Multiple Copy Simultaneous Search and Construction of Ligands in Binding Sites: Application to Inhibitors of HIV-1 Aspartic Proteinase

Amedeo Caflisch, Andrew Miranker, and Martin Karplus*

Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Received December 1, 1992

Rational ligand design is a complex problem that can be divided into three parts: the search for optimal positions and orientations of functional groups in the binding site, the connection of such positions to form candidate ligands, and the estimation of their binding constants. Approaches for addressing the first two parts of the problem are described in the present work. They are applied to the construction of peptide ligands in the binding site of the human immunodeficiency virus 1 (HIV-1) proteinase. The primary objective is to test the method by comparison of the results with the MVT-101 complex structure for which coordinates are available; the results obtained with the liganded and unliganded proteinase structure are used to examine the utility of the latter for binding studies. A secondary objective is to show how to find new inhibitor candidates. The multiple copy simultaneous search (MCSS) method is utilized to search for optimal positions and orientations of a set of functional groups. For peptide ligands, functional groups corresponding to the protein main chain (*N*-methylacetamide) and to protein side chains (e.g., methanol, ethyl guanidinium) are used. The resulting *N*-methylacetamide minima are connected to form hexapeptide main chains with a simple pseudoenergy function that permits a complete search of all possible ways of connecting the minima. Side chains are added to the main-chain candidates by application of the same pseudoenergy function to the appropriate functional group minima. A set of 15 hexapeptides with the sequence of MVT-101 is then minimized by a Monte Carlo scheme, which allows for escape from local minima. Comparison of the MCSS results with the structure of MVT-101 in the HIV-1 binding site showed that all of its functional group positions correspond (within 2.4 Å) to some (usually more than one) MCSS minima. There were also many other low-energy MCSS minima which do not appear in any known inhibitors, e.g., methyl ammonium minima in the neighborhood of the catalytic aspartates. Among the 15 lowest minima are seven hexapeptides with the same main-chain orientation as the one found by X-ray crystallography for the inhibitor MVT-101 in the binding site and eight with the main chain oriented in the opposite direction; the latter tend to be more stable. [*Addendum*: These results are in agreement with recent high-resolution crystallographic data provided after the study was completed. They show that the MVT-101 binds in two orientations and that the published orientation represents the minor conformer. (M. Miller *et al.* Private communication.)] A set of terminal blocked dipeptides were constructed from low-energy MCSS minima at one open end of the HIV-1 aspartic proteinase binding site and their interactions with the protein were analyzed. It was shown that some of the dipeptides can be connected to known hexapeptide ligands. The paper demonstrates that the combination of a method for an exhaustive search of the binding site for functional group minima (MCSS) with a highly efficient method for constructing molecules from them provides a novel and effective approach to the theoretical design and docking of candidate peptide ligands. The results of the present analysis suggest several modifications of MVT-101 that may have increased affinity and/or specificity for the HIV-1 aspartic proteinase binding site.

1. Introduction

Simulation-based drug design is concerned with the prediction of ligands that bind strongly to key regions of biologically important molecules (e.g., enzyme active sites, receptor proteins) so as to inhibit or alter their activity. An immense effort has been and continues to be dedicated to developing methods by which drug design (or more properly, ligand design, since whether or not a ligand is a drug involves factors beyond the present concerns) can be made a more rational process. However, the field is still in its infancy as is evident from recent reviews,[1] as well as from earlier evaluations.[2,3] Thus, new ideas and methods for approaching the ligand design problem are needed.

The strategy for rational ligand design that is described here has three parts. The first is an efficient method for the search of known binding sites (or more generally receptor surfaces) for positions that interact strongly with a range of functional groups. The importance of functional groups has been stressed recently in an assessment of binding modes of inhibitors for trypsanosomal triosephosphate isomerase.[4] Second, given a set of such positions and orientations for functional groups, it is necessary to connect the functional groups to form molecules that are candidates for synthesis. Finally, a method is needed to estimate which of the resulting molecules are likely to have the strongest binding constants. In the present paper the first two problems are addressed; the third problem will be considered in a subsequent paper. The stepwise procedure introduced here is more efficient than doing everything at once; i.e., the first two steps, which need to be as inclusive as possible, can make use of relatively simple representations of the interactions; evaluation of the resulting candidates in the third step requires a more

0022-2623/93/1836-2142$04.00/0

sophisticated and time-consuming treatment of the interactions that should be applied to only a limited set of molecules.

To solve the first problem, the multiple copy simultaneous search (MCSS) method was developed.[5] It is based on a combination of Monte Carlo and energy minimization/quenched dynamics techniques and makes possible the efficient determination of energetically favorable positions and orientations of functional groups in the binding site of a protein whose three-dimensional structure is known or can be modeled. Of particular importance in the MCSS method is its use of chemically realistic functional groups that permit determination of their expected orientations, as well as their positions. It provides CH_3–X bond vectors that can be used to connect the minima to form candidate ligands. Also, the MCSS method with a quenched dynamics protocol allows for flexibility of the binding site. The method can aid in the design of molecules that incorporate the functional groups by data base searches, modification of known ligands, and/or *de novo* construction.

Preliminary tests of the MCSS method were made on the influenza coat protein hemagglutinin, and it was demonstrated that certain of the functional group minima correspond to those of the bound sialic acid ligand.[5] To provide an in-depth test of the method, this paper examines peptide binding to the HIV-1 proteinase. The problem was chosen for its intrinsic interest and because the structures and binding constants of a series of peptide or peptide–analogue inhibitors are available for comparison. The primary purpose of the MCSS test presented here is to compare functional group binding sites found by the method with known functional group sites from inhibitors whose crystal structures are available. In addition, new functional group sites and ligands are suggested. Reference to other modeling studies of HIV-1 proteinase are given in the review of Kuntz.[6]

The HIV-1 proteinase binding site obtained from the complex with MVT-101[7,8] was used to generate functional group sites with the MCSS method. For comparison, a more limited analysis based on the unliganded proteinase structure was made.[9] Since the present test is concerned only with peptide ligands, the functionality maps obtained with the MCSS method were used directly to construct the molecules of interest; i.e., the backbone was synthesized simply by joining *N*-methylacetamide (NMA) minima and the side chains by joining various functional groups to the backbone. All possible ways of connecting the NMA minima were evaluated with a simple pseudoenergy function that leads to reasonable peptide bonds and eliminates bad contacts. The resulting main chains were clustered by a root-mean-square (RMS) deviation criterion and a representative of each of the clusters was selected. The MCSS minima for functional groups that correspond to the side chains of MVT-101 were then connected to these representatives by use of the same pseudoenergy function. Monte Carlo minimizations (MCM)[10–14] were then performed with a polar hydrogen potential on the resulting peptides to find energy minima. These were compared with the structure of the ligand MVT-101 in the binding site. Comparisons were also made with the coordinates of a number of other ligands supplied by Drs. J. Erickson and A. Wlodawer. In addition, some dipeptides that bind very differently from the known inhibitors were constructed and analyzed.

