, where the four peptidomimetic inhibitors taken from the bound PDB structures that were used for mapping are added in green. Figure 3D is a side view of the binding region. Both Figures 3C and 3D show that while each of the peptidomimetic inhibitors makes significant contacts in the S2 and S4 regions neither aliskiren nor the CS-Map probes do so to a visible extent. Interestingly, one of the densest regions of probe molecules is found in the S3SP subsite, a region of the binding pocket that was described in the Novartis publication as being unique to the binding modality of aliskiren versus other renin inhibitors.

Table 1. Rankings of the Consensus Sites Present in the Subsites of the Binding Pocket of Renin for Each of the Five Structures Mapped[a]

PDB/subsite	S4	S3	S2	S1	S1′	S2′
1RNE	NP[b]	1 (28)	NP[b]	1 (28)	6 (7)	2 (15)
1HRN	NP[b]	1 (24)	4 (12)[c]	1 (24)	4 (12)[c]	2 (15)
1BIL	NP[b]	1 (19)	NP[b]	1 (19)	NP[b]	2 (15)
1BIM	NP[b]	1 (24)	NP[b]	1 (24)	2 (14)	2 (14)
2REN	NP[b]	1 (11)	NP[b]	2 (8)	NP[b]	NP[b]

[a] The number in parentheses indicates the number of probe clusters used to create the consensus site. [b] NP = not present. [c] Both consensus sites have the same number of clusters.

The rank(s) of the consensus site(s) occupying the different subsites of the binding pocket are summarized in Table 1, where the assignment of residues to a subsite was utilized from a previous publication.[35] With the exception of the unbound structure, the top ranked consensus sites occupy both the S1 and S3 subsites of the binding pocket for each structure; in the case of the bound structure, the consensus site in the S3 pocket is first. Conformational changes undergone by aspartyl proteases upon ligand binding may account for the difference in mapping results existing between the unbound and the bound structures, in particular the change in shape of the S3 region of the active site.[36]

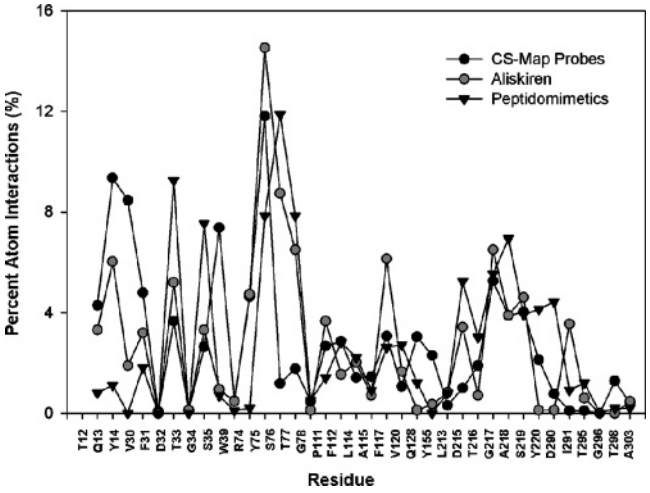

Figure 4. Distribution of atom-based residue interactions in the binding pocket of renin for the CS-Map probes, aliskiren, and three different peptidomimetics. In the case of the peptidomimetics, the average number of interactions at each residue was utilized. The Pearson correlation coefficient between the CS-Map probes and aliskiren is 0.72, while that between the CS-Map probes and the peptidomimetics is 0.19.

Conversely, no significantly populated consensus site is present in the S4 pocket, and only a single, low-ranked consensus site is found in the S2 subsite. This analysis suggests that the S2 and S4 subsites bind drug-like functional groups with lower affinity than the other subsites of the peptide binding pocket. Based on these results we can conclude that the S1, S3, and S2′ subsites of the binding pocket are hot spots for fragment binding and, within the S3 subsite, the S3SP region displays particularly high affinity for small molecules.

Residue-Based Analysis of Hot Spots in the Renin Binding Pocket. Subsequent to the characterization of hot spots within the binding pocket of renin, we applied CS-Map to the identification of specific residues that are crucial for ligand affinity within the hot spot regions (see Computational Methods). Residues were defined as part of the binding pocket if any of their atoms were within 6 Å of an atom of aliskiren. In addition, calculating interactions for the probe molecules, we determined interactions for both aliskiren and the four peptidomimetic inhibitors shown in Figure 3C,D for comparison. The resulting residue-based interaction distributions are shown in Figure 4, with the residues composing the binding pocket placed in sequence order on the horizontal axis. Residue interactions were calculated separately for each peptidomimetic inhibitor and then averaged to create one value. A high level of agreement exists between the distributions for aliskiren, shown in dark gray, and the CS-Map probes, shown in black; in particular, atom interactions are enriched in both distributions for residues Gly13 and Tyr75, located in the S3SP and S1 subsites, respectively. However, while both aliskiren and the peptidomimetic inhibitors interact significantly with the catalytic Asp32, located in the S1 subsite, the residue was not highly interactive with CS-Map probes; because both aliskiren and the mapping probes make the highest level of interactions in the S1 subsite with Tyr75, this may suggest that while Asp32 is necessary for catalysis, it may contribute less to the binding affinity of ligands than surrounding residues.

