JCTC Journal of Chemical Theory and Computation

Bound Ligand Conformer Revealed by Flexible Structure Alignment in Absence of Crystal Structures: Indirect Drug Design Probed for HIV-1 Protease Inhibitors

Alok Juneja,[†] Henning Riedesel,[†] Milan Hodoscek,[‡] and E. W. Knapp*,[†]

Institute of Chemistry & Biochemistry, Freie Universität Berlin, Fabeckstr. 36a, D-14195 Berlin, Germany, and National Institute of Chemistry, Hajdrihova 19, SI-1001 Ljubljana, Slovenia

Received November 11, 2008

Abstract: In the absence of structural knowledge on the target protein, the bound ligand conformer (BLC) can be constructed approximately by an indirect drug-design approach that uses a set of ligands binding to the same target. Once the bound ligand conformer (BLC) is known, different strategies of drug design can be pursued. The indirect drug-design approach of the present study is based on the assumption that a set of ligands with chemically different architecture binding to the same target protein carry hidden information of their corresponding true BLCs. It is shown how this information can be extracted by pairwise flexible structure alignment (FSA) using molecular dynamics (MD) simulations with attractive intermolecular interactions that derive from the molecular similarity of the ligands and allow the ligands to adopt the same space. The FSA approach is performed with a newly designed module overlap in the experimental CHARMM-29a1, which soon will become publicly available. Combining the conformations obtained from FSA of different ligand pairs yields consensus ligand conformers (CLCs) that should be similar to the BLCs. This procedure was validated on HIV-1 protease (HIV-P), where at present 44 crystal structures with bound ligands of sufficiently diverse chemical composition are available. The FSA approach identifies four different clusters of HIV-P BLCs. These clusters are consistent with the H-bond patterns of the ligands bound to HIV-P in the crystal structures exhibiting four different binding modes. The cluster-specific CLCs are indeed very similar (rmsd $\approx$ 2 Å) to the corresponding BLCs from the crystal structure, demonstrating the feasibility of the present approach.

Introduction

Fighting diseases more effectively requires an increased effort on the design of new drugs. Rational drug design, a major tool in the development of new drugs, is a challenging task for structural, chemical, and computational biology. Most of the known theoretical approaches to drug design are structure-based direct methods, such as docking,[1−6] which employ knowledge of the target structure and the binding pocket. When the target structure is unknown, purely ligand-based indirect drug design is still possible in case several ligands of chemically different architecture are known to bind to the same target.[6−11] One indirect drug design approach is the quantitative structure−activity relationship (QSAR).[12−17] QSAR relates variations in drug activity to changes in chemical composition, characterizing the drugs by suitable molecular descriptors.[18−20] Specific descriptors characteristic for active molecules, the so-called pharmacophores,[21−24] are used in virtual screening studies to classify drugability of molecules from a given database.[25−31] A variant is 3D-QSAR, which employs information of the spatial arrangement of the ligand components[32−35] and may also use

* Corresponding author phone: 0049-30-838-54387; fax: 0049-30-838-56921; e-mail: knapp@chemie.fu-berlin.de.

[†] Freie Universität Berlin.

[‡] National Institute of Chemistry.

10.1021/ct8004886 CCC: $40.75
Published on Web 02/23/2009

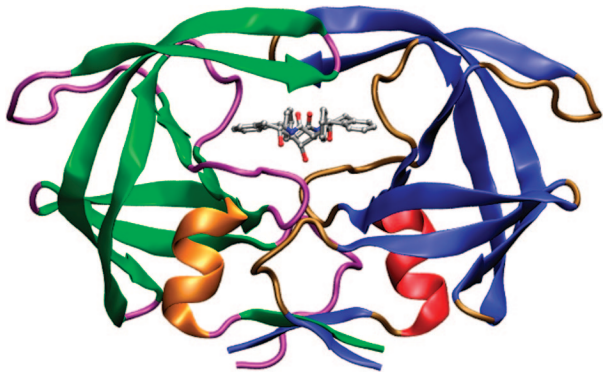


Figure 1. Crystal structure of HIV-1 protease[55] with the HIV-P ligand L10 (see Table 1) as a ball and stick model. The homodimer is represented as rubber band model possessing a *C*2 rotational symmetry with the rotation axis oriented vertically in the drawing plane. The binding pocket extends over both monomers.

information of the binding pocket.[33,36] However, in the latter case the QSAR approach is no longer indirect. Another 3D-QSAR approach is comparative molecular field analysis (CoMFA),[33,34,37,38] which in addition to steric aspects is also using electrostatic information of the ligands. More recently, in the comparative molecular similarity indices analysis (CoMSIA) steric and electrostatic aspects of ligands were complemented by H-bond and hydrophobic patterns.[38−41] However, in contrast to the present approach, these methods generally do not yield coordinates of the target bound ligand conformer (BLC).

Here, we report an indirect drug design approach, which yields approximate BLC coordinates. We strictly use information of the chemical architecture and composition of ligands that bind to the same target (protein binding pocket). Hence, we employ an approach of indirect drug design. Structures cocrystallized with ligands are only used after the study is completed to compare with the true BLC. In this study, we exploit the conformational variability of a set of ligands that bind to the same target protein to gain information on the BLC. Flexible structure alignment (FSA) of these ligands leads to consensus ligand conformers (CLC), which need to be compatible with the diverse chemical architectures of the ligands. Since these ligands bind in the same pocket of the target protein, it is expected that they assume a common geometry in the binding pocket (the true BLC) which should be reflected by the CLCs obtained with FSA. For a sufficiently large set of ligands with different chemical architectures, the CLCs obtained by FSA should be similar to the BLCs. Consequently, one can employ an ensemble of these CLCs as an integral part of the pharmacophore to design new drugs or to perform virtual screening on molecular libraries.[25−31] In earlier work we used a similar FSA approach to align two proteins with equivalent function.[42]

The purpose of the present study is to test this FSA approach by considering a protein−ligand model system for which a large number of BLCs are known. Presently, one of the best studied protein−ligand systems is HIV-1 protease (HIV-P) for which crystal structures with 44 chemically diverse bound ligands are available (see Table 1). The human immunodeficiency virus (HIV) is derived from immature polyproteins that contain HIV-P.[43] During viral replication, HIV-P cleaves the peptide bonds of the Gag and Pol polyproteins yielding the appropriately sized active protein components of the mature virus. HIV-P is proteolytically active as a homodimer possessing *C*2 rotational symmetry (Figure 1)[44,45] with the binding cavity extending symmetrically over both subunits of the homodimer.

The FSA approach in this study is performed with a newly designed module overlap in the modified CHARMM-29a1, which is not yet publicly available. It employs MD simulations of ligand pairs using a modified energy function. This energy function employs conventional intramolecular interactions, but instead of conventional intermolecular interactions it uses energy terms based on molecular similarity allowing molecular groups from different ligands to attract each other in proportion to their similarities and to adopt the same space. The measure of similarity is based on three features: (i) volume/shape, (ii) charge, and (iii) electrostatic potential. Although the latter two features both refer to electrostatics, they differ in a practical sense, since charges focus on short-range while the electrostatic potential describes also long-range aspects of electrostatic interactions. With this similarity-based energy function, MD simulations of a ligand pair with subsequent energy minimization yield superimposed structures of ligand pairs in low energy CLCs, which can be compared with the BLCs of the crystal structures.

In short, our working hypothesis is that in a pairwise FSA ligands that bind to the same target protein but have different intrinsic flexibility adopt CLCs close to the BLCs of the crystal structures. Hence, where the true BLC is not available, we can use the ensemble of these CLCs as integral part of pharmacophores to find new drugs for specific targets by virtual screening.[25−31]

Methods

Abbreviations. FSA, flexible structure alignment; RSA, rigid structure alignment; HIV-P, HIV-1 protease; MD, molecular dynamics; rmsd, root-mean-square deviation; BLC, bound ligand conformer; CLC, consensus ligand conformer.

Overview. In parts A1 and A2 we describe how molecules of completely different chemical composition can be characterized and their similarity measured. In part B we introduce the set of test data for HIV-P, for which the structures of 44 cocrystallized ligands are available. In part C we describe the general methodology and the modified energy function which is the basis of the FSA approach. In part D the procedure for structure alignment is described. It comprises preparation of ligand structures for MD simulation (part D1), performance of the FSA (part D2), and generation of the CLCs (part D3). Technical details are given respectively in parts S2, S5, and S9 of the Supporting Information.

A1: Estimating Molecular Similarity. The concept of molecular similarity, originally introduced by Carbo to compare charge densities $\rho_a(\vec{r})$ and $\rho_b(\vec{r})$ of molecules a and b, respectively, and known as the Carbo Index (CI),[46,47] is defined by

$$\mathrm{CI(a,b)} = \frac{\int \rho_a(\vec{r})\rho_b(\vec{r})\mathrm{d}\vec{r}}{\left(\int \rho_a^2(\vec{r})\mathrm{d}\vec{r}\right)^{1/2}\left(\int \rho_b^2(\vec{r})\mathrm{d}\vec{r}\right)^{1/2}} \quad (1)$$

If charge densities of molecules a and b are identical in shape but differ in magnitude, then $\rho_a(\vec{r}) = c\rho_b(\vec{r})$, where c being an arbitrary positive constant, yields CI(a,b) = 1. Hence, CI quantifies similarities in shape but not in magnitude. An alternative definition of molecular similarity, the Hodgkin Index (HI)

$$\mathrm{HI(a,b)} = \frac{2\int \rho_a(\vec{r})\rho_b(\vec{r})\mathrm{d}\vec{r}}{\int \rho_a^2(\vec{r})\mathrm{d}\vec{r} + \int \rho_b^2(\vec{r})\mathrm{d}\vec{r}} \quad (2)$$

was proposed by Hodgkin and Richard.[47−49] In contrast to the CI, the HI is sensitive to both shape and magnitude of the density distributions ρ_a and ρ_b. Both similarity indices vary between +1 and −1, extremes which correspond to identity and complementarity, respectively. The values of the similarity indices CI(a,b) and HI(a,b) are identical if $\int \rho_a^2 \mathrm{d}\vec{r} = \int \rho_b^2 \mathrm{d}\vec{r}$ but generally |CI(a,b)| > |HI(a,b)|.

A2: Features Characterizing Molecules. We use a combination of three different features to characterize the structure of molecules. These are shape and volume, distribution of atomic partial charges, and electrostatic potential. The latter two features are related, since they both are based on electrostatics, but charges focus on the short-range aspect of electrostatics only, while the electrostatic potential also considers long-range contributions.

