

Use of MCSS to Design Small Targeted Libraries: Application to Picornavirus Ligands

Diane Joseph-McCarthy,^{*,†,‡,§} Simon K. Tsang,^{||} David J. Filman,[†] James M. Hogle,^{*,†,||} and Martin Karplus^{*,‡}

Contribution from the Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115, Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, Biological Chemistry Department, Wyeth Research, 87 Cambridge Park Drive, Cambridge, Massachusetts 02140, and Committee on Higher Degrees in Biophysics, Harvard University, Cambridge, Massachusetts 02138

Received November 15, 2000

Abstract: Computational methods were used to design structure-based combinatorial libraries of antipicornavirus capsid-binding ligands. The multiple copy simultaneous search (MCSS) program was employed to calculate functionality maps for many diverse functional groups for both the poliovirus and rhinovirus capsid structures in the region of the known drug binding pocket. Based on the results of the MCSS calculations, small combinatorial libraries consisting of 10s or 100s of three-monomer compounds were designed and synthesized. Ligand binding was demonstrated by a noncell-based mass spectrometric assay, a functional immunoprecipitation assay, and crystallographic analysis of the complexes of the virus with two of the candidate ligands. The P1/Mahoney poliovirus strain was used in the experimental studies. A comparison showed that the MCSS calculations had correctly identified the observed binding site for all three monomer units in one ligand and for two out of three in the other ligand. The correct central monomer position in the second ligand was reproduced in calculations in which the several key residues lining the pocket were allowed to move. This study validates the computational methodology. It also illustrates that subtle changes in protein structure can lead to differences in docking results and points to the importance of including target flexibility, as well as ligand flexibility, in the design process.

Introduction

Structure-based ligand design is the objective of many computer programs that are now available.^{1–4} However, there are relatively few published cases of the successful design of ligands.^{5–7} One difficulty is that the computational results have errors due to the approximate methods being used. These can be overcome by combinatorial chemical methods that require only that suitable ligand fragments be proposed so that a biased library can be designed. The multiple copy simultaneous search (MCSS) method^{8–10} is ideally suited for such biased library design.

The structure-based combinatorial design of picornavirus capsid-binding ligands was undertaken as a test case of this approach. Polioviruses and the related rhinoviruses, which are the major cause of the common cold, are both in a subgroup of picornaviruses that are known to bind a number of ligands that prevent infectivity by inhibiting structural rearrangements that are required for cell entry. This subgroup also contains a number of less known viruses (including the Coxsackie A and B viruses, the echoviruses, and several enterovirus serotypes) that are significant human pathogens. Poliovirus was used for the experiments because of the relative ease of working with this well characterized picornavirus that is largely noninfective to immunized individuals. It is likely that the same approach would work for rhinoviruses or any of the other closely related viruses listed above. Maps of the preferred sites for various functional groups in the binding site of the poliovirus structure were calculated with MCSS and a series of candidate ligands were suggested.¹¹ The proposed bisbenzimidazole ligands differ from existing ligands¹² in that they consist of a unique pattern of

* Corresponding authors.

[†] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School.

[‡] Department of Chemistry and Chemical Biology, Harvard University.

[§] Biological Chemistry Department, Wyeth Research.

^{||} Committee on Higher Degrees in Biophysics, Harvard University.

(1) Joseph-McCarthy, D. Computational approaches to structure-based ligand design. *Pharmacol. Ther.* **1999**, *84*, 179–191.

(2) Bohacek, R. S.; McMartin, C. Modern computational chemistry and drug discovery: structure generating programs. *Curr. Opin. Chem. Biol.* **1997**, *1*, 157–161.

(3) Caflisch, A.; Karplus, M. Computational combinatorial chemistry for de novo ligand design: Review and assessment. *Perspect. Drug Discovery Des.* **1995**, *3*, 51–84.

(4) Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. Structure-Based Molecular Design. *Acc. Chem. Res.* **1994**, *27*, 117–123.

(5) Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med.* **2000**, *78*, 269–281.

(6) Murcko, M. A.; Caron, P. R.; Charifson, P. S. *Struct.-Based Drug Des.* **1999**, *34*, 297–306.

(7) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: A major success of structure-assisted drug design. *Ann. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249–284.

(8) Miranker, A.; Karplus, M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins* **1991**, *11*, 29–34.

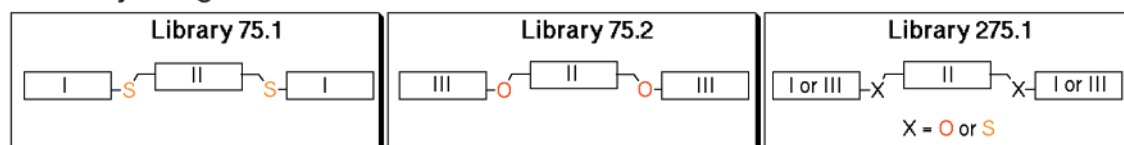
(9) Caflisch, A.; Miranker, A.; Karplus, M. Multiple copy simultaneous search and construction of ligands in binding sites. *J. Med. Chem.* **1993**, *36*, 2142–2167.

(10) Evensen, E.; Joseph-McCarthy, D.; Karplus, M. *MCSSv2*; Harvard University: Cambridge, MA, 1997.

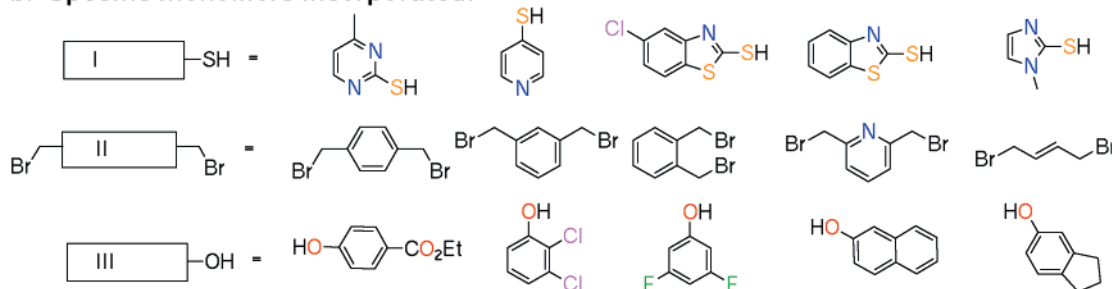
(11) Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Use of the multiple copy simultaneous search (MCSS) method to design a new class of picornavirus capsid binding drugs. *Proteins* **1997**, *29*, 32–58.

(12) Zhang, A.; Nanni, R. G.; Oren, D. A.; Rozhon, E. J.; Arnold, E. Three-dimensional structure–activity relationships for antiviral agents that interact with picornavirus capsids. *Virology* **1992**, *3*, 453–471.

a. Library Design:



b. Specific Monomers Incorporated:



c. Combinatorial Synthesis:

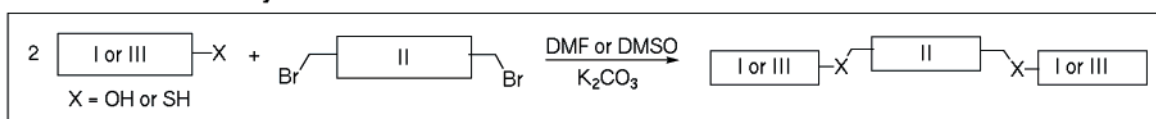


Figure 1. Schematic showing (a) the library design, (b) the specific monomers incorporated at each of the three monomer positions in the library ligands, and (c) the synthesis route. Because binders were found in library 75.1, libraries 75.2 and 275.1 were not tested.

