Journal of Computer-Aided Molecular Design, **14:** 161–179, 2000.
KLUWER/ESCOM

Design of dimerization inhibitors of HIV-1 aspartic proteinase: A computer-based combinatorial approach

Amedeo Caflisch[a,*], Hans J. Schramm[b] & Martin Karplus[c,*]
[a]*Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland;* [b]*Max-Planck-Institut für Biochemie, D-8033 Martinsried, Germany;* [c]*Laboratoire de Chimie Biophysique, Institut Le Bel, Université Louis Pasteur, F-6700 Strasbourg, France*

Received 15 February 1999; Accepted 30 June 1999

Key words: de novo design, finite-difference Poisson–Boltzmann, HIV-1 aspartic proteinase, inhibitors of dimerization, MCSS, structure-based drug design

Summary

Inhibition of dimerization to the active form of the HIV-1 aspartic proteinase (HIV-1 PR) may be a way to decrease the probability of escape mutations for this viral protein. The Multiple Copy Simultaneous Search (MCSS) methodology was used to generate functionality maps for the dimerization interface of HIV-1 PR. The positions of the MCSS minima of 19 organic fragments, once postprocessed to take into account solvation effects, are in good agreement with experimental data on peptides that bind to the interface. The MCSS minima combined with an approach for computational combinatorial ligand design yielded a set of modified HIV-1 PR C-terminal peptides that are similar to known nanomolar inhibitors of HIV-1 PR dimerization. A number of N-substituted 2,5-diketopiperazines are predicted to be potential dimerization inhibitors of HIV-1 PR.

Introduction

The acquired immunodeficiency syndrome (AIDS) is caused by the human immunodeficiency virus (HIV) which encodes, among other indispensable enzymes, an aspartic proteinase (HIV-1 PR) whose function is essential for viral maturation and replication [1]. A large number of inhibitors that bind to the active site of HIV-1 PR have been developed and some of them are being used as drugs against AIDS [2, 3]. One problem with most of these, as well as those in clinical trials, is that rapid mutation of the virus has led to drug resistance [2, 4–10]. Although multidrug therapy has had some success in overcoming this problem [11], it is important to search for inhibitors that are less likely to be susceptible to 'escape' by random mutation. Since HIV-1 PR is a homodimer, one possibility is to design inhibitors that prevent dimerization or favor dissociation. Each chain consists of 99 amino acids and contributes one catalytic aspartic acid (Asp25, Figure 1a). The main reason for targeting the dimerization interface is that the virus is less likely to be able to find mutant residues at the interface that prevent inhibitor binding and still permit docking of the two subunits, which is essential for forming the composite active site. The fact that the proteinase is a homodimer means that the mutation(s) appear in both monomers. Thus, dimerization inhibitors are expected to be less sensitive to mutagenesis. The two chains are stabilized by noncovalent interactions and the largest part of the interface is formed by a four-stranded antiparallel β-sheet consisting of the four N- and C-terminal segments of the proteinase (residues 1–4 and 96–99, respectively). A series of oligopeptides derived from the C-terminal segment of the HIV-1 PR monomer have shown inhibition of HIV-1 PR by interfering with dimer formation and stability [12–17]. Recently, a series of synthetic compounds consisting of two tripeptidic strands attached to an aromatic scaffold were shown to inhibit dimerization of HIV-1 PR in the submicromolar range [18].

*To whom correspondence should be addressed. E-mail: caflisch@bioc.unizh.ch

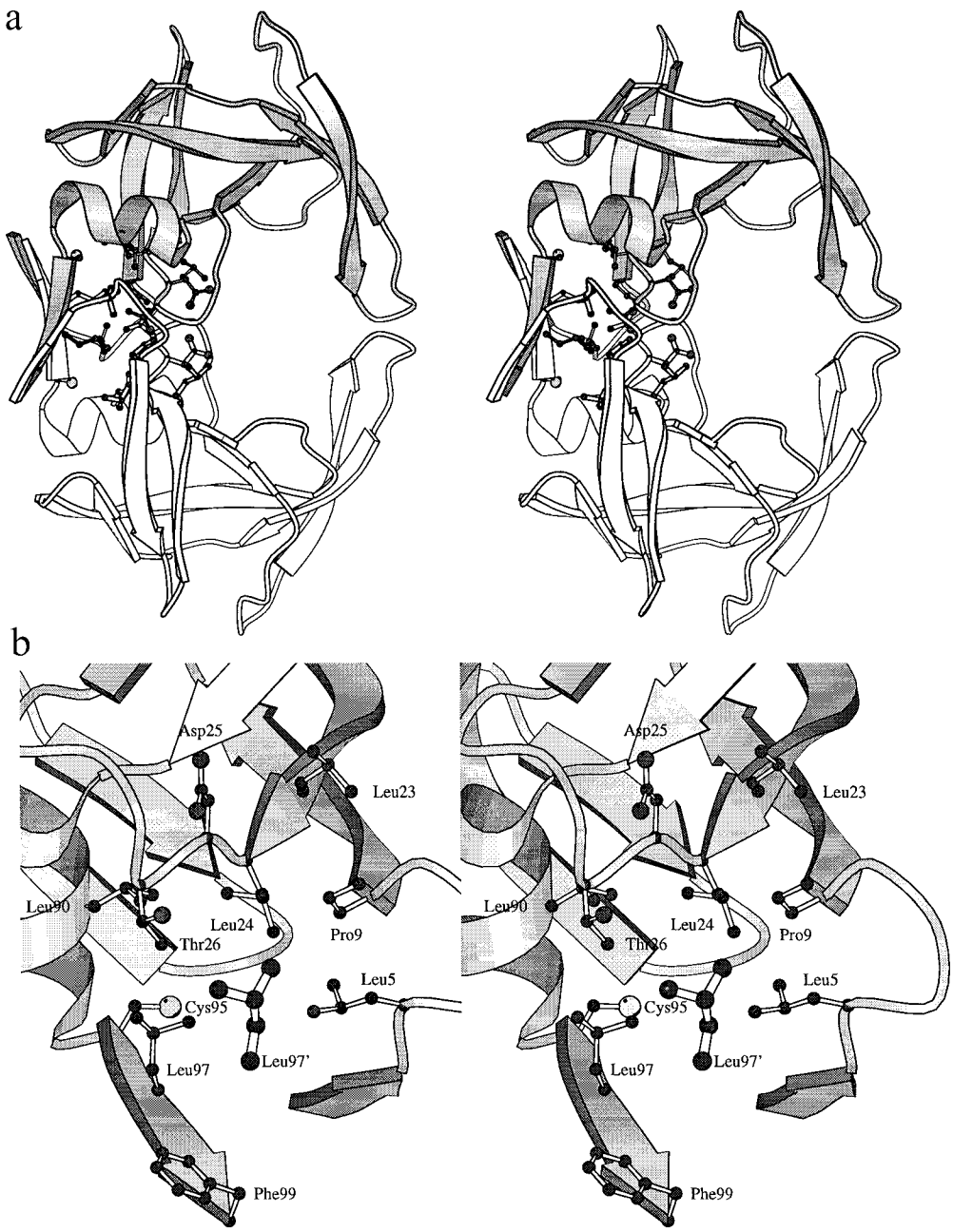


Figure 1. (a) Stereoview of the HIV-1 PR homodimer (PDB code 3HVP). All of the stereoviews in this paper are wall-eyed. The backbone is shown by a ribbon diagram and the residues of the dimerization interface are in a ball and stick representation. One of the two polypeptide chains is shaded. (b) Zoom image of the hydrophobic pocket in the HIV-1 PR monomer after rotation by 70 degrees around the *z*-axis and 90 degrees around the *y*-axis. All the labeled residues (but Phe99) are sensitive to mutation [27]. The Leu97′ side chain is drawn with larger balls and sticks. In all figures primed residues refer to the second monomer. Figures were made with the program MOLSCRIPT [56]. (c) Color stereoview of the hydrophobic pocket in the HIV-1 PR monomer. Atoms and bonds are colored by element type: carbon white, oxygen red, nitrogen blue, and sulfur yellow. The orientation is the same as in (b). The Thr26′ and Leu97′ side chains are drawn with larger cylinders. Hydrogen bonds between the Thr26′ side chain and the unprimed monomer are in green dashed lines. The molecular surface of the unprimed monomer is shown with dots. The figure was made with the program WITNOTP (A. Widmer, unpublished).

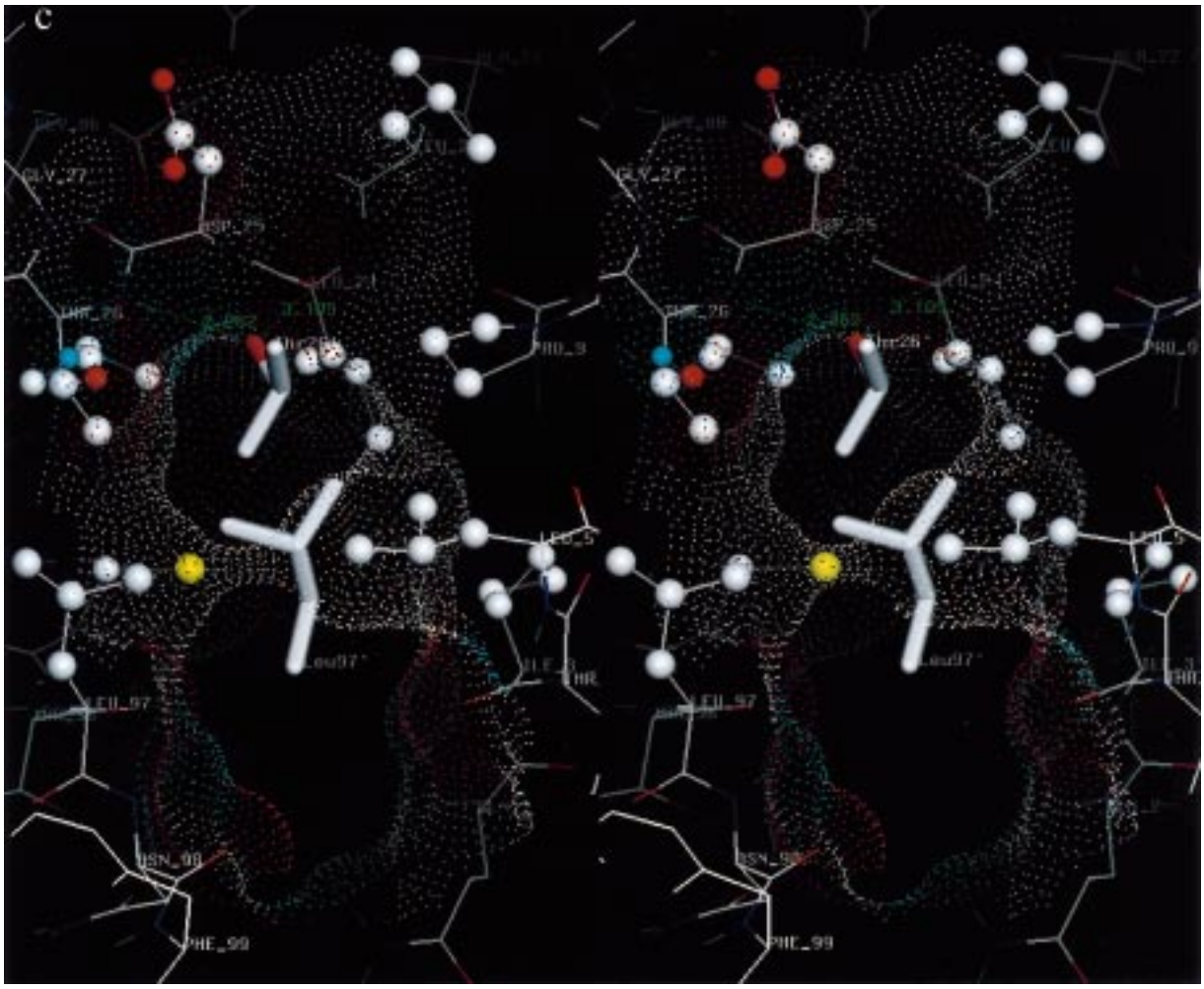


Figure 1 (continued).