To describe shape and volume of a molecule, we represent its atoms (m) by three-dimensional Gaussian distribution functions $g_m(\vec{r})$[50,51]

$$g_m(\vec{r}) = \frac{\sqrt{V_m}}{\sigma_m^{3/2}\pi^{3/4}} \exp\left(\frac{-(\vec{r} - \vec{r}_m)^2}{2\sigma_m^2}\right) \quad (3)$$

where V_m is the van der Waals (vdW) volume of atom m, $\vec{r}_m$ is the position of the center of atom m, σ_m is the width of the Gaussian accounting approximately for the atomic radius $R_m \approx \sigma_m$. We normalized $g_m(\vec{r})$ to yield for the self-overlap the volume of atom m, i.e., $\int g_m^2(\vec{r})\mathrm{d}\vec{r} = V_m$. The volume common to atoms m and n is given by the overlap integral

$$S_{mn} = \int (g_m(\vec{r})g_n(\vec{r}))\mathrm{d}\vec{r} = \sqrt{V_m V_n}\left(\frac{2\sigma_m\sigma_n}{\sigma_m^2 + \sigma_n^2}\right)^{3/2} \exp\left(\frac{-(\vec{r}_m - \vec{r}_n)^2}{2(\sigma_m^2 + \sigma_n^2)}\right) \quad (4)$$

A molecule is represented as the sum of $g_m(\vec{r})$ corresponding to its individual atoms. The similarity of a pair of molecules can then be calculated as the overlap integral of the molecular distribution functions. The approach to estimate molecular similarity based on atom-centered Gaussians is not restricted to volume and shape similarity[52] but can be applied to any atom based molecular property. Accordingly, a molecular distribution function referring to property P of molecule can be defined as the sum of $g_m^{(P)}(\vec{r})$ referring to the corresponding atomic property P. Hence, for property P (vol, charge, epot for volume/shape, charge, electrostatic potential, respectively) a molecule a, which consists of n_a atoms, is described by the distribution function

$$\rho_a^{(P)}(\vec{r}) = \sum_{m=1}^{n_a} g_m^{(P)}(\vec{r}) \quad (5)$$

Estimation of the similarly between two molecules a and b with respect to property P requires computation of the overlap integral of the two corresponding molecular distribution functions $\rho_a^{(P)}(\vec{r})$ and $\rho_b^{(P)}(\vec{r})$

$$S^{(P)}(\mathrm{a,b}) = \int \rho_a^{(P)}(\vec{r})\rho_b^{(P)}(\vec{r})\mathrm{d}\vec{r} = \sum_m^{n_a}\sum_n^{n_b} S_{mn}^{(P)} \quad (6)$$

Hence, the atom-based form of HI that measures the similarity with respect to property P for a pair of molecules (a, b) is given by

$$\mathrm{HI}^{(P)}(\mathrm{a,b}) = \frac{2S^{(P)}(\mathrm{a,b})}{S^{(P)}(\mathrm{a,a}) + S^{(P)}(\mathrm{b,b})} \quad (7)$$

where $S^{(P)}(a,b)$ is the overlap integral for the molecule pair (a, b) and $S^{(P)}$(a,a) and $S^{(P)}$(b,b) are the self-overlaps of molecules a and b, respectively.

The result of volume overlap of two atoms n and m yields, in analogy to eq 4

$$S_{mn}^{(\mathrm{vol})} = w_m^{(\mathrm{vol})} w_n^{(\mathrm{vol})}\left(\frac{2R_m^{(\mathrm{vol})}R_n^{(\mathrm{vol})}}{(R_m^{(\mathrm{vol})})^2 + (R_n^{(\mathrm{vol})})^2}\right)^{3/2} \exp\left(\frac{-(\vec{r}_m - \vec{r}_n)^2}{2(kR_m^{(\mathrm{vol})})^2 + 2(kR_n^{(\mathrm{vol})})^2}\right) \quad (8)$$

where the weighting factors $w_n^{(\mathrm{vol})}$ replace the volume factors in eq. (4), $R_n^{(\mathrm{vol})}$ is the effective atomic radius for the volume of atom n and k is a factor that scales all atomic radii (i.e., k determines the degree of localization of the Gaussians). The default values of k and $w_n^{(\mathrm{vol})}$ are unity. Likewise, for charge overlap between atoms m and n with atomic partial charges q_m and q_n (given in units of the elementary charge) and $R_n^{(\mathrm{charge})}$, the effective atomic radii for the charges, we have

$$S_{mn}^{(\mathrm{charge})} = q_m q_n\left(\frac{2R_m^{(\mathrm{charge})}R_n^{(\mathrm{charge})}}{(R_m^{(\mathrm{charge})})^2 + (R_n^{(\mathrm{charge})})^2}\right)^{3/2} \exp\left(\frac{-(\vec{r}_m - \vec{r}_n)^2}{2(kR_m^{(\mathrm{charge})})^2 + 2(kR_n^{(\mathrm{charge})})^2}\right) \quad (9)$$

where the charges of the atom pairs enter as first power instead of square root to enhance the similarity measure. In the present application we use $R_n^{(\mathrm{charge})} = R_n^{(\mathrm{vol})}$.

As similarity measure for the electrostatic potentials of two charges q_m and q_n at positions $\vec{r}_m$and $\vec{r}_n$, respectively, we use

$$S_{mn}^{(\mathrm{epot})} = q_m q_n \exp(-\gamma^{(\mathrm{epot})}|\vec{r}_m - \vec{r}_n|) \quad (10)$$

where $\gamma^{(\mathrm{epot})}$ is a range parameter governing the length scale over which the electrostatic potentials of atoms m and n are compared (a large $\gamma^{(\mathrm{epot})}$ corresponds to short distances, the default value is unity). A derivation of the expression for the electrostatic potential overlap, eq 10, is given in the part

Table 1. HIV-1 Protease (HIV-P) with the 44 Ligands Considered in This Study

no.	ligand (inhibitor/substrate)[a]	PDB id	resol (Å)
L1	TMC114 or UIC-94017 (017)	1T3R	1.20
L2	AQ148 (ARQ)	3AID	2.50
L3	ABT-378 (AB1)	1MUI	2.80
L4	SB203386 (IM1)	1SBG	2.30
L5	BOC-PHM-TYR-ILE-GLY	1MTR	1.75
L6	BMS-182193	1ODW	2.10
L7	L-739,622 (3IN)	2BPZ	2.50
L8	L-738,317 (1IN)	2BPW	2.80
L9	L-735,524 (MK1)	2BPX	2.80
L10	L-700,417 (VAC)	4PHV	2.10
L11	KNI-272 (KNI)	1HPX	2.00
L12	SB203238 (GAN)	1HBV	2.30
L13	AHA455 (A1A)	2BQV	2.10
L14	UCSF8 (THK)	2AID	1.90
L15	AHA006 (NMB)	1AJV	2.00
L16	U100313 (U02)	2UPJ	3.00
L17	TPV	1D4S	2.50
L18	AKC (AKC4P_133A)	2BB9	1.35
L19	HOE/BAY 793 (BAY)	1VIJ	2.40
L20	TS-126 (IPF)	2A1E	1.30
L21	A-74704	9HVP	2.80
L22	A-84538 (RIT)	1HXW	1.80
L23	A79285 (A85)	1DIF	1.70
L24	CGP 53820 (C20)	1HIH	2.20
L25	GR137615 (G37)	1HTG	2.00
L26	ROC	1FB7	2.60
L27	KI2-PHE-GLU-GLU- NH2	1NH0	1.03
L28	SD146 (146)	1QBT	2.10
L29	MSA367 (MS3)	1EC3	1.80
L30	p1-p6 substrate	1KJF	2.00
L31	MA-CA substrate	1KJ4	2.90
L32	CA-p2 substrate	1F7A	2.00
L33	p2-NC substrate	1KJ7	2.00
L34	RT-RH substrate	1KJG	2.00
L35	U89360E (U0E)	1AXA	2.00
L36	MVT-101	4HVP	2.30
L37	JG-365	7HVP	2.40
L38	U-85548E	8HVP	2.50
L39	LP-130 (LP1)	1ODY	2.00
L40	acetyl-pepstatin	5HVP	2.00
L41	CH_2-CBG-Asn-Tyr-CH_2-Pro-ILE-Val-NH_2	1CPI	2.05
L42	SKF108738 (HEF)	1HEF	2.20
L43	U75875	1HIV	2.00
L44	SDZ283-910	1A8G	2.50

[a] Five ligands (L30−L34) are natural substrate oligopeptides. All other ligands listed in the table are inhibitors.

S1 of Supporting Information. Although charge distribution and electrostatic potential are derived from the same set of atomic partial charges, they focus on different aspects of these charges. The charge distribution considers the similarity at short distances accounting for direct interactions of the ligands with atoms in the binding pocket like hydrogen bonds and salt bridges, while the electrostatic potential considers also the long-range aspect of the charge distribution.

A similarity measure considering contributions from all three properties (volume/shape, charge, and electrostatic potential) is the weighted sum of the individual HI yielding the combined HI

$$\mathrm{HI}_{\mathrm{combined}} = \frac{w_{\mathrm{vol}}}{w_{\mathrm{sum}}}\mathrm{HI}_{\mathrm{vol}} + \frac{w_{\mathrm{charge}}}{w_{\mathrm{sum}}}\mathrm{HI}_{\mathrm{charge}} + \frac{w_{\mathrm{epot}}}{w_{\mathrm{sum}}}\mathrm{HI}_{\mathrm{epot}} \quad (11)$$

where $w_{\mathrm{sum}} = w_{\mathrm{vol}} + w_{\mathrm{charge}} + w_{\mathrm{epot}}$; $\mathrm{HI}_{\mathrm{vol}}$, $\mathrm{HI}_{\mathrm{charge}}$, and $\mathrm{HI}_{\mathrm{epot}}$ are the individual Hodgkin indices and w_{vol}, w_{charge}, w_{epot} (default values are unity) are the relative weights for volume/shape, charge, and electrostatic potential overlap, respectively. The resulting similarity index $\mathrm{HI}_{\mathrm{combined}}$ can vary formally between −1 and +1, but in practice it will not come close to −1. As the value of $\mathrm{HI}_{\mathrm{combined}}$ approaches unity the similarity of the two molecules increases.

B: HIV-1 Protease Ligand Test Set. To test whether FSA of ligands with the same function can provide information on the ligand conformers in the binding pocket of the target protein, we consider HIV-1 protease (HIV-P), for which X-ray structures of 44 different HIV-P ligand complexes are available (listed in Table 1). Of these 39 are HIV-P inhibitors (less flexible than peptides), and the remainder (L30−L34) are natural substrate oligopeptides[53] that are cleaved by HIV-P during viral replication. Although the HIV-P binding cavity possesses *C*2 rotational symmetry (Figure 1),[44,45] only one inhibitor (L19) exhibits the same symmetry while five of the inhibitors (L6, L10, L15, L21, L28) possess mirror symmetry. All the remaining HIV-P ligands, including the natural substrate oligopeptides, possess no symmetry. These inhibitor substrates, cocrystallized with HIV-P, are available in the protein data bank (PDB).[54]

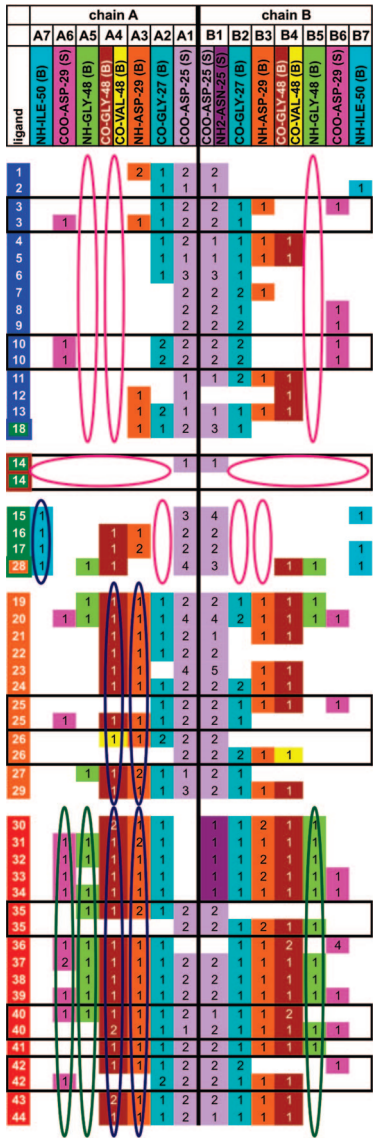
Because of the *C*2 symmetry of the HIV-P binding pocket, several ligands (L3, L10, L14, L25, L26, L35, L40, L42) appear in the crystal structures in two opposite orientations with slight variations in the conformation. For the analysis of the H-bond pattern of the ligand with the protein, we used both crystal structures (see Table 2, discussed in part F of Results and Discussion). For analyses of HIV-P ligand coordinates, we used only the first of the two alternative ligand orientations listed in the PDB data file. This is justified, since the rmsd between alternate conformers of the same HIV-P ligand overlaid with the Kabsch algorithm[56,57] are generally small (see Table SIV).