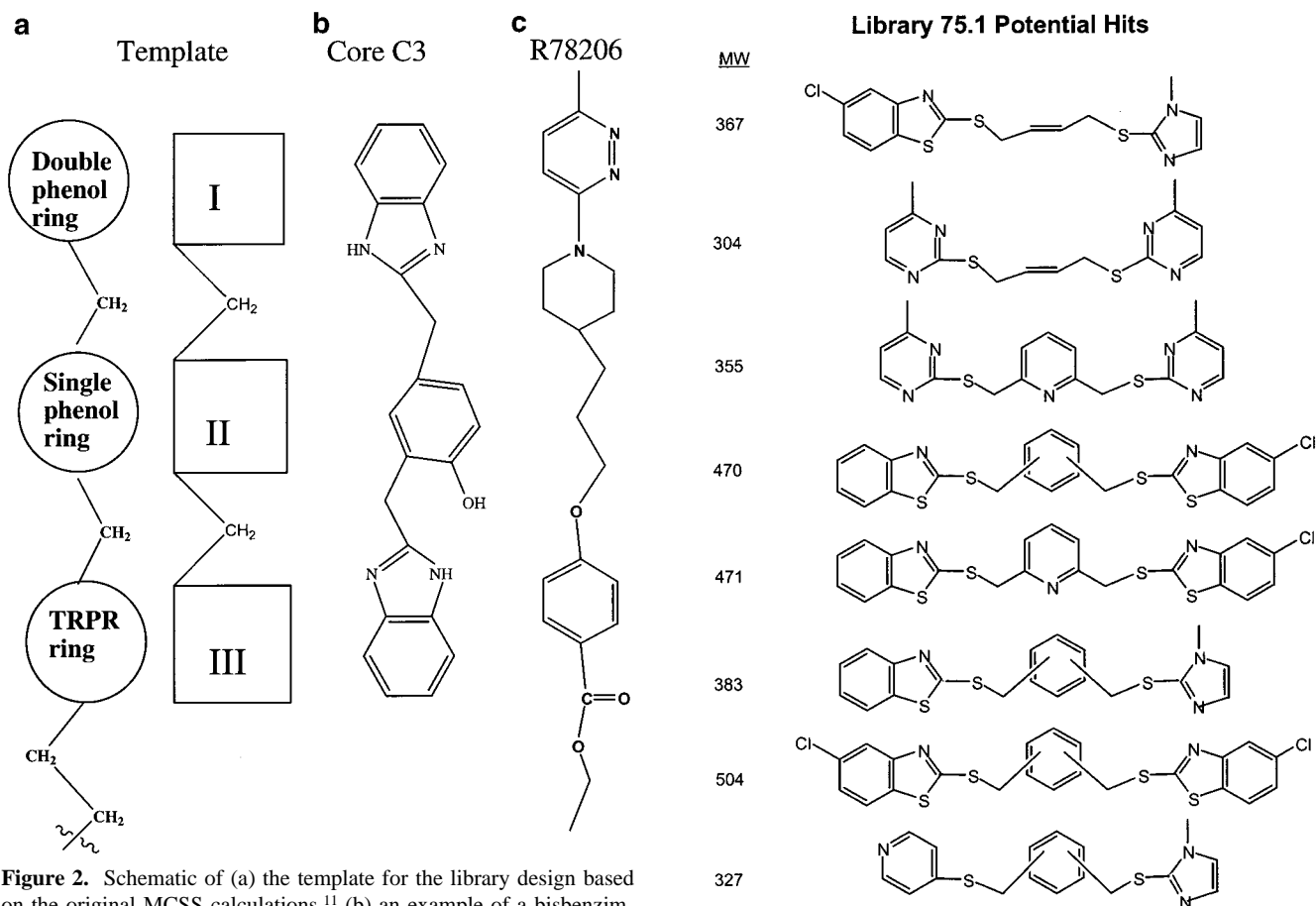


Figure 2. Schematic of (a) the template for the library design based on the original MCSS calculations,¹¹ (b) an example of a bisbenzimidazole compound from the original proposed series, and (c) the best of the existing Janssen compounds.¹² The Janssen compounds were not used in the library design but were used as standards in the experiments.

ring groups positioned to optimally fill the binding pocket and include heteroatoms expected to hydrogen bond to polar side

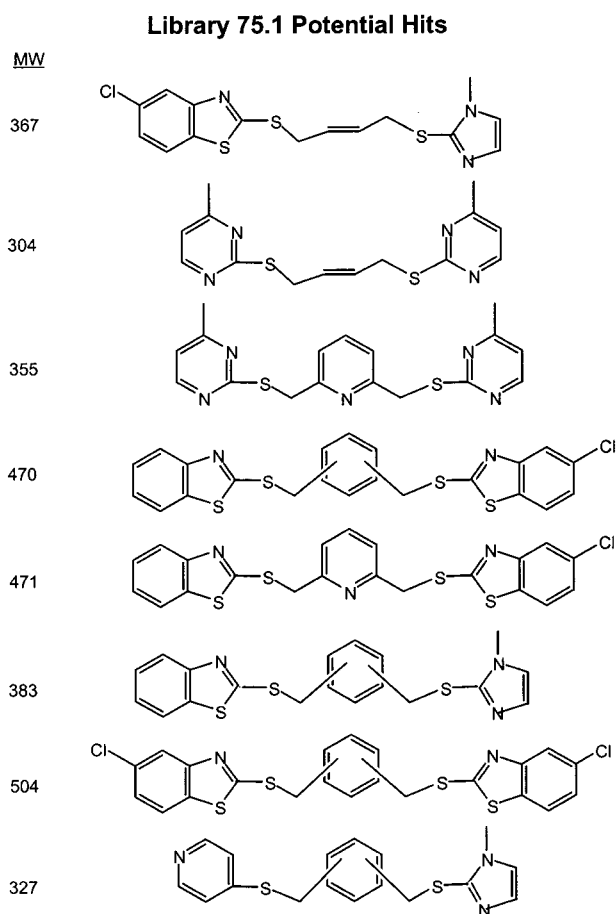
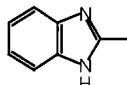
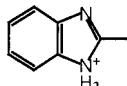
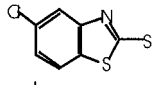

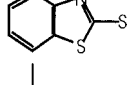





Figure 3. Potential hits from the preliminary screening of library 75.1 (see Figure 1) using the MS assay. The molecular weight is given for each compound and is used to identify it.

chains in the mainly hydrophobic binding pocket. Early attempts at synthesizing and testing individual compounds failed because of a combination of technical problems, including compound

Table 1. Functional Group Maps for P1/Mahoney Poliovirus

structure	chemical name	energy range ^a	number of minima	initial number of copies ^b	select minima in pocket ^c
	methylbenzimidazole (polarH) ^d	-42.2, -0.7	68	750	-14.8, 35th (deep in pocket) -12.2, 44th (near entrance)
	protonated methylbenzimidazole (polarH) ^d	-68.4, -0.1	58	500	-11.3, 49th (near entrance)
	2-methylmercapto-5-chlorobenzothiazole	-37.9, -0.7	127	4500	-18.1, 72nd (deep in pocket) Figure 9
	2-methylmercapto-1-methylimidazole	-43.1, -0.5	319	4500	-0.2, 242th (near entrance) Figure 9
	2-methylmercaptobenzothiazole	-30.8, -1.0	208	4500	-17.5, 59th Figure 10
	2-methylmercapto-4-methylpyrimidine	-40.3, -0.5	320	4500	-6.6, 290th Figure 10
	2,6-dimethylpyridine	-31.4, -0.3	181	4500	-9.6, 118th (main center)
	benzene	-10.9, -2.4	124	4500	-7.5, 33rd (main center)
	benzene, L396 complex protein (polarH) ^{d,e}	-9.4, -1.8	25	1000	-8.9, 5th (main center) -5.7, 18th (L396 center ring)
	benzene, flexible protein ^f	-10.9, -7.1	94	1200	-8.6, 9th (main center) -5.9, 64th L396 center ring)

^a Energies are in kcal/mol, and only minima with interaction energies less than 0 kcal/mol are saved. ^b The number of group copies initially randomly distributed at least 1 Å from the protein in the approximately 20 × 20 × 30 Å³ box surrounding the binding site pocket in the poliovirus structure. ^c Select minima shown in Figures 9 and 10, benzene minima at the center of the main binding pocket and at the position of the central ring in the L396 structure shown in Figure 11, and minima used in the initial design of the proposed bisbenzimidazole compounds and the libraries shown in Figure 1. ^d Minima were calculated using the polar hydrogen representation (a united-atom polar-hydrogen representation for the group copies as well as the protein using CHARMM²⁴ PARAM19²⁵ parameters). All other functional group maps were calculated using the hybrid representation (the same polar hydrogen representation for the groups²⁵ and an all atom representation for the protein²⁸). If parameters for a group were missing from the standard CHARMM sets, they were taken from the Quanta 4.0 PARM.PRM CHARMM set (Molecular Simulations, Inc., 1998). Partial charges for all groups in this table except for benzene were taken from the MMFF force field.²⁹ ^e The protein structure from the L396-P1/Mahoney complex was used. All other calculations used the native P1/Mahoney structure with the sphingosine removed from the main binding pocket. ^f Residues W1108, I1110, M1132, L1134, and L1261 of the P1/Mahoney poliovirus structure were also allowed to move during the minimizations of the group copies. For all other MCSS calculations, the entire protein was held fixed.

purity and reproducibility of assays and the protonation state of the bisbenzimidazole ligands (i.e., the benzimidazole groups are probably protonated at physiological pH). Two new assays to test for specific binding in the pocket of this viral system were developed recently: a mass spectroscopy (MS) assay which can be used as an initial screen of libraries of up to 100 or so compounds¹³ and a more definitive immunoprecipitation (IP) assay.¹⁴ To circumvent potential problems with the protonation state of the benzimidazole-containing compounds, small libraries of thiol-containing analogues were designed and synthesized. Details on the assays and their use in screening the combinatorial libraries are given in a separate paper.¹³

On the basis of the preliminary screening by MS of a 75-compound library designed by the use of the MCSS results, a series of six-compound libraries were synthesized and tested using both the MS and immunoprecipitation assays. Two of the six-compound libraries showed binding in both assays. The

(13) Tsang, S. K.; Cheh, J.; Isaacs, L.; Joseph-McCarthy, D.; Choi, S.-K.; Pevear, D. C.; Whitesides, G. M.; Hogle, J. M. A structurally biased combinatorial approach for discovering new anti-picornaviral compounds. *Chem. Biol.* **2001**, *8*, 33–45.

(14) Tsang, S. K.; Danthi, P.; Chow, M.; Hogle, J. M. Stabilization of poliovirus by capsid-binding antiviral drugs is due to entropic effects. *J. Mol. Biol.* **2000**, *296*, 335–340.

individual compounds making up these libraries were synthesized and tested. The best binding compounds from each of the libraries were crystallized with poliovirus and the structures determined by X-ray crystallography. One ligand bound as predicted by the MCSS calculations, and the other bound in a somewhat different mode. Calculations demonstrated that small conformational changes in a few side chains of the target protein are required to obtain the correct binding mode.