In this paper a computer-based combinatorial approach is used to design dimerization inhibitors of HIV-1 PR. First, the Multiple Copy Simultaneous Search (MCSS) method [19] is employed to exhaustively sample the energetically favorable positions and orientations of functional groups in the dimerization interface. For all of the MCSS minima an approximated binding free energy is calculated by supplementing the *vacuum* force field with electrostatic and nonpolar solvation contributions calculated in the continuum approximation. An advance in the present study with respect to most previous ligand design works is the use of a partially flexible protein structure for the replication of the MCSS minima with favorable binding free energy. This is important since monomeric HIV-1 PR might have a conformation different from the one in the dimer [20]. The replication is performed with an approach based on quenched molecular dynamics with very weak harmonic constraints on the protein region close to the MCSS fragment while protein regions far from the binding site are kept fixed. The multiple copy sampling with a partially flexible protein had been used previously (based on minimization alone, i.e., without molecular dynamics) to study the side chain placement of Phe14 in the zinc-finger motif which contains the first 31 residues and the coordinating zinc ion of the protein Zif268 [21, 22].

Finally, the MCSS minima are connected by the Computational Combinatorial Ligand Design (CCLD) program which allows the fast and automatic generation of diverse candidate ligands [23].

A number of modified C-terminal peptides designed with the help of the MCSS/CCLD approach

have most of their functional groups in common with known inhibitors of dimerization of HIV-1 PR.

Methods

Dimerization interface

To limit the high mutation rate which causes the development of resistance to drugs in the viral proteinase, the amino acid side chains buried at the dimer interface were selected as target regions for ligand design. The solvent accessible surface (SAS) of the amino acid side chains in the native structure of HIV-1 PR was calculated with the CHARMM [24] implementation of the Lee and Richards algorithm [25] using a probe sphere of 1.4 Å radius. The calculation was repeated for the structure of the monomer. By subtracting the two one obtains the loss of side chain SAS upon dimerization, which is a measure of their contribution to the dimerization interface. For each monomeric chain of 99 residues the total side chain SAS buried on dimerization is 1157 $Å^2$ and the backbone contribution is 402 $Å^2$; this gives a total buried surface of 3118 $Å^2$ upon dimerization. These values are larger than the ones of the V_L–V_H interface in antibodies, although antibody F_V fragments have about 10% more residues than HIV-1 PR (i.e., about 110 amino acids in V_L and 110 in V_H). For the antisteroid antibody DB3 (PDB code 1DBA), the side chain SAS buried at the interface is 704 $Å^2$ for V_L and 709 $Å^2$ for V_H while the backbone SAS buried is only 105 $Å^2$ for V_L and 131 $Å^2$ for V_H. This yields a total of 1649 $Å^2$ which is close to the average value found in the known F_V structures [26]. Burial of main chain atoms is about 4 times larger in HIV-1 PR than in antibody F_V fragments mainly because of the presence of the four-stranded β-sheet in the dimerization interface of the former.

Visual analysis with the help of the molecular modelling program WITNOTP (A. Widmer, unpublished) revealed that some of the interface residues which become buried on dimerization [Leu5 (90.2 $Å^2$ loss of side chain SAS upon dimerization), Pro9 (23.6 $Å^2$), Leu23 (17.8 $Å^2$), Leu24 (11.9 $Å^2$), Asp25 (18.3 $Å^2$), Thr26 (45.7 $Å^2$), Leu90 (5.8 $Å^2$), Cys95 (25.4 $Å^2$), Leu97 (89.9 $Å^2$), and Phe99 (171.8 $Å^2$)] form a pocket of mainly hydrophobic character which is occupied by Thr26′ and Leu97′ in the dimer (Figure 1b, c); primed residues refer to the second monomer. The Leu97′ side chain is involved in favorable van der Waals contacts with the Leu5, Pro9, Leu24, Thr26, Cys95, and Leu97 side chains while the side chain hydroxyl of Thr26′ donates to the main chain CO of Leu24 and accepts from the NH of Thr26 (Figure 1c). Mutation of any of these residues except for Phe99 reduces activity [27]; it is likely that Phe99 is unimportant because it is at the end of the pocket. The hydrophobic pocket at the dimerization interface is adjacent to the interface β-sheet formed by the four terminal segments of the dimer (Figure 1a, b) which has already been used as a target region for the development of dimerization inhibitors [12–18]. Other surface side chains whose SAS loss upon dimerization exceeds 40 $Å^2$ are: Trp6 (62.6 $Å^2$), Ile50 (98.5 $Å^2$), Arg87 (90.3 $Å^2$), Thr96 (75.9 $Å^2$), and Asn98 (57.8 $Å^2$). Ile50 is on the flap while Trp6, Thr96, and Asn98 are on the exposed surface of the interface β-sheet. Substitution of Arg87, which is involved in a surface salt bridge with Asp29, results in the complete loss of proteolytic activity [28]. Because of their high solvent exposure in the dimer, these amino acid side chains are not a suitable target for ligand design.

System setup

The crystal structure of synthetic unliganded HIV-1 PR at 2.8 Å resolution refined to an *R* factor of 0.184 [29] was used; it is available from the Brookhaven Protein Data Bank [30] (entry code 3HVP). The WITNOTP program was employed to modify the synthetically more accessible α-amino-N-butyric acid at positions 67 and 95 into cysteines [29] as in the wild type enzyme; Cys95 is part of the interface (see Figure 1b). The coordinates of the hydrogen atoms were generated with the HBUILD [31] option of the CHARMM program [24] which was utilized for all minimizations performed in this work. For both the protein and the functional groups, the parameters from the polar hydrogen set (PARAM19) were used [32]. Polar hydrogens are treated explicitly, whereas aliphatic and aromatic hydrogens are considered part of the extended carbon atom to which they are bonded. Standard protonation states were used for the titratable residues; in particular, Asp25 was not protonated because Asp25′ from the primed chain is responsible for the high pK_a, i.e., only in the dimeric form is the pK_a shifted. The positions of the hydrogen atoms in the monomer were first minimized by 200 steps of the conjugate gradient algorithm with fixed heavy atoms. This was followed by a minimization with harmonic constraints on non-hydrogen atoms (force constant of 10 kcal/mol $Å^2$) until the root mean square of the

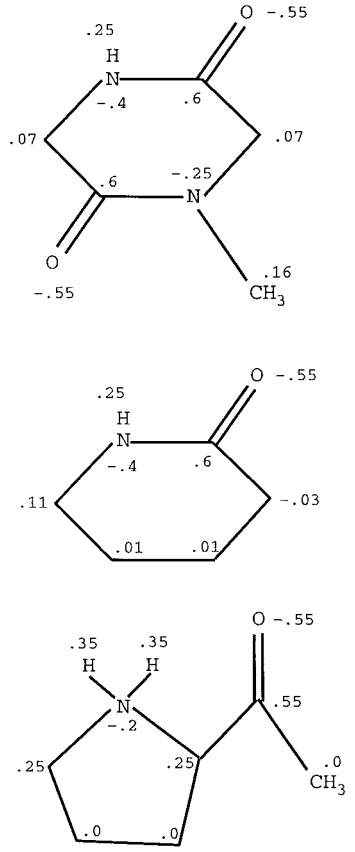


Figure 2. Chemical structure and partial charges of DKPI, δ-lactam, and 2-acylpyrrolidine.

gradient reached a value of 0.0098 kcal/mol Å (20 steps of steepest descent followed by 125 iterations of conjugate gradient). The root mean square deviation of the side chains lining the hydrophobic pocket was 0.2 Å. The minimized structure was used as rigid body in MCSS and CCLD and as starting structure for the quenched molecular dynamics.

Multiple copy simultaneous search (MCSS)

The MCSS method determines energetically favorable positions and orientations (local minima of the potential energy) of functional groups on the surface of a protein or receptor of known three-dimensional structure [19]. It has been applied to generate functionality maps of the sialic acid binding site of the influenza coat protein hemagglutinin [19], the active site of HIV-1 PR [33], FKBP [34], thrombin [23, 35, 36], and picornavirus capsid proteins [37].

For the present application the library of standard MCSS groups was supplemented with N-methyl-2,5-diketopiperazine (DKPI), δ-lactam, and 2-acylpyrrolidine (Figure 2). These fragments can form hydrogen bonds with one or two β-strands and are expected to fit in the uncomplete β-sheet at the dimerization interface of the HIV-1 PR monomer. Parameters for the new groups were derived in analogy to the corresponding PARAM19 parameters of N-methylacetamide and proline.

For each of the functional groups listed in Table 1, 5000 replicas were randomly distributed in a sphere of 12 Å radius with its center located between the side chains of Thr26′ and Leu97′. This sphere contains the hydrophobic pocket of the dimerization interface.

The force field used for the MCSS procedure in this study differs in two ways from the standard CHARMM PARAM19 force field used in the earlier applications [19, 33]. First, the vacuum Coulombic potential is screened by a distance-dependent dielectric constant [35, 38]. Second, a simple but useful approximation to take account of desolvation is used for nonpolar fragments, which do not bear any partial charge in the polar-hydrogen approximation. MCSS determines their optimal position in the protein binding site exclusively by van der Waals interactions. Hence, minima of nonpolar fragments occur in hydrophilic, as well as hydrophobic, pockets because of the lack of an energy penalty for protein desolvation. A representative example of a cluster of propane minima in a mainly hydrophilic region of the HIV-1 aspartic proteinase binding site was shown in Figure 2 of Reference 35. To approximate desolvation of polar regions of the protein, the attractive contribution of the van der Waals interaction energy is switched off between atoms of fully nonpolar MCSS fragments (propane, cyclopentane, cyclohexane, and benzene) and all protein polar hydrogen, nitrogen, oxygen, and sp^2 carbon atoms. In addition, the van der Waals radius of nitrogen and oxygen atoms was increased from the PARAM19 default value of 1.6 Å to 2.2 Å and the van der Waals radius of the aliphatic carbons was reduced by 0.1 Å to avoid the large van der Waals distance between carbons often produced by PARAM19. This modified force field has been tested on the P3-P2′ region of the thrombin active site and on the surface of the A peptide chain of the leucine zipper [39], i.e., residues 249-281 of the yeast transcriptional activator protein GCN4 [40, 41].