As compared to the high level of agreement existing between aliskiren and the CS-Map probes with respect to the distribution of residue interactions, the comparison of interaction distributions between the CS-Map probes and the peptidomimetics yielded a very low level of correlation. Residues predicted to be hot spots based on the analysis of interactions made with the peptidomimetics, shown in Figure 4 as the light gray distribution, were located primarily in the S2 subsite, such as residues Ser76 and Ala218. Agreement between the peptidomimetics and aliskiren was only seen in the S1 region with the two catalytic aspartic acids, Asp32 and Asp215. As a quantitative measure of similarity, we calculated Pearson correlation coefficients (R) in a pairwise fashion for the three distributions, where values can range from -1 (perfectly anti-correlated) to 1 (perfectly correlated). Assuming that there are 20 residues in the entire binding site, the threshold for a correlation coefficient to be significant with a p-value of less than 0.01 is $R = 0.52$. The calculated R-value between the CS-Map probes and aliskiren was 0.72, significantly higher than the R-value of 0.19 existing between the CS-Map probes and the peptidomimetic inhibitors. The correlation between aliskiren and the peptidomimetics was an intermediate value of 0.53; the main reason for the differences in correlation existing between the two different inhibitor types when compared to the mapping results is due to the affinity of the CS-Map and aliskiren for the S3SP pocket as compared to peptidomimetic inhibitors. Strong hydrophobic interactions in this region allow aliskiren to exhibit high affinity despite its decrease in peptide-like character. The residues predicted by CS-Map as being highly interactive can serve as starting points for the development of high-affinity, drug-like inhibitors.

Characterization of Druggable Binding Pockets Using CS-Map. Using those proteins comprising the druggable training group from the Fesik study,[3] we sought to identify hot spots and specific residue interactions important for ligand binding for a variety of drug targets. Consisting of both druggable and

Table 2. Summary of Results on the Mapping of Proteins Used as the Training Set in the Druggability Study Published by Fesik and Co-Workers[a]

protein	site	PDB ID	druggable[3]	consensus site rank(s) (number of clusters)	consensus site avg cluster energy rank
survivin	Bir3	1E31	N	not found	not found
Akt-PH	IP3	1H10	N	6 (6)	6
PDZ-PSD95	peptide	1IU0[b]	N	6 (11)	5
Pin1	peptide	1I8H[b]	N	6 (14)	5
FKBP	FK506	1FKJ, 1FKT[b]	Y	1 (26), 1 (23),[b] 2 (15)[b]	1
UK	peptide	1FV9	Y	1 (20)	1
SCD	substrate	1G4K, 3USN[b]	Y	1 (26), 2 (18)[b]	1
ErmAm	SAH	1QAM	Y	1 (27)	1
LFA	IDAS	1RD4, 1DGQ[b]	Y	1 (18), 1 (18)[b]	3
MurA	UDPNAG	1UAE	Y	2 (21)	3
MDM2	P53	1YCR	Y	1 (33), 2 (19)	1
PTP1B	cat. ptyr	1PTY,[c] 1PH0[d]	Y	5 (10),[c] 5 (9)[d]	3,[c] 6[d]
PTP1B	noncat. ptyr	1PTY,[c] 1PH0[d]	Y	9 (4),[c] 9 (2)[d]	7

[a] The PDB accession code indicates the protein structure used for mapping. [b] NMR structure. [c] Closed conformation of WPD loop. [d] Open conformation of WPD loop.

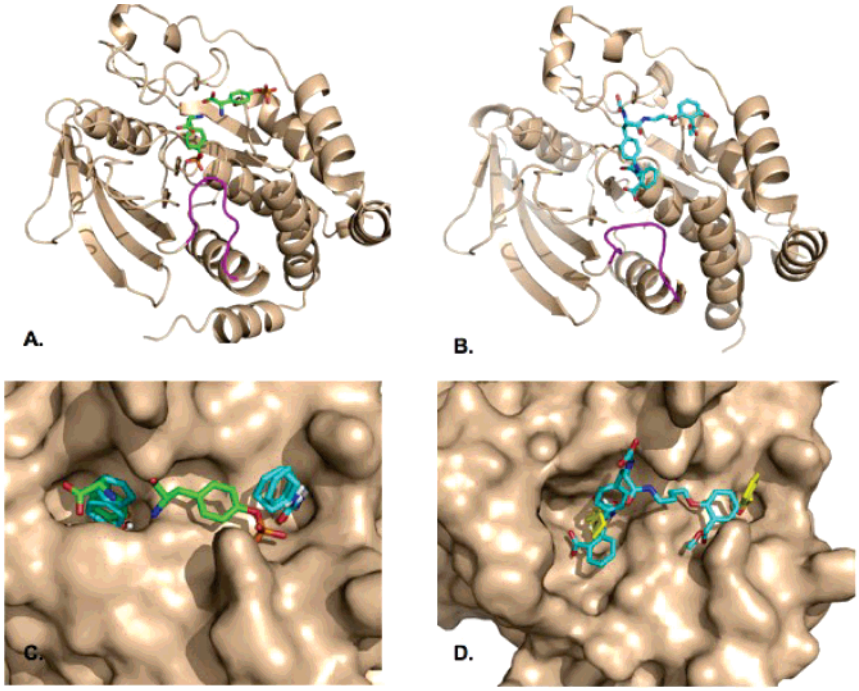

Figure 5. Mapping analysis of open and closed conformations of PTP1B. (A) Cartoon depiction of the open conformation of PTP1B (PDB ID 1PTY). Also shown in green are the bound catalytic (left most) and noncatalytic (right side) phosphotyrosyl (pTyr) groups. The closed conformation of the WPD loop is highlighted in magenta. (B) Structure of the "open" conformation of PTP1B (PDB ID 1PH0), with the WPD loop in magenta and the bound active site inhibitor in blue. (C,D) Surface representations of the active site of PTP1B derived from (A) and (B), respectively. Shown bound in (C) are the pTyr groups in green and CS-Map probes in blue. In (D) the active site inhibitor is colored in blue and the CS-Map probes are colored in yellow. The small, deep nature of the catalytic pTyr site in the closed structure (C) as compared to that of the open conformation (D) is evident from these representations. Mapping results of the closed conformation (A,C) reveal a low-energy consensus site in the catalytic pTyr binding pocket and a less favorable region on the far end of the second pTyr binding location. The consensus sites formed in the same regions of the open conformation of PTP1B are not as energetically favorable, suggesting that a larger molecule may be needed to gain affinity from both binding regions.