There exist several methods that are related to various aspects of the approach used here. One is the GRID program[15] which is the best known and most widely used method for functional group searches. GRID determines favorable binding sites with an interaction grid based on an empirical energy function. It uses simple spherical ligands and has emphasized positional information, although very recently specific hydrogen-bonding interactions have been considered.[15b–d] Another is the GROW program of Moon and Howe. Moon and Howe[16] utilized a template set of *in vacuo* generated amino acid conformations (constructed as *N*-acetyl-*N'*-methylamide) and iteratively pieced them together by amide end group superposition in a model of the target receptor. Their growth algorithm (GROW) corresponds to a tree process in which each library template is attached to the seed (or to each actual construct) and the search space is kept under control by pruning according to an energy evaluation. The main difference between the method of Moon and Howe and the present approach for constructing peptides is the use in the latter of predetermined optimal positions and orientations of functional groups in the receptor binding site.

Another related approach is that embodied in the program LUDI.[17,18] It makes use of statistical data from small-molecule crystal structures to determine possible binding geometries for a ligand that interacts with hydrogen bonding and hydrophobic sites of the receptor. These are then linked together with organic fragments in a manner that is stated to be similar to that used in the GROW program. The present paper differs from LUDI in that the functional group positions and orientations in the binding site are obtained directly by the MCSS method.

Section 2 describes the methods used for the binding site search and for the construction and minimization of the peptides. The results are presented and analyzed in section 3. The conclusions are outlined in section 4.

2. Methods

In this section we first summarize the MCSS method as implemented in the present study. We then outline the approach employed to construct peptides from the MCSS minima, and finally, the procedure used for obtaining locally minimized peptide positions that can be compared with known peptide ligands.

2.1. MCSS Method. The multiple copy simultaneous search (MCSS) method[5] determines energetically favorable positions and orientations (local minima of the potential energy) of functional groups on the surface of a protein of known three-dimensional structure. The present implementation is very similar to that used in the original description.[5] Several thousands replicas of a given group are randomly distributed inside a sphere whose radius is chosen large enough to cover the entire region of interest; in the present case this is the active site of the HIV-1 proteinase. Between 2000 and 5000 replicas of each group were distributed in a spherical region of 12-Å radius centered on the side chain of one of the two active site aspartate residues (see below). Groups characterized by an interaction energy less favorable than a given cutoff were discarded; a cutoff of 5 kcal/mol was used. In a test with an initial distribution of 50 000 NMA groups and an energy cutoff of 500 kcal/mol no additional sites were found after MCSS minimization. Thus, the number of replicas and the energy cutoff values used in this study provide sufficient coverage of the binding site. Details of the

Table I. Partial Charges for Atoms in MCSS Groups[a]

NMA		acetamide		methanol		acetate		methyl ammonium		ethyl guanidinium	
CH_3	0.05	CH_3	0.00	CH_3	0.25	CH_3	−0.16	CH_3	0.25	CH_3	0.00
C	0.55	C	0.55	O	−0.65	C	0.36	N	−0.30	CH_2	0.10
O	−0.55	O	−0.55	H	0.40	O1	−0.60	H1	0.35	N	−0.40
N	−0.35	N	−0.60			O2	−0.60	H2	0.35	H	0.30
H	0.25	H1	0.30					H3	0.35	C	0.50
CH_3	0.05	H2	0.30							N1	−0.45
										H1	0.35
										H2	0.35
										N2	−0.45
										H1	0.35
										H2	0.35
total	0.00		0.00		0.00		−1.00		1.00		1.00

[a] All charges are in unit of e.

method are given in the original paper.[5] The CHARMM[19,20] program was utilized for all minimizations performed in this work.

The MCSS can also be performed by energy minimization or quenched molecular dynamics of replicas in the force field of a flexible protein, which could have the essential parts (e.g. certain side chains) replicated, as well. However, the main aim of the present study was to find all optimal positions and orientations (energy minima) of a range of functional groups in a known receptor site and to connect such minima to form specified ligands. Hence, use of a fixed protein permits a more direct interpretation of the computed binding modes and their comparison with the known structure. Further, this is a good approximtion for the HIV-1 proteinase because its binding site is very similar for different ligands; the RMS deviation of the 198 C_α atoms for the complex with MVT-101 is 0.65 and 0.37 Å relative to the complex with inhibitors A-74704[21] and JG-365,[22] respectively.

The total energy (i.e., intrareplica and replica–protein) was included in the minimization. At the end of the minimizations, positions characterized by an interaction energy larger than one-half of the solvation enthalpy of the particular functional group were discarded (cf. refs 5 and 23).

The MCSS methodology was applied to *N*-methylacetamide (NMA), acetamide, methanol, acetate, methyl ammonium, ethyl guanidinum, propane, and isobutane. Values of the parameters for NMA were derived from that of the peptide group; for acetamide, methanol, acetate, ethyl guanidinium, and methyl ammonium they were derived from those for asparagine, threonine, aspartate, arginine, and lysine, respectively; and for propane and isobutane they were derived from isoleucine and valine, respectively; the charges used are given in Table I.

The major difference between the present study and the original MCSS application[5] consists of the extension to larger functional groups and to more complex aliphatic molecules, whose introduction was motivated by the structure of the ligands under consideration. In the original study, no flexible dihedral angles were present in the functional groups to eliminate sampling problems. The largest functional group used here, ethyl guanidinum, consists of 11 atoms and has one soft internal dihedral in the polar hydrogen approximation. Tests with two different starting conformations of the ethyl tail yielded the same minimized positions. This indicates that the MCSS minimization procedure is able to treat the flexible part of the ethyl guanidinum groups; i.e., the computed minima are independent of the choice of the starting conformation.

With the eight groups listed above it is possible to define the peptide backbone (NMA) and 12 among the 19 standard amino acid side chains (within an eventual addition or deletion of a CH_2 or CH_3 group, see next subsection and Figure 1). Most of these are needed to construct the inhibitor MVT-101, whose formula is *N*-acetyl-Thr-Ile-Nle-ψ[CH_2-NH]-Nle-Gln-Arg-amide. Although not needed for MVT-101, acetate (Asp, Glu), methyl ammonium (Lys), and isobutane (Val) were included for constructing dipeptide ligands.

2.2. Construction of Peptides. Since the present test application of the ligand design methodology is concerned with generating peptide ligands, the MCSS positions and orientations for the chosen functional groups in the binding site can be connected directly to build complete molecules. The backbone is formed by connecting NMA minima, and the side chains are generated by attaching the various functional groups to the main chain. In constructing the peptide ligands, the functional groups were kept fixed in their respective minimized positions and a simple pseudoenergy function, E_{ps}, was used to evaluate all possible ways of connecting the groups. The pseudoenergy, E_{ps}, is made up of two parts,

$$E_{ps} = E_b + E_{nb} \quad (1)$$

Here E_b represents a "bonding" interaction and is used to determine whether two groups have relative positions that permit them to be joined together; E_{nb} eliminates interacting groups with bad steric contacts. The use of fixed functional group positions makes possible the rapid construction of possible ligands. However, because the groups are fixed, a rather permissive pseudoenergy criterion has to be used in joining them together. Subsequent minimization (see below) serves to rationalize the internal structure of the ligands with satisfactory bonding geometry.