Materials and Methods

Viral System. All picornaviruses have a roughly spherical viral capsid that consists of 60 copies each of the viral proteins VP1 (306 amino acids for poliovirus), VP2 (272 amino acids), VP3 (238 amino acids), and VP4 (69 amino acids), where VP1–3 are eight-stranded, wedge-shaped antiparallel β-barrels with two flanking helices and VP4 is a short peptide with no regular structure. In the middle of the β-barrel core of VP1, there is a site that is normally occupied by a lipid-like ligand.¹⁵ This ligand, sometimes called pocket factor, has been modeled as a sphingosine in most poliovirus structures. This pocket has also been shown to be the binding site for a group of antiviral compounds.^{12,16–18}

(15) Filman, D. J.; Syed, R.; Chow, M.; Macadam, A. J.; Minor, P. D.; Hogle, J. M. Structural factors that control conformational transitions and serotype specificity in type 3 poliovirus. *EMBO J.* **1989**, *8*, 1567–1579.

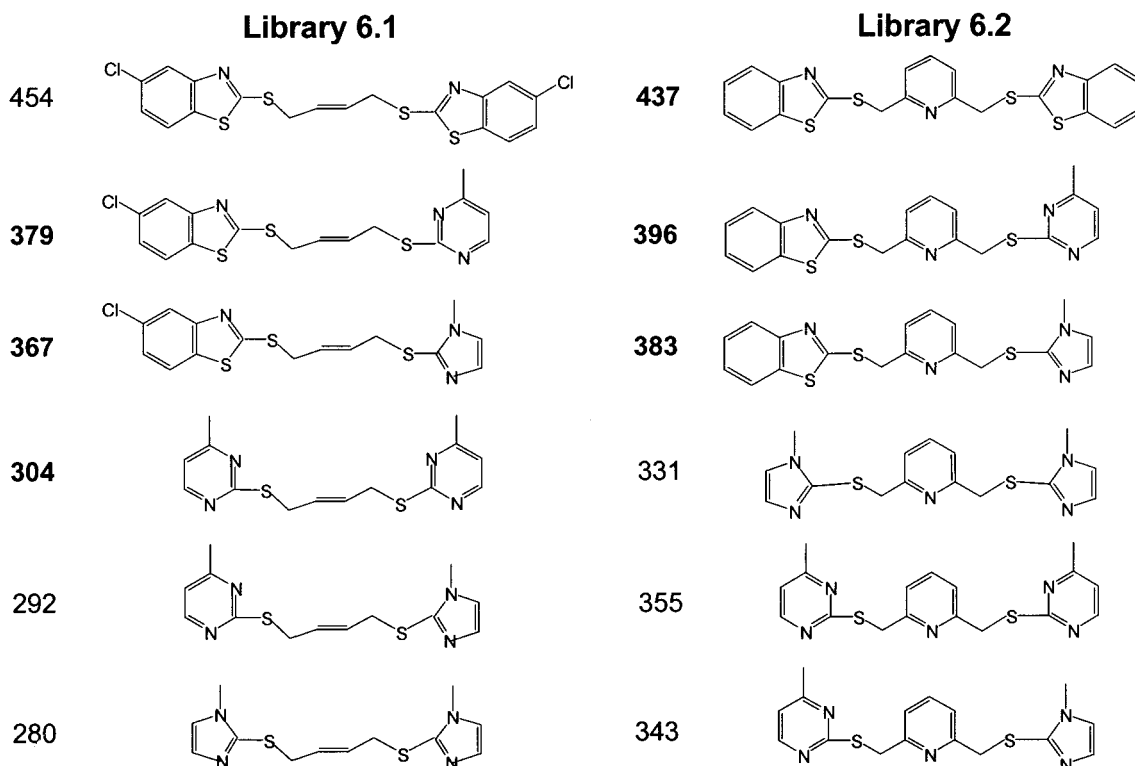


Figure 4. Two of the small six-compound libraries synthesized and tested. The molecular weights are given for each compound, and the compounds shown to bind to the poliovirus capsid protein are indicated in bold print. For library 6.1, L379, L367, and L304 showed binding; for library 6.2, L437, L396, and L383 showed binding.

MCSS Calculations and Library Design. The MCSS program^{8,10} calculates energetically favorable binding site positions (potential energy minima) for a given functional group in the target region of a macromolecule structure. Typically, several thousand copies of the functional group are randomly distributed in a box or sphere surrounding the binding region of the target structure and energy minimized simultaneously; each copy of the group experiences the full force field of the macromolecule but the group copies do not interact with each other. In one of its first published applications to design nonpeptidyl ligands, MCSS was used to calculate functional group maps for diverse chemical groups of the poliovirus and rhinovirus capsid structures in the region of the known drug binding pocket.¹¹ Linker carbon atoms were placed visually between selected minima, and the positions of the linker atoms were optimized to form chemically sensible candidate ligands. These calculations led to the design of combinatorial libraries for the synthesis of candidate ligands. A schematic of the library design is shown in Figure 1.

In the original study,¹¹ MCSS maps were calculated for P3/Sabin poliovirus and Rhinovirus 14 structures for the default set of functional groups in the program; they are *N*-methylacetamide, methanol, water, acetic acid, methylammonium, magnesium ion, magnesium ion plus water, 3-methylindole, 5-methylimidazole, phenol, benzene, toluene, cyclohexane, propane, and isobutane. These calculations were used to partition the main binding site into three contiguous regions and design a template (Figure 2). Although the P1/Mahoney poliovirus was used in the present experiments, the results calculated for P3/Sabin poliovirus

were expected to be useful because the binding sites in P3/Sabin and P1/Mahoney poliovirus are highly conserved. The P3/Sabin structure was chosen for the initial calculations because it was slightly higher resolution. A series of bisbenzimidazole compounds were designed to overlap with the proposed template and make interactions similar to those of the optimally placed functional groups (see Figure 2 for a summary). Benzimidazole was chosen for ease of synthesis.

Additional functional group maps have since been calculated for methylbenzimidazole and protonated methylbenzimidazole, as well as all for the library monomers shown in Figure 1a, for the capsid structures of the P3/Sabin¹⁵ and P1/Mahoney¹⁹ strains of poliovirus. The same region of the protein was used for all calculations; that is, copies of each group were randomly distributed in the same binding site box. The polar hydrogen representation was used for benzene as well as for methylbenzimidazole, and protonated methylbenzimidazole, while the hybrid hydrogen representation was used for the other groups. It has previously been shown that for benzene the polar and hybrid representations give essentially the same results with comparable normalized energies.¹¹ All energies given in Table 1 have been normalized by subtraction of a reference energy for each group type which is now standard in MCSS. Table 1 contains a complete list of the functional group maps calculated for the P1/Mahoney poliovirus structure. On the basis of the results of these and the original MCSS calculations, as well as the first series of proposed ligands in the original paper, three small combinatorial libraries were designed and subsequently synthesized.

Each library ligand is composed of three connected monomers. While in the original template (Figure 2a) a single atom connected each monomer, examination of the functional group minima in the pocket indicated that a two-atom connector might also be allowed. Because it provided the more synthetically accessible route, the two-atom connector was tried. The specific monomers that were allowed at each of the three monomer positions of the ligands (Figure 1b) were determined to a large extent by the MCSS calculations. 2-Methylmercaptobenzothiazole was selected instead of methylbenzimidazole to circumvent

(16) Grant, R. A.; Hiremath, C. N.; Filman, D. J.; Syed, R.; Andries, K.; Hogle, J. M. Structure of poliovirus complexes with anti-viral drugs: Implications for viral stability and drug design. *Curr. Biol.* **1994**, *4*, 784–797.

(17) Badger, J.; Minor, I.; Kremer, M. J.; Oliveira, M. A.; Smith, T. J.; Griffith, J. P.; Guerin, D. M. A.; Krishnaswamy, S.; Luo, M.; Rossmann, M. G.; McKinlay, M. A.; Diana, G. D.; Dutko, F. J.; Fancher, M.; Rueckert, R. R.; Heinz, B. A. Structural analysis of a series of antiviral agents complexed with human rhinovirus 14. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 3304–3308.

(18) Badger, J.; Minor, I.; Oliveira, M. A.; Smith, T. J.; Rossmann, M. G. Structural analysis of antiviral agents that interact with the capsid of human rhinoviruses. *Proteins* **1989**, *6*, 1–19.

(19) Hogle, J. M.; Chow, M.; Filman, D. J. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science* **1985**, *229*, 1358–1365.