non-druggable proteins, this training set was used by Fesik and co-workers to create a druggability index. In addition to hot spot identification, we applied CS-Map toward the computational prediction of druggability based solely on protein structure, using those structures from the druggability study for which complete structural coordinates were available from the PDB.

Mapping results for the 12 proteins analyzed, consisting of 13 binding sites, are summarized in Table 2, including the rank of the consensus site found in a binding region and the number of probe clusters that were used to create the consensus site. The last column of Table 2 indicates the average energy ranking values of probes comprising the consensus sites found in the binding region, where the highest ranked cluster of each probe type was used to determine the average value and then estimated to the nearest whole number.

The protein binding pockets determined to be non-druggable in the experimental study conducted at Abbott include the bir3 binding region of survivin, the phosphatidylinositol (3,4,5)-trisphosphate binding domain of protein kinase B/AKT, the PZD domain of protein-95 (PSD95), and the WW domain of Pin1. As shown by Table 2, the highest ranking consensus site determined for any of the non-druggable proteins was fourth, suggesting that a low consensus site ranking is predictive of binding regions that are not druggable. Conversely, for each druggable binding pocket from the study, the consensus sites found in those regions were top ranked based on size and, in most cases, top ranked based on average cluster energy. To assess the sensitivity of our results to structure resolution, we mapped NMR structures available for druggable proteins in the training set shown in Table 2. Within binding regions, all mapping results derived from NMR data are in good agreement with those determined using crystallographic data. Analysis of this set of proteins suggests that CS-Map can be used to predict the druggability of a binding region when structural information is available; that is, those binding regions with top-ranked or near top-ranked consensus sites in terms of cluster membership and/or average cluster energy, are characteristic of druggable binding regions.

The sole exception in our study to both the above-defined druggability criteria is PTP1B, where two distinct conformations of the binding region were analyzed using CS-Map. These two conformations of the binding region result primarily from the positition of the WPD loop; the loop can either be "closed", where the catalytic phosphotyrosyl (pTyr) binding region forms a small, deep pocket, or "open", resulting in a shallower catalytic pTyr binding region that is continuous with the secondary phosphotyrosyl binding location. Representative structure of each conformation was chosen from the PDB for our mapping analysis. A closed conformation structure of PTP1B, PDB 1PTY, is shown in cartoon and surface representations in Figure 5A and C, respectively; the WPD loop is highlighted in magenta in 5A and the bound phosphotyrosyl groups are outlined in green in both 5A and 5C. Mapping results for 1PTY are shown superimposed with the phosphotyrosyl groups in Figure 5C; although this consensus site is low-ranked in terms of cluster membership, the small and relatively polar probes, namely, dimethyl ether, urea, ethanol, acetonitrile, acetone, and acetaldehyde, bind preferentially in the catalytic phosphotyrosyl binding site with cluster energies that are always either first or second ranked, where the top-ranked cluster is defined as that with the lowest average free energy of its members. These

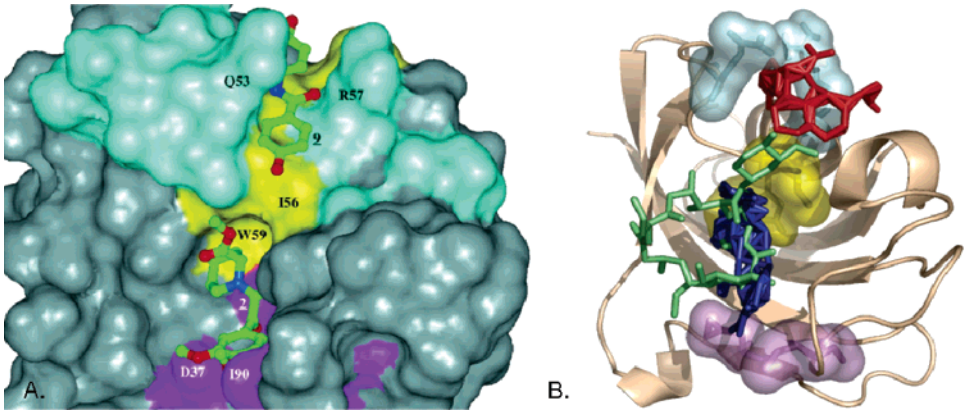

Figure 6. Mapping results for the FK-506 binding protein (FKBP). (A) Previously published NMR data where the presence of three hot spots in the FK-506 binding pocket was determined. Reprinted with permission from *Science* (http://www.sciencemag.org) (Volume 274; S. B. Shuker et al.; Discovering High-Affinity Ligands for Proteins: SAR by NMR; pages 1531−1534), ref 38. Copyright 1996 AAAS. The residues comprising these hot spots are highlighted in cyan, yellow, and purple. The locations of the consensus sites correspond to the hot spots determined by the NMR experiment. (B) Comparison of top two ranked consensus sites, shown in blue and red, respectively, resulting from the CS-Map algorithm. The bound structure of FK-506 is shown in green.

probes are located deep in the catalytic pTyr binding site and, consequently, are not visible in Figure 5C. Larger polar probes, such as benzaldehyde and *t*-butanol, bind at slightly higher energies. Additionally, larger hydrophobic probes, that is, cyclohexane and benzene, are either low-ranked or absent in the active site. Our analysis suggests that the size of a consensus site is not always enough to label a binding region as druggable or non-druggable, particularly when the pocket is small and exhibits a strong preference for probes of a singular type.