The bonding energy, E_b, is defined as

$$E_b = \sum_{i<j} k_b |r_{ij}^2 - r_0^2| \quad (2)$$

where the sum runs over pairs of atoms (i,j), that are superposed or bonded together in the ligand; the atoms are separated by a distance r_{ij}, with i and j in different functional groups. For the connection of different NMA minima (main-chain construction), i represents the acetyl carbon of one NMA and j represents a *N*-methyl carbon of an adjacent NMA (see Figure 1a); accordingly the atoms are superposed (r_0 equals 0 Å). During side-chain connection, if i and j represent covalently bonded atoms, an r_0 value of 1.5 Å is used (as for the C_α atom of a main chain and the extended carbon atom (C_β) of a methanol group

Figure 1. (a) Schematic representation of the connection of three NMA minima; the C_α of group i and the *N*-methyl carbon of group $i+1$ are superimposed. (b) Schematic representation of the connection of MCSS minima to construct the inhibitor MVT-101; aliphatic hydrogens are omitted. The bonds between MCSS minima are dashed. Extended atom carbon positions, which were added to complete the side chains are indicated by a "g" (see text).

in the case of a threonine sidechain (see P_3 in Figure 1b)); whereas if i and j represent atoms separated by two covalent bonds, an r_0 value of 2.5 Å is utilized (as for the C_α atom of a main chain and the extended carbon atom (C_γ) of an acetamide group in the case of a glutamine side chain (see P_2' in Figure 1b)).

The nonbond energy E_{nb} is defined as

$$E_{nb} = \sum_{i<j} k_{nb}(r_{ij,0}^2 - r_{ij}^2) \qquad (3)$$

where the sum runs over all pairs of atoms (i,j), with $r_{ij} < r_{ij,0}$ with i and j belonging to different functional groups and separated by at least three covalent bonds in the full ligand. Contact distances were obtained by adding the van der Waals radii of atoms i and j, taken from PARAM 19; i.e., the values 0.8, 1.9, 1.6, and 1.6 Å were used for H, C, N, and O atoms, respectively. The $r_{ij,0}$ values were determined from the pair parameters by scaling them down by a factor of $2^{1/6}$, which gives the distance where repulsion balances attraction in the Lennard-Jones potential.[24] This softens the pseudoenergy function. It is used for determining possible connections between fixed groups, whose relative positions are improved in subsequent minimization steps.

For the main-chain evaluations, values for the force constants k_b and k_{nb} were chosen equal to 1.0 kcal/(mol Å^2). This yielded a total of 8034 hexapeptide mainchains characterized by a pseudoenergy smaller than E_{cut} = 50 kcal/mol. Tests with a k_b value of 2 kcal/(mol Å^2)(k_{nb} of 1.0 kcal/(mol Å^2)) yielded 2933 main chains with E_{cut} of 75 kcal/mol, while a k_{nb} value of 2.0 kcal/(mol Å^2) (k_b of 1.0 kcal/(mol Å^2)) generated 58 501 main chains with E_{cut} of 75 kcal/mol. The first set of pseudoenergy parameters was selected because the number of main chains was appropriate for further clustering. For the side-chain connections (made after clustering of the mainchains), a k_b value of 2.0 kcal/(mol Å^2) was used (k_{nb} of 1.0 kcal/(mol Å^2) and E_{cut} of 50 kcal/mol for each side chain) since minimization of structures characterized by too large separations between the side chain and the corresponding C_α (as in the case of a k_b value of 1.0 kcal/(mol Å^2)) yielded conformations with bad internal geometries.

Given the pseudoenergy function, the construction of peptides was performed in three steps; main-chain generation, clustering of backbone structures, and side-chain attachment. Since the function (eqs 1–3) contains only positive contributions, it is possible to discard a partially built structure if its pseudoenergy E_{ps} is larger than the cutoff value, E_{cut}. The fixed functional group positions, the simple energy function, and the use of a cutoff for eliminating incomplete ligands makes the calculation very

fast. It is feasible, for example, to evaluate all possible ways of building terminal blocked hexapeptide backbones from the 83 NMA-minimized positions obtained from the MCSS method (see section 3.2.1); the time required on a single processor of an SGI 340 GTX is about 20 s (see also section 2.5). This would be an impossible task if all 83!/(83−7)! (about 2×10^{13}) complete hexapeptide main chains would have to be evaluated.

After possible main chains had been constructed, a clustering procedure based on their RMS deviations from each other was performed to reduce the number of peptides to be studied in detail. First, the RMS deviations between all pairs of structures are computed and for each structure the number of other structures within a given RMS is determined. The structure with the largest number of neighbors is selected as the representative of the first cluster, i.e., the cluster and its representative are selected at the same time by the same criterion. The structure with the next largest number of neighbors is chosen as the representative of the second cluster; none of the structures included in the first cluster can be a representative of the second or any subsequent cluster. The procedure is continued until all of the structures have been assigned to some cluster and have as a representative a similar structure. Because structures may belong to more than one cluster a set of disjoint clusters is formed by assigning a structure to the largest set in which it appears. Main-chain clustering reduced the number of plausible main chains to the set of their representatives (from 8034 to 104; see Results). The reduced set was used for side-chain attachment and minimization. A RMS deviation of 2.6 Å for the whole backbone was chosen as the similarity criterion in the clustering phase. Values of 2.0, 2.6, and 3.0 Å yielded 245, 104, and 65 main-chain representatives (i.e., clusters), respectively. The choice of 2.6 Å balances CPU time requirements for successive minimization with a maximal number of significantly different main chains. The same RMS value was utilized in the analysis of the final ligand conformations. The clustering method was also used in the analysis of the MCSS minima themselves, although all minima were included during main-chain construction and side-chain attachment.

Side chains for the selected set of backbone hexapeptide representatives were constructed in a manner similar to that used for the backbone. For each backbone C_α, the MCSS minimized position of the appropriate group with the lowest value of the pseudoenergy (eqs 1–3) after attachment was selected. The RMS distance values (3.0 Å) used for assessing MCSS minimized position as side-chain candidates are somewhat arbitrary; tests with a maximum distance of 3.5 Å in the Nle, Gln, and Arg side-chain evaluation yielded the same minima in most cases, and only three additional peptide structures.