C: Maximizing the Similarity of Molecules. In the search for CLCs of pairs of chemically disparate molecules, we monitored the similarity measured by $\mathrm{HI}_{\mathrm{combined}}$, eq 11. As the conformers of the pair of molecules are being adjusted to maximize their similarity, each of the two molecules must at the same time assume a reasonable low-energy structure. To achieve a compromise between these two criteria, we constructed an effective potential energy function

$$E(\mathrm{a,b}) = E_{\mathrm{intra}}(\mathrm{a}) + E_{\mathrm{intra}}(\mathrm{b}) - e_{\mathrm{S}}\mathrm{HI}_{\mathrm{combined}}(\mathrm{a,b}) \quad (12)$$

that simultaneously accounts for both. The intramolecular contributions to E [E_{intra}(a), E_{intra}(b)] are taken to be the true physical ones (i.e., bonded interactions depending on bond length, bond angle, torsions; and nonbonded interactions depending on van der Waals and Coulomb terms within the same molecule). The true intermolecular contribution is replaced by a nonphysical term proportional to the combined Hodgkin index, where the parameter e_{S} is set so that all three terms in eq 12 are of comparable magnitude. Different values of e_{s} are used during the FSA approach as given in part S2, S5, and S9 of Supporting Information. The gradient of this effective potential energy function can be used for MD simulations, which will yield CLCs of the molecule pair under consideration. In these combined CLCs both molecules are structurally aligned in a low energy conformation such that similar parts of the two molecules are superimposed.

Table 2. HIV-P H-Bond Pattern



Column 1: ligand number; Row 1: polypeptide chain A or B of HIV-P; Row 2: formal numbering of the H-bond partner in HIV-P (an H-bond is formed if the participating non-hydrogen atoms are closer than 3.5 Å); Row 3: specification of H-bond partners in format (atom_name)-(amino_acid_type)-(amino_acid_number)-(backbone(B)/side_chain(S)). For mutant HIV-Ps, the residues 1 (Asn replaces Asp) and 4 (Val replaces Gly) possess different side chains. From these, B1, A4, and B4 are H-bonds with some HIV-P ligands as indicated. Each column has a color code referring to the H-bond partner in HIV-P. Each line characterizes the H-bond pattern for a specific HIV-P ligand labeled in column 1. For the ligands (L3, L10, L14, L25, L26, L35, L40, L42), alternate conformations were found in the HIV-P crystal structures. For these ligands, both H-bond patterns are given in subsequent lines. These cases are highlighted by black boxes as for instance for ligand L3. A filled cell denotes an H-bond of the ligand with the particular group of HIV-P; the digits (1 or 2) count the number of H-bonds formed between the ligand and this group. Empty cells denote absence of possible H-bonds. H-bond patterns persistent for a group of ligands are highlighted by ellipses. Red (blue) ellipses denote strict absence (presence) of specific H-bonds. Green ellipses indicate dominant occurrences of a particular H-bond. The five substrate HIV-P ligands (L30−L34) belong to the red cluster. The clustering of the HIV-P ligands is based on the similarity of the H-bond pattern and indicated by the color code in column 1. A detailed description of H-bond pattern is given in Table SV.

D: Computational Procedures.

D1: Model Preparation of Ligands and Generation of Initial Conformers. To avoid the bias of crystal structure information of the ligand conformers used in this study rigorously, we were choosing the initial conformers for the MD simulations to differ as much as possible from the conformers they adopt in the HIV-P binding pocket. The average rmsd of the initial conformers relative to the BLCs is 4.36 Å. More details are given in part S2 and Figure S1 and S2 of the Supporting Information.

D2: Pairwise Flexible Structure Alignment (FSA). The alignment of a pair of flexible molecules is performed by MD simulation using the effective potential energy function, eq 12. Seven hundred twenty different starting arrangements of the ligand pairs were used. The most similar ligand pair conformer is extracted from the final 80 ps of 105 ps trajectory runs at room temperature and used for further analysis, as detailed in part S5 of the Supporting Information. The CPU time for single 105 ps trajectory is 85 min on AMD Athlon 2.2 GHz.

D3: Generating Consensus Ligand Conformers (CLCs). Since the pairwise FSA of the 44 HIV-P ligands yields for each ligand 43 different conformers, it is necessary to determine consensus ligand conformers (CLCs) from them. Alternatively we could have used multiple FSA considering all HIV-P ligands in a single alignment attempt and thus obtained CLCs directly. However, in such an approach one would assume that all 44 HIV-P ligands obey the same binding mode and give equivalent conformers. But, as a result of this study we can show that the 44 HIV-P ligands actually exhibit four different binding modes. Hence, the direct use of a multiple FSA approach would obscure this information and yield less than optimum results. Pairwise FSA keeps information on different binding mode and allows one to cluster the ligands according to mutual similarity of their conformers. After this analysis, CLCs are computed for each of the similarity clusters, as explained in part S9 of the Supporting Information.

Results and Discussions

Overview. Before using the concept of FSA to generate conformers of HIV-P ligands that can be related to the corresponding bound ligand conformers (BLCs) fitting in the HIV-P binding pocket, we carried out a number of studies for each of the 44 HIV-P ligands independently. In part A below, we discuss (i) how much ligand conformers can deviate from the corresponding BLCs, (ii) how close ligand conformers can come by chance to their corresponding BLCs through equilibrium fluctuations during a short MD simulation at room temperature, and (iii) how much the ligand conformers of such an MD simulation can deviate on average from the corresponding BLCs. In part B, we analyze the similarity of ligand conformers adopted in the HIV-P binding pocket of the crystal structure. As a typical result we compared reference ligand L1 with all other ligands L*x* in the main text. The comparison of the other ligand pairs is found in Figure S4 of the Supporting Information. In part C, we examine in detail the conformers of reference ligand

L1 resulting from FSA with all other HIV-P ligands and compared the computed ligand conformers with the conformer adopted by ligand L1 in the crystal structure. Part D provides an overview of the pairwise FSA results considering all combinations of ligand pairs, and part E contains an analysis of the degrees of similarities between the different pairs of HIV-P ligands computed by the FSA approach. These results are clustered to assign the HIV-P ligands to different classes (part E) using solely the FSA data without consideration of the crystal-structure information. The binding modes of the HIV-P ligand conformers in the crystal structures based on the H-bond pattern were analyzed in part F and compared with the results of the cluster analysis based on the FSA results. In part G we use the results of ligand classifications based on the FSA to construct consensus ligand conformers (CLC) for each ligand cluster separately, as described in part D3 of Methods. Finally, in part H we visualize the CLCs of the four obtained HIV-P ligand clusters.

A: Comparing HIV-P Ligand Conformers Obtained by MD Simulations with BLCs from Crystal Structures. First, we examined how close and how distant HIV-P ligand conformers may come by chance to the corresponding crystal structure conformers if conventional MD simulations are performed for each ligand separately in vacuo using the CHARMM energy function. For this purpose we performed MD simulations of 225 ps (including 5 ps for heating) at constant energy ($T \approx 300$ K), employing only the last 100 ps of the trajectory to allow for sufficient equilibration. In Figure 2a the RMSDs of these ligand conformers relative to the crystal structures are displayed. The RMSDs of ligands 1 to 18 (displayed in the left half of Figure 2a) are smaller than that of 19 to 44. This correlates with the size of the ligands (see Table SII of the Supporting Information), since larger ligands generally lead to larger RMSDs. For the energy-minimized starting structures of the MD simulation, the conformers with the largest deviation from the crystal structures were selected from 100 randomly generated structures (see part D1 in Methods). The RMSDs of these selected conformers are in the range 4−5.5 Å, larger than the time-averaged RMSDs (about 4 Å) and the minimum RMSDs (about 3 Å) occurring during the MD simulation runs. This range of rmsd values clearly shows that the HIV-P ligands spontaneously adopt conformers close to the conformer in the HIV-P binding pocket only rarely. Hence, MD simulations of HIV-P ligands performed independently from information of the HIV-P binding pocket generally do not yield conformers close to the BLCs. The actual variation in ligand conformers between crystal structure, average structure of an MD simulation of the individual ligand, and the CLC from the FSA approach as described later in detail in connection with Figure 7 is displayed for ligand L1 in Figure 2b. Analogue figures for all other 43 HIV-P ligands are shown in Figure S9 of the Supporting Information.

B: Analysis of the Similarity of HIV-P Ligand Conformers in the Crystal Structure. For the FSA method to be successful, there must be significant similarities among the BLCs from the 44 different HIV-P ligands. To meet this condition, we first performed pairwise rigid structure alignments (RSA) of the BLCs. This was done in two different ways: (i) Protein structure-based RSA, in which the ligands were ignored and HIV-Ps from the 44 crystal structures were aligned pairwise by the Kabsch algorithm.[56,57] The resulting translational and rotational transforms were then used to superimpose the ligands. For these RSA, the combined

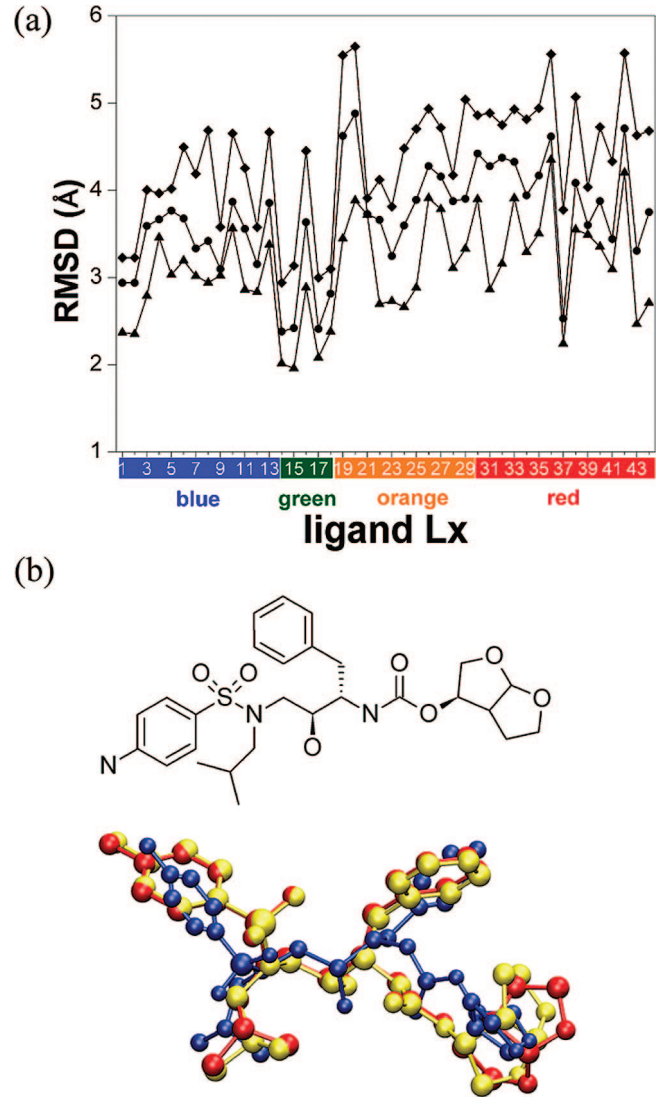


Figure 2. Part a: HIV-P ligand conformers obtained by MD simulations for all 44 HIV-P ligands considered using the last 100 ps of a 225 ps trajectories in vacuo at constant energy ($T \approx 300$ K). The RMSDs of the ligand conformers are given relative to the crystal structures. They refer to the starting conformers of the MD simulations (top line, which coincides with middle line of Figure S2), the conformers with minimum deviation to the crystal structures obtained by MD simulation (bottom line), and the average conformers (middle line). The ligand numbers L*x* at the abscissa are defined in Table 1; their background colors refer to the four clusters of ligands classified according to structure similarity as discussed later in connection with Figure 6. Part b: Chemical architecture of HIV-P ligand L1 (top) and three-dimensional structures (bottom) of the bound ligand conformer from the crystal structure (yellow), average structure from MD simulation of ligand L1 alone (blue) with 2.95 Å rmsd from crystal structure, and consensus ligand conformer (CLC) after FSA (red) as described in connection with Figure 7 with 1.65 Å rmsd.