A structure of the open conformation of PTP1B, PDB 1PH0, is shown in cartoon representation in Figure 5B, where a surface of the active site is depicted in Figure 5D. Highlighted in blue in both representations is the bound active site inhibitor of 1PH0. Mapping of 1PH0, illustrated in yellow in Figure 5D, reveals consensus sites in locations proximal to those determined for 1PTY; however, while the number of probe clusters comprising these consensus sites is equivalent to those derived for 1PTY, the energy ranking decreases significantly, with average cluster energy rank in the catalytic pTyr site falling from third to sixth (see Table 2). Thus, our analysis of the open conformation reveals only relatively weak hot spots, located at opposite ends of the binding region as shown in Figure 5D. The change in hot spot affinity upon conformational shift supports the finding that larger inhibitors are necessary to bridge both affinity regions. However, as a result of their size, these inhibitors typically exhibit poor ADMET properties.[37] Our study of PTP1B demonstrates that, unlike the analysis of renin, CS-Map is sensitive to large changes in conformation of the binding region; however, these changes may be reflective of a change in compound properties necessary for inhibition.

Identification of Hot Spots within Druggable Binding Pockets. For each druggable binding pocket defined in Table 2, we identified groups of highly interacting residues for hot spot formation. Hot spots resulting from our mapping analyses were then compared to published data to validate our findings; we considered our results to be in agreement with experimental data if residues that we predicted to be in hot spots were implicated as being important in ligand binding events by crystallographic, spectroscopic, or other biophysical methods. Table 3 lists the residues that we predicted to be highly interactive in ligand binding events using CS-Map; an asterisk (*) indicates those residues for which we found literature evident for our findings. For the majority of residues, we were able to validate our predictions. Of those residues for which we could not find reports of involvement in ligand binding, a few were discovered to mitigate the binding process by hydrogen bonding with water, such as H299 in MurA.[41]

Table 3. Hot Spots Predicted by CS-Map for the Training Set of Druggable Proteins Listed in Table 2[a]

protein	druggable pocket(s)[3]	PDB ID	predicted hot spot residues
FKBP	FK-506	1FKJ	Y26*, F36*, F46*, Q53*, V55*, I56*, R57*, W59*[38]
Akt-PH	IP3	1H10	K14*, R15*, G16*, E17*, Y18*, I19*, R23*, R25*[39,40]
ErmAM	SAH	1QAM	Q10*, N11*, F12*, I37, G38*, I102*[41,42]
MDM2	P53	1YCR	L54*, L57*, G58*, I61*, Y67*, Q72*, H73*[43]
MurA	UDPNAG	1UAE	R120*, V122*, V161, S162*, V163*, E188*, H299, T304, V327*[44]
PTP1B	cat. pTyr, noncat. pTyr	1PH0	R24*, A27, Y46*, C215*, S216*, A217*, I219*, R221*, Q262*[45,46]
SCD	substrate	1G4K	N162, V163*, L164*, A165*, L197*, V198*, H199*, Y220*[47−49]
UK	peptide	1FV9	D191*, S192*, C193*, W217*, G218*, G220*, C221*[50]
LFA	IDAS	1RD4	L132*, F153*, Y166*, K232, I235*, Y257*, I259*, V286*, K287*, L302*[51,52]

[a] An * indicates a residue for which existing literature supports the prediction.

A pharmaceutical target of particular interest in Table 2 is the FK-506 binding protein (FKBP). Inhibition of FKBP leads to the inhibition of calcineurin, serving as a mechanism for the activation of immunosuppression in organ transplant patients. An NMR-based screening study performed on FKBP, published in Science a decade ago, identified several residues for which chemical shifts were detected upon inhibitor binding;[38] a figure from that publication is reproduced here as Figure 6A, where the residues colored in cyan, yellow, and purple compose the three hot spot regions uncovered in their analysis.

Figure 6B is our mapping analysis of FKBP performed using structure 1FKJ from the PDB, with the first- and second-ranked consensus sites colored blue and red, respectively. Superimposed in green on the mapping results in Figure 6B is the bound conformation of FK-506 for comparison. To illustrate the similarities existing between the hot spots defined by the NMR

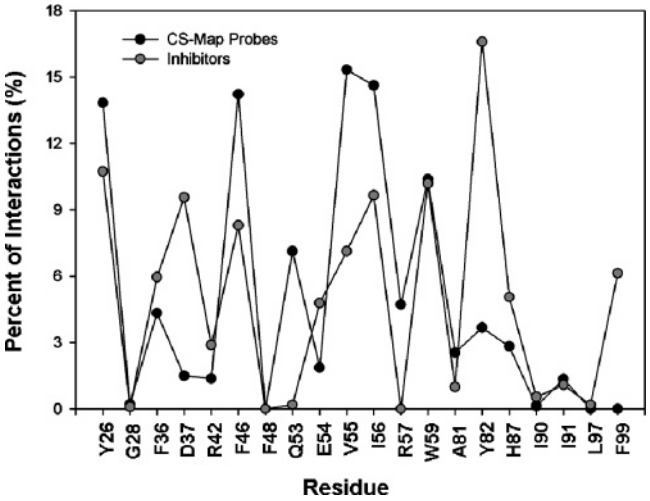

Figure 7. Distribution of the atom-based residue interactions in the FK-506 binding pocket for CS-Map probes and four FKBP inhibitors. The interactions of four different inhibitors of FKBP, including FK-506, were also calculated on the same residues for comparison.