Table 1. Functional groups used for MCSS

Group	Electrostatic solvation free energy[c]	CHARMM energy[a]		No. of minima found	$\Delta G_{\text{binding}}$[b]				No. of minima with $\Delta G_{\text{binding}} < 0$
		lowest	highest		lowest	2nd	3rd	highest	
Nonpolar groups:									
Propane	0.0	−5.9	−0.0	136	−11.1	−11.0	−10.8	4.7	124
Cyclopentane	0.0	−7.3	−0.0	102	−12.5	−12.2	−11.7	3.8	94
Cyclohexane	0.0	−7.7	−0.0	81	−13.6	−12.2	−11.9	3.7	77
Benzene	0.0	−7.5	−0.0	58	−13.6	−13.4	−11.8	1.8	55
Polar groups:									
Methanol	−7.4	−17.5	3.9	112	−7.2	−6.7	−6.5	9.7	67
2-Propanone	−5.3	−19.0	−0.0	87	−10.0	−9.2	−8.3	8.2	27
N-methylacetamide	−9.1	−22.2	−4.1	192	−11.1	−11.0	−10.8	9.2	105
Pyridine	−0.4	−11.7	−2.5	141	−13.7	−13.6	−12.7	7.5	98
Pyrrole	−3.2	−13.5	2.4	141	−10.6	−10.0	−9.4	−2.0	29
Imidazole	−6.0	−17.7	−0.0	175	−11.5	−11.4	−11.1	9.0	79
Phenol	−6.8	−20.1	−3.2	224	−14.1	−13.6	−13.1	8.8	146
N-methyl-2,5- diketopiperazine	−11.2	−25.5	−2.7	336	−13.8	−12.9	−11.9	6.7	165
δ-Lactam	−6.7	−20.3	−0.3	317	−14.5	−13.5	−13.1	6.3	162
Charged groups:									
Acetate ion	−71.5	−37.3	−2.0	88	−9.4	−8.3	−7.8	8.6	9
Methylammonium	−99.0	−41.0	−2.7	71	−0.8	0.6	1.7	9.1	1
Methylguanidinium	−84.5	−44.1	−5.7	224	−7.0	−6.4	−6.2	7.4	88
Pyrrolidine	−82.2	−41.0	−9.1	89	−2.3	−1.9	1.3	10.1	2
2-Acylpyrrolidine	−84.3	−45.0	−3.5	170	−9.8	−9.2	−7.6	10.9	99
Benzamidine	−80.6	−38.7	−5.6	163	−10.1	−7.6	−7.2	7.4	42

All energy values are in kcal/mol.
[a]The CHARMM energy is the sum of intermolecular and intraligand energies.
[b]Calculated by use of Equation 1.
[c]Calculated by numerical solution of the LPB equation.

Binding free energy estimate

For every protein–MCSS minimum complex the binding free energy is estimated by use of the following approximation [23]:

$$\Delta G_{\text{binding}} = \Delta E^{\text{fragm}} + \Delta E^{\text{interm}}_{\text{vdW}} + \Delta G^{\text{interm}}_{\text{elect}} + \Delta G^{\text{protein}}_{\text{elect,desolv}} + k\Delta G^{\text{fragm}}_{\text{elect,desolv}} + \Delta G^{\text{complex}}_{\text{np}}. \quad (1)$$

The fragments are flexible during the MCSS minimization procedure and the first term on the right side represents the difference in the energy of the fragment upon binding; it can be written as

$$\Delta E^{\text{fragm}} = \Delta E^{\text{fragm}}_{\text{bonding}} + \Delta E^{\text{fragm}}_{\text{vdW}} + \Delta E^{\text{fragm}}_{\text{elect}}, \quad (2)$$

and is a sum of the bonding (bonds, angles, and torsions) energy terms ($\Delta E^{\text{fragm}}_{\text{bonding}}$), the intrafragment van der Waals energy ($\Delta E^{\text{fragm}}_{\text{vdW}}$), and the intrafragment *vacuum* Coulombic energy ($\Delta E^{\text{fragm}}_{\text{elect}}$). At the end of the minimization, the CHARMM force field is used to compute ΔE^{fragm} and $\Delta E^{\text{interm}}_{\text{vdW}}$, which is the van der Waals intermolecular (interm) energy between protein and fragment. $\Delta E^{\text{fragm}}_{\text{vdW}}$ is calculated with the standard PARAM19 force field (cutoff of 7.5 Å and VSHIFT) while $\Delta E^{\text{fragm}}_{\text{elect}}$ is evaluated with a constant dielectric of 1.0 and infinite cutoff. The solvation free energy is expressed as a sum of electrostatic and nonpolar contributions [42, 43]. The electrostatic contribution to the free energy of binding consists of shielded intermolecular interaction ($\Delta G^{\text{interm}}_{\text{elect}}$), protein desolvation ($\Delta G^{\text{protein}}_{\text{elect,desolv}}$), and desolvation of the fragment ($\Delta G^{\text{fragm}}_{\text{elect,desolv}}$); the latter two contain the self (or Born) term and the intramolecular shielding. The total electrostatic free energy of the fragment is the sum of $\Delta E^{\text{fragm}}_{\text{elect}}$ and $\Delta G^{\text{fragm}}_{\text{elect,desolv}}$. The electrostatic solvation energy values are calculated by solving the linearized Poisson-Boltzmann (LPB) equation by the

finite-difference procedure [44–46] with the UHBD program [47]. A scaling factor (k) for the electrostatic desolvation of the fragment is introduced (see Equation 1) to take into account the fact that when a fragment is part of a larger ligand its desolvation on binding is smaller. For all MCSS minima, a value of $k = 0.4$ was used; this is based on the comparison of the electrostatic desolvation energy upon binding of N-methylacetamide and the dipeptide N-acyl-Gly-NH-CH_3 to HIV-1 PR (dimeric form) [23]. $\Delta G^{\text{interm}}_{\text{elect}}$ and $\Delta G^{\text{fragm}}_{\text{elect,desolv}}$ are zero for fully apolar fragments like benzene and cyclohexane which have no partial charges in the PARAM19 approximation.

On the basis of experimental data on alkane-water partition coefficients [48], the nonpolar contribution to the free energy of binding ($\Delta G^{\text{complex}}_{\text{np}}$, i.e., the sum of cavitation energy and solute–solvent van der Waals energy) is assumed to be proportional to the loss in SAS [25]:

$$\Delta G^{\text{complex}}_{\text{np}} = \gamma \left(\text{SAS}^{\text{complex}} - (\text{SAS}^{\text{protein}}_{\text{isolated}} + \text{SAS}^{\text{fragm}}_{\text{isolated}}) \right). \quad (3)$$

The constant γ, which may be interpreted as the *vacuum*–water microscopic surface tension, is assigned a value of 0.025 kcal/mol Å^2 [49]. For the structure of the complex and its isolated components, the total area, i.e., area of polar and nonpolar groups [50], is computed by the CHARMM implementation of the Lee and Richards algorithm [25] by using a probe sphere of 1.4 Å radius.

Since the protein structure is fixed for the MCSS runs, no strain is introduced by the binding of the functional groups. The entropic contributions of the solute to the free energy of binding are expected to be similar for different ligands and are neglected.

Replication of minima with partial protein flexibility

Some of the MCSS minimized positions with good binding free energy were used as starting structures for conformational sampling with a partially flexible binding site. The reasons for this are twofold. Firstly the structure of the HIV-1 PR monomer is unknown and it is likely that the lack of two interdigitating strands from the second monomer might confer some flexibility to the interface region, particularly to the residues in the terminal regions of the monomeric chain [20]. Secondly it has recently been recognized in using CCLD on other protein targets that additional ligands with alternative binding can be generated by locally replicating the minima with a good binding free energy [39]; corresponding results have been obtained with MCSS and the DLD method for constructing ligands [34]. The additional sampling was done separately for each of the MCSS minimum/HIV-1 PR monomer complex that was chosen for replication. The MCSS fragment and the protein residues within 7 Å of it were constrained with a harmonic force constant of 0.01 kcal/mol Å^2. More distant residues were kept fixed to prevent excessive displacements, which could significantly affect the functionality maps of other fragments. The sampling procedure consisted of 100 cycles of a short (0.4 ps) high temperature molecular dynamics (MD) run followed by energy minimization. For each MD run the initial and final temperatures were about 1200 K and 600 K, respectively. Bond oscillations involving the hydrogen atoms were frozen by the SHAKE algorithm [51] and a timestep of 2 fs was used. The energy minimization consisted of a maximum of 1000 steps of the conjugate gradient algorithm and a convergence criterion of 0.09 kcal/mol Å which was reached in more than 90% of the minimizations. A distance-dependent dielectric constant was used during both dynamics and minimization. For each of the MCSS minima listed in Table 5 the quenched MD procedure was repeated 100 times for a total of 40 ps MD. At the end of each quenched MD cycle, the minimized structure was used as starting conformation for the next cycle and a new random seed was used for the assignment of velocities. Redundant conformations among the 100 structures generated by quenched MD were identified by a similarity function consisting of a double summation over all possible atom pairs in two conformations [52] and discarded. This yielded a set of different orientations of the MCSS fragment and its protein surroundings (see Results section). In most cases the root mean square deviation of the heavy atoms of the fragment and the flexible protein side chains were smaller than 1.0 Å and 1.5 Å, respectively. During the ligand construction phase (see below) all of the different positions of the fragments were employed while the different protein surroundings were not taken into account; this approximation is valid because of the small protein distortion in the replication phase.

Computational combinatorial ligand design (CCLD)

Candidate inhibitors of HIV-1 PR dimerization were designed with the CCLD program [23]. The mini-

mized coordinates of the HIV-1 PR monomer (see section 'System setup') were used as the receptor structure. For the MCSS minima and the fragments replicated with the partially flexible protein the carbon atoms with free valences were defined as linkage atoms. Minimal and maximal values in Å for the distance between linkage atoms were: $d < 0.43$, 0-bond; $1.2 < d < 1.8$, 1-bond; $2.2 < d < 2.7$, 2-bond (methylene or keto); $3.4 < d < 4.5$, 3-bond (ethylene or amide). Before starting the combinatorial search CCLD generates a list of bonding fragment pairs, a list of overlapping fragment pairs and a list of overlapping linker-fragment pairs. Overlapping fragment pairs are defined as having one or more bad contacts, and in the case of only one bad contact, no geometrically correct way for converting the bad contact into a linkage. If a linker l between the fragments $f1$ and $f2$ has one or more bad contacts with the fragment $f3$ the pair $l-f3$ is added to the list of overlapping linker-fragment pairs and during ligand construction the fragment $f3$ will not be included whenever a growing ligand already contains the fragments $f1$ and $f2$ linked by l. The use of this list represents an improvement with respect to the original method [23]. The three lists are used during the combinatorial search and the two overlap lists prevent the generation of molecules with bad contacts.