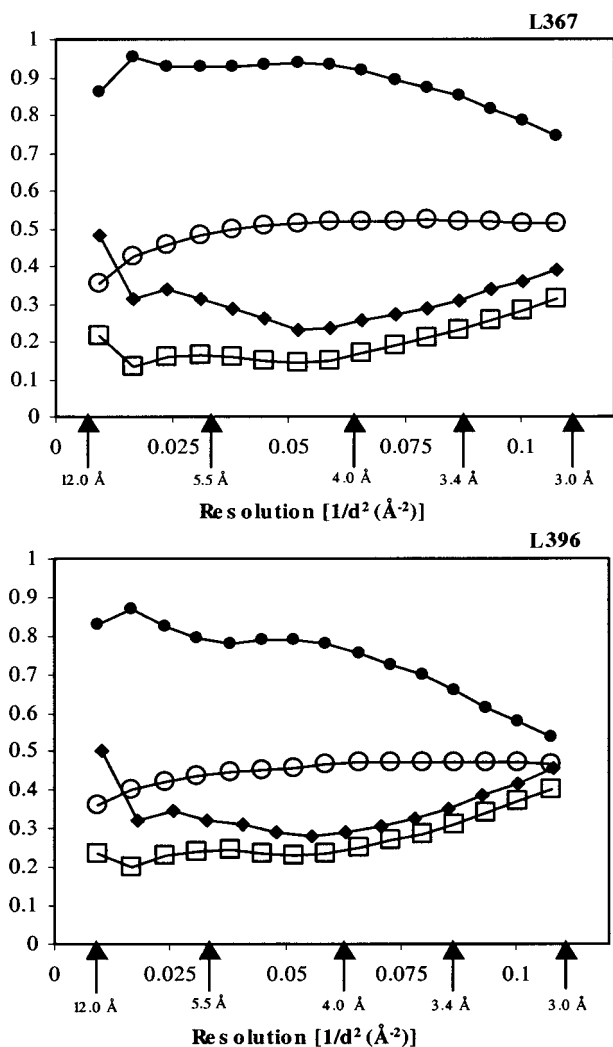


Figure 5. Crystallographic data processing statistics plotted as a function of resolution for the L367 complex structure (top) and for L396 (bottom). Closed circles represent the *R* factor for the correlation coefficient. Open circles correspond to the percent completeness for each resolution bin. Closed diamonds denote the crystallographic *R* factor. Open squares denote the noncrystallographic *R* factor.

issues with the protonation state of the monomer and because a relatively simple synthesis was devised with this type of monomer. Also, at this stage, during the design of the libraries, all of the MCSS maps were reviewed in succession (i.e., those for methylbenzimidazole and protonated methylbenzimidazole, in addition to the default MCSS groups). Each of the three regions of the binding site was examined successively, and optimal functional groups at each site were noted. Group positions were considered optimal on the basis of their interaction energy with the protein and their orientation relative to the next contiguous binding site region; only those with favorable energies which could also be readily attached to the next monomer were selected. Within the framework of the desired chemistry (see Figure 1c) and the availability of reagents in the laboratory, similar functional groups were selected for inclusion in the library. Specific groups were selected on the basis of shape and hydrogen bonding capabilities. So, for example, for the linker monomer, only single ring groups, rings with an atom capable of accepting a hydrogen bond (on the basis of benzene and imidazole minima positions), and chains were selected (overlapping propane and *N*-methylacetamide and other minima). As the library synthesis and testing progressed, the specific groups incorporated were modeled and MCSS maps calculated to confirm that the minima overlapped with the previously modeled, similar functional groups, as expected. For example, the assumption that the position of 2-methylmercaptobenzothiazole in the binding site would overlap with the

position of methylbenzimidazole was confirmed. Knowledge of existing ligands¹² (see Figure 2c) was not used in the library design.

Biological Assays. Briefly, the MS assay is a noncell-based assay which can rapidly screen a mixture of compounds for specific binding to the viral target.¹³ In the MS assay, a library of up to 100s of compounds is incubated with virus for several hours, compounds which do not bind are washed away, and the compounds that have bound to the virus are extracted using an organic solvent which loosens up the viral protein. A mass spectrum is taken of the resulting solution to identify the bound compounds, assuming no fragmentation has occurred and only singly charged species are formed in the MS.

In the functional IP assay, the rate of conversion of the native poliovirus particle to the 135S particle (based on sedimentation) is measured; the existing antiviral compounds inhibit the transition to this infectious intermediate.¹⁴ For poliovirus, antibodies exist that are directed against a peptide corresponding to residues 21–40 of VP1. In the native virus particle, the peptide recognized by these antibodies is buried inside the virus capsid; upon conversion to the 135S particle, the N-terminus of VP1, including this peptide, is externalized. To determine the rate of conversion, radio-labeled virus and the compound or a mixture of compounds to be tested are simultaneously added to antibodies that are attached to beads. The amount of converted virus is measured at set time intervals. Known capsid-binding agents slow the rate of this thermally induced 160S to 135S conversion, and there is a rough correlation between the rate of conversion at a given temperature and the MIC (minimum inhibitory concentration required to inhibit viral replication) for the compound.¹⁴ The MS assay is used for the initial screening of libraries, and the more definitive IP assay is carried out to confirm apparently positive results.

Library Synthesis and Testing. The synthesis was done combinatorially using solution phase chemistry (the chemical reactions are shown in Figure 1c).¹³ The libraries are expected to consist of 75, 75, and 275 compounds each, assuming that all of the reactions occurred as planned; that is, if compounds consisting of all possible combinations of the allowed monomers have been synthesized. No tests were made to verify this, so the libraries may have contained a smaller number of compounds. The first 75-compound library was screened for binding to the P1/Mahoney poliovirus capsid using the MS assay, and eight peaks potentially corresponding to ligands were identified (Figure 3). Four of the peaks (304, 367, 343, and 471) could be unambiguously assigned to individual library compounds. The other four peaks (327, 383, 470, and 504) could only correspond to the specific library compounds with the benzyl linkers depicted in Figure 3. Because the three dimethylbenzene linkers (Figure 1) all have the same molecular weight, however, the precise connectivity of the benzyl linker to monomers 1 and 3 in these compounds cannot be determined by the MS screen. The compounds corresponding to these four peaks are likely to have connectivities similar to those of L343 and L471 (1,3 connectivities) as the modeling predicts, but this was not confirmed directly.

On the basis of this preliminary MS result, two small six-compound libraries were synthesized (Figure 4). The first six-compound library, library 6.1, consists of monomers 2-mercapto-5-chlorobenzothiazole, 2-mercapto-4-methylpyrimidine, and 2-mercapto-1-methylimidazole at positions 1 and 3 and monomer *trans*-2-butene at position 2. The second six-compound library, library 6.2, consists of the same three monomers at positions 1 and 3, except without the Cl substituent on the benzothiazole, which was expected to improve the solubility, but with a more bulky group (2,6-dimethylpyridine) at position 2. The central ring with the nitrogen was chosen for the linker monomer instead of a simple benzyl ring to eliminate the ambiguity associated with the benzyl linkers.

The compounds in library 6.2 were of more interest because they represent a novel class of compounds for the picornavirus system, and they were predicted by the calculations to better fill the binding pocket and, therefore, to bind more tightly. Both six-compound libraries showed overall protection against conversion of the native particle to the 135S particle in the IP assay. As anticipated, library 6.2 overall showed greater protection than library 6.1. The smaller libraries were also examined using the MS assay, and peaks that correspond to L379, L367, and L304 from library 6.1 and L437, L396, and L383 from library

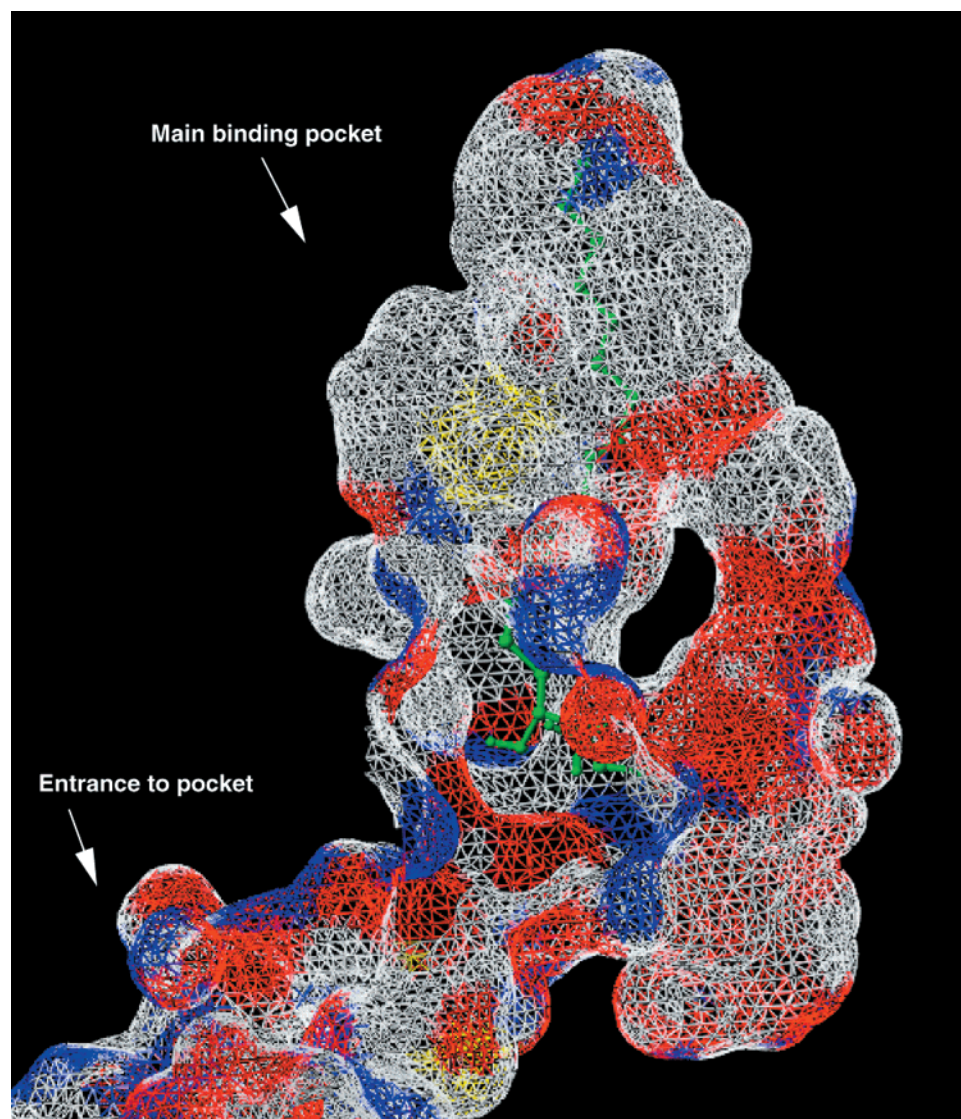


Figure 6. Surface showing the binding pocket in the capsid protein of the native P1/Mahoney poliovirus structure. The MOLCAD multichannel surface, as calculated in Sybyl 6.6 (Tripos, Inc., 2000), is colored by element according to the surrounding protein atoms. Spingosine is shown in green in ball-and-stick representation to indicate the location of the main binding pocket.