analysis and the CS-Map-based approach, the residue color scheme is kept the same for between Figure 6A and B; the first-ranked consensus site lies within the large region defined by the purple and yellow residues from the NMR study, and the second-ranked consensus site occupies the hot spot outlined by the cyan residues. A more detailed residue distribution is shown in Figure 7, where the *R*-value between the CS-Map probes, shown in black, and four inhibitors of FKBP, including FK-506 and shown in gray, is 0.51. This correlation value was determined between the probes and an average residue interaction percentage from the four inhibitors; these inhibitors were extracted from four bound structures of FKBP from the PDB, namely, 1FKD, 1FKJ, 1QPF, and 1QPL. The highest percentages of interactions for the CS-Map probes occur at residues Tyr26, Val55, Ile56, and Trp59; the latter two are validated by the NMR study, while the former two are highly interactive with known inhibitors.

Interestingly, the distribution of residue interactions for the known inhibitors is not in perfect agreement with the NMR study; three of the residues identified in the NMR screen, Gln53, Arg57, and Ile90, have very little to no interaction with the four bound ligands. Both Gln53 and Arg57 make significant contacts with the mapping probes; these differences suggest that the identification of hot spots through high-throughput experimental or computational methods provides insight into binding affinities that cannot be surmised from crystallographic contact data alone.

Comparison of CS-Map to Experimental Fragment-Based Approaches for Hot Spot Identification: A Case Study with *E. coli* KPR. Through a series of NMR-based screening studies of KPR with NADPH-derived analogues in conjunction with calorimetric analysis of two single point mutants, Ciulli et al. uncovered two hot spots in the NADPH binding region of KPR, located in the regions that bind the 2′-phosphate and reduced nicotinamide groups of NADPH.[4] Two point mutations of residues within these regions, Arg31 and Asn98, were used to confirm their findings. The fragmentation of NADPH performed by the authors to create analogues is reproduced here as Figure 8A; the 2′-phosphate and reduced nicotinamide groups, highlighted in red and blue, respectively, have significantly higher ligand efficiency values than the β-phosphate ribose fragment, shown in white. Ligand efficiency was defined as the change in the dissociation constant, K_D, relative to the change in the number of heavy atoms upon addition of the fragment to the starting fragment AMP, shown in green. Our goal was to confirm the presence of the two hot spots that were described in the study, as well as to predict the differences in ligand efficiencies existing among different regions of the binding pocket.

The three structures of KPR made available by the PDB were used for a mapping-based prediction of hot spots in the NADPH binding region; these structures represent both the unbound (1KS9) and cofactor/cofactor-analogue (1YJQ and 1YON, respectively) bound conformations of KPR. Consensus sites located in the NADPH binding region for structure 1YJQ are shown in Figure 8B; the sites, colored in cyan, are superimposed with the structure of NADPH, in green, to illustrate the concentration of probe clusters at either end of the binding

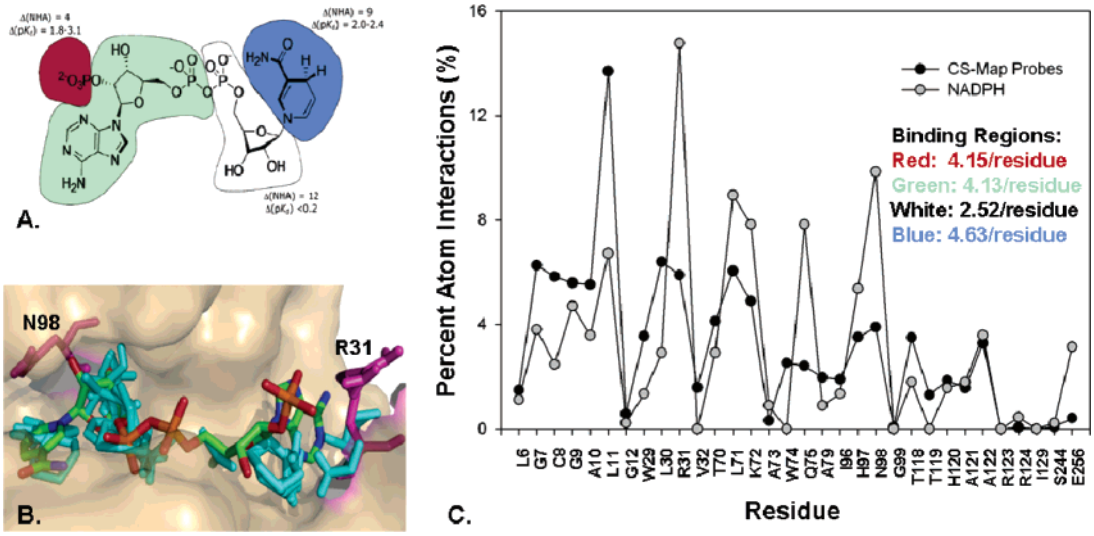

Figure 8. Mapping analysis of the NADPH binding region of KPR from *E. coli*. (A) Previous experimental work revealed a range of ligand efficiency values corresponding to distinct fragments of NADPH. The 2′-phosphate region, circled in red, and the reduced nicotinamide group, highlighted in blue, exhibit significantly higher ligand efficiencies than the B-phosphate ribose group. (B) Mapping analysis of the NADPH-bound structure of KPR from the PDB is predictive of two hot spots regions, indicate by the consensus sites in blue, located on opposite sides of the binding region. (C) The atom-based residue distribution of CS-Map probes in the binding pocket, colored in black, corresponds to the ligand efficient regions of the binding region shown in (A). Atom interactions are enriched among those residues that bind the red and blue regions of NADPH defined in (A), whereas the region that binds the white fragment in (A) has a decreased average value of interactions per residue.