In the present study, the combinatorial generation of putative ligands was performed in a 'restrictive' mode; i.e., CCLD was required to generate candidate ligands each containing one copy of a user-specified fragment (e.g., benzene minimum 1). The fact that the orientations of the fragments were sampled in slightly different protein surroundings does not affect the outcome of the CCLD procedure which uses the protein structure only for the correct placement of the linkage atoms. The latter are affected only marginally by the small displacements of the protein regions around the MCSS minima allowed during the quenched MD protocol.

Results and analysis

Functionality maps of the rigid protein

In considering the MCSS results obtained with the rigid protein both structural (e.g., intermolecular hydrogen bonds, van der Waals contacts) and energetic properties of the minima are analyzed. In addition, a comparison of the functional group sites with the HIV-1 PR interchain interactions is given for nonpolar, polar, and charged fragments. Only the unprimed chain was used as the receptor structure for the MCSS, LPB, and CCLD calculations. The primed chain is an excellent 'inhibitor' and its interactions are compared with the simulation results. In particular the 'consensus sequence' Thr_{94}-Val,Ile_{95}-Ser,Thr_{96}-Tyr_{97}-Asp,Glu,Asn_{98}-Leu,Tyr,Trp_{99} of potent peptidic inhibitors [16] derived from the C-terminal 94-99 segment of the proteinase (Gly_{94}-Cys_{95}-Thr_{96}-Leu_{97}-Asn_{98}-Phe_{99}) is used as a basis of comparison.

Minima of nonpolar groups with the rigid protein

There are 124 propane, 94 cyclopentane, 77 cyclohexane, and 55 benzene minima with favorable binding free energy (Table 1) and they are distributed in the apolar regions of the dimerization interface. Since the apolar groups have similar functionality maps and energy values only the positions of benzene are described in detail. Minima 1 (Figure 3a) and 2 (not shown) overlap the Leu97′ side chain and have an approximated binding free energy of -13.6 and -13.4 kcal/mol, respectively (Table 2). This suggests the replacement of Phe for Leu at position 97, in agreement with experimental data which show that the tetrapeptide inhibitor Thr-Phe-Asn-Phe has a slightly better inhibitory activity (IC_{50} value of 450 μM) than the 'wild-type' C-terminal peptide Thr_{96}-Leu_{97}-Asn_{98}-Phe_{99} (IC_{50} of 660 μM) [14].

Minima 3 (Figure 3a) and 4 (not shown) overlap the Leu5′ side chain. Since the N-terminal 'wild-type' pentapeptide Ac-Pro_1-Gln_2-Ile_3-Thr_4-Leu_5 has shown moderate inhibition [14], it would be useful to test the Ac-Pro-Gln-Ile-Thr-Phe peptide.

Minima 5 (Figure 3a), 6, and 9 interact with the Phe99 side chain and overlap the side chain of Cys95′. The 'wild-type' C-terminal peptide Cys_{95}-Thr_{96}-Leu_{97}-Asn_{98}-Phe_{99} has an IC_{50} of 150 μM for HIV-1 PR [15]. A mutated C-terminal peptide with a larger aromatic or aliphatic side chain instead of Cys_{95} is predicted to have stronger affinity. The 'consensus sequence' of potent peptidic inhibitors derived from the C-terminus has Val or Ile at position 95 [16].

Minimum 8 matches the benzene ring of Phe99′ (RMSD of 0.7 Å, Figure 3a) and is slightly more buried. This implies that a larger side chain might fit in agreement with the 'consensus sequence' which has Leu, Tyr, or Trp at position 99 [16]. For HIV-2 PR the 'wild-type' peptide Cys_{95}-Thr_{96}-Leu_{97}-Asn_{98}-Phe_{99} has an IC_{50} of 400 μM [15]. Peptides with a bulky C-terminal amino acid (Phe99 or Trp99) have a weaker activity for HIV-2 PR than for HIV-1 PR [16],

Table 2. Minima of benzene

Rank[a]	Intermolecular v.d. Waals[b]	Desolvation		$\Delta G_{binding}$[e]	MCSS rank[f]	Site occupied by
		nonpolar[c]	elect[d]			
1	−7.4	−8.2	2.0	−13.6	2	Leu97′
2	−7.1	−8.1	1.9	−13.4	1	Leu97′
3	−7.4	−7.0	2.6	−11.8	5	Leu5′
4	−7.5	−7.0	3.3	−11.2	6	Leu5′
5	−5.5	−6.3	1.4	−10.4	8	Cys95′
6	−4.0	−6.2	0.8	−9.4	15	Cys95′
7	−5.2	−6.6	2.8	−9.0	16	Ala28,Ile84[g]
8	−6.8	−7.7	5.9	−8.6	3	Phe99′
9	−6.8	−8.1	6.4	−8.5	14	Cys95′
10	−4.0	−7.0	2.6	−8.5	17	Ala28,Ile84[g]
11	−3.8	−5.4	1.2	−8.0	21	Leu5, Pro9[g]
20	−3.3	−4.6	0.9	−7.0	35	Ile3′

Energy values in kcal/mol are listed for the 10 benzene minima with the lowest binding free energy and for other benzene minima discussed in the text.
[a]Ranked among the minima of benzene according to binding free energy $\Delta G_{binding}$. Minima with rank in bold are shown in Figure 3a.
[b]Calculated with CHARMM with VSHIFT option and cutoff of 7.5 Å.
[c]Calculated by use of Equation 3.
[d]Only protein desolvation since fragment is apolar. Calculated by numerical solution of the LPB equation as shown in Figure 1a of Reference 23.
[e]Calculated by use of Equation 1, i.e., sum of columns 2, 3, 4, and ΔE^{fragm}. The latter is always unfavorable and smaller than 0.003 kcal/mol for all benzene minima.
[f]Ranked among the minima of the same functional group according to total CHARMM energy which consists of the sum of vacuum intermolecular (modified van der Waals parameter, see text) and intraligand energies.
[g]The residues of the nonprimed chain closest in space to benzene are listed for the sites not occupied by any group of the second monomer.

probably because of the Cys95Met mutation which reduces the size of the pocket available in the HIV-2 PR monomer for the side chain of residue 99. Thus, to improve activity for the HIV-2 PR one should modify the 'wild-type' C-terminal peptide at Phe_{99} into Leu_{99}, i.e., Phe_{95}-Thr_{96}-Leu_{97}-Asn_{98}-Leu_{99}.

Minima 12 to 19 are close to the minima 1 to 10 while minimum 11 (Figure 3a and Table 2) interacts with the Leu5 side chain and minimum 20 interacts with the Leu97 and Phe99 side chains and partially overlaps the Ile3′ side chain (Figure 3a and Table 2).

Analogous results are found for propane, cyclopentane, and cyclohexane. As an example, cyclohexane minima 1–3, 4–6, and 7–8 overlap the Leu97′, Leu5′, and Cys95′ side chains, respectively. All of the apolar cavities filled by hydrophobic residues of the primed chain are occupied by minima of nonpolar fragments with very favorable estimated binding free energies (Figure 3a and Table 2).

Minima of polar groups with the rigid protein

The polar fragments used in this study can be divided in three sets: simple (methanol, 2-propanone, N-methylacetamide), aromatic (pyrrole, imidazole, pyridine, phenol), and cyclic nonaromatic (N-methyl-2,5-diketopiperazine (DKPI), δ-lactam). Fragments belonging to the same set (e.g., pyrrole and imidazole) show similar functionality maps. To simplify the analysis only some of the low free energy minima are discussed and in particular those which match similar moieties of the primed chain. There are 112 MCSS minima of methanol and the lowest free energy minimum donates to the main chain carbonyl of Thr96 (not shown). Minima 20 and 32 match the side chain hydroxyl of Thr26′ and Thr96′, respectively (Figure 3b). A number of N-methylacetamide minima match the main chain of the C-terminal strand of the primed monomer (residues 95′ to 99′) and are involved in the same hydrogen bonds as the corresponding peptide groups. As an example, minimum 5, which over-