Extended carbon atoms, representing side-chain C_β's and C_γ's, were generated by simple steric considerations when necessary to complete the side chain. Figure 1b shows how the MCSS functionalities define the side chains needed in the present application. To generate the structure of inhibitor MVT-101, only functional group minima corresponding to the given side chain were used; e.g., for the first amino acid, which is a Thr, only the methanol MCSS minima were considered and for the third and fourth side chain (Nle P1 and P1′), only the propane minima were used. During the generation of optimal dipeptide ligands at one end of the binding site, the MCSS minima of all functional groups were taken into account for each C_α (see section 3.4).

2.3. Minimization of Peptides. The method for connecting MCSS minima yields ligands structures that contain some stretched bonds and a few bad contacts. To obtain ligands with more realistic internal coordinates, the ligands were minimized in the force field of the complete static proteinase. Because of the relatively short nonbonded cutoff (7.5 Å) and the fact that the protein is rigid, the minimizations are very efficient. Five-hundred iterations of the steepest descent algorithm followed by 3000 conjugate gradient[25] steps were performed. The convergence criterion was a gradient of 0.01 kcal/(mol Å). Both the strain in the bonds and the bad contacts were relieved in the first (steepest descent) stage of the minimization, but the convergence criterion was reached only during the second stage (conjugate gradient). At the end of the minimization, a subset of ligand conformations characterized by a total energy (intraligand plus ligand–protein interaction) lower than a given cutoff value was chosen for further study; a somewhat arbitrary cutoff value of −250 kcal/mol was employed.

Given this subset of ligands (there were 15), the best minima were determined by use of the MCM procedure,[10-14] which combines the power of the Metropolis Monte Carlo method[26] in global optimization and that of the conjugate gradient in local optimization. Each MCM cycle begins by random changes ($-180° < \tau < 180°$) involving n torsion angles, τ, where n is chosen with a probability 2^{-n}, (cf. refs 13 and 14). The n affected torsion angles are randomly selected from among all the variable dihedral angles. The resulting structure is then minimized by 100 conjugate gradient steps. The Metropolis criterion[26] at a temperature of 310 K[13] is then applied to this new local minimum. A set of such cycles yields several local energy minima and the lowest lying one is selected for further analysis. The utility of such an approach for oligopeptide docking to proteins has been demonstrated.[13] First, 100 cycles (each cycle consists of a random perturbation, 100 conjugate gradient minimization steps, and comparison of the minimized structure with the last one accepted by the Metropolis criterion) involving random changes in all variable dihedrals of the ligand were performed. Then, 200 additional cycles involving only random perturbations in the side-chain dihedrals were used. The conformation characterized by the lowest total energy (intraligand plus ligand–protein) was saved for final analysis from each of the runs started from the 15 ligand conformations characterized by a total energy smaller than −250 kcal/mol. The intraligand energies of the selected subset of conformations ranged from −119.6 to −44.1 kcal/mol; furthermore, their bonding energy terms were low (i.e., they had acceptable geometries). To test the choice of the cutoff value, eight among the 48 MVT-101 conformations characterized by a total energy value larger than −250 kcal/mol were also minimized by the MCM scheme. At the end of the minimization their total energy was still higher than the minima obtained from all of the 15 original conformations.

The X-ray structure of MVT-101 in the rigid protein was submitted to the same minimization scheme as a basis for comparison.

2.4. Structures Used for Analysis. The crystal structures of synthetic free HIV-1 proteinase and synthetic HIV-1 proteinase complexed with the inhibitor MVT-101 were used. The former is a 2.8-Å resolution structure refined to an R factor of 0.184[9] (Brookhaven Protein Data Bank[8] entry 3HVP) and the latter is a 2.3-Å resolution

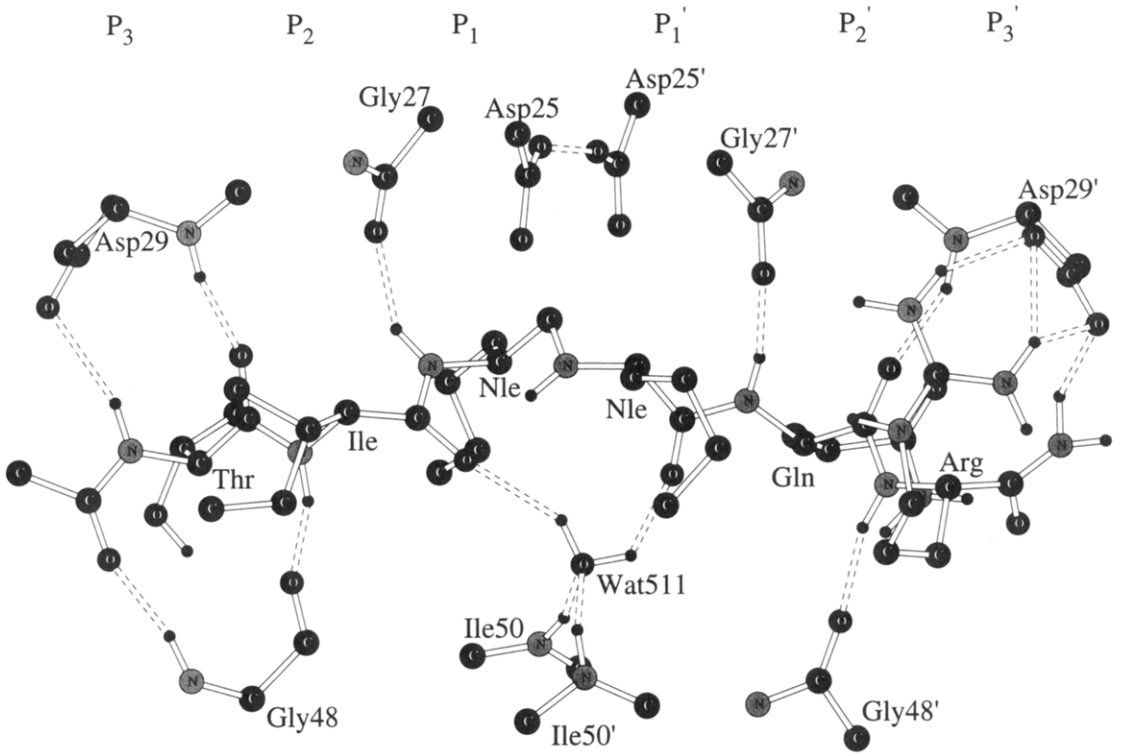


Figure 2. Schematic representation of the hydrogen-bond interactions (broken lines) between the inhibitor MVT-101 and HIV-1 aspartic proteinase.[7]