Hodgkin indices $HI_{combined}$ were generally between 0.4 and 0.5 (red circles in Figure 3) for all HIV-P ligand pairs involving the reference ligand L1 and ligands L2−L44. (ii) Ligand conformer-based RSA, in which the torsion angles of the ligands were fixed by use of stiff torsion potentials (as described in the CHARMM script $Script_{FSA}$ in part S8 of the Supporting Information) to obtain approximately rigid ligand models. With these rigid ligand models, the FSA procedure was applied to all HIV-P ligand pairs involving the reference ligand L1 and all other ligands L2−L44. The resulting $HI_{combined}$ values (green diamonds in Figure 3) (essentially between 0.5 and 0.6) are larger than the values obtained by the protein structure-based RSA. Results of RSA for all other HIV-P ligand pairs, which are qualitatively similar, are given in Figure S4 of the Supporting Information.

C: Pairwise Flexible Structure Alignment of Reference Ligand L1 with all HIV-P Ligands. FSA of HIV-P ligands were performed for all 946 = (44 × 43)/2 possible ligand pairs. The results for 43 ligand pairs involving the reference ligand L1 and all other ligands are displayed in Figure 3. These results are typical of those for all the other combinations of ligand pairs, which are shown in 43 additional graphs in Figure S4 of the Supporting Information. The $HI_{combined}$ for the initial conformers of the ligand pairs used in the MD simulations to perform the FSA are displayed as gold stars in Figure 3. These low $HI_{combined}$ values, generally between 0.3 and 0.4, correspond to ligand conformers that are significantly dissimilar from the BLCs. During the MD simulation the similarity increases, as indicated by $HI_{combined}$ values that vary between 0.5 and 0.65 (blue triangles in Figure 3) and are often considerably larger than those of the starting conformers and also slightly larger than those obtained by ligand conformer-based RSA of the HIV-P ligand conformers of the crystal structures (green diamonds in Figure 3). It is encouraging that the dependence of the $HI_{combined}$ on the ligand L*x* is similar for the conformer pairs (L1, L*x*) obtained by FSA (blue triangles) and from the crystal structures (green diamonds), although for the latter the $HI_{combined}$ are slightly smaller.

The open symbols in Figure 3 refer to comparisons of the L1 conformer from the crystal structure with L1 conformers obtained by FSA of L1 with the 43 other HIV-P ligands. Since different conformers of the same L1 ligand were considered in this comparison, both rmsd (open brown triangles, right scale in Figure 3) and $HI_{combined}$ (open cyan diamonds, left scale in Figure 3) were evaluated. It is noteworthy that the RMSDs relative to the L1 crystal-structure conformer are all close to 2 Å (open brown triangles, right scale), which is about as small as the minimum rmsd (= 2.3 Å) obtained from the last 100 ps of a vacuum MD simulation of L1 at room temperature (see Figure 2a) and is considerably smaller than the rmsd (= 3.9 Å) from the initial conformer used for the FSA approach (black solid line in Figure 3). Parallel with small RMSDs we observed $HI_{combined}$ values generally greater than 0.5 when comparing the L1 conformer from the crystal structure with the L1 conformers obtained by FSA with the 43 other HIV-P ligands (open cyan diamonds). These results, which are typical of all ligand pairs (L*x*, L*y*) (see Figure S4 in the Supporting Information), indicate the success of the FSA approach in generating ligand conformers that are close to the corresponding BLCs.

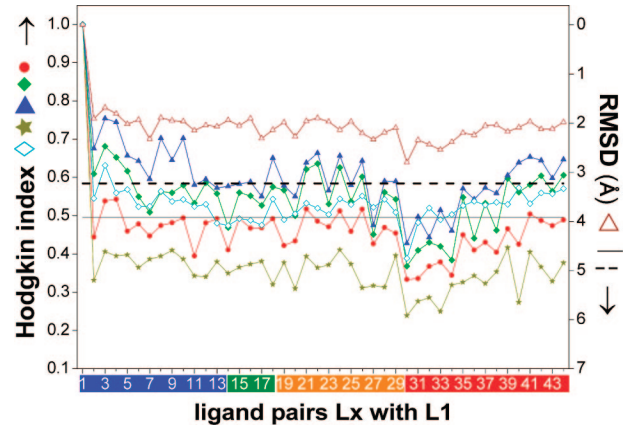


Figure 3. Results of pairwise structure alignment of the HIV-P reference ligand L1 with 43 other HIV-P ligands L*x*, *x* = 2, 3, ... 44, listed on the abscissa. Ligands are defined in Table 1. The left scale refers to $HI_{combined}$, eq 11, the right scale refers to RMSDs (in reverse direction) given relative to the crystal structures. For the sake of completeness, we formally assigned $HI_{combined}$ = 1 and rmsd = 0.0 aligning ligand L1 with itself. $HI_{combined}$ (left scale) is displayed for ligand conformer-based RSA using ligand conformers from the crystal structures (green diamonds), for protein structure-based RSA in which the HIV-P crystal structures were aligned by the Kabsch algorithm[56,57] (red circles), for initial conformers of the HIV-P ligands used for the MD simulation of FSA (gold stars), for FSA results of ligand pairs (L1, L*x*) consisting of the reference L1 and all other ligands L*x* of HIV-P (blue triangles), for L1 ligand conformer in the crystal structure with the L1 conformers obtained by pairwise FSA with all other 43 HIV-P ligands L*x* (open cyan diamonds), and the corresponding rmsd (right scale) (open brown triangles). The rmsd relative to the crystal structure of the L1 conformer used as starting conformer in pairwise FSA (solid black line); rmsd of the same L1 conformer after energy minimization (dashed black line). *N* numbers on the abscissa refer to ligands designated in Table 1; their background colors refer to the four clusters of ligands classified according to structural similarity as discussed later in connection with Figure 6.

Analyzing the dependencies of $HI_{combined}$ in Figure 3 with respect of the different types of clusters labeled by the background color code of the ligand numbers at the abscissa, we observe significant lower similarities of L1 with the red labeled ligands. This is typical for all blue labeled ligands like L1, which are members of the same ligand cluster. As will be seen subsequently (in part E), this deviation can be explained by the large dissimilarity of the ligands belonging to the blue cluster relative to the ligands of the red cluster. The cluster analysis shows that these two ligand clusters are the most dissimilar and adopt different binding modes in the HIV-P binding cavity.

D: Pairwise Alignment Overview of Data from all HIV-P Ligand Pairs. The data for all 946 = 44 × 43/2 pairs of HIV-P ligands yield 44 diagrams like the one of Figure 3, which displays the data of all 43 ligand pairs involving L*y* = L1 as reference ligand. Since these diagrams provide too many details, they are deferred to the Supporting

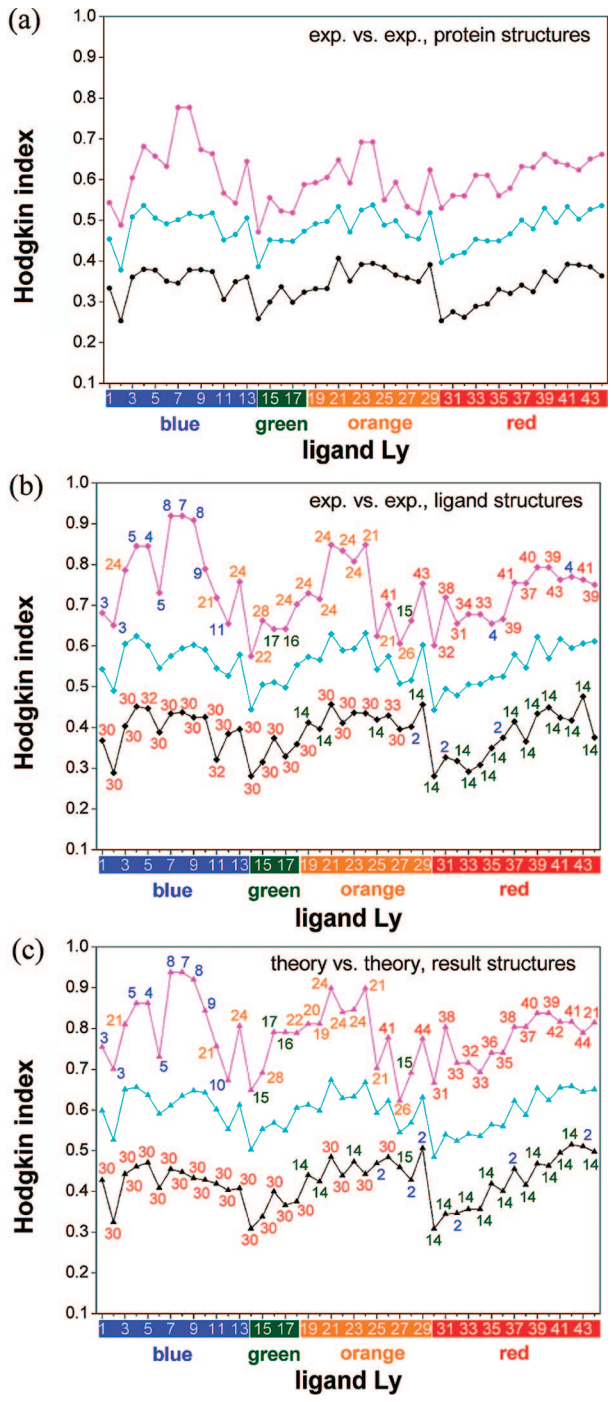


Figure 4. Maxima (magenta), minima (black), and averages (cyan) of $HI_{combined}$, eq 11, from HIV-P ligand pairs formed between specific reference ligands L*y* displayed on the abscissa and other 43 HIV-P ligands. *N* numbers on the abscissa refer to ligands listed in Table 1. (a) $HI_{combined}$ obtained by protein structure-based RSA of HIV-P ligand conformers form the crystal structures (corresponding to red circles in Figure 3 and Figure S4); (b) $HI_{combined}$ obtained by ligand conformer-based RSA of HIV-P ligand conformers from the crystal structures (corresponding to green diamonds in Figure 3 and Figure S4); (c) $HI_{combined}$ obtained by FSA of HIV-P ligand pairs (corresponding to blue triangles in Figure 3 and Figure S4). The numbers at the maxima (magenta) and minima (black) of $HI_{combined}$ in parts b and c refer to the ligand numbers L*x* which, combined with the reference ligand L*y*, yield the maximum and minimum of $HI_{combined}$, respectively.