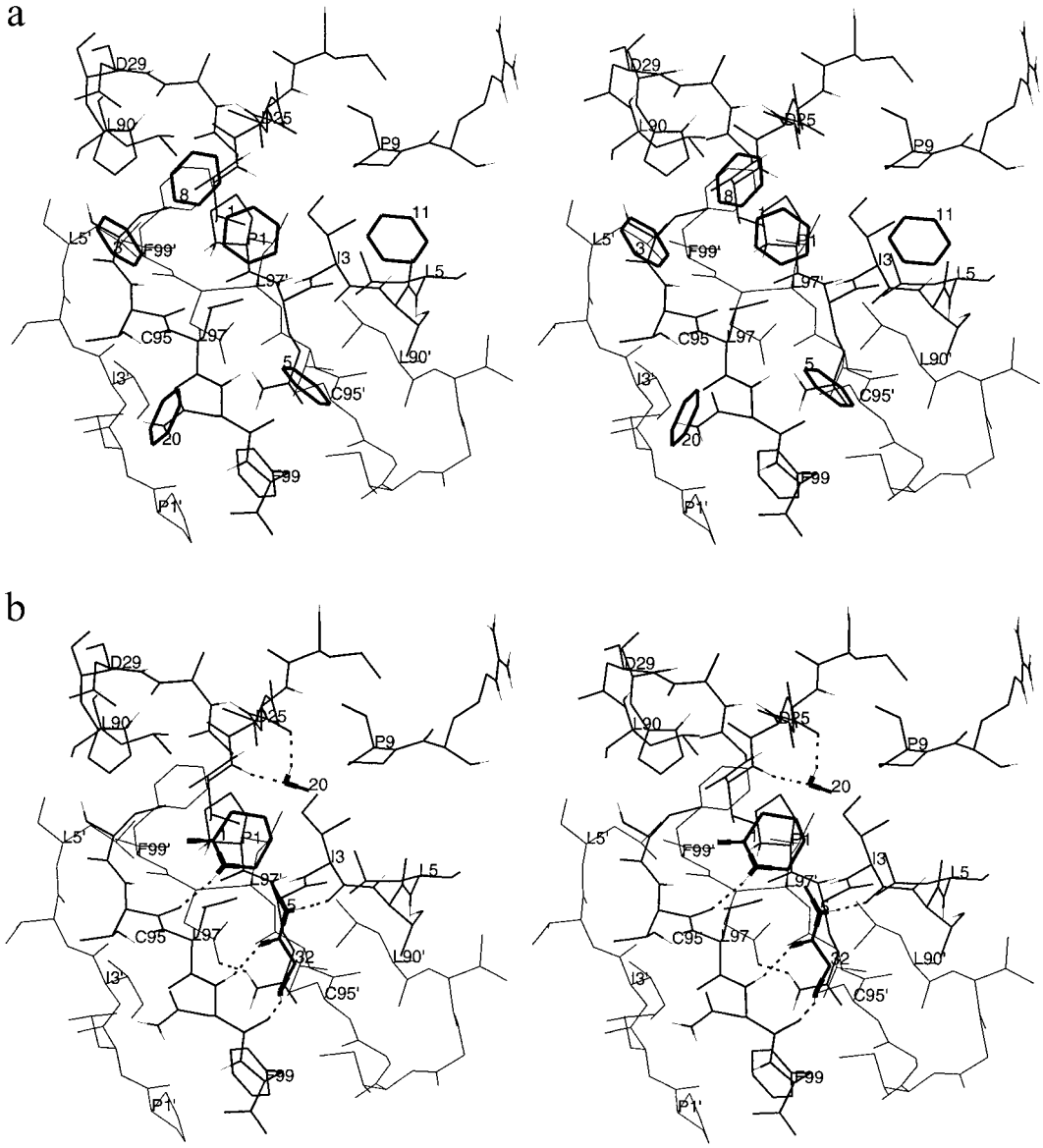


Figure 3. Stereoviews of interesting MCSS minima (thick lines for heavy atoms and thin lines for polar hydrogens) at the HIV-1 PR dimerization interface. The polypeptide chain used as receptor structure is shown in medium lines (nonprimed residues); the second chain is shown in thin lines (primed residues) though it was removed during the MCSS procedure. Some C_α atoms of HIV-1 PR are labeled. The MCSS minima are labeled according to their binding free energy rank within minima of the same type. Hydrogen bonds between protein and MCSS minima are shown in dashed lines. (a) Representatives of the 20 best benzene minima, (b) δ-lactam minimum 1, N-metylacetamide minimum 5, methanol minima 20 and 32, (c) N-methyl-2,5-diketopiperazine (DKPI) minimum 2 and phenol minimum 3, (d) best minima of acetate, methylammonium, methylguanidinium, and benzamidine, acetate minimum 7, and 2-acylpyrrolidine minimum 12.

laps the 96′-97′ amide group, donates to the C=O of Ile3 and accepts from the main chain NH of Asn98 (Figure 3b).

The minimized positions of polar aromatic fragments reflect a balance between optimal van der Waals interactions of the aromatic ring and favorable electrostatic interactions of their polar groups. The best minimum of imidazole is close to the side chain of Leu5′ and donates a hydrogen bond to the C=O of Cys95. Imidazole minima 2 to 4 donate to the main chain carbonyl of Thr96 and have their aromatic ring close to the side chain of Leu97′. The NH and N atoms of imidazole minimum 12 make the same hydrogen bonds as the NH and O atoms of N-methylacetamide

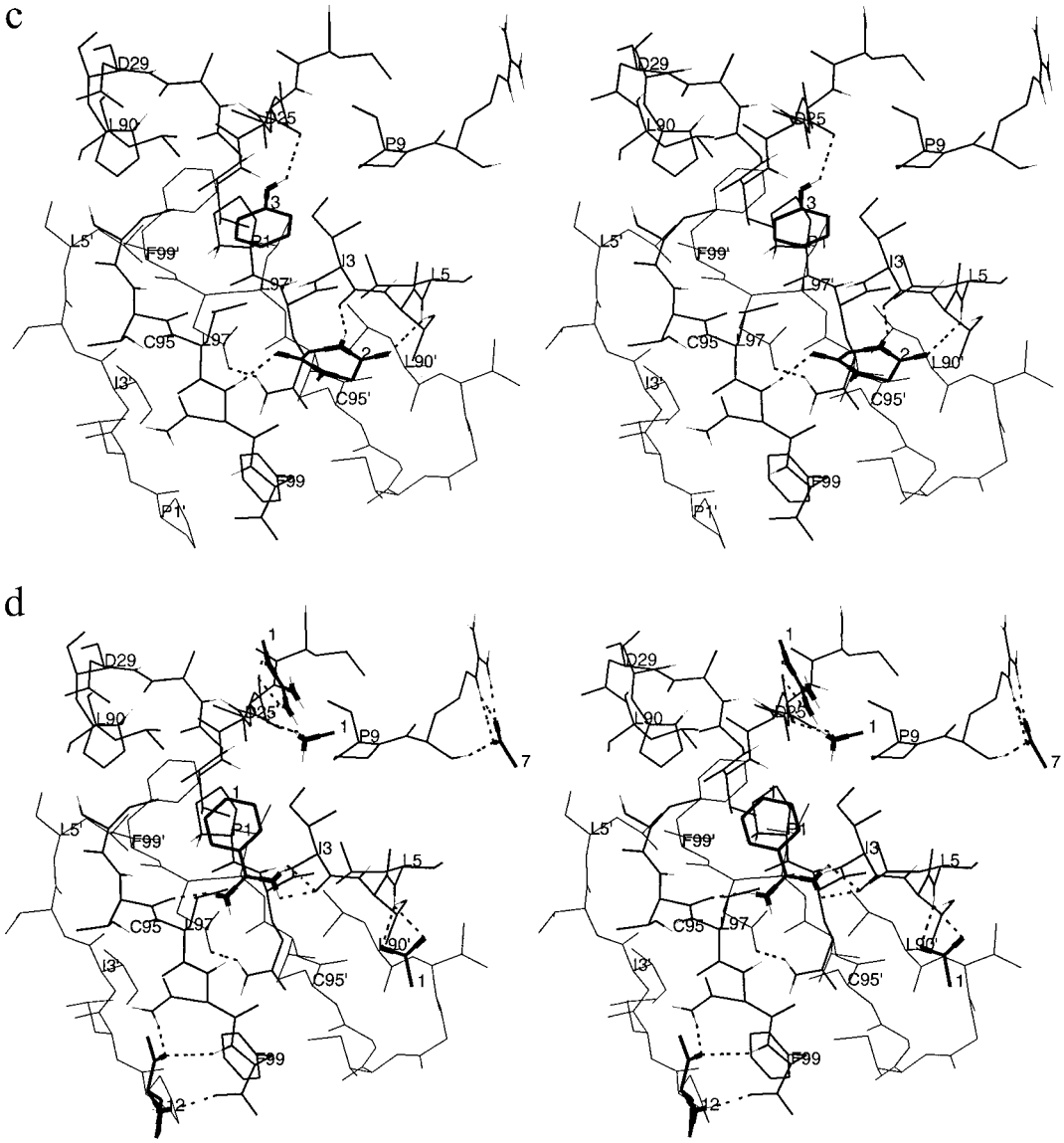


Figure 3 (continued).

minimum 5. The best three minima of phenol overlap the side chain of Leu97′ (Table 3) and donate a hydrogen bond to the main chain C=O of Thr96 (minima 1 and 2) and to the C=O of Leu24 (minimum 3, Figure 3c). The position and interactions of phenol minimum 3 suggest the replacement Leu97Tyr for dimerization inhibitors derived from the C-terminal segment in agreement with the aforementioned 'consensus sequence' which has only Tyr at position 97 [16].

The importance of using the free energy functional (Equation 1) for ranking MCSS minima is illustrated by phenol minimum 3, which was ranked as 101 by the MCSS energy. Other cases of the role of the solvation correction are evident from comparing the free energy ranking and the MCSS ranking in Tables 2–4. The effect for a nonpolar group like benzene is much less important because of the modified potential used for the fully apolar groups in MCSS which prevents them from appearing in hydrophilic pockets of the protein.

The map of pyridine is similar to the benzene map. Pyridine minima 1–3, 4–8, and 9 overlap the Leu97′, Leu5′, and Cys95′ side chains, respectively.

Table 3. Minima of polar groups

Rank[a]	Strain[b]	Intermolecular		Desolvation			$\Delta G_{binding}$ [h]	MCSS	Site and
		v.d.Waals[c]	Elect[d]	Nonpolar[e]	Electrostatic			rank[i]	H-bond partners
					Protein[f]	Ligand[g]			
Phenol:									
1	0.0	−7.1	−2.7	−8.6	2.6	1.6	−14.1	74	Leu97′; 96CO
2	0.1	−7.6	−2.6	−8.6	2.9	2.2	−13.6	61	Leu97′; 96CO
3	0.3	−8.4	−2.0	−8.5	3.1	2.3	−13.1	101	Leu97′; 24CO
4	0.0	−9.6	−0.7	−7.3	2.2	2.3	−13.0	141	Leu5′
5	0.6	−7.6	−2.3	−8.5	3.0	2.0	−12.8	115	Leu97′; 24CO
6	0.1	−8.5	−3.0	−7.2	3.5	2.4	−12.6	38	Leu5′; Thr91 $O_{\gamma 1}$
7	0.0	−6.9	−3.0	−8.8	4.6	1.6	−12.5	76	Leu97′; 96CO
8	0.1	−8.4	−2.7	−7.2	5.1	1.0	−12.1	97	Leu5′; 97NH
9	0.1	−6.8	−3.2	−7.3	3.1	2.4	−11.7	68	Leu5′; 95CO
10	0.1	−8.4	−3.0	−8.7	7.1	1.6	−11.3	91	Thr96′; 98NH
DKPI:									
1	0.1	−9.1	−4.6	−10.6	7.7	2.7	−13.8	33	Leu97′; 3NH, 3CO
2	0.1	−8.8	−7.5	−9.5	9.7	3.1	−12.9	1	95′-96′; 5NH, 98NH, 3CO
3	0.1	−11.6	−3.2	−10.1	10.7	2.2	−11.9	30	Thr96′; 5NH, Gln2 $N_{\epsilon 2}$
4	0.7	−10.6	−3.7	−9.7	8.9	2.5	−11.8	40	Thr96′; 5NH, Gln2 $N_{\epsilon 2}$
5	0.4	−10.2	−5.2	−10.7	11.5	2.3	−11.8	38	Leu97′; 3NH, 98NH
6	0.3	−7.5	−6.4	−8.9	8.9	2.2	−11.5	23	Cys95′; 5NH, 98NH
7	0.1	−8.7	−6.5	−9.9	11.1	2.4	−11.4	15	Cys95′; 5NH, 98NH
8	0.4	−10.9	−5.9	−10.8	13.3	2.6	−11.3	2	Thr96′; 98NH, 98CO, Gln2 $N_{\epsilon 2}$
9	1.0	−10.5	−5.4	−10.9	12.4	2.3	−11.0	28	Leu97′; 3NH, 98NH
10	0.1	−6.8	−5.2	−5.7	5.3	1.8	−10.5	47	Pro1’; 99NH, 97CO, Asn98 $N_{\delta 2}$

Energy values in kcal/mol are listed for the 10 MCSS minima of phenol and DKPI with the lowest binding free energy.
[a]Ranked among the minima of the same functional group according to binding free energy. Minima with rank in bold are shown in Figure 3c.
[b]The sum of intraligand energy terms is calculated with CHARMM (Equation 2) by using infinite cutoff for the Coulombic term and constant dielectric.
[c]Calculated with CHARMM with VSHIFT option and cutoff of 7.5 Å.
[d]Calculated by numerical solution of the LPB equation.
[e]Calculated by use of Equation 3.
[f]Calculated as shown in Figure 1a of Reference 23.
[g]Calculated as shown in Figure 1c of Reference 23. Values are scaled by k=0.4 (see text).
[h]Calculated by use of Equation 1, i.e., sum of columns 2 to 7.
[i]Ranked among the minima of the same functional group according to total CHARMM energy, i.e., sum of vacuum intermolecular (unmodified van der Waals parameters) and intraligand energies.