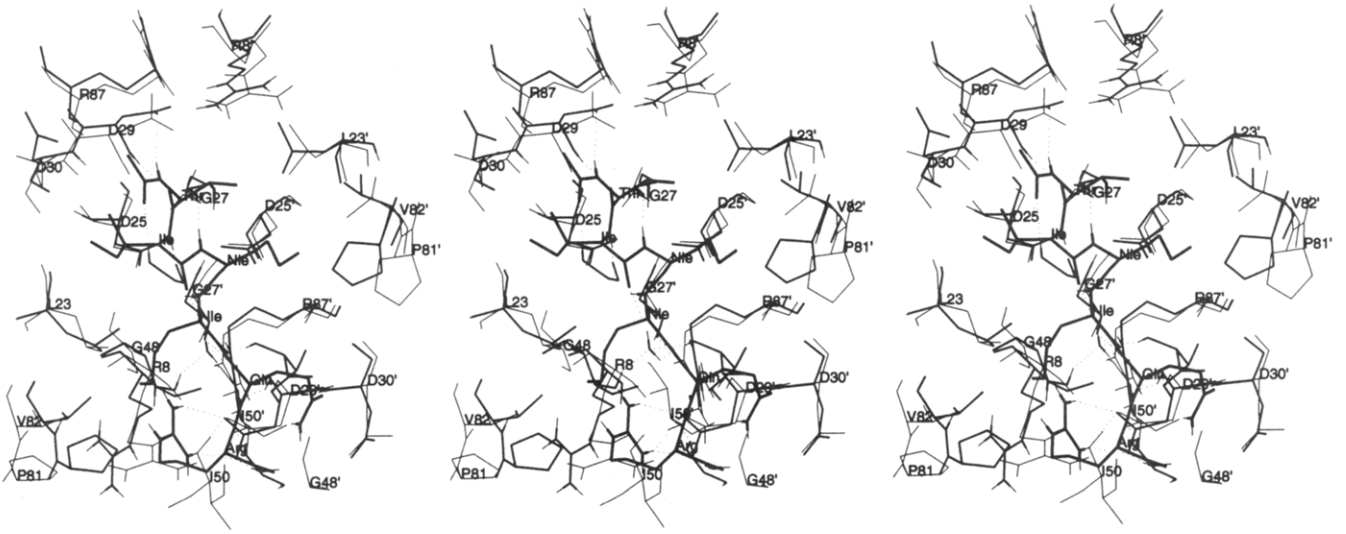


Figure 3. Stereoview of MVT-101 (thick lines) in complexed proteinase structure (thin lines). The native structure of the proteinase (medium lines) is also shown superimposed on the complexed conformation. C_α atoms of the native enzyme (one-letter code) and of the inhibitor (three-letter code) are labeled. (In all stereofigures, the view is cross-eyed for the left pair and wall-eyed for the right pair.)

structure with an R factor of 0.176[7] (Brookhaven Protein Data Bank entry 4HVP). Inhibitor MVT-101, which has a K_i value of 780 nM, is the reduced analogue of a hexapeptide substrate having the sulfur of the Met side chain replaced by a methylene group in Nle for synthetic simplification.[7] A stereoview of the inhibitor in the binding site is shown in Figure 3 and the hydrogen-bonding interactions of the inhibitor are shown schematically in Figure 2. There are seven main-chain–main-chain hydrogen bonds between the inhibitor and the proteinase binding site; here and throughout the present study, a cutoff of 3.5 Å for the heavy atoms distance and of 90° for the donor–hydrogen–acceptor angle are used as hydrogen-bond criteria. It is interesting that the polar Thr side chain at S3 does not have any specific interactions and appears disordered in the crystal.[7] As to the Gln side chain at S2′, it makes no hydrogen bonds; the closest polar groups of the proteinase are NH of Asp 30′ and CO of Gly 48′.

No minimization was performed on the crystal structures utilized for this study. Coordinates of the polar hydrogen atoms were generated with the HBUILD[27] option of the CHARMM program. Also, atom $O_{\delta 2}$ of the catalytic residue Asp 25′, which was chosen as binding site center for the MCSS simulations, was protonated to form a hydrogen bond with the side chain of Asp 25; i.e., a hydrogen bond is introduced in the plane of the four Asp 25 and Asp 25′ oxygens. This is motivated by the proximity (2.6 Å) between the carboxyl groups of the two catalytic aspartates.[7,28]

The distance between the two terminal blocking groups of MVT-101 from the active site center is approximately

Table II. Minima Found by MCSS

group	lowest energy[a]	highest energy[a]	cutoff energy[b]	no. accepted	no. discarded	no. of clusters[c]	largest cluster dimension	no. isolated minima[d]
NMA	−51.8	−10.3	−9.6	83	0	31	9	7
acetamide	−49.8	−12.3	−9.6	91	0	30	11	7
methanol	−38.9	−6.7	−5.1	72	0	36	7	13
acetate	−78.1	−0.9	−46.5	4	26	4	1	4
methyl ammonium	−107.7	−20.5	−37.5	20	8	8	5	2
ethyl guanidinium	−101.5	−18.1	−37.5	163	6	37	21	2
propane	−7.8	−2.6	−2.4	79	0	15	19	0
isobutane	−8.9	−3.7	−2.4	67	0	16	19	2

[a] Interaction energy with the HIV-1 aspartic proteinase binding site. All energy values are in kcal/mol. [b] The energy cutoff corresponds to one-half the solvation enthalpy of a particular group (see text); values were derived from refs 49 and 50. The values for the acetamide and ethyl guanidinium were set equal to the ones of NMA and methyl ammonium, respectively. [c] The clustering criterion is based on the all-atom RMS deviation (see text); only accepted minima were clustered. [d] Isolated minima do not have any other minima within 2.6-Å all-atom RMS deviation.

Table III. Aspartic Proteinase Side Chains-MCSS Minima/Inhibitor Polar Interactions[a]

enzyme residue	MCSS minima						inhibitors				
	acet[b]	mamm	eguan	NMA	acm	meth	JG-365	U-85548	A-74704	Ace-pepstatin	MVT-101
Arg 8	2[c]			7	7	6					**Arg P3′ (5)**
Asp 25		7	37	4	7	5	Phe P1 (2) Pro P1′ (1)	Leu P1 (2)	Phe P1 (3)	Sta P1 (2)	Nle P1′ (1)
Asp 29		3	36	16	18	11	Leu P3 (4) **Ser P4 (1)** **Asn P2 (1)**	**Gln P3 (4)** **Asn P2 (1)**	Cbz P3 (4)	Ala P1′ (3) Sta P2′ (2)	**Thr P3 (4)** Ace P4 (1)
Asp 30		1	16	11	15	5	**Asn P2 (2)** **Ser P4 (2)**	**Asn P2 (2)** **Ser P4 (1)**		Sta P2′ (3)	Ace P4 (1)
Arg 8′	2			13	15	12	Leu P3 (1)	**Gln P3 (3)**			
Asp 25′		8	44	12	16	2	Phe P1 (2)	Leu P1 (2)	Phe P1 (2) Phe P1′ (1)	Sta P1 (3)	Nle P1′ (1)
Asp 29′		4	43	17	15	15	Ile P2′ (3) Val P3′ (1)	Ile P2′ (2) Val P3′ (1)	Cbz P3′ (1)	Val P3 (3)	**Arg P3′ (5)** **Gln P2′ (2)**
Asp 30′		2	18	11	11	5			Cbz P3′ (1)	Ace P4 (1)	**Gln P2′ (3)**
Lys 45′				3	2	1					
Arg 87′	1			1		4					

[a] Charged binding side residues that have at least one N or O atom within 4.2 of any N or O atom of the ligand. Only accepted MCSS minima are considered. Polar and charged side chains of the inhibitors are in bold. Number of contacts with inhibitor residues are in parentheses. Inhibitors are listed from left to right in order of descending affinity, i.e., increasing inhibition constant. [b] Abbreviations used: acet, acetate; mamm, methyl ammonium; eguan, ethyl guanidinium; NMA, *N*-methylacetamide; acm, acetamide; meth, methanol; Ace, acetyl; Nle, norleucine; Sta, statine; Cbz, carbobenzyloxy. [c] Number of MCSS minima having at least one N or O atom within 4.2 Å of any N or O atom of the proteinase residue.