6.2 were observed in the spectra. The individual compounds from both six-compound libraries were then synthesized and tested using both assays, and the best binding compound from each was crystallized with the capsid protein. The rate constant (10^{-5} s^{-1}) for the P1/Mahoney 160S to 135S transition is 643 ± 23 in the presence of 0.1% DMSO, and it is 261 ± 27 for L367, 140 ± 2 for L383, and 115 ± 18 for L396.¹³ L304 and L379 only marginally reduced the rate constant for the conversion, and L379 had limited solubility in the conversion buffer.¹³ The individual synthesis of L437 did not produce a pure compound, and so, while the product was at least as active as L396, it was not pursued. In subsequent cell-based assays, L367, L396, and L383 were tested and shown to have MIC values (minimum inhibitory concentration required to inhibit viral replication) of $3.2 \mu\text{M}$, $11 \mu\text{M}$, and $0.26 \mu\text{M}$ against P1/Mahoney poliovirus.¹³

Virus Growth and Purification. P1/Mahoney was grown in HeLa cells propagated in suspension and purified by differential centrifugation and CsCl density gradient fractionation as described previously.²⁰ Purified virus was dialyzed into PBS and concentrated to 7.85 mg/mL using a microconcentrator.

Crystallization of Virus–Compound Complexes. Crystallization of P1/Mahoney has been previously described.¹⁹ All work was done at

$4 \text{ }^\circ\text{C}$. Concentrated virus was pipetted into buttons and dialyzed against 2 mL of a reservoir solution of PMC7 (10 mM Pipes, 5 mM MgCl_2 , 1 mM CaCl_2 , pH 7.0), 1% PEG 400, and 0–100 mM NaCl. Crystals appeared within a matter of days and took roughly one week to grow to full size ($\sim 0.3 \text{ mm}^3$). The crystals were transferred to a vial of synthetic mother liquor, which contained an additional 25% ethylene glycol as cryoprotectant. The crystals were allowed to stabilize for 3 days in the synthetic mother liquor. Subsequently, these crystals were soaked in a solution identical to the synthetic mother liquor, except that it contained $2.2 \times 10^{-4} \text{ M}$ of compound 367 or $8 \times 10^{-5} \text{ M}$ of compound L396 at a final DMSO concentration of 0.1%. These crystals were allowed to soak in the presence of compound for a week. Then, the crystals were transferred to a freezing solution, identical to the compound-containing synthetic mother liquor, with the exception of the presence of 1% glucose. The crystals were immediately frozen in liquid nitrogen and stored in a liquid nitrogen Dewar flask until used.

Data Collection and Processing. Diffraction data were collected on a Mar 345 detector mounted on an Enraf-Nonius GX-13 rotating anode operating at 40kV, 60mA, with a $100 \mu\text{m}$ focus and modified Franks mirror optics. Only one crystal was used for each data set. The crystals were kept in a liquid nitrogen stream at all times. Sixty images were collected for each crystal. Each image covered 0.3° oscillation with an exposure time of 1 h. The crystals diffracted to about 2.5 \AA . A suitably complete data set consisting of over 1 million observations

(20) Yeates, T. O.; Jacobson, D. H.; Martin, A.; Wychowski, C.; Girard, M.; Filman, D. J.; Hogle, J. M. Three-dimensional structure of a mouse-adapted type 2/type 1 poliovirus chimera. *EMBO J.* **1991**, *10*, 2331–2341.

Table 2. Summary of Crystallographic Processing Data

data processing statistics	L367	L396
no. of crystals	1	1
maximum resolution	3.0 Å	3.0 Å
no. of total reflections	682 253	635 138
no. of unique reflections	471 537	425 994
unit cell parameters (<i>a</i> , <i>b</i> , <i>c</i>) (Å)	320.3, 355.4, 377.5	319.4, 355.2, 377.8
<i>c</i> -translation relative to native virus	0.25	0.25
rotation about <i>c</i> -axis relative to native virus	2.56°	3.88°
R_{sym}^a	0.165	0.190
R_{cryst}^b	0.306	0.361
R_{nonx}^c	0.212	0.291
R_{corr}^d	0.887	0.760
bond length RMS deviation	0.016	0.011
bond angle RMS deviation	2.640	2.475
occupancy		
structure factor comparison	81%	84%
XPLOR group refinement	53%	76%

$$R_{\text{sym}} = \frac{\sum_{hj} |I_{hj} - \langle I_h \rangle|}{\sum_h \langle I_h \rangle}$$

$$R_{\text{cryst}} = \frac{\sum_h ||F_{\text{obs}}| - |F_{\text{calc}}||}{\sum_h |F_{\text{obs}}|}$$

^c R_{nonx} is the same as R_{cryst} except that the calculated amplitudes are derived from the transform of the converged averaged electron density. ^d R_{corr} is the linear correlation coefficient measuring the agreement between F_{obs} and the transform of the converged averaged electron density.

was collected to 3 Å resolution in 3–4 days. Reliable measurements of a large percentage of the observations could be obtained, in a reasonable time frame, only to 3 Å on a rotating anode source. The “icosahedrally unique” data are well sampled, and the resulting electron density maps (submitted as Supporting Information) are readily interpretable at this resolution because of the extensive (30-fold) noncrystallographic symmetry (NCS) of the viral system. The digitized data were processed using DENZO²¹ and merged with SCALEPACK.²¹ Data collection and refinement statistics are summarized in Table 2 and Figure 5.

Structure Determination. The structure of the complexes was determined by molecular replacement using the structure of native P1/Mahoney as the molecular replacement model. To reduce phase bias due to ligand binding, the atoms in the sphingosine ligand, the atoms in the native model around the VP1 pocket, and a portion of the GH loop (residues 1230–1240) whose structure changes upon ligand binding were removed. Reciprocal space trial searches were used to determine the position of the virus along the crystallographic *c* axis and the rotation of the virus about the *c* axis. Model-based starting phases were refined by cycles of NCS-based averaging to convergence. The density map generated from the initial phases produced density in the pocket region that was clearly distinguishable as the compound of interest and not the natural ligand or “pocket factor”. Subsequent refinements of the phases and of the atomic model were executed by alternating cycles of noncrystallographic symmetry refinement, interactive model building, and pseudoreal space refinement in a local region around the protomer (the icosahedral asymmetric unit consisting of one copy each of VP1, VP2, VP3, and VP4) using XX12.²²

(21) Otwinowski, Z.; Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **1997**, *276A*, 307–326.

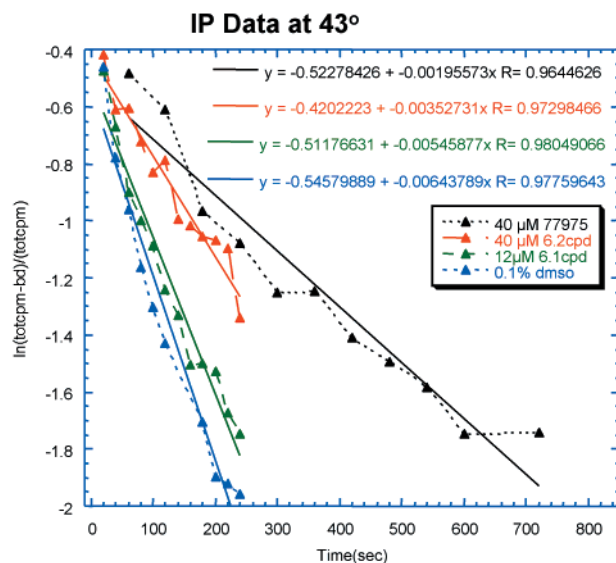


Figure 7. Immunoprecipitation data for the rate of conversion of P1/Mahoney from the 160S to the 135S particle in the presence of various compounds. The log of the percent 160S remaining is plotted versus time, and the first-order rate constant for the conversion is determined from the slope of each plot.

Compound Occupancy Estimation. Two methods were used to determine the compound occupancy of the VP1 binding pocket. In the first method, linear combinations of pseudostructure factors derived from Fourier transformation of protomer box models²² of the native (sphingosine containing) and virus–compound complexes were compared with the transform of the refined protomer box; with protomer box sampling of the averaged map, the total occupancy is implicitly assumed to be 1.0. In the second method, the compound atoms were subjected to group occupancy refinement using X-PLOR²³ using the XX12²² package. For L396, the occupancy of the compound relative to the native ligand sphingosine was estimated to be 84%. For L367, two alternate conformations of the five-membered ring appeared equally likely; two conformations had equal occupancies of 26% as estimated with X-PLOR.

Ligand–Protein Interaction Energies. The interaction energies of L367 and L367 with the two linker sulfur atoms removed, L367–2S, were calculated with the native P1/Mahoney and P3/Sabin protein structures. The conformation of L367 in its complex with P1/Mahoney was manually docked into each native protein structure. For each calculation, the protein was held fixed and the ligand subjected to up to 500 steps of ABNR minimization using CHARMM²⁴ and the polar hydrogen force field.²⁵ A 13 Å cutoff was applied to the nonbonded energy terms using the SHIFT truncation function, and the minimizations were stopped when the potential energy gradient dropped below 0.01 kcal/mol·Å.