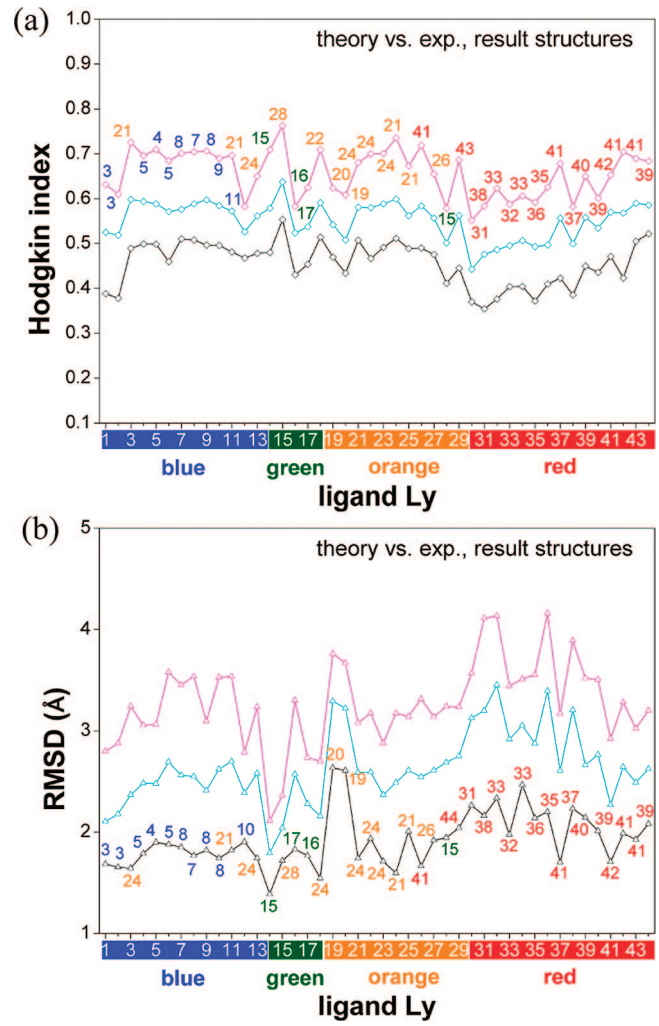


Figure 5. Comparison of ligand conformers obtained by pairwise FSA with conformers from crystal structures. Maxima (magenta), minima (black), and averages (cyan) of $HI_{combined}$ (a) and rmsd (b) for ligand conformers formed between specific reference ligand conformers L*y* (the BLCs) displayed on the abscissa and other 43 HIV-P ligands L*x* (obtained by FSA). *N* numbers at maxima (magenta) of $HI_{combined}$ in part a and minima (black) of rmsd in part b refer to ligand number L*x* which by pairwise FSA yields the conformer of ligand L*y* (marked on the abscissa) most similar to the crystal structure. Note that the data points at L*y* = L1 use the information from the curve (open cyan diamonds) in Figure 3. *N* numbers on the abscissa identify ligands listed in Table 1.

Information (Figure S4). Figures 4 and 5 contain an overview and summary of these results displaying maxima, minima, and averages of $HI_{combined}$ and rmsd for HIV-P ligand pairs formed between a particular reference ligand L*y* (denoted at the abscissa) and the 43 other ligands (L*x*). Parts a and b of Figure 4 show maxima, minima, and averages [referring to ligand pairs (L*y*, L*x*) with the average running over L*x*] of $HI_{combined}$ obtained from the BLC of the HIV-P crystal structures using protein structure-based (Figure 4a) and ligand conformer-based (Figure 4b) RSA (see part B of Results and Discussion).

The maxima of $HI_{combined}$ obtained by RSA, which are around 0.6 and 0.8 for the protein structure-based and ligand conformer-based alignment of the BLC, respec-

tively, indicate a large degree of similarity. On the other hand, the minima of $HI_{combined}$ around 0.3 and 0.4 for protein structure-based and ligand conformer-based alignment, respectively, indicate low similarity. The average of $HI_{combined}$ is typically the mean of the maxima and minima, which suggests that for a given HIV-P reference ligand there are generally ligands with high and with low degrees of similarity. As we will see later, this large variability in HIV-P ligand conformers is a trait of HIV-P ligands that possess different binding modes.

In parts a and b of Figure 4 only experimental data on the similarity of the HIV-P ligand conformers of the crystal structures are analyzed, while in part c only computational data on ligand conformers obtained by FSA are compared. Comparing the dependencies of $HI_{combined}$ on the reference ligand L*y* in parts b and c, one observes a nearly quantitative agreement in particular for the maximal and the average $HI_{combined}$. This suggests a close relationship between the ligand conformers obtained by FSA and the corresponding ligand conformers in the HIV-P crystal structures, although the former did not use any information from the crystal structures. The numbers at the maximal (minimal) $HI_{combined}$ in parts b and c of Figure 4 label the HIV-P ligand L*x* that is most similar (distant) to the reference ligand L*y* on the abscissa. That these numbers for the experimental (Figure 4b) and computational data (Figure 4c) often agree suggests again close similarity between experimental and computational ligand conformers. Furthermore, with few exceptions these numbers possess the same color as the background color of the reference ligand L*y* from the abscissa which shows that the ligand pairs with maximal (minimal) similarity belong to the same (most distant) cluster and therefore possess the same (very different) binding mode in the HIV-P binding pocket.

The ligand pairs with minimum similarity in Figures 4b and 4c are the ligands L30, L2, L14. This agrees with the results of the cluster analysis (Figure 6, see part E), which demonstrates that L30 is most distant from ligands of the blue and green cluster, L2 is most distant from ligands of the orange cluster, and L14 is most distant from ligands of the red cluster. This is a clear demonstration that the FSA approach can indeed be used to generate ligand conformers which are close to the BLCs.

A direct comparison between theory and experiment for the HIV-P ligand conformers from crystal structures and from pairwise FSA is given in Figure 5. Here, $HI_{combined}$ (part a) and rmsd (part b) are displayed as functions of the reference ligand L*y* for which ligand pairs with all other HIV-P ligands were considered to evaluate averages, maxima, and minima. The maxima of $HI_{combined}$ vary roughly between 0.6 and 0.75, and the minima of rmsd are generally below 2.0 Å, which indicates high similarity within subsets of HIV-P ligands. Simultaneously, the minima of $HI_{combined}$ generally vary between 0.4 and 0.5, and the maxima of rmsd between 3.0 Å and 4.0 Å, which demonstrates a low similarity between subsets of HIV-P ligands possessing different binding modes. The $HI_{combined}$ (rmsd) are slightly higher (lower) for the ligands of the first two clusters (on the left side of Figure 5, marked by blue and green background colors) compared to

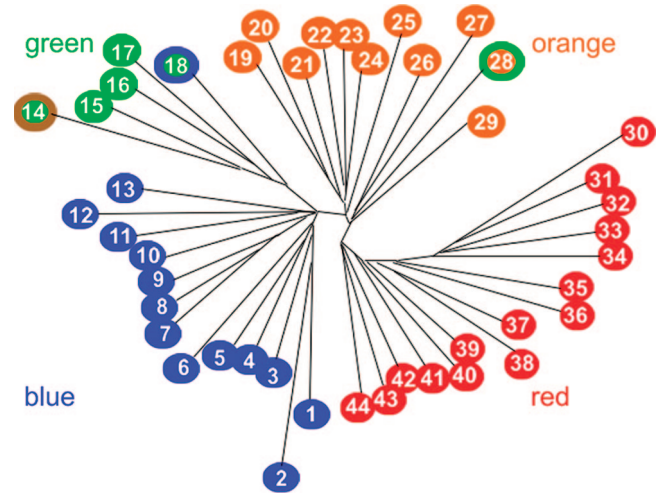


Figure 6. Phylogenetic tree of HIV-P ligands generated by T-REX[59] based on the similarity measure $HI_{combined}$, eq 11, obtained by FSA. Besides the tree structures the closeness of the ligand symbols in the diagram is also a qualitative measure of structure similarity. On the basis of similarity, four clusters can be clearly discriminated. The background color code reflects the results of an H-bond pattern analysis of the crystal structures (Table 2). Small differences in the FSA based cluster analysis appeared for the ligands L14 and L18 that appear to belong to the green cluster and for L28 that belongs to the orange cluster. The five natural substrate oligopeptides (L30−L34) belong to the red cluster. Note that the color code denoting the ligands at the abscissa of Figure 2−5, 7 uses the results of the classification based on FSA as displayed above.

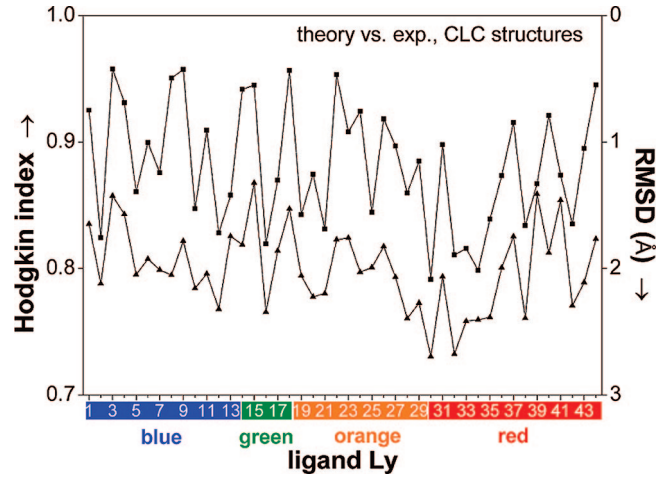


Figure 7. Cluster specific HIV-P ligand CLCs obtained by FSA using the cluster result from Figure 8. Combined Hodgkin indices, $HI_{combined}$, eq 11, (left scale, squares) and RMSDs (right scale, in reverse direction, triangles) of the CLCs displayed relative to the ligand conformers in the crystal structures. CLCs were generated without crystal structure information, as explained in the Method part D3.

ligands of the other two clusters (orange and red background colors). This behavior is likely due to the generally smaller size of the HIV-P ligands in the first two clusters compared with the ligands of the other two clusters. Table SII in the Supporting Information, which exhibits the HIV-P ligand structures, indicates the size differences. The numbers at the