12 Å. Accordingly, this value was chosen as the radius of the sphere for the initial random distribution of groups for the MCSS. Such a sphere includes the whole enzyme active site, which can be approximated by a cylinder about 24 Å in length with a radius of 5 Å. All heavy atoms and polar hydrogens of the proteinase residues in a 20 Å spherical region centered on the $O_{\delta 2}$ of the Asp 25′ side chain were taken into account during the MCSS minimizations, while the field of the entire enzyme was considered when the MCM procedure was applied to the ligand.

2.5. Time Requirements. The MCSS minimizations for the eight functional groups took about 50 h of CPU time on a single processor of a Silicon Graphics Iris 340 GTX. The connecting and clustering of the groups found by MCSS procedure required less than 1 h of CPU time for the hexapeptide ligands. The Monte Carlo minimizations for 15 starting conformations of the hexapeptide inhibitor required about 45 h of CPU time. Thus, the whole procedure required about 4 days of CPU time. Since, for a static protein, the time required for MCSS and MCM depends almost linearly on the number of atoms in the ligand, the determination of the structures of large peptides (up to 12–14 residues) bound to receptors of nearly the same size as in the present application (about 200 residues) could be performed in less than 10 days of CPU time on single processor of a SGI 340.

3. Results and Analysis

3.1. MCSS Results. 3.1.1. Distribution of Energy Minima. Table II lists the number of minima and the range of their interaction energies with the HIV-1 proteinase. All minimized positions of the neutral groups were accepted in terms of the enthalpy cutoff, while 29%, 4%, and 87% of the methyl ammonium, ethyl guanidinium, and acetate minima, respectively, were discarded. The particularly low acceptance ratio of the acetate minima is due in part to the large solvation enthalpy and in part to the presence of several negatively charged residues in the proteinase binding site.

3.1.2. Polar and Charged Group Minima. The MCSS minimized positions were first clustered by the iterative algorithm described in section 2.2. Values of all-atom RMS deviations ranging from 2.0 to 3.0 were tested. For the 83 NMA minima, an RMS of 2.6 Å yielded clusters that had the best correlation with one obtained by an analysis of the replica-protein hydrogen bonds. Furthermore, a plot of the number of clusters versus the RMS deviation criterion (from 0 to 12 Å) showed a sigmoidal character for all of the MCSS groups, the region of maximal slope is located between 2.4 and 2.8 Å (results not shown). The clustering was performed only to facilitate the analysis of the results, since all accepted MCSS minima were included in the generation of ligand molecules (see section 2.2).

Table IV. NMA Clusters with More than One Member in the Complexed Proteinase Structure

cluster[a]	no.	minima in a given cluster[b]	interaction energy [kcal/mol]	HB-acc[c] from	HB-don[c] to	site
1	9	12, 16, 21–24, 26, 28–29	−34.7 to −29.6	Wat 511	Asp 25 or Asp 25′	S1–S1′
2	8	55–56, 64, 67, 69–70, 72, 74	−18.4 to −13.0	B 29 or B 30	Asp 29 or B 48	S3–S2
3	7	60–61, 65–66, 78, 80–81	−17.0 to −11.7	B 29′	B 48′	S2′–S3′
4	5	11, 13–14, 27, 31	−36.3 to −28.8	Wat 511	Asp 25 or B 27	S1
5	5	58–59, 62, 75–76	−17.4 to −13.0		B 27	S2
6	4	20, 34–36	−30.6 to −27.7	Wat 511	B 27′	S1′
7	4	46, 48, 51, 53	−25.2 to −21.4		Asp 29′	S3′
8	3	1–3	51.8 to −48.5	Lys 45′	Asp 30′	S4′
9	3	8–9, 32	−39.0 to −28.4		Asp 29′	S3′
10	3	15, 17, 30	−33.9 to −28.9	B 48	Asp 29 or Asp 30	S4
11	3	18–19, 25	−31.3 to −30.3	Arg 8′		S4
12	3	33, 38, 41	−28.5 to −27.4	Arg 8′		S4
13	3	63, 68, 82	−15.7 to −10.9	B 29	Asp 29	S3
14	2	4–5	−42.2 to −41.9	Arg 8	Asp 29′	S4′
15	2	6–7	−41.2 to −40.5	Arg 8	B 81	S4′
16	2	37, 39	−27.8 to −27.7	Arg 8′		S4
17	2	44, 47	−26.5 to −25.0	Arg 8′	B 48	S4
18	2	50, 52	−23.2 to −22.9		B 27	S2

[a] The clusters are ordered in decreasing size. [b] Minima are numbered in order of increasing interaction energy; i.e., no. 1 has the best interaction energy, no. 2 the next best, and so on. [c] Most of the NMA minima in a given cluster act as acceptors and/or as donors for one or more hydrogen bonds (HB) with the proteinase residues listed in the corresponding row. The letter B specifies a backbone hydrogen bond.