Results and Discussion

An initial series of bisbenzimidazole compounds had been proposed on the basis of calculated MCSS functional group maps.¹¹ However, calculations showed that the deepest end of the capsid ligand-binding pocket (see Figure 6) is not expected to accommodate a positive charge, and no methylammonium

(22) Jacobson, D. H.; Hogle, J. M.; Filman, D. J. A pseudo-cell based approach to efficient crystallographic refinement of viruses. *Acta Crystallogr.* **1996**, *D52*, 693–711.

(23) Brünger, A. T. *X-PLOR Manual Version 3.1*; Yale University Press: New Haven, CT, 1997.

(24) 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.

(25) Neria, E.; Fischer, S.; Karplus, M. Simulation of activation free energies in molecular systems. *J. Chem. Phys.* **1996**, *105*, 1902–1921.

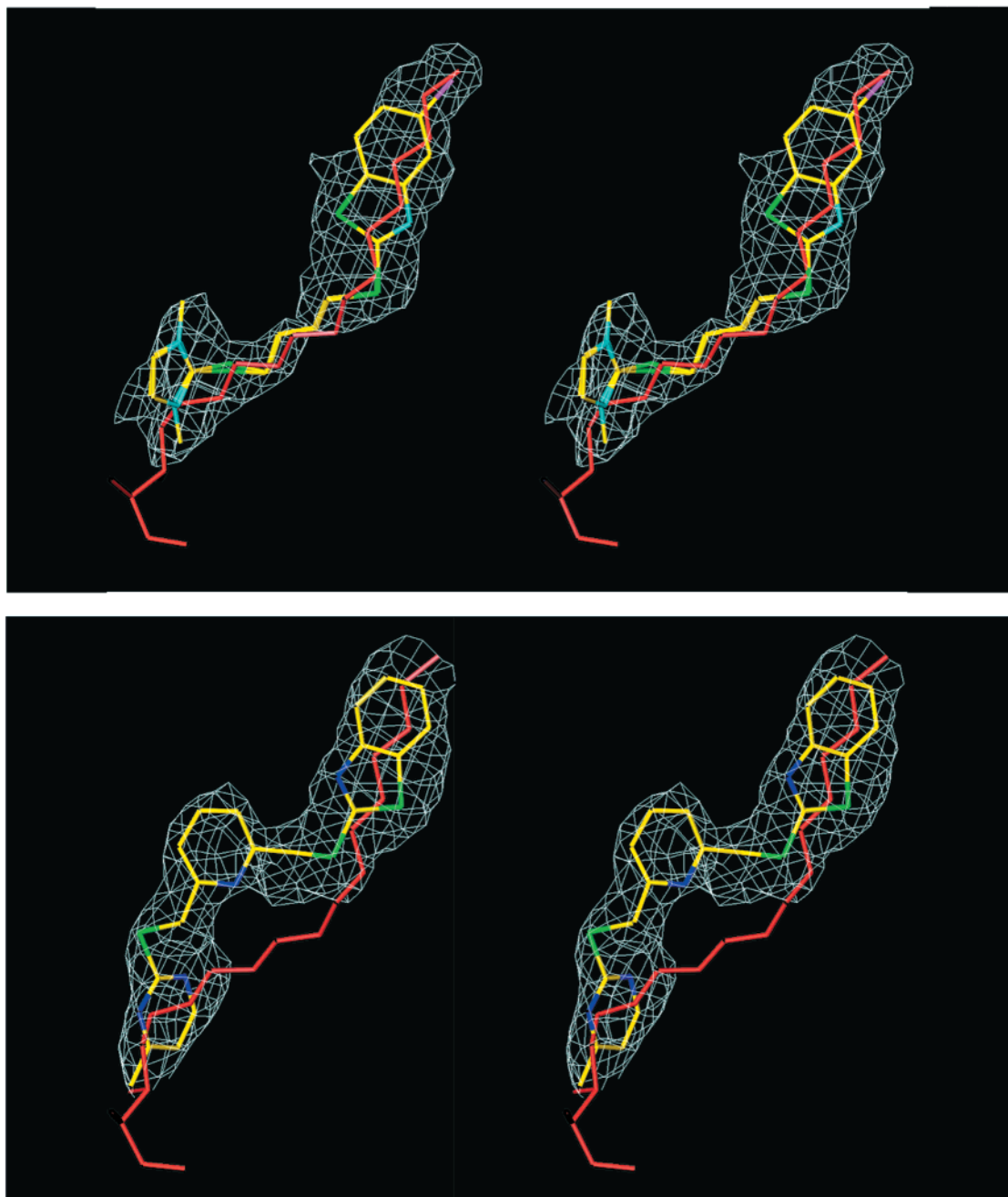


Figure 8. Stereoview of averaged F_o electron density maps contoured about the ligand in (top) the L367 complex structure and (bottom) the L396 structure. For comparison, the natural ligand sphingosine as it is bound in the native P1/Mahoney structure is indicated in red.

minima were found in the binding site pocket. Subsequent calculation of protonated methylbenzimidazole maps (see Table 1) confirmed this prediction. Because the bisbenzimidazole compounds are expected to have one of the nitrogens in this region and the pK_a of methylbenzimidazole is 6.1, thiol analogues of these compounds were synthesized. To avoid the need to predict accurately the binding of specific ligands and to decrease the time required for synthesis, small combinatorial libraries of compounds were synthesized instead of highly pure individual compounds.

The design of the libraries is shown in Figure 1, and the specific monomers included at each of the three monomer positions were largely determined by the earlier MCSS calculations for P3/Sabin poliovirus and subsequent MCSS calculations for P1/Mahoney poliovirus. Functional group minima were examined in each of the three binding site partitions (corre-

sponding to the proposed binding sites for the three monomers of a library ligand), and groups with favorable interaction energies and positions suitable for connection to the next monomer in the ligand were selected. Some modifications to the selected monomers were made for ease of synthesis. The results of the MCSS calculations carried out specifically for the P1/Mahoney poliovirus structure are summarized in Table 1. They include the library monomers as well as benzene, methylbenzimidazole, and protonated methylbenzimidazole.

Three small libraries containing a maximum of 75, 75, and 275 compounds, respectively, were synthesized as described in the Materials and Methods section and then rapidly screened using the newly developed MS assay.¹³ On the basis of these preliminary screening results for the first library, library 75.1, two smaller six-compound libraries were synthesized. These libraries, library 6.1 and library 6.2, were then assayed for

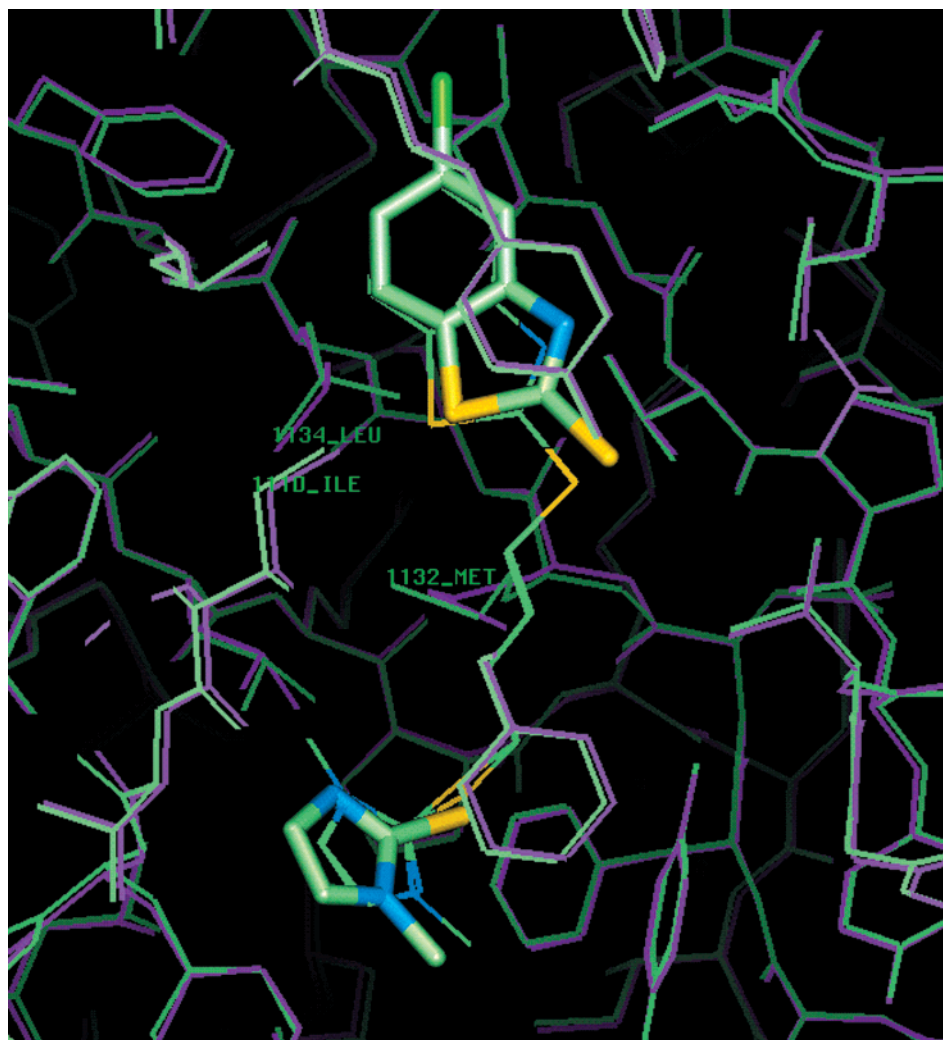


Figure 9. Superposition of the X-ray structure of L367 bound to P1/Mahoney poliovirus and the native P1/Mahoney structure with selected MCSS minima. The L367-poliovirus complex structure protein is shown in light green. The native poliovirus structure is in purple, and MCSS minima for 2-methylmercapto-5-chlorobenzothiazole and 2-methylmercapto-1-methylimidazole are shown in a licorice model colored by element.