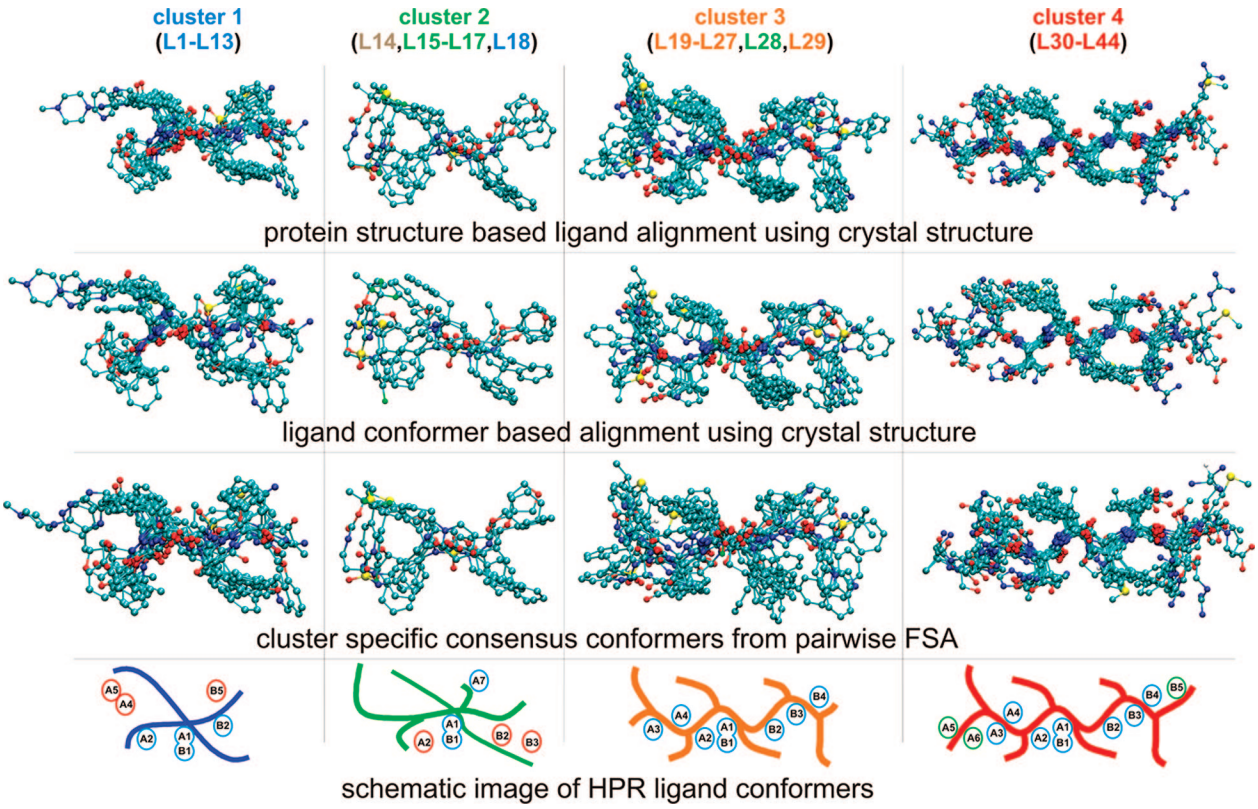


Figure 8. Structure alignments of HIV-P ligands for four clusters found from pairwise similarity based on FSA (see Figure 6). First row: protein structure-based RSA of ligand conformers in crystal structures. Second row: ligand conformer-based RSA using ligand conformers from crystal structures. Third row: structure alignment of consensus conformers from pairwise FSA. Last row: schematic representation of ligand conformers from protein structure-based RSA of first row. Circles indicate dominant H-bond patterns found from analysis of crystal structures. Labels in circles refer to residue identified in Table 2. Red circles indicate strict absence; blue circles strict presence of H-bonds; green circles dominant occurrences of H-bonds (circles obey same color code as ellipses in Table 2).

maximal $HI_{combined}$ in Figure 5a label the HIV-P ligand L*x*, which by FSA yields the L*y* conformer (on the abscissa) most similar to the crystal structure. Evidently FSA with ligand pairs belonging to the same cluster yield the highest similarity with the BLCs (see color code of ligand numbers in Figure 5). The same applies to the minimal RMSDs in Figure 5b.

E: Clustering HIV-P Ligands Based on Similarity Obtained by FSA. The preceding analysis of HIV-P ligand similarity suggests a heterogeneous set of ligands that may involve different binding modes. To purse this idea, we applied a cluster algorithm to explore the molecular similarities which consider functional groups and the scaffold of the ligands. The combined Hodgkin indices $HI_{combined}$, eq 11, of all $43 \times 44/2 = 946$ ligand pairs were evaluated by the FSA approach. We used the inverse of the $HI_{combined}$ as a measure of the distance between ligands (see Table SVI of the Supporting Information) to classify the ligands in different clusters by the neighbor-joining cluster algorithm.[58] We then used the results of this cluster algorithm to construct a phylogenetic tree by the program T-REX.[59] The resulting tree, displayed in Figure 6, shows four distinct ligand clusters. The smallest cluster (green, L14−L18) consists of five members. Together with the second largest cluster with 13 members (blue, L1−L13), it harbors the HIV-P ligands of smaller size. The third cluster contains 11 members (orange, L19−L29), while the largest cluster, containing the natural substrate oligopeptides, has 15 members (red, L30−L44). The colored circles around L14, L18, and L28 highlight differences in ligand classification derived from the H-bond patterns of the HIV-P ligands in the crystal structures (see part F, Table 2).

F: Classification of HIV-P Ligands Based on H-Bond Patterns from Crystal Structures. We can use the information from the crystal structures to classify HIV-P ligands. For this purpose, we analyzed the H-bond patterns of the HIV-P ligands, which have been discussed in numerous publications.[53,55,60−98] The binding affinities of these HIV-P ligands are also available[53,55,60−102] (see Table SV, Supporting Information) and their correlations with the H-bond patterns have been studied (Table SV), but no clear relations were found.

In the present study H-bonds are assumed to be present if the distances between the participating non-hydrogen atoms are smaller than 3.5 Å. The HIV-P ligands are listed in Table 2 in four groups (blue, green, orange, red, except for the outlier L14, which possesses a single H-bond and very low binding affinity, Table SV) corresponding to the H-bond patterns to which they belong. Using the same color code as in Figure 6, which displays the clustering results based on the FSA approach, Table 2 provides an abbreviated list of H-bonds that are significant for the clustering of the ligands. A complete list of the H-bonds formed between the ligands and the HIV-P homodimer is given in Table SV of

the Supporting Information. The H-bonds, which are most relevant for discrimination between different groups of ligands, are highlighted by elongated ellipses (red ellipses for absence of possible H-bonds, blue ellipses for presence of H-bonds, and green ellipses for dominant, but not exclusive, occurrences of specific H-bonds).

For eight ligands (L3, L10, L14, L25, L26, L35, L40, L42) crystal structures with two different ligand orientations are available. In these cases, both H-bond patterns are given in Tables 2 and SV. Except for L14 the H-bond pattern of these ligands possesses mirror symmetry, which is due to the *C*2 symmetry of the HIV-P homodimer. L14 binds poorly, exhibiting one or two H-bonds in the two orientations.

Comparing the four classes of HIV-P ligands obtained by an analysis of the H-bond patterns (Table 2) with the clusters resulting from the similarities of pairwise FSA, one finds that the results of the two classifications are nearly identical (Figure 6). If we exclude HIV-P ligand L14, which is a poor binder, only two other ligands fall into similarity clusters that differ from the results of H-bond classification. Ligand L18 (L28) belongs to the blue (green) cluster according to the H-bond pattern but belongs to the green (orange) cluster based on FSA (Figure 6). Note that the ligands L15−L17, L28 of the green cluster have a polar ring structure in the center of the ligand scaffold in common, which induces a specific H-bond pattern. In L18 the corresponding polar ring in the ligand scaffold is off-center, giving rise to in a different H-bond pattern, which accordingly places the ligand in the blue cluster. The similarity of L18 to ligands of the blue or green cluster is nearly equal (see Figure S6 of the Supporting Information) with a slight preference for the green cluster. This contrasts with L28, which, based on similarity, is too large to fit into the green cluster of small ligands. Therefore, according to FSA it is placed in the orange cluster, while according to the H-bond pattern it should be placed in the green cluster.

G: Cluster-Specific Consensus Ligand Conformers Based on Pairwise FSA. For each of the 44 HIV-P ligands considered, the pairwise FSA method yields 43 different conformers. From these conformers one can derive CLCs without using crystal structure information. Since four different clusters of ligands were found based on pairwise similarity (see Figure 6), a CLC can be derived for each ligand L*y* using only the L*y* conformers obtained by pairwise FSA with the other ligands belonging to the same cluster. The procedure to generate such CLCs is described in part D3 of Methods. The resulting values of $HI_{combined}$ and RMSDs of the CLCs with the BLC of the crystal structures are given in Figure 7. The $HI_{combined}$ based on the CLCs vary now between 0.8 and 0.95. They are considerably larger than the corresponding maximal $HI_{combined}$ varying between 0.6 and 0.75 based on conventional FSA that disregards the results of ligand classification (Figure 5a). The RMSDs of the CLCs are around 2.0 Å. This value is as small as the minimal rmsd taken from the 43 ligand conformers obtained by conventional FSA (Figure 5b). However, in the latter case we do not know which one of the 43 ligand conformers yields the minimal rmsd. This demonstrates clearly how important the information of the ligand-cluster analysis is for the construction of cluster-specific CLCs.

H: Visualizing the HIV-P Ligand Clusters. Structure alignments of HIV-P ligands were performed for the BLCs and for conformers obtained by pairwise FSA. Two different pairwise rigid-structure alignment (RSA) methods were used for the ligand conformers from the crystal structures: (i) ligand conformer-based RSA using the pairwise FSA machinery with rigid torsion angles to align the BLCs; (ii) protein structure-based RSA using the Kabsch algorithm[56,57] to align the HIV-P crystal structures (without considering the ligands) by minimizing the rmsd, applying the resulting translation and rotation transforms to superimpose the corresponding ligands on their bound conformers. The results of the protein structure-based RSA are depicted in the first row of Figure 8 which shows the overlaid ligand structures for the four different ligand clusters found by analyzing the FSA results. The blue and green clusters contain 13 and 5 small ligands, while the orange and red clusters contain 11 and 15 large ligands, respectively.

For the ligand conformer-based RSA, the transforms (translations and rotations) necessary to perform the structural overlay of ligands are obtained by the same method used to generate CLCs for a set of ligand conformers as described in part S9 of the Supporting Information. The results of this structure alignment are shown in the second row of Figure 8.

The first two rows in Figure 8 display structure overlays of the HIV-P ligand conformers as they appear in the crystal structures obtained through the two different overlay techniques discussed above. The third row of Figure 8 shows the results of structural overlay of HIV-P ligands, where the ligand conformers were generated by the FSA method without any information from the HIV-P crystal structures. Here, we consider the cluster-specific CLCs obtained by the FSA method as described in preceding part G. The results of these structure overlays are strikingly similar to the RSA of the ligand conformers of the crystal structures of the first two rows of Figure 8. This finding demonstrates again the feasibility of the FSA approach determining the ligand conformers in the protein binding pocket without using knowledge on the protein structure.

The last row of Figure 8 summarizes schematically the similarity pattern of the HIV-P ligand conformers in the four different clusters derived by the FSA procedure. The most relevant H-bonds, which characterize the four different binding modes of the HIV-P ligands, are also indicated. It is interesting that the natural substrate peptides, which are cleaved by HIV-P, all belong to the red ligand cluster (right column in Figure 8), which possesses the most complex binding mode. The other HIV-P ligands belonging to the red cluster have the same binding mode found by an analysis of the H-bond pattern of the crystal structures (Table 2) and thus possess the same conformation. Moreover, ligands from the same cluster often possess chemically similar scaffolds, as evident from the chemical composition of the HIV-P ligands displayed in Table SII of the Supporting Information. Hence, for HIV-P ligands, differences in ligand

binding modes also correlate with differences in ligand scaffolds. If this is a general feature of ligands, it will be more difficult to find new drugs possessing different scaffolds.[103]

Conclusions

In the present study we used MD simulations on ligand pairs from a set of ligands binding to the same target protein to perform flexible structure alignment (FSA). To make this procedure feasible, we used a specialized energy function. While it includes conventional intramolecular interactions, the true intermolecular interactions are replaced by attractive interactions derived from similarity measures accounting for volume/shape, atomic partial charges, and electrostatic potential of the ligands. These attractive terms allow equivalent molecular groups from a pair of flexible ligands to be superimposed, yielding consensus ligand conformers (CLC) that are most similar to each other. The assumption of the FSA approach is that the CLCs contain information on the true bound ligand conformer (BLC). Thus, if a sufficient number of ligands of different chemical architecture binding to the same target are available, it should be possible to obtain approximate BLCs without using any experimental structural information of the target protein.