region. Residues Arg31 and Asn98, used for mutational confirmation of the hot spots determined in the experimental study, are highlighted in magenta in Figure 8B to highlight the presence of the probe clusters in these regions. Probe interactions in the NADPH binding region were calculated and summed across all three structures to create the residue distribution shown in Figure 8C, where the distribution for the CS-Map probes is outlined in black and the distribution for NADPH is outlined in gray; an *R*-value of 0.6 exists between the two distributions, suggesting that several key residues are predicted using CS-Map.

To test this hypothesis, we assigned each residue in the binding region to one of four groups based on the fragment of NADPH shown in Figure 8A with which it primarily interacts; this was determined using HBPLUS, where a residue was assigned to a group if the majority of its interactions with NADPH occurred primarily with the fragment representing that group. The percentage of interactions for each residue comprising the group was then normalized by the number of group members to create a value analogous to ligand efficiency, where higher values indicate a higher concentration of ligand interactions in that region. As shown in Figure 8C, residues comprising the groups that interact primarily with the 2′-phosphate and the reduced nicotinamide moieties of NADPH, labeled as the red and blue groups, respectively, in Figure 8C, have a drastically increased average number of interactions per residue, 4.15 and 4.63%, than those residues comprising the group that interacts primarily with the β-phosphate ribose moiety of NADPH, the white group. Residues interacting with the AMP analogue, the green group, have an average interaction value comparable to that of the red and blue groups. These residue-based efficiency values derived from interactions between CS-Map probes and KPR are consistent with the ligand efficiency values for the NADPH fragments shown in Figure 8A; this analysis suggests that in addition to providing spatial information concerning the location of hot spots, CS-Map can also be used to predict ligand efficiencies of fragments of larger molecules.

Conclusions

We have described the application of a computational method for binding site prediction, the CS-Map algorithm, to the identification of hot spots within druggable binding sites. The success of this approach was demonstrated for a variety of proteins on which various experimental analyses were performed to arrive at similar results. As illustrated by the analysis of renin, our method is capable of distinguishing regions that bind drug-like molecules with high-affinity from those that bind nondrug-like molecules, such as peptidomimetics. Based on our study of druggable proteins from the NMR screen performed by Fesik and co-workers, we determined that in addition to the identification of residues that play crucial roles in ligand affinity, CS-Map can also be used to assess the druggability of a binding pocket based on the rank of consensus sites located within the pocket. Interestingly, as noted in the Results and Discussion, our analyses were consistent for structures derived either via crystallographic or NMR methods. However, as was made apparent by the analysis of open and closed conformations of PTP1B, mapping results are not as robust to large changes in the conformation of the binding site as they are to minor ones, as was the case with renin. Given these results, the question of whether CS-Map can be used to analyze homology models will require further investigation. Last, the prediction of hot spots within the NADPH binding pocket of KPR illustrated the usefulness of CS-Map toward the prediction of ligand-efficient binding regions as well as validating hot spots derived from biophysical methods. Given these successes, we conclude that this approach can be applied to the prediction of hot spots for protein targets where only general druggability features are currently known, as well as validate new protein targets where druggability is not known. This, in turn, could provide a wealth of information for fragment-based drug design efforts.

Acknowledgment. We gratefully thank Frank Guarnieri and colleagues at SolMap Pharmaceuticals for numerous insightful conversations regarding this work. This work was made possible through Grants R01GM064700 and R41GM075473 from the National Institute of General Medical Sciences.