binding using the MS assay as well as a novel IP assay. Overall, both six-compound libraries showed protection against conversion of the native P1/Mahoney poliovirus to the 135S particle relative to the DMSO control (Figure 7). As expected, library 6.2 showed improved protection relative to library 6.1. Peaks were identified in the MS assay that are likely to correspond to L379, L367, and L304 from library 6.1 and L437, L396, and L383 from library 6.2. In subsequent IP assays of the individual compounds, L396 appeared to be a better binder than L367, and both were better binders than L304. Two ligands, L367 and L396, were selected for further studies on the basis of their antiviral activities (3.2 μM and 11 μM MIC, respectively, in a standard infectivity assay¹³) and solubility. Crystal structures of L367 and L396 bound to the P1/Mahoney capsid protein were determined. Attempts were made to obtain a crystal structure of a third compound, L304, in complex with poliovirus, but electron density in the ligand binding pocket was consistent only with sphingosine, the natural ligand, and not L304; this result is in accord with the fact that L304 is a weaker ligand. The results for L367 and L396 are shown in Figure 8.

The MCSS calculations correctly identified the positions of the functional groups in L367 bound to P1/Mahoney as potential energy minima. Specifically, minima for 2-methylmercapto-5-chlorobenzothiazole and 2-methylmercapto-1-methylimidazole overlap with the position of those functional groups in the complex structure (see Table 1 and Figure 9). Thus, the L367-

Table 3. Ligand-Protein Interaction Energies

protein structure	ligand	interaction energy (RMS gradient) ^a
P3/Sabin	L367	-17.6 (0.0083)
	L367 - 2S	-10.5 (0.0043)
P1/Mahoney	L367	-4.3 (0.0066)
	L367 - 2S	7.7 (0.0073)

^a The potential energy is in kcal/mol, and the root-mean-square of the energy gradient is in kcal/mol·Å.

poliovirus structure is consistent with the MCSS results and confirms the binding predictions. Furthermore, retrospectively, the X-ray position for L367 in P1/Mahoney was energy minimized in the native P3/Sabin structure as well as in the native P1/Mahoney structure. The interaction energies for the L367 and L367 - 2S ligands are given in Table 3 and indicate that not only is the two atom linker between monomers acceptable but also it is preferred for both proteins.

For L396, minima for 2-methylmercaptobenzothiazole and 2-methylmercapto-4-methylpyrimidine overlap with the positions of those functional groups in the complex structure (see Table 1 and Figure 10), but there is no benzene or 2,6-dimethylpyridine minimum that overlaps with the position of the central ring in L396. The central ring in L396 has a somewhat different binding mode from what was predicted (Figure 11). The central ring was predicted to lie in the main binding pocket as defined by the position of sphingosine in the

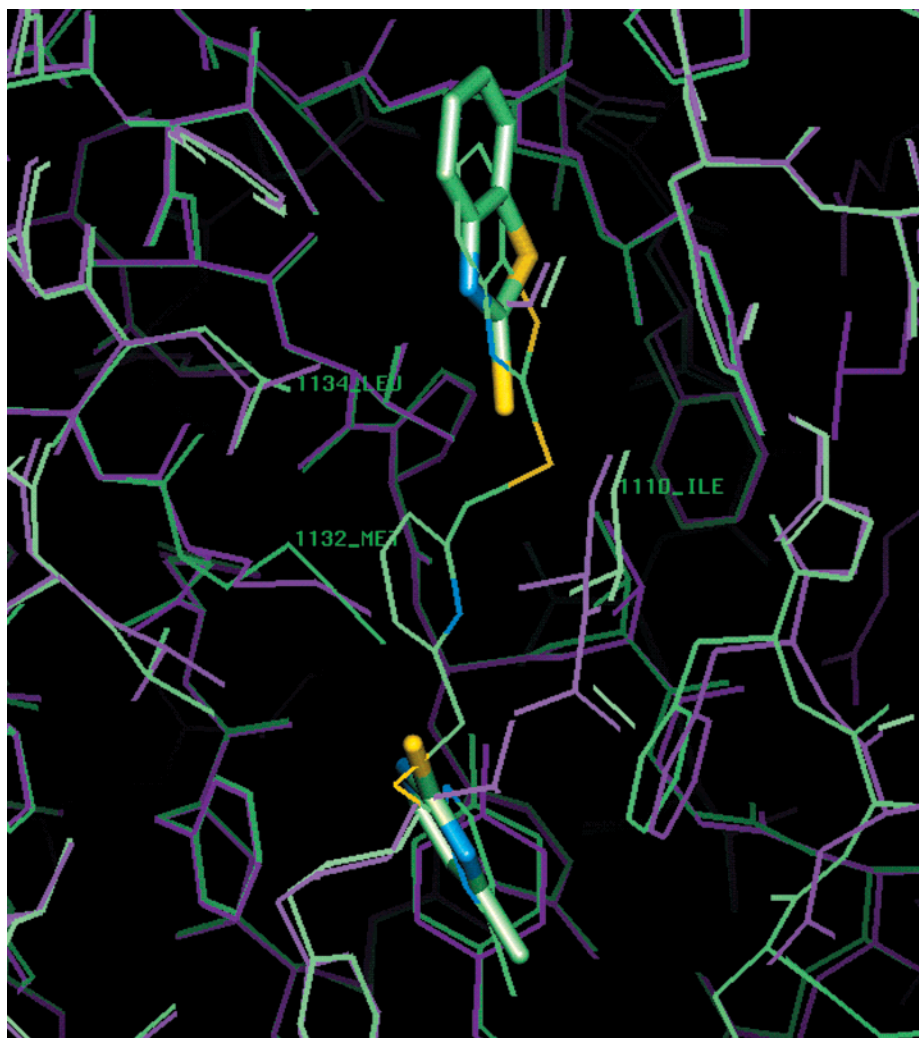


Figure 10. Superposition of the X-ray structure of L396 bound to P1/Mahoney poliovirus and the native P1/Mahoney structure with selected MCSS minima. The L396-poliovirus complex structure protein is shown in light green. The native poliovirus structure is in purple, and MCSS minima for 2-methylmercaptobenzothiazole and 2-methylmercapto-4-methylpyrimidine are shown in a licorice model colored by element.

native structure (see the benzene functional group map in blue in Figure 11), but instead, it bulges out into a small side pocket. For both the native P1/Mahoney and P3/Sabin poliovirus structures, there is no ring minimum at this location. The small difference in the actual and predicted binding may be because of the greater length of the spacers connecting the rings in L396, relative to the proposed structures. Having a two atom linker instead of the one atom linker between each ring may allow the position of the central ring to shift out of the main binding pocket somewhat with consequent small changes in a few protein side chain positions. Overall, the root-mean-squares difference between main chain atoms in the P1/Mahoney native structure and the L396 complex structure is only 0.58 Å for the cutout region used in the calculations. Thus, the two protein structures are very similar to one another (as can also be seen in Figure 10). When the L396 complex protein structure is used for the MCSS calculations, there is a benzene minimum that overlaps with the central ring position of L396 (shown in purple in Figure 11). When this benzene group is placed in exactly the same position in the native P1/Mahoney structure, it makes bad contacts (closer than 2.65 Å between a hydrogen and a heavy atom) with the side chains of Trp 1108, Ile 1110, Leu 1134, and Leu 1261 (see Figure 12a). Similarly, when the L396 ligand is placed in the native P1/Mahoney structure by superposition of main chain atoms between the two protein

structures, the central ring in L396 makes bad contacts with Trp 1108, Ile 1110, and Leu 1134 (Figure 12b). The side chains surrounding that benzene group that have shifted most in the L396 complex structure are Ile 1110, Met 1132, and Leu 1134 (see Figure 10).