The lowest free energy minimum of DKPI overlaps the side chain of Leu97′ and is involved in hydrogen bonds with the NH and C=O groups of Ile3 (Table 3). DKPI minima 2, 6, and 7 act as acceptor for the NH of Leu5 and Asn98. In addition, minimum 2 donates a hydrogen bond to the C=O of Ile3 and its N-methyl group is close to the C_β of Cys95′ (Figure 3c). The three best minima of δ-lactam overlap the side chain of Leu97′ and donate a hydrogen bond to the main chain carbonyl of Thr96 (Figure 3b).

Minima of charged groups with the rigid protein

Charged groups have few favorable positions because of the high desolvation penalty upon binding [23]. Three of the four lowest energy minima of acetate ion are involved in hydrogen bonds with the NH of Leu5 (Figure 3d) and have an estimated binding free energy ranging from −9.4 kcal/mol to −5.4 kcal/mol (Table 4). Acetate minimum 7 is close to the Asp29′ side chain; it makes a salt bridge with Arg8 and accepts also from the main chain NH of Arg8 (Figure 3d). Yet, its estimated free energy of binding is only −3.8 kcal/mol. This is due to the position of the Arg8 side chain which is on the

Table 4. Minima of acetate ion

Rank[a]	Strain[b]	Intermolecular		Desolvation			$\Delta G_{binding}$ [h]	MCSS rank[i]	Site and H-bond partners
		v.d.Waals[c]	Elect[d]	Nonpolar[e]	Electrostatic				
					Protein[f]	Ligand[g]			
1	0.1	−8.6	−8.9	−4.7	5.3	7.4	−9.4	10	5NH
2	0.0	−6.7	−6.7	−4.6	4.9	4.7	−8.3	12	50NH, 51NH
3	0.0	−7.9	−6.1	−5.0	5.6	5.6	−7.8	14	5NH
4	0.0	−7.3	−7.2	−4.5	5.9	7.4	−5.8	13	5NH
5	0.0	−6.5	−7.3	−7.0	10.0	6.0	−4.8	16	3NH, Gln2 $N_{\epsilon 2}$
6	0.0	−6.0	−7.3	−6.8	9.7	5.9	−4.6	17	3NH, Gln2 $N_{\epsilon 2}$
7	0.1	−4.1	−11.5	−4.0	10.0	5.7	−3.8	3	8NH, Arg8 N_{ϵ}
8	0.1	−2.7	−7.4	−4.5	6.7	4.7	−2.9	11	Arg87 N_{η}
9	0.1	−3.8	−5.6	−4.7	5.9	5.8	−2.5	15	Arg87 N_{η}
10	0.1	−0.3	−10.4	−3.1	10.5	3.2	0.0	5	Arg8 N_{η}

Energy values in kcal/mol are listed for the 10 MCSS minima of phenol and DKPI with the lowest binding free energy.
[a]Ranked among the minima of the same functional group according to binding free energy. Minima with rank in bold are shown in Figure 3d.
[b]The sum of intraligand energy terms is calculated with CHARMM (Equation 2) by using infinite cutoff for the Coulombic term and constant dielectric.
[c]Calculated with CHARMM with VSHIFT option and cutoff of 7.5 Å.
[d]Calculated by numerical solution of the LPB equation.
[e]Calculated by use of Equation 3.
[f]Calculated as shown in Figure 1a of Reference 23.
[g]Calculated as shown in Figure 1c of Reference 23. Values are scaled by k=0.4 (see text).
[h]Calculated by use of Equation 1, i.e., sum of columns 2 to 7.
[i]Ranked among the minima of the same functional group according to total CHARMM energy, i.e., sum of vacuum intermolecular (unmodified van der Waals parameters) and intraligand energies.

protein surface; hence, the electrostatic intermolecular energy between acetate minimum 7 and the protein (−11.5 kcal/mol calculated with UHBD), intermolecular van der Waals energy (−4.1 kcal/mol), and nonpolar solvation (−4.0 kcal/mol) are balanced by the electrostatic desolvation of the protein (10.0 kcal/mol) and the electrostatic desolvation of the ligand (5.7 kcal/mol). The best minima of methylammonium, methylguanidinium, and pyrrolidine are involved in hydrogen bonds with the side chain of Asp25, in agreement with the results obtained for the dimeric form [33] while the best benzamidine minimum has its ring on top of the Leu97′ side chain and donates to the main chain carbonyl of Ile3 and Thr96 (Figure 3d). There are two interesting minima of 2-acylpyrrolidine. In minimum 1 the aliphatic part of the ring overlaps the Leu97′ side chain while its polar groups accepts from the NH of Ile3 and donate to the C=O of Thr96. Minimum 12 (−6.5 kcal/mol of binding free energy) matches the Pro1′ residue and is involved in hydrogen bonds with the NH and COO^- of the C-terminal residue Phe99 and the side chain amide of Asn98 (Figure 3d).

Replication of minima with partial protein flexibility

The quenched MD protocol was run for the lowest free enegy minima of the apolar fragments and some of the polar fragments. The number of distinct conformations sampled ranges from two (benzene minimum 5) to forty-six (methanol minimum 32, see Table 5). The total CHARMM energy (distance dependent dielectric and 7.5 Å nonbonding cutoff) of the quenched MD conformations does not vary by more than 20% for most fragments (Table 5). In most quenched MD runs the total CHARMM energy improves; e.g., in the run started from the cyclohexane free energy minimum 4 the CHARMM energy changes from −286.4 kcal/mol for the first quenched minimum to −308.8 kcal/mol for the 76th minimum. Only for the quenched MD runs started from cyclopentane free energy minimum 1 and DKPI free energy minimum 2, does the first minimum sampled have the most favorable total CHARMM energy; i.e., there was no improvement during the quenched MD run. The solvation energy contributions were not calculated for the conformations obtained by quenched MD since it is not clear how to introduce the

Table 5. Conformations sampled with flexible protein

Corresponding residue in the prime chain[a]	MCSS fragment	MCSS minima		Quenched MD minima	
		Rank[b]	$\Delta G_{binding}$[c] [kcal/mol]	Number of distinct conformations[d]	CHARMM energy range[e] [kcal/mol]
Leu97′					
	benzene	1	−13.6	14	−237.0 −222.2
	cyclohexane	1	−13.6	11	−250.3 −233.9
	cyclopentane	1	−12.5	21	−238.1 −221.0
Leu5′					
	benzene	3	−11.8	14	−301.8 −284.4
	cyclohexane	4	−11.3	21	−306.7 −286.4
	cyclopentane	5	−10.4	36	−281.8 −262.6
Cys95′					
	benzene	5	−10.4	2	−74.0 −72.5
	cyclohexane	7	−10.2	5	−73.5 −58.4
	cyclopentane	12	−10.4	30	−67.4 −55.2
C_β and C_γ of Arg87′					
	benzene	11	−8.0	8	−161.4 −146.4
	cyclohexane	16	−8.0	5	−143.7 −131.7
	cyclopentane	20	−10.4	12	−135.6 −128.2
Phe99′					
	benzene	8	−8.6	7	−219.2 −201.5
95′-96′ main chain					
	DKPI	2	−12.9	15	−144.8 −109.3
	DKPI	7	−11.4	11	−122.3 −106.6
Thr26′					
	methanol	20	−4.5	16	−215.1 −191.1
Thr96′					
	methanol	32	−3.7	46	−100.6 −79.7

[a] Residues closest in position in the primed chain.
[b] Ranked among the minima of the same functional group according to binding free energy ($\Delta G_{binding}$).
[c] Calculated by use of Equation 1.
[d] Clustered with an exponential similarity function.
[e] Total CHARMM energy with distance dependent dielectric and nonbonding energy cutoff of 7.5 Å. Energy ranges for the representatives of the clusters, i.e., the conformations used by CCLD.

energy changes originating from the flexible regions of the protein which vary for the different runs.

Figure 4 shows the 11 conformations obtained by quenched MD starting from DKPI minimum 7. In 10 of the 11 the carbonyl groups of DKPI are involved in hydrogen bonds with the main chain NH of Leu5 and Asn98 as in the MCSS orientation used as starting structure; in the 11th the hydrogen bond to the NH of Asn98 is lost. For some of the protein side chains at the binding surface a significant number of different rotameric states are sampled. The rotation of the Leu5 side chain is almost free, while the Gln2/Thr4, Thr96/Gln98, and Leu97/Phe99 side chain pairs rotate in a concerted way (Figure 4). On the other hand, Ile3 and Trp6 do not assume significantly different orientations.

De novo design of dimerization inhibitors

The CCLD program was run using as input fragments the MCSS minima with $\Delta G_{binding} < -5$ kcal/mol and all the conformations obtained with quenched MD (Table 5). CCLD used the same protein conformation as in MCSS. As mentioned above, the orientations of the fragments sampled in slightly different protein surroundings do not affect the outcome of the CCLD procedure which uses the protein structure only for the correct placement of the linkage atoms.