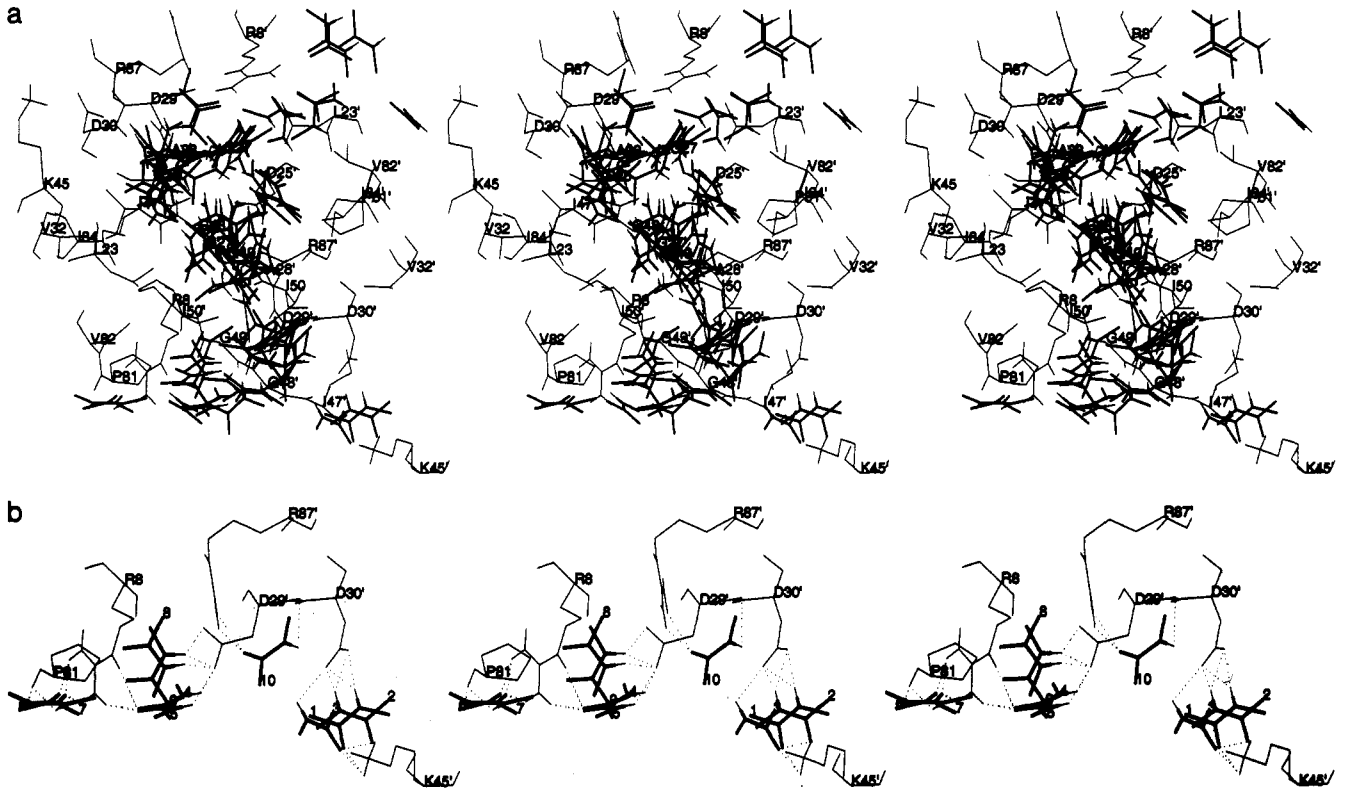


Figure 4. (a) Stereoview of the 83 MCSS-minimized positions of NMA (thick lines for the heavy atoms and thin lines for polar hydrogens) in the HIV-1 proteinase binding site of the liganded structure (thin lines). C_α atoms of the binding site residues are labeled. (b) Residues at one open end of the binding site (near S4′) and the ten lowest minima of NMA; the intermolecular hydrogen bonds between the NMA groups and the protein are dotted. These positions were utilized to design dipeptides with many strong interactions (see section 3.4).

Table III summarizes the polar and charged functional group interactions with charged amino acids in the binding site and at both open ends. There are low-energy minima in the neighborhood of all of the charged side chains. This contrasts with the known inhibitor structures, which have good interactions with only some of the charged groups; e.g., none of the inhibitors have charged interactions with the catalytic Asp 25 and 25′.

Polar groups are distributed over much of the binding site. They tend to form small clusters with less than three members; i.e., 30, 18, and 17 small clusters for methanol, NMA, and acetamide, respectively. Since the NMA functionality is the main-chain peptide group model, we consider it in detail. The 18 NMA clusters consisting of more than 1 member are listed in Table IV. It can be seen from Tables II and IV that the NMA minima have a wide range of interaction energies; the interaction energy can be as large in magnitude as 5 times the cutoff value. For a given cluster the interactions tend to fall in a relatively narrow range. Most, but not all, of the NMA minima in a given cluster have the same dipole orientation and the same polar interactions; for the partial charges, see Table I. Figure 4a shows all 83 NMA minima in the HIV-1 binding site; this figure, like the others which include a large number of minima, is meant to illustrate the distribution of groups in the site, rather than the position

Table V. NMA Clusters with More than One Member in the Free Proteinase

cluster[a]	no.	minima[b]	interaction energy [kcal/mol]	HB-acc[c] from	HB-don[c] to	site
1	7	17, 20, 22–23, 25, 27, 29	−27.1 to −26.8	Arg 8′		S4
2	7	18–19, 21, 24, 26, 28, 30	−27.1 to −26.8	Arg 8		S4′
3	6	73, 80, 88, 96, 99, 102	−14.0 to −10.7		B 48′	S3′
4	6	72, 81, 95, 100–101, 103	−14.0 to −10.7		B 48	S3
5	5	40, 43, 49–51	−21.0 to −16.8	Asp 25		S1
6	5	61, 63, 70, 86, 91	−15.8 to −11.5		B 48	S3
7	5	62, 64, 71, 77, 90	−15.8 to −11.5		B 48′	S3′
8	5	79, 89, 97–98, 104	13.3 to −9.9	Asp 25′		S1′
9	4	78, 92–94	−13.4 to −11.2		B 27	S2
10	3	6, 8, 32	−30.0 to −26.0	Arg 8		S4′
11	3	7, 9, 31	−30.0 to −26.0	Arg 8′		S4
12	2	1–2	−53.4 to −52.9	Lys 45′	Asp 30′	S4′
13	2	3–4	−35.9 to −34.0	Arg 87′		S4′
14	2	10, 16	−28.3 to −27.2	Arg 8′		S4

[a] The clusters are ordered in decreasing size. Only three among the 14 clusters consisting of two minima are listed. [b] Minima are numbered with respect to their interaction energy. [c] Most of the NMA minima in a given cluster act as acceptor or as donor for one or more hydrogen bonds (HB) with the proteinase residues listed in the corresponding row. The letter B specifies a backbone hydrogen bond.