We explored the FSA approach using a test set of 44 HIV-P crystal structures. For HIV-P, we have nowadays one of the largest databases of protein−ligand structures comprising many chemically diverse ligands. Since this database was 10 years ago only half as large, this study could not have been performed much earlier (see Figure S7).

Analyzing the similarity of the HIV-P ligand conformers obtained from the pairwise FSA, we found four different ligand clusters. There are two clusters of small and two clusters of large ligands. Ligands belonging to the same cluster exhibit similar backbone architecture. The five natural substrate oligopeptides all belong to the same cluster of large ligands. The four clusters agree well with a classification based on H-bond pattern of the ligands in the HIV-P binding pocket.

RMSDs of HIV-P ligands averaged over MD trajectories of independent ligands are 4 Å, if measured relative to the bound ligand conformers (BLCs) in the crystal structure. Even for closest approach to the BLCs in these MD trajectories the RMDSs are still 3 Å. In contrast, the consensus ligand conformers (CLCs) agree well with the BLC of the crystal structures with an rmsd of about 2 Å. Hence, the present approach of indirect drug design offers a possibility to generate coordinates of the BLCs without using any structural information from the target of the ligand. The application to 44 HIV-P ligands demonstrated that the procedure works well although the ligands possess four different binding modes. The ensemble of coordinates of CLCs belonging to the same binding mode that are generated by the present FSA approach approximate the true BLCs. In order that the FSA approach of indirect drug design presented in this study can work, one needs several (say five or more) ligands of different chemical architecture binding to the same target that possess the same binding mode as for instance the case for the considered 44 HIV-P ligands. In combination with the chemical composition (i.e., electrostatics, H-bond pattern, and hydrophobicity) of the ligands, the ensemble of CLCs can be used to define pharmacophores, thus opening alternative ways for drug design.

Acknowledgment. We thank Dr. Martin Karplus for providing the program CHARMM. We also acknowledge preliminary work from Srdjan Pokorni. We are grateful for useful suggestions and proofreading by Dr. Dennis Diestler. This work was supported by the Deutsche Forschungsgemeinschaft (Sfb 765 Project C1).

Supporting Information Available: Derivation of the electrostatic potential overlap. Details of the preparation of initial ligand conformers and generating consensus ligand conformers. Procedure for pairwise FSA by MD simulation with schematic demonstration. The RMSDs of 100 random structures generated for each of the 44 ligands. CHARMM scripts to perform structural overlap and generate consensus ligand conformers. 2D structures of 44 HPR ligands. $HI_{combined}$ and rmsd between alternate ligand conformers. Detailed hydrogen bonding pattern of bound HPR ligands. Pairwise structure alignment of the HPR reference ligands L*y* (L1, L2, ... L44) with all other 43 HPR ligands L*x* (x=1, 2, ... 44). $HI_{combined}$ from the initial conformers of the HPR ligands. Visualization of HPR ligand clusters on the basis of H-bond pattern. Distance matrix to generate the HPR ligand clusters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kuntz, I. D. *Science* **1992**, *257*, 1078–1082.

(2) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comput. Chem.* **2004**, *13*, 505–524.

(3) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. *J. Med. Chem.* **1994**, *37*, 1035–1054.

(4) DesJarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Montellano, P. R. O. d.; DeCamp, D. L.; Babe, L. M.; Craik, C. S. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644–6648.

(5) Hol, W. G. J. *Angew. Chem., Int. Ed. Engl.* **1986**, *25*, 767–778.

(6) Richards, W. G. *Pure Appl. Chem.* **1994**, *66*, 1589–1596.

(7) Loew, G. H.; Villar, H. O.; Alkorta, I. *Pharm. Res.* **1993**, *10*, 475–486.

(8) Sheridan, R. P.; Rusinko, A.; Nilakantan, R.; Venkataraghavan, R. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 8165–8169.

(9) Marshall, G. R.; Cramer, R. D. *Trends Pharmacol. Sci.* **1988**, *9*, 285–289.

(10) Bender, A.; Glen, R. C. *Org. Biomol. Chem.* **2004**, *2*, 3204–3218.

(11) Willett, P.; Barnard, J. M.; Downs, G. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996.

(12) Lewis, R. A. *J. Med. Chem.* **2005**, *48*, 1638–1648.

(13) Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*; VCH: Weinheim, 1993; Vol. 1.

(14) Yang, G.-F.; Huang, X. *Curr. Pharm. Des.* **2006**, *12*, 4601–4611.

(15) Tropsha, A.; Golbraikh, A. *Curr. Pharm. Des.* **2007**, *13*, 3494–3504.

(16) Leach, A. R., *Molecular Modelling: Principles and applications*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, 2001.

(17) Waterbeemd, H. v. d.; Rose, S. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C.-G., Ed.; Academic Press: New York, 2008; Chapter 23, pp 491−513.

(18) Labute, P. *J. Mol. Graph. Model.* **2000**, *18*, 464–477.

(19) Labute, P. *Methods Mol. Biol.* **2004**, *275*, 261–278.

(20) Gozalbes, R.; Doucet, J. P.; Derouin, F. *Curr. Drug Targets Infect. Disord.* **2002**, *2*, 93–102.

(21) Mason, J. S.; Good, A. C.; Martin, E. J. *Curr. Pharm. Des.* **2001**, *7*, 567–597.

(22) Güner, O. F. *Curr. Top. Med. Chem.* **2002**, *2*, 1321–1332.

(23) Sippl, W. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C.-G., Ed.; Academic Press: New York, 2008; Chapter 28, pp 572−586.

(24) Langer, T.; Hoffmann, R. D. *Expert Opin. Drug Discovery* **2006**, *1*, 261–267 .

(25) Kontoyianni, M.; Madhav, P.; Suchanek, E.; Seibel, W. *Curr. Med. Chem.* **2008**, *15*, 107–116.

(26) Mclnnes, C. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494–502.

(27) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–351.

(28) Villoutreix, B. O.; Renault, N.; Lagorce, D.; Sperandio, O.; Montes, M.; Miteva, M. A. *Curr. Protein. Pept. Sci.* **2007**, *8*, 381–411.

(29) Cavasotto, C. N.; Orry, A. J. W. *Curr. Top. Med. Chem.* **2007**, *7*, 1006–1014.

(30) Jain, A. N. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 396–403.

(31) Klebe, G. *Drug Discovery Today* **2006**, *11*, 580–594.

(32) Martin, Y. C. In *3D QSAR in Drug Design: Recent Advances*; Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds. Kluwer Academic: The Netherlands, 1998; Vol. 3, pp 3−23.

(33) Langer, T.; Bryant, S. D. In *The Practice of Medicinal Chemistry*, 3rd ed.; Wermuth, C.-G., Ed. Academic Press: New York2008; Chapter 29, pp 587−604.

(34) Akamatsu, M. *Curr. Top. Med. Chem.* **2002**, *2*, 1381–94.

(35) Mor, M.; Rivara, S.; Lodola, A.; Lorenzi, S.; Bordi, F.; Plazzi, P. V.; Spadoni, G.; Bedini, A.; Duranti, A.; Tontini, A.; Tarzia, G. *Chem. Biodivers.* **2005**, *2*, 1438–1451.

(36) Sippl, W. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 825–830.

(37) Cramer, R. D.; Patterson, D. E.; Bunce, J. D. *J. Am. Chem. Soc.* **1988**, *110* (18), 5959–5967.

(38) Böhm, M.; Stürzebecher, J.; Klebe, G. *J. Med. Chem.* **1999**, *42*, 458–477.

(39) Klebe, G. In *3D QSAR in Drug Design: Recent Advances*; Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds. Kluwer Academic: The Netherlands, 1998; Vol. 3, pp 87−104.

(40) Klebe, G.; Abraham, U. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 1–10.

(41) Suh, M.-E.; Park, S.-Y.; Lee, H.-J. *Bull. Korean Chem. Soc.* **2002**, *23*, 417–422.

(42) Ullmann, G. M.; Hauswald, M.; Jensen, A.; Kostic, N. M.; Knapp, E. W. *Biochemistry* **1997**, *36*, 16187–16196.

(43) Gandhi, R.; Bartlett, J. G.; Linkinhoker, M. Johns Hopkins University Division of Infectious Diseases and AIDS Service, 1999.

(44) Toh, H.; Ono, M.; Saigo, K.; Miyata, T. *Nature* **1985**, *315*, 691–692.

(45) Pearl, L. H.; Taylor, W. R. *Nature* **1987**, *329*, 351–354.

(46) Carbo, R.; Leyda, L.; Arnau, M. *Int. J. Quantum Chem.* **1980**, *17*, 1185–1189.

(47) Good, A. C. *J. Mol. Graphics* **1992**, *10*, 144–151.

(48) Hodgkin, E. E.; Richards, W. G. *Int. J. Quantum Chem.* **1987**, 105–110.

(49) Good, A. C.; Richards, W. G. In *3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity*, Kubinyi, H.; Folkers, G.; Martin, Y. C., Eds. Kluwer Academic: The Netherlands, 1998; Vol. 2, pp 321−338.

(50) Good, A. C.; Hodgkin, E. E.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 188–191.

(51) Good, A. C.; Richards, W. G. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 112–116.

(52) Meyer, A. Y.; Richards, W. G. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 427–439.

(53) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. *Structure* **2002**, *10*, 369–381.

(54) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(55) Bone, R.; Vacca, J. P.; Anderson, P. S.; Holloway, M. K. *J. Am. Chem. Soc.* **1991**, *113*, 9382–9384.

(56) Kabsch, W. *Acta Crystallogr.* **1976**, *A32*, 922–923.

(57) Kabsch, W. *Acta Crystallogr.* **1978**, *A34*, 827–828.

(58) Saitou, N.; Nei, M. *Mol. Biol. Evol.* **1987**, *4*, 406–425.

(59) Makarenkov, V. *Bioinformatics* **2001**, *17*, 664–668.

(60) Fitzgerald, P.; McKeever, B.; VanMiddlesworth, J.; Springer, J.; Heimbach, J.; Leu, C.; Herber, W.; Dixon, R.; Darke, P. *J. Biol. Chem.* **1990**, *265*, 14209–14219.

(61) Erickson, J.; Neidhart, D. J.; VanDrie, J.; Kempf, D. J.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, M.; Wideburg, N.; Kohlbrenner, W. E.; Simmer, R.; Helfrich, R.; Paul, D. A.; Knigge, M. *Science* **1990**, *249*, 527–533.

(62) Swain, A.; Miller, M.; Green, J.; Rich, D.; Schneider, J.; Kent, S.; Wlodawer, A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.

(63) Miller, M.; Schneider, J.; Sathyanarayana, B.; Toth, M.; Marshall, G.; Clawson, L.; Selk, L.; Kent, S.; Wlodawer, A. *Science* **1989**, *246*, 1149–1152.