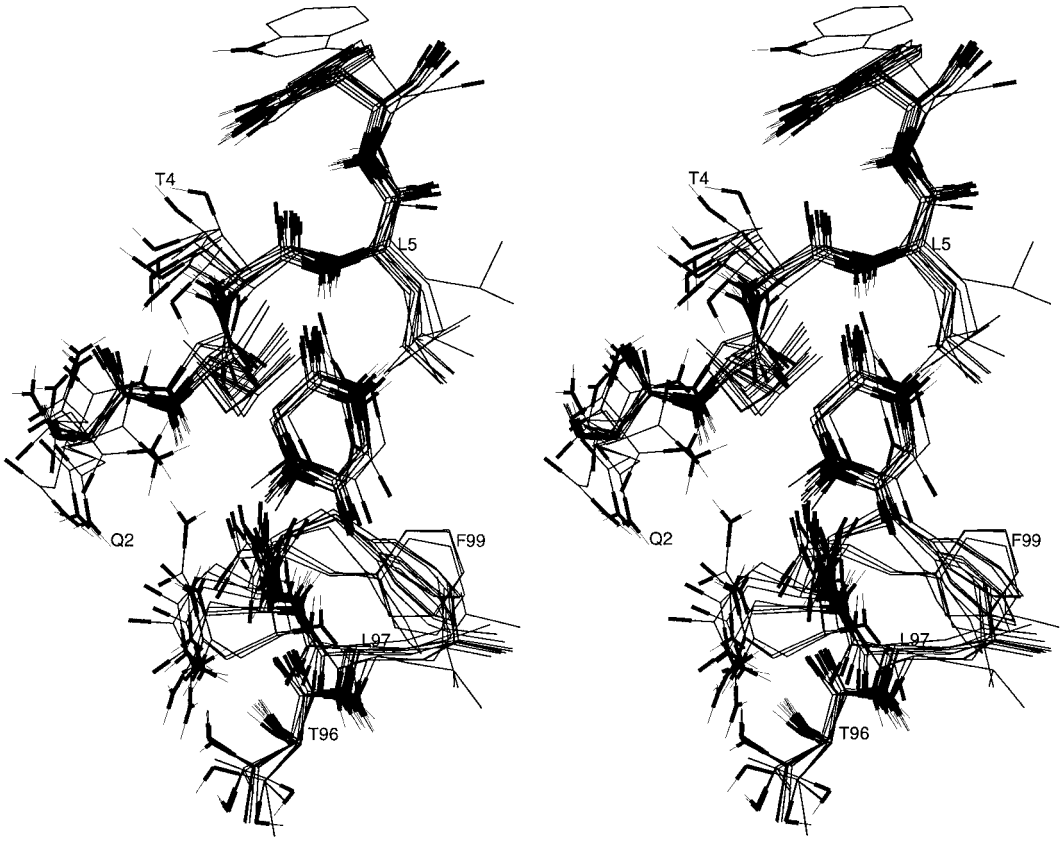
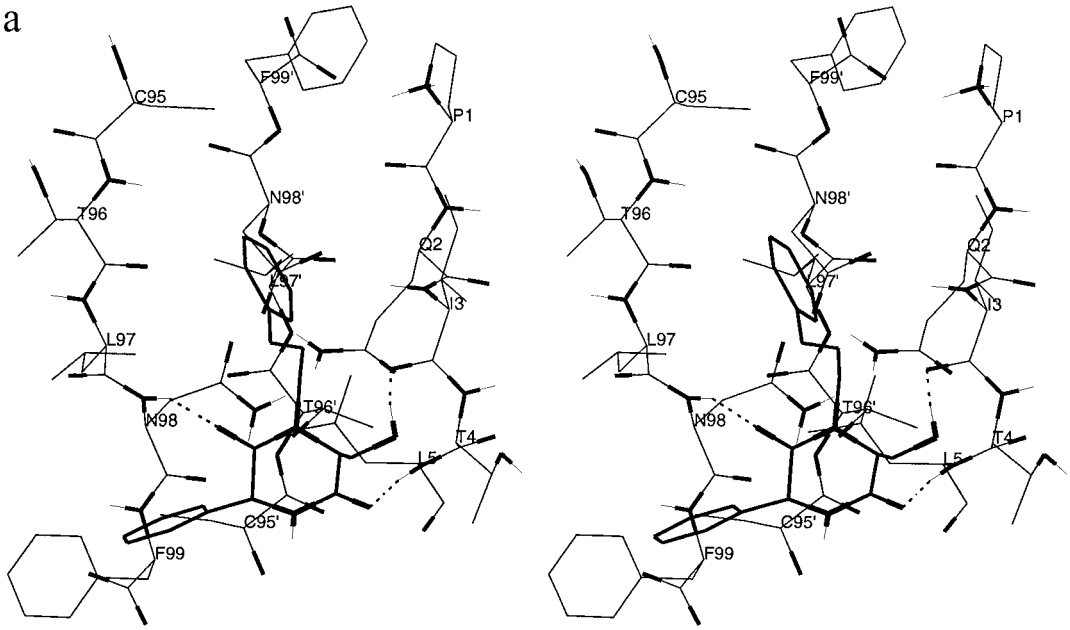


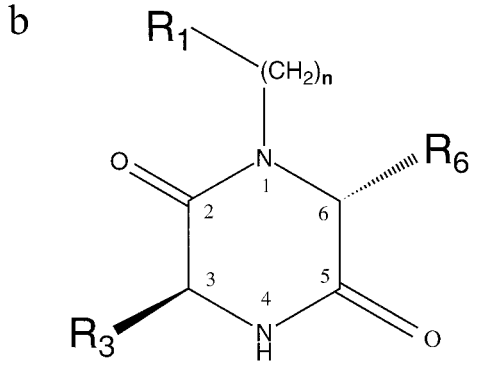
Figure 4. Stereoview of the conformational space sampled by the molecular dynamics and minimization procedure. The figure shows the 11 conformations of DKPI minimum 7 and the flexible region of the protein around it.

N-substituted diketopiperazines

DKPI is an interesting fragment that can be used for a combinatorial chemistry approach. Furthermore, it can dock between two β-strands (separated by one strand) by making up to three hydrogen bonds. DKPI minima 2 and 7 are docked between the N-terminal and C-terminal strands of the proteinase and can be substituted at their α-carbons and at the N-methyl group. In a restricted CCLD run only hits that contained one of the 25 quenched MD conformations of DKPI minima 2 and 7 were saved. This yielded a set of putative ligands with interesting binding motifs. A preliminary restricted run which utilized only the DKPI minimum positions 2 and 7 and did not use the 25 quenched MD conformations did not generate any hit. Thus, the conformations sampled by the quenched MD procedure were essential for the generation of hits in the restricted runs. Figure 5a shows an interesting compound (**I**) which makes use of the DKPI scaffold to connect the benzene rings matching the Cys95′ and Leu97′ side chains and the methanol group close to the Thr96′ side chain. The carbonyl groups of DKPI accept hydrogen bonds from the main chain NH of Leu5 and Asn98 (see above), while its NH group is accessible to solvent. Visual analysis of other CCLD hits suggested a set of putative ligands which might be synthesized by a combinatorial approach starting from a series of standard and N-substituted amino acids. As shown in the schema in Figure 5b the DKPI scaffold might be substituted at the C_α atom in position 3 with an aromatic side chain (R_3), and at the C_α atom in position 6 with a short polar side chain (R_6). The nitrogen substituent (R_1) can be either propylbenzene (as in Figure 5a) or an aliphatic moiety which would occupy the hydrophobic pocket of the Leu97′ side chain.

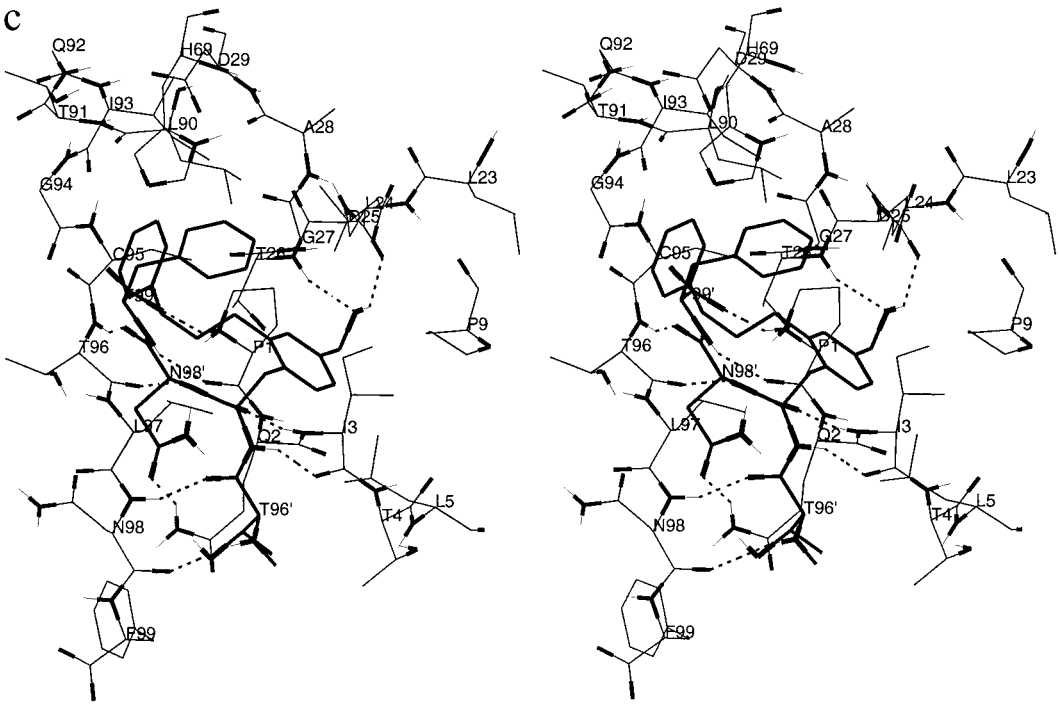
Alternatively, it is clear from Figure 5a that the CCLD hit **I** without the R_1 extension could be attached as an N-terminal fragment to the C-terminal peptide Leu_{97}-Asn_{98}-Phe_{99} or the modified peptide Tyr_{97}-Glu_{98}-Leu_{99} from the consensus sequence [16].



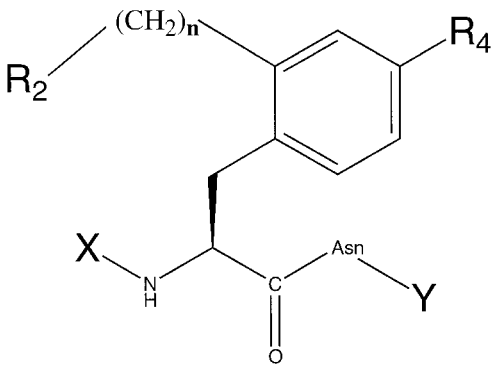


R_3 =	R_6 =	R_1 =
CH_2-benzene (Phe)	**CH_2OH (D-Ser)**	**benzene**
CH_2-phenol (Tyr)	$CHCH_3OH$ (D-Thr)	cyclopentane
CH_2-imidazole (His)		cyclohexane
CH_2-indole (Trp)		norbornane
benzene		adamantane
phenol		
imidazole		n = 2
indole		**3**

Figure 5. Putative ligands suggested by CCLD (thick lines) in the HIV-1 PR monomeric structure (thin lines). (a) Ligand **I** is a 2,5-diketopiperazine derivative. It is shown in the minimized structure of the HIV-1 PR monomer obtained by the quenched MD procedure started from DKPI minimum 7. The C-terminal pentapeptide of the second chain is shown as a basis of comparison. (b) 2,5-Diketopiperazine derivatives with trans substituents at C-3, C-6, which could be synthesized by a combinatorial approach. The substituents in bold characters are those of ligand **I**. (c) Ligand **II** is a modified C-terminal tetrapeptide. It is shown in the native structure of the HIV-1 PR monomer. (d) Series of ligands of type **II**. As a basis of comparison the four C-terminal residues of wild-type HIV-1 PR are Thr_{96}-Leu_{97}-Asn_{98}-Phe_{99}. The substituents in bold characters are those of ligand **II**.



d



X = Ser **Thr**

Y = Leu **Phe**

R_2 = **benzene**
cyclopentane
cyclohexane
norbornane
adamantane

R_4 = **CH_2OH**
CHCH3OH
$CHOH(CH_2)_2$-benzene
$CHOH(CH_2)_2$-cyclopentane
$CHOH(CH_2)_2$-cyclohexane

n=2
3

Figure 5 (continued).

Modified C-terminal peptides

A restricted CCLD run used the 14 quenched MD conformations of benzene minimum 1 and generated a series of polysubstituted benzenes. Some of these are interesting candidates for modifications of the C-terminal tetrapeptide with Phe at position 97. The putative ligand shown in Figure 5c (**II**) is a modification of the Thr_{96}-Phe_{97}-Asn_{98}-Phe_{99} peptide in which the aromatic ring of Phe_{97} is substituted in the ortho and para positions. A propylbenzene at the ortho position reaches the hydrophobic region occupied by Leu5′. The CH_2OH moiety at the para position participates in hydrogen bonds with the main chain C=O of Leu24 and the NH of Thr26.