An MCSS benzene map was calculated for the native P1/Mahoney structure with the five residues (Trp 1108, Ile 1110, Met 1132, Leu 1134, and Leu 1261) allowed to move during the minimization. The mobile residues move in the average field of the functional group copies, and the remainder of the protein structure is held fixed during the calculations. In addition to allowing these residues to move, the initial energy cutoff for discarding group copies during the calculation was increased from the default value of 250 to 500 kcal/mol (with a final value of 10 kcal/mol). This energy cutoff was used to discard group copies whose interaction energies were too high after the first cycle of minimization, and it was changed in decrements geometrically as the calculation proceeded to the final value. The increased energy cutoff allows enough minimization of the side chain positions to occur before the group copies are automatically discarded because their interaction energy is too high. Using this procedure, a benzene minimum is found at the location of the central ring in the L396 complex structure (Figure 11). All five of the mobile protein residues made similar adjustments (Figure 13), with their final positions intermediate

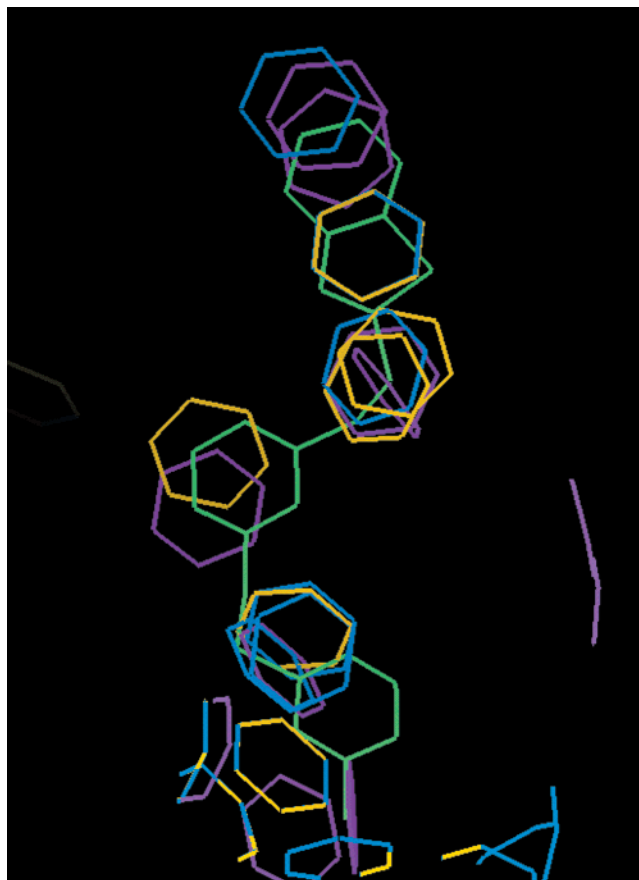


Figure 11. Superposition of L396 (in light green) as bound to P1/Mahoney poliovirus and benzene minima calculated for different capsid protein structures. Only the protein backbone atoms were superimposed. BENZ minima calculated for the native P1/Mahoney poliovirus structure (in blue), the protein structure from the L396–poliovirus complex (in purple), and the native P1/Mahoney poliovirus structure with five protein residues also allowed to move during the minimization (in yellow) are shown.

between the P1/Mahoney and L396 complex conformations; overall, the RMSD for the five residues before and after the MCSS minimization was 0.64 Å (0.29 Å for Trp 1108, 0.40 Å for Ile 1110, 1.34 Å for Met 1132, 0.36 Å for Leu 1134, and 0.37 Å for Leu 1261). The same type of calculation with only protein residues 1110 and 1132 free to move, with only residues 1110, 1132, and 1134 free, or with residues 1108, 1110, 1134, and 1261 free did not identify the observed position as a minimum. Even though Met 1132 would not directly make bad contacts with the ligand group in the native structure, its position must shift in concert with the other residues for the ring to be accommodated in this side pocket. Previous studies of complexes of poliovirus with a series of candidate drugs from Janssen Pharmaceuticals and Sterling-Winthrop laboratories had implicated residues 1132, 1134, and 1261 as critical determinants of binding affinity and demonstrated that drug binding was often accompanied by significant shifts in Met 1132 and smaller linked changes in the positions of 1134 and 1261.¹⁶

Conclusions

The MCSS method has been used to design small combinatorial libraries of capsid-binding ligands for poliovirus. The binding of the proposed ligands has been confirmed using newly developed MS and IP assays. For two of the best binding compounds, crystal structures were determined showing that the compounds displace the natural ligand sphingosine and bind in

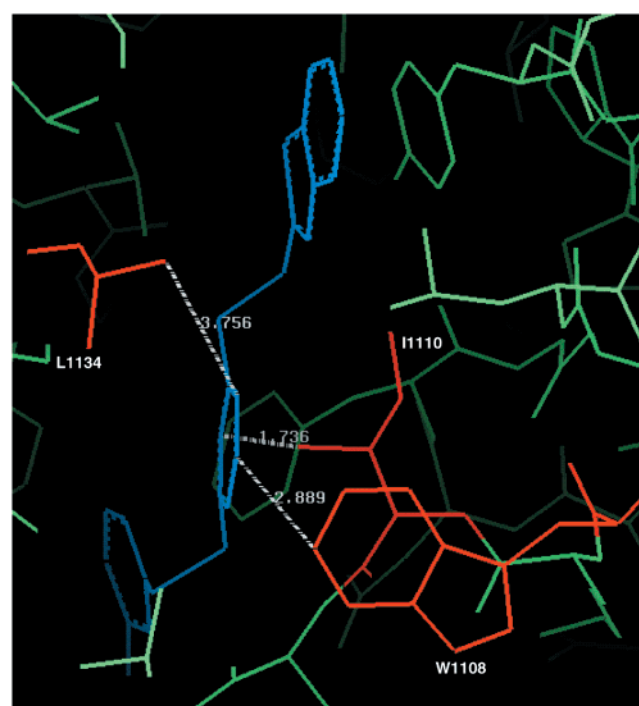
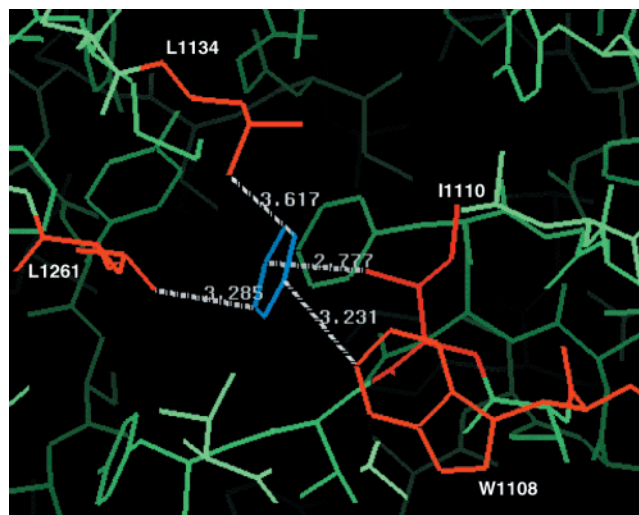


Figure 12. Close contacts indicated by dashed white lines (a) between a benzene minimum calculated for the L396 protein structure in blue and residues W1108, I1110, L1134, and L1261 of P1/Mahoney in red and (b) between the central ring of L396 in blue and residues W1108, I1110, and L1134 of P1/Mahoney in red. Main chain protein atoms were superimposed.

the known “drug binding pocket” of the capsid. It is likely that the library compounds will also bind to other picornaviruses and similarly inhibit viral replication (L383 has been shown to be active against rhinoviruses 3 and 14, and L379 against rhinovirus 14¹³).

The present study represents the first published crystallographic validation of the MCSS-based methodology for the design of effective libraries of small molecule ligands. It also illustrates a case in which very small changes in protein structure can lead to subtle differences of the bound structure relative to the prediction. However, in the present case, this does not matter because the focused combinatorial library generates the necessary ligands. In future applications of the computational methodology, it may be desirable to include some degree of flexibility for the protein,²⁶ as well as a solvation correction to

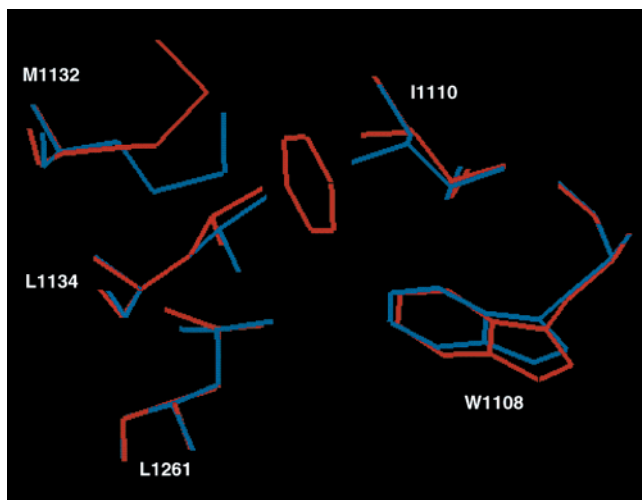


Figure 13. Displacement of flexible protein residues during MCSS minimization of benzene. W1108, I1110, M1132, L1134, and L1261 are shown in blue for the native P1/Mahoney structure and in red after the MCSS benzene search.

rank the final results. With the current implementation, it is possible to allow key active site residues or active site hydrogens to be flexible, for example. Within a very local region of the binding site, such as one of the three partitions of the main binding pocket in poliovirus, only the lowest energy minima for each group were examined. A solvation correction could be added as a postprocessing step to aid in ranking the minima (e.g., see ref 27).

(26) Stultz, C. M.; Karplus, M. MCSS functionality maps for a flexible protein. *Proteins* **1999**, *37*, 512–529.

(27) Caffisch, A. Computational combinatorial ligand design: Application to human alpha-thrombin. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 372–396.

The advantage of designing ligand libraries, instead of individual compounds, is that the scoring function for selecting ligands does not need to be highly accurate. That is, the selected monomers need to be predicted to bind well and in roughly the correct positions (relative to one another) to be connected by the intended chemistry, but then, the exact molecules formed (once the chemistry is carried out) are tested experimentally. The computational methods, such as MCSS, when used in conjunction with combinatorial chemistry, require only that enough information be provided to guide (focus) the library design. In concluding, we stress that combining MCSS or related methods for combinatorial ligand design with chemical combinatorial libraries is a powerful approach for obtaining lead compounds. It makes “focusing” of the chemical library possible without requiring highly accurate scoring of the computational results.

Acknowledgment. This work is supported in part by grants from the National Science Foundation (M.K.) and the National Institutes of Health (J.M.H.).

Note Added after ASAP Posting

The values in rows 3 and 4 of Table 2 were incorrect in the version released ASAP on 11/28/01. The correct values have been input in this version.

JA003972F

(28) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D. M. K. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(29) Halgren, T. A. Merck Molecular Force Field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.