References

(1) DeLano, W. L. Unraveling hot spots in binding interfaces: Progress and challenges. *Curr. Opin. Struct. Biol.* **2002**, *12*, 14−20.
(2) Kortemme, T.; Baker, D. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14116−14121.
(3) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability indices for protein targets derived from NMR-based screening data. *J. Med. Chem.* **2005**, *48*, 2518−2525.
(4) Ciulli, A.; Williams, G.; Smith, A. G.; Blundell, T. L.; Abell, C. Probing hot spots at protein−ligand binding sites: A fragment-based approach using biophysical methods. *J. Med. Chem.* **2006**, *49*, 4992−5000.
(5) 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*, 4290−4295.
(6) Kortvelyesi, T.; Dennis, S.; Silberstein, M.; Brown, L., III; Vajda, S. Algorithms for computational solvent mapping of proteins. *Proteins* **2003**, *51*, 340−351.
(7) Silberstein, M.; Dennis, S.; Brown, L.; Kortvelyesi, T.; Clodfelter, K.; Vajda, S. Identification of substrate binding sites in enzymes by computational solvent mapping. *J. Mol. Biol.* **2003**, *332*, 1095−1113.
(8) Mattos, C.; Bellamacina, C. R.; Peisach, E.; Pereira, A.; Vitkup, D.; Petsko, G. A.; Ringe, D. Multiple solvent crystal structures: Probing binding sites, plasticity, and hydration. *J. Mol. Biol.* **2006**, *357*, 1471−1482.
(9) Mattos, C.; Ringe, D. Locating and characterizing binding sites on proteins. *Nat. Biotechnol.* **1996**, *14*, 595−599.
(10) Mattos, C.; Ringe, D. Proteins in organic solvents. *Curr. Opin. Struct. Biol.* **2001**, *11*, 761−764.
(11) Bohm, H. J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61−78.
(12) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.
(13) Lawrence, M. C.; Davis, P. C. CLIX: A search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins* **1992**, *12*, 31−41.
(14) Miranker, A.; Karplus, M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins* **1991**, *11*, 29−34.
(15) Laurie, A. T.; Jackson, R. M. Q-SiteFinder: An energy-based method for the prediction of protein−ligand binding sites. *Bioinformatics* **2005**, *21*, 1908−1916.
(16) Cavasotto, C. N.; Orry, A. J.; Abagyan, R. A. Structure-based identification of binding sites, native ligands, and potential inhibitors for G-protein coupled receptors. *Proteins* **2003**, *51*, 423−433.
(17) Raman, V. K.; Lee, Y. A.; Lindpaintner, K. The cardiac renin−angiotensin−aldosterone system and hypertensive cardiac hypertrophy. *Am. J. Cardiol.* **1995**, *76*, 18D−23D.
(18) Stanton, A. Potential of renin inhibition in cardiovascular disease. *JRAAS* **2003**, *4*, 6−10.
(19) Staffileno, B. A. Treating hypertension with cardioprotective therapies: The role of ACE inhibitors, ARBs, and beta-blockers. *J. Cardiovasc. Nurs.* **2005**, *20*, 354−364.
(20) Simoneau, B.; Lavallee, P.; Anderson, P. C.; Bailey, M.; Bantle, G.; Berthiaume, S.; Chabot, C.; Fazal, G.; Halmos, T.; Ogilvie, W. W.; Poupart, M. A.; Thavonekham, B.; Xin, Z.; Thibeault, D.; Bolger, G.; Panzenbeck, M.; Winquist, R.; Jung, G. L. Discovery of non-peptidic P2-P3 butanediamide renin inhibitors with high oral efficacy. *Bioorg. Med. Chem.* **1999**, *7*, 489−508.
(21) Tong, L.; Pav, S.; Lamarre, D.; Simoneau, B.; Lavallee, P.; Jung, G. Crystallographic studies on the binding modes of P2-P3 butanediamide renin inhibitors. *J. Biol. Chem.* **1995**, *270*, 29520−29524.