(64) Rutenber, E.; Fauman, E.; Keenan, R.; Fong, S.; Furth, P.; Montellano, P. O. d.; Meng, E.; Kuntz, I.; DeCamp, D.; Salto, R.; Rose, J. R.; Craik, C. S.; Stroud, R. M. *J. Biol. Chem.* **1993**, *268*, 15343–15346.

(65) Priestle, J.; Fässler, A.; Rösel, J.; Tintelnot-Blomley, M.; Strop, P.; Grütter, M. *Structure* **1995**, *3*, 381–389.

(66) Munshi, S.; Chen, Z.; Li, Y.; Olsen, D. B.; Fraley, M. E.; Hungate, R. W.; Kuo, L. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *D54*, 1053–1060.

(67) Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. *Biochemistry* **1991**, *30*, 1600–1609.

(68) Ringhofer, S.; Kallen, J.; Dutzler, R.; Billich, A.; Visser, A. J. W. G.; Scholz, D.; Steinhauser, O.; Schreiber, H.; Auer, M.; Kungl, A. J. *J. Mol. Biol.* **1999**, *286*, 1147–1159.

(69) Brynda, J.; Rezacova, P.; Fabry, M.; Horejsi, M.; Stouracova, R.; Sedlacek, J.; Soucek, M.; Hradilek, M.; Lepsik, M.; Konvalinka, J. *J. Med. Chem.* **2004**, *47*, 2030–2036.

(70) Bäckbro, K.; Löwgren, S.; Österlund, K.; Atepo, J.; Unge, T.; Hultén, J.; Bonham, N. M.; Schaal, W.; Karlén, A.; Hallberg, A. *J. Med. Chem.* **1997**, *40*, 898–902.

(71) Hong, L.; Hartsuck, J. A.; Foundling, S.; Ermolieff, J.; Tang, J. *Protein Sci.* **1998**, *7*, 300–305.

(72) Hong, L.; Treharne, A.; Hartsuck, J. A.; Foundling, S.; Tang, J. *Biochemistry* **1996**, *35*, 10627–10633.

(73) Abbenante, G.; March, D. R.; Bergman, D. A.; Hunt, P. A.; Garnham, B.; Dancer, R. J.; Martin, J. L.; Fairlie, D. P. *J. Am. Chem. Soc.* **1995**, *117*, 10220–10226.

(74) Geremia, S.; Demitri, N.; Wuerges, J.; Benedetti, F.; Berti, F.; Tell, G.; Randaccio, L. *ChemMedChem* **2006**, *1*, 186–188.

(75) Thaisrivongs, S.; Skulnick, H. I.; Turner, S. R.; Strohbach, J. W.; Tommasi, R. A.; Johnson, P. D.; Aristoff, P. A.; Judge, T. M.; Gammill, R. B.; Morris, J. K.; Romines, K. R.; Chrusciel, R. A.; Hinshaw, R. R.; Chong, K.-T.; Tarpley, W. G.; Poppe, S. M.; Slade, D. E.; Lynn, J. C.; Horng, M.-M.; Tomich, P. K.; Seest, E. P.; Dolak, L. A.; Howe, W. J.; Howard, G. M.; Schwende, F. J.; Toth, L. N.; Padbury, G. E.; Wilson, G. J.; Shiou, L.; Zipp, G. L.; Wilkinson, K. F.; Rush, B. D.; Ruwart, M. J.; Koeplinger, K. A.; Zhao, Z.; Cole, S.; Zaya, R. M.; Kakuk, T. J.; Janakiraman, M. N.; Watenpaugh, K. D. *J. Med. Chem.* **1996**, *39*, 4349–4353.

(76) Silva, A. M.; Cachau, R. E.; Sham, H. L.; Erickson, J. W. *J. Mol. Biol.* **1996**, *255*, 321–340.

(77) Andersson, H. O.; Fridborg, K.; Löwgr; en, S.; Alterman, M.; Mühlman, A.; Björsne, M.; Garg, N.; Kvarnström, I.; Schaal, W.; Classon, B.; Karlén, A.; Danielsson, U. H.; Ahlsén, G.; Nillroth, U.; Vrang, L.; Öberg, B.; Samuelsson, B.; Hallberg, A.; Unge, T. *Eur. J. Biochem.* **2003**, *270*, 1746–1758.

(78) Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C. A. *J. Mol. Biol.* **2000**, *301*, 1207–1220.

(79) Hong, L.; Zhang, X.; Hartsuck, J.; Tang, J. *Protein Sci.* **2000**, *9*, 1898–1904.

(80) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. *J. Med. Chem.* **1991**, *34*, 3340–3342.

(81) Hoog, S. S.; Zhao, B.; Winborne, E.; Fisher, S.; Green, D. W.; DesJarlais, R. L.; Newlander, K. A.; Callahan, J. F.; Abdel-Meguid, S. S.; Moore, M. L.; Huffman, W. F. *J. Med. Chem.* **1995**, *38*, 3246–3252.

(82) Murthy, K.; Winborne, E.; Minnich, M.; Culp, J.; Debouck, C. *J. Biol. Chem.* **1992**, *267*, 22770–22778.

(83) Thanki, N.; Rao, J.; Foundling, S. I.; Howe, W. J.; Moon, J. B.; Hui, J. O.; Tomasselli, A. G.; Heinrikson, R. L.; Thaisrivongs, S.; Wlodawer, A. *Protein Sci.* **1992**, *1*, 1061–1072.

(84) Smith, A. B.; Charnley, A. K.; Harada, H.; Beiger, J. J.; Cantin, L.-D.; Kenesky, C. S.; Hirschmann, R.; Munshi, S.; Olsen, D. B.; Stahlhut, M. W.; Schleif, W. A.; Kuo, L. C. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 859–863.

(85) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. *Structure* **1995**, *3*, 581–590.

(86) Jhoti, H.; Singh, O. M. P.; Weir, M. P.; Cooke, R.; Murray-Rust, P.; Wonacott, A. *Biochemistry* **1994**, *33*, 8417–8427.

(87) Ekegren, J. K.; Unge, T.; Safa, M. Z.; Wallberg, H.; Samuelsson, B.; Hallberg, A. *J. Med. Chem.* **2005**, *48*, 8098–8102.

(88) Thaisrivongs, S.; Watenpaugh, K. D.; Howe, W. J.; Tomich, P. K.; Dolak, L. A.; Chong, K.-T.; Tomich, C.-S. C.; Tomasselli, A. G.; Turner, S. R.; Strohback, J. W.; Mulichak, A. M.; Janakiraman, M. N.; Moon, J. B.; Lynn, J. C.; Horng, M.-M.; Hinshaw, R. R.; Curry, K. A.; Rothrock, D. J. *J. Med. Chem.* **1995**, *38*, 3624–3637.

(89) Kempf, D.; Marsh, K.; Denissen, J.; McDonald, E.; Vasavanonda, S.; Flentge, C.; Green, B.; Fino, L.; Park, C.; Kong, X.; Wideburg, N.; Saldivar, A.; Ruiz, L.; Kati, W.; Sham, H.; Robins, T.; Stewart, K.; Hsu, A.; Plattner, J.; Leonard, J.; Norbeck, D. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 2484–2488.

(90) Lange-Savage, G.; Berchtold, H.; Liesum, A.; Budt, K.-H.; Peyman, A.; Knolle, J.; Sedlacek, J.; Fabry, M.; Hilgenfeld, R. *Eur. J. Biochem.* **1997**, *248*, 313–322.

(91) March, D. R.; Abbenante, G.; Bergman, D. A.; Brinkworth, R. I.; Wickramasinghe, W.; Begun, J.; Martin, J. L.; Fairlie, D. P. *J. Am. Chem. Soc.* **1996**, *118*, 3375–3379.

(92) Stoll, V.; Qin, W.; Stewart, K. D.; Jakob, C.; Park, C.; Walter, K.; Simmer, R. L.; Helfrich, R.; Bussiere, D.; Kao, J.; Kempf, D.; Sham, H. L.; Norbeck, D. W. *Bioorg. Med. Chem.* **2002**, *10*, 2803–2806.

(93) Kervinen, J.; Thanki, N.; Zdanov, A.; Tino, J.; Barrish, J.; Lin, P. F.; Colonno, R.; Riccardi, K.; Samanta, H.; Wlodawer, A. *Protein Pept. Lett.* **1996**, *3*, 399–406.

(94) Kervinen, J.; Lubkowski, J.; Zdanov, A.; Bhatt, D.; Dunn, B. M.; Hui, K. Y.; Powell, D. J.; Kay, J.; Wlodawer, A.; Gustchina, A. *Protein Sci.* **1998**, *7*, 2314–2323.

(95) Rutenber, E. E.; McPhee, F.; Kaplan, A. P.; Gallion, S. L.; Hogan, J. C., Jr.; Craik, C. S.; Stroud, R. M. *Bioorg. Med. Chem.* **1996**, *4*, 1545–1558.

(96) Jadhav, P. K.; Ala, P.; Woerner, F. J.; Chang, C.-H.; Garber, S. S.; Anton, E. D.; Bacheler, L. T. *J. Med. Chem.* **1997**, *40*, 181–191.

(97) Abdel-Meguid, S. S.; Metcalf, B. W.; Carr, T. J.; Demarsh, P.; DesJarlais, R. L.; Fisher, S.; Green, D. W.; Ivanoff, L.; Lambert, D. M.; Murthy, K. H. M.; Petteway, S. R., Jr.; Pitts, W. J.; Tomaszek, T. A., Jr.; Winborne, E.; Zhao, B.; Dreyer, G. B.; Meek, T. D. *Biochemistry* **1994**, *33*, 11671–11677.

(98) Surleraux, D. L. N. G.; Tahri, A.; Verschueren, W. G.; Pille, G. M. E.; Kock, H. A. d.; Jonckers, T. H. M.; Peeters, A.; Meyer, S. D.; Azijn, H.; Pauwels, R.; Bethune, M.-P. d.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B. T. P. *J. Med. Chem.* **2005**, *48*, 1813–1822.

(99) Rutenber, E. E.; Voss, J. J. D.; Hoffman, L.; Stroud, R. M.; Lee, K. H.; Alvarez, J.; McPhee, F.; Craik, C.; Montellano, P. R. O. d. *Bioorg. Med. Chem.* **1997**, *5*, 1311–1320.

(100) Ermolieff, J.; Lin, X.; Tang, J. *Biochemistry* **1997**, *36*, 12364–12370.

(101) Barrish, J. C.; Gordon, E.; Alam, M.; Lin, P.-F.; Bisacchi, G. S.; Chen, P.; Cheng, P. T. W.; Fritz, A. W.; Greytok, J. A.; Hermsmeier, M. A.; Humphreys, W. G.; Lis, K. A.; Marella, M. A.; Merchant, Z.; Mitt, T.; Morrison, R. A.; Obermeier, M. T.; Pluscec, J.; Skoog, M.; Slusarchyk, W. A.; Spergel, S. H.; Stevenson, J. M.; Sun, C.-q.; Sundeen, J. E.; Taunk, P.; Tino, J. A.; Warrack, B. M.; Colonno, R. J.; Zahler, R. *J. Med. Chem.* **1994**, *37*, 1758–1768.

(102) Ghosh, A. K.; Sridhar, P. R.; Leshchenko, S.; Hussain, A. K.; Li, J.; Kovalevsky, A. Y.; Walters, D. E.; Wedekind, J. E.; Grum-Tokars, V.; Das, D.; Koh, Y.; Maeda, K.; Gatanaga, H.; Weber, I. T.; Mitsuya, H. *J. Med. Chem.* **2006**, *49*, 5252–5261.

(103) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894–2896.

CT8004886