A series of compounds analogous to ligand **II** is shown in Figure 5d. Among these, the compounds with the CH_2OH para-substitution at Phe are similar to the modified C-terminal peptide Ac-Ser-Tyr-Glu-Leu which has an inhibition constant of HIV-1 PR dimerization of 290 nM [16]. They have the additional ortho substitution at Phe which might improve binding affinity.

Discussion

When HIV-1 PR active-site inhibitors are administered to patients there is a rapid onset of resistance by point mutations [8, 9]. Dimerization or dissociative inhibitors are expected to induce less mutations in the viral genome than active-site directed ligands since the association of the two HIV-1 PR subunits is essential to form the composite active site. The dissociation constant of HIV-1 PR depends on both pH and temperature. Lower pH and temperature stabilize the dimer. Values of the dissociation constant range from 39 pM at pH 6.0 and 25 °C [53] to 130 nM at pH 7.5 and 37 °C [54]. Because of the very strong dimerization and the shifting of the dissociation equilibrium toward the dimer state by substrate stabilization, competitive substrate inhibition is easier to achieve [54].

In this study the MCSS/CCLD approach was used to design candidate ligands for the dimerization interface of HIV-1 PR. The binding free energy of each protein-MCSS minimum complex was estimated by supplementing the CHARMM energy with electrostatic and nonpolar solvation contributions calculated in a continuum approximation. For the MCSS minima with good binding free energy a procedure based on quenched molecular dynamics and energy minimization was used for additional sampling with a partially flexible protein binding site. The actual flexibility of the monomeric chain is expected to be larger than in the present molecular dynamics protocol. On the other hand, peptidic ligands designed on the basis of a rigid conformation of the monomer have been shown to inhibit dimerization in the submicromolar range [16–18].

A set of apolar fragments were predicted to fill the hydrophobic pockets which are occupied only partially by the Cys95′ and the Leu97′ side chains. This suggested two sets of putative ligands some of which are similar to a modified C-terminal peptide with nanomolar inhibition constant of HIV-1 PR dimerization.

Because of their synergistic action [16], dimerization inhibitors could be used as components of therapeutic ‘cocktails’ [55].

Acknowledgements

We thank Dr Claus Ehrhardt for helpful discussions. We gratefully acknowledge A. Widmer (Novartis, Basel) for providing the molecular modelling program WITNOTP and Prof. J. A. McCammon for the UHBD program (version 5.1), which was used for all finite-difference Poisson–Boltzmann calculations.

This work was supported by the Swiss Federal Office of Public Health (Nationales Aids-Forschungs-Programm, grant nr. 3139-043652.95 to A.C.), the Swiss National Science Foundation (grant No. 31-53604.98 to A.C.), the Helmut Horten Foundation (grant to A.C.), the U.S.A. National Science Foundation (grant to M.K.), and the E.C. program BIOMED-2 (grant to H.J.S. and M.K.).

References

1. Kohl, N.E., Emini, E.A., Schleif, W.A., Davis, L.J., Heimbach, J.C., Dixon, R.A.F., Scolnick, E.M. and Sigal, I.S., Proc. Natl. Acad. Sci. USA, 85 (1988) 4686.
2. Hoetelmans, R.M., Meenhorst, P.L., Mulder, J.W., Burger, D.M., Koks, C.H. and Beijnen, J.H., Pharm. World Sci., 19 (1997) 159.
3. Patick, A.K., Boritzki, T.J. and Bloom, L.A., Antimicrob. Agents Chemother., 41 (1997) 2159.
4. Sardana, V.V., Schlabach, A.J., Graham, P., Bush, B.L., Condra, J.H., Culberson, J.C., Gotlib, L., Graham, D.J., Kohl, N.E. and LaFemina, R.L., Biochemistry, 33 (1994) 2004.
5. Erickson, J.W., Nat. Struct. Biol., 2 (1995) 523.
6. Baldwin, E.T., Bhat, T.N., Liu, B., Pattabiraman, N. and Erickson, J.W., Nat. Struct. Biol., 2 (1995) 244.

7. Gulnik, S.V., Suvorov, L.I., Liu, B., Yu, B., Anderson, B., Mitsuya, H. and Erickson, J.W., Biochemistry, 34 (1995) 9282.
8. Ala, P.J., Huston, E.E., Klabe, R.M., MCabe, R.M., Duke, J.L., Rizzo, C.J., Korant, B.D., DeLoskey, R.J., Lam, P.Y., Hodge, C.N. and Chang, C.H., Biochemistry, 36 (1997) 1573.
9. Ermolieff, J., Lin, X. and Tang, J., Biochemistry, 36 (1997) 12364.
10. Rayner, M.M., Cordova, B. and Jackson, D.A., Virology, 236 (1997) 85.
11. Brodt, H.R., Kamps, B.S., Gute, P., Knupp, B., Staszewski, S. and Helm, E.B., AIDS, 11 (1997) 1731.
12. Zhang, Z.Y., Poorman, R.A., Maggiora, L., Heinrikson, R.L. and Kézdy, F.J., J. Biol. Chem., 266 (1991) 15591.
13. Schramm, H.J., Nakashima, H., Schramm, W., Wakayama, H. and Yamamoto, N., Biochem. Biophys. Res. Commun., 179 (1991) 847.
14. Schramm, H.J., Billich, A., Jaeger, E., Rücknagel, K.P., Arnold, G. and Schramm, W., Biochem. Biophys. Res. Commun., 194 (1993) 595.
15. Babé, L.M., Rosé, J. and Craik, C.S., Protein Sci., 1 (1992) 1244.
16. Schramm, H.J., Boetzel, J., Büttner, J., Fritsche, E., Göhring, W., Jaeger, E., König, S., Thumfart, O., Wenger, T., Nagel, N.E. and Schramm, W., Antiviral Res., 30 (1996) 155.
17. Zutshi, R., Franciskovich, J., Shultz, M., Schweitzer, B., Bishop, P., Wilson, M. and Chmielewski, J., J. Am. Chem. Soc., 119 (1997) 4841.
18. Bouras, A., Boggetto, N., Benatalah, Z., de Rosny, E., Sicsic, S. and Reboud-Ravaux, M., J. Med. Chem., 42 (1999) 957.
19. Miranker, A. and Karplus, M., Proteins Struct. Funct. Genet., 11 (1991) 29.
20. Venable, R.M., Brooks, B.R. and Carson, F.W., Proteins Struct. Funct. Genet., 15 (1993) 374.
21. Zheng, Q. and Kyle, D.J., Proteins Struct. Funct. Genet., 19 (1994) 324.
22. Zheng, Q. and Kyle, D.J., Biorg. Med. Chem., 4 (1996) 631.
23. Caflisch, A., J. Comput.-Aided Mol. Design, 10 (1996) 372.
24. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S. and Karplus, M., J. Comput. Chem., 4 (1983) 187.
25. Lee, B. and Richards, F.M., J. Mol. Biol., 55 (1971) 379.
26. Padlan, E.A., Mol. Immun., 31 (1994) 169.
27. Loeb, D.D., Swanstrom, R., Everitt, L., Manchester, M., Stamper, S.E. and Hutchinson III, C.A., Nature, 340 (1989) 397.
28. Louis, J.M., Smith, C.A., Wondrak, E.M., Mora, P.T. and Oroszlan, S., Biochem. Biophys. Res. Commun., 164 (1989) 30.
29. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L., Clawson, L., Schneider, J. and Kent, S.B.H., Science, 245 (1989) 616.
30. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
31. Brünger, A. and Karplus, M., Proteins Struct. Funct. Genet., 4 (1988) 148.
32. Neria, E., Fischer, S. and Karplus, M., J. Chem. Phys., 105 (1996) 1902.
33. Caflisch, A., Miranker, A. and Karplus, M., J. Med. Chem., 36 (1993) 2142.
34. Miranker, A. and Karplus, M., Proteins Struct. Funct. Genet., 23 (1995) 472.
35. Caflisch, A. and Karplus, M., Persp. Drug Discov. Design, 3 (1995) 51.
36. Grootenhuis, P.D.J. and Karplus, M., J. Comput.-Aided Mol. Design, 10 (1996) 1.
37. Joseph-McCarthy, D., Hogle, J.M. and Karplus, M., Proteins Struct. Funct. Genet., 29 (1997) 32.
38. Gelin, B.R. and Karplus, M., Proc. Natl. Acad. Sci. USA, 72 (1975) 2002.
39. Caflisch, A. and Ehrhardt, C., In Veerapandian, P. (Ed.), Structure-Based Combinatorial Ligand Design. Structure-Based Drug Design, Marcel Dekker, Inc., New York, NY, 1997, pp. 541–558.
40. O'Shea, E.K., Rutkowski, R. and Kim, P.S., Science, 243 (1989) 538.
41. O'Shea, E.K., Klemm, J.D., Kim, P.S. and Alber, T., Science, 254 (1991) 539.
42. Still, W.C., Tempczyk, A., Hawley, R.C. and Hendrickson, T., J. Am. Chem. Soc., 112 (1990) 6127.
43. Sitkoff, D., Sharp, K.A. and Honig, B., J. Phys. Chem., 98 (1994) 1978.
44. Warwicker, J. and Watson, H.C., J. Mol. Biol., 157 (1982) 671.
45. Gilson, M.K. and Honig, B.H., Proteins Struct. Funct. Genet., 4 (1988) 7.
46. Bashford, D. and Karplus, M., Biochemistry, 29 (1990) 10219.
47. Davis, M.E., Madura, J.D., Luty, B.A. and McCammon, J.A., Comput. Phys. Commun., 62 (1991) 187.
48. Hermann, R.B., J. Phys. Chem., 76 (1972) 2754.
49. Chothia, C., Nature, 248 (1974) 338.
50. Honig, B. and Nicholls, A., Science, 268 (1995) 1144.
51. Ryckaert, J.P., Ciccotti, G. and Berendsen, H.J.C., J. Comput. Phys., 23 (1977) 327.
52. Klebe, G., Mietzner, T. and Weber, F., J. Comput.-Aided Mol. Design, 8 (1994) 751.
53. Grant, S.K., Deckman, I.C., Culp, J.S., Minnich, M.D., Brooks, I.S., Hensley, P., Debouck, C. and Meek, T.D., Biochemistry, 31 (1992) 9491.
54. Darke, P.L., Jordan, S.P., Hall, D.L., Zugay, J.A., Shafer, J.A. and Kuo, L.C., Biochemistry, 33 (1994) 98.
55. Reijers, M.H., Weverling, G.J., Jurriaans, S., Wit, F.W., Weigel, H.M., Ten-Kate, R.W., Mulder, J.W., Frissen, P.H., van Leeuwen, R., Reiss, P., Schuitemaker, H., de Wolf, F. and Lange, J.M., Lancet, 352 (1998) 185.
56. Kraulis, P., J. Appl. Crystallogr., 24 (1991) 946.