(22) Wood, J. M.; Maibaum, J.; Rahuel, J.; Grutter, M. G.; Cohen, N. C.; Rasetti, V.; Ruger, H.; Goschke, R.; Stutz, S.; Fuhrer, W.; Schilling, W.; Rigollier, P.; Yamaguchi, Y.; Cumin, F.; Baum, H. P.; Schnell, C. R.; Herold, P.; Mah, R.; Jensen, C.; O'Brien, E.; Stanton, A.; Bedigian, M. P. Structure-based design of aliskiren, a novel orally effective renin inhibitor. *Biochem. Biophys. Res. Commun.* **2003**, *308*, 698−705.
(23) Allikmets, K. Aliskiren (Speedel). *Curr. Opin. Invest. Drugs* **2002**, *3*, 1479−1482.
(24) Rahuel, J.; Rasetti, V.; Maibaum, J.; Rueger, H.; Goschke, R.; Cohen, N. C.; Stutz, S.; Cumin, F.; Fuhrer, W.; Wood, J. M.; Grutter, M. G. Structure-based drug design: The discovery of novel nonpeptide orally active inhibitors of human renin. *Chem. Biol.* **2000**, *7*, 493−504.
(25) Ciulli, A.; Abell, C. Biophysical tools to monitor enzyme−ligand interactions of enzymes involved in vitamin biosynthesis. *Biochem. Soc. Trans.* **2005**, *33*, 767−771.
(26) Begley, T. P.; Kinsland, C.; Strauss, E. The biosynthesis of coenzyme A in bacteria. *Vitam. Horm.* **2001**, *61*, 157−171.
(27) Dennis, S.; Vajda, S. Semiglobal simplex optimization and its application to determining the preferred solvation sites of proteins. *J. Comput. Chem.* **2002**, *23*, 319−334.
(28) Gilson, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* **1988**, *4*, 7−18.
(29) Zhang, C.; Cornette, J. L.; Delisi, C. Consistency in structural energetics of protein folding and peptide recognition. *Protein Sci.* **1997**, *6*, 1057−1064.
(30) Schaefer, M.; Karplus, M. A Comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(31) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm−A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(32) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.
(33) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.
(34) McDonald, I. K.; Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*, 777−793.
(35) Rahuel, J.; Priestle, J. P.; Grutter, M. G. The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. *J. Struct. Biol.* **1991**, *107*, 227−236.
(36) Sali, A.; Veerapandian, B.; Cooper, J. B.; Moss, D. S.; Hofmann, T.; Blundell, T. L. Domain flexibility in aspartic proteinases. *Proteins* **1992**, *12*, 158−170.
(37) Liu, G. Protein tyrosine phosphatase 1B inhibition: Opportunities and challenges. *Curr. Med. Chem.* **2003**, *10*, 1407−1421.
(38) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531−1534.
(39) Hu, Y.; Qiao, L.; Wang, S.; Rong, S. B.; Meuillet, E. J.; Berggren, M.; Gallegos, A.; Powis, G.; Kozikowski, A. P. 3-(Hydroxymethyl)-bearing phosphatidylinositol ether lipid analogues and carbonate surrogates block PI3-K, Akt, and cancer cell growth. *J. Med. Chem.* **2000**, *43*, 3045−3051.
(40) Thomas, C. C.; Deak, M.; Alessi, D. R.; van Aalten, D. M. High-resolution structure of the pleckstrin homology domain of protein kinase b/akt bound to phosphatidylinositol (3,4,5)-trisphosphate. *Curr. Biol.* **2002**, *12*, 1256−1262.
(41) Hajduk, P. J.; Dinges, J.; Schkeryantz, J. M.; Janowick, D.; Kaminski, M.; Tufano, M.; Augeri, D. J.; Petros, A.; Nienaber, V.; Zhong, P.; Hammond, R.; Coen, M.; Beutel, B.; Katz, L.; Fesik, S. W. Novel inhibitors of Erm methyltransferases from NMR and parallel synthesis. *J. Med. Chem.* **1999**, *42*, 3852−3859.
(42) Schluckebier, G.; Zhong, P.; Stewart, K. D.; Kavanaugh, T. J.; Abad-Zapatero, C. The 2.2 A structure of the rRNA methyltransferase ErmC' and its complexes with cofactor and cofactor analogs: Implications for the reaction mechanism. *J. Mol. Biol.* **1999**, *289*, 277−291.
(43) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **1996**, *274*, 948−953.
(44) Skarzynski, T.; Mistry, A.; Wonacott, A.; Hutchinson, S. E.; Kelly, V. A.; Duncan, K. Structure of UDP-*N*-acetylglucosamine enolpyruvyl transferase, an enzyme essential for the synthesis of bacterial peptidoglycan, complexed with substrate UDP-*N*-acetylglucosamine and the drug fosfomycin. *Structure* **1996**, *4*, 1465−1474.
(45) Liu, G.; Xin, Z.; Liang, H.; Abad-Zapatero, C.; Hajduk, P. J.; Janowick, D. A.; Szczepankiewicz, B. G.; Pei, Z.; Hutchins, C. W.; Ballaron, S. J.; Stashko, M. A.; Lubben, T. H.; Berg, C. E.; Rondinone, C. M.; Trevillyan, J. M.; Jirousek, M. R. Selective protein tyrosine phosphatase 1B inhibitors: Targeting the second phosphotyrosine binding site with noncarboxylic acid-containing ligands. *J. Med. Chem.* **2003**, *46*, 3437−3440.
(46) Szczepankiewicz, B. G.; Liu, G.; Hajduk, P. J.; Abad-Zapatero, C.; Pei, Z.; Xin, Z.; Lubben, T. H.; Trevillyan, J. M.; Stashko, M. A.; Ballaron, S. J.; Liang, H.; Huang, F.; Hutchins, C. W.; Fesik, S. W.; Jirousek, M. R. Discovery of a potent, selective protein tyrosine phosphatase 1B inhibitor using a linked-fragment strategy. *J. Am. Chem. Soc.* **2003**, *125*, 4087−4096.
(47) Dunten, P.; Kammlott, U.; Crowther, R.; Levin, W.; Foley, L. H.; Wang, P.; Palermo, R. X-ray structure of a novel matrix metalloproteinase inhibitor complexed to stromelysin. *Protein Sci.* **2001**, *10*, 923−926.
(48) Dhanaraj, V.; Ye, Q. Z.; Johnson, L. L.; Hupe, D. J.; Ortwine, D. F.; Dunbar, J. B., Jr.; Rubin, J. R.; Pavlovsky, A.; Humblet, C.; Blundell, T. L. X-ray structure of a hydroxamate inhibitor complex of stromelysin catalytic domain and its comparison with members of the zinc metalloproteinase superfamily. *Structure* **1996**, *4*, 375−386.
(49) Pavlovsky, A. G.; Williams, M. G.; Ye, Q. Z.; Ortwine, D. F.; Purchase, C. F., II; White, A. D.; Dhanaraj, V.; Roth, B. D.; Johnson, L. L.; Hupe, D.; Humblet, C.; Blundell, T. L. X-ray structure of human stromelysin catalytic domain complexed with nonpeptide inhibitors: Implications for inhibitor selectivity. *Protein Sci.* **1999**, *8*, 1455−1462.
(50) Hajduk, P. J.; Boyd, S.; Nettesheim, D.; Nienaber, V.; Severin, J.; Smith, R.; Davidson, D.; Rockway, T.; Fesik, S. W. Identification of novel inhibitors of urokinase via NMR-based screening. *J. Med. Chem.* **2000**, *43*, 3862−3866.
(51) Crump, M. P.; Ceska, T. A.; Spyracopoulos, L.; Henry, A.; Archibald, S. C.; Alexander, R.; Taylor, R. J.; Findlow, S. C.; O'Connell, J.; Robinson, M. K.; Shock, A. Structure of an allosteric inhibitor of LFA-1 bound to the I-domain studied by crystallography, NMR, and calorimetry. *Biochemistry* **2004**, *43*, 2394−2404.
(52) Liu, G.; Huth, J. R.; Olejniczak, E. T.; Mendoza, R.; DeVries, P.; Leitza, S.; Reilly, E. B.; Okasinski, G. F.; Fesik, S. W.; von Geldern, T. W. Novel *p*-arylthio cinnamides as antagonists of leukocyte function-associated antigen-1/intracellular adhesion molecule-1 interaction. 2. Mechanism of inhibition and structure-based improvement of pharmaceutical properties. *J. Med. Chem.* **2001**, *44*, 1202−1210.

JM061134B