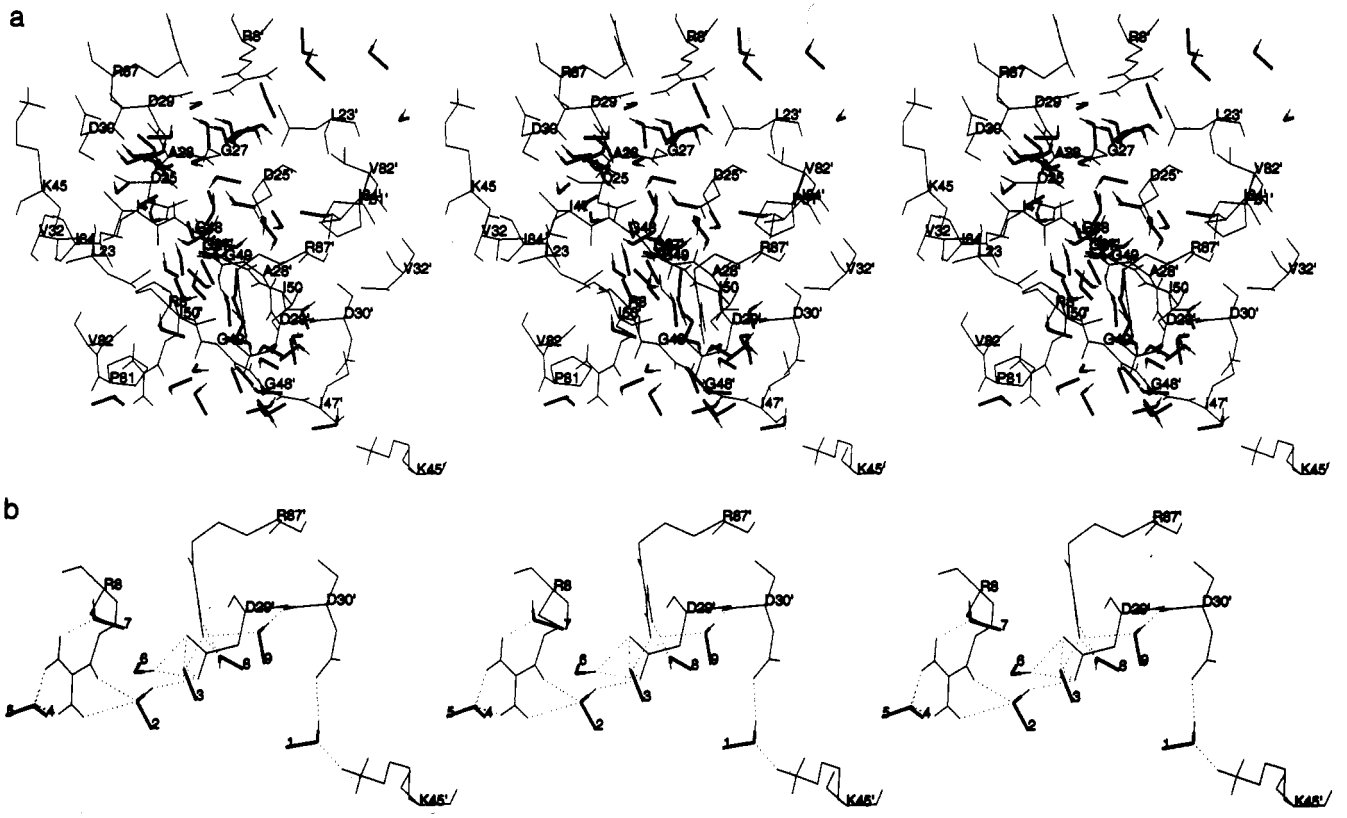


Figure 5. (a) The 72 MCSS-minimized positions of methanol. (b) The nine lowest energy methanol minima. (The same conventions as in Figure 4 are used.)

of any given minimum. Most of the minima are in regions of the binding site that correspond to the MVT-101 (see Figure 3). However, the 10 lowest energy NMA minima (Table IV) are at one open end of the binding site (near S4′), which consists of residues Arg 8, Asp 29′, Asp 30′, Lys 45′, Pro 81, and Arg 87′ (Figure 4b). They are involved in optimal polar interactions mainly with these charged protein side chains. The largest cluster is formed by 9 NMA minima interacting with both the catalytic aspartates (see Tables III and IV) and Wat 511. It is interesting that the inhibitor does not interact with the catalytic aspartates; it does hydrogen bond to Wat 511.

The acetamide functionality has approximately the same structure and charge distribution as NMA (see Table I). Accordingly, the minima are in essentially the same positions and make the same polar interactions with the protein. Even though acetamide was run with the MCSS method because it better represents a Gln or Asn side chain than NMA, the results are such that NMA could have been used.

Methanol minima can be found over much of the binding site and also at both open ends. They participate in interactions with most of the charged and polar side chains (see Table III), as well as with many proteinase main-chain NH and CO groups (Figure 5, parts a and b). In fact, the distribution of methanol minima is similar to that of the NMA functionality, as can be seen by comparing Figures 4 and 5. Most of the 20 lowest methanol minima are at the open ends of the binding site. Like NMA and acetamide, the methanol minima include a wide range of energies. The nine lowest energy minima are shown in Figure 5b, which indicates the hydrogen bonds. There are several minima for methanol that interact strongly with Asp 25 and Asp 25′. The lowest energy minima of NMA, acetamide, and methanol (i.e., the three polar groups tested) occupy the same site and act as acceptor for the

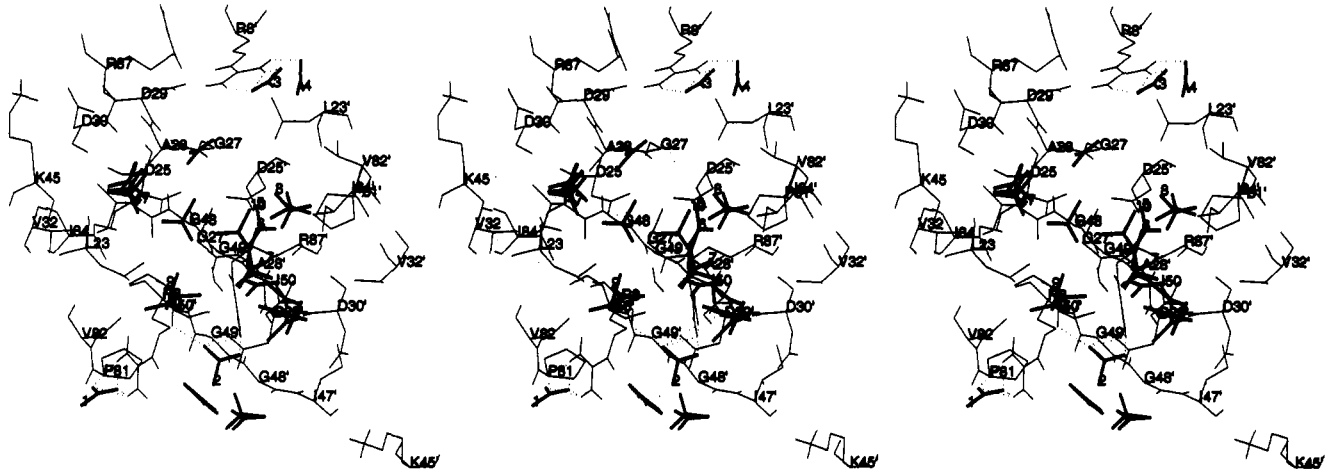


Figure 6. The 30 MCSS-minimized positions of acetate. C_α atoms of the binding site residues and methyl carbons of the 10 lowest energy acetate minima are labeled. Hydrogen bonds of the four accepted acetate minima are dotted. (The same conventions as in Figure 4 are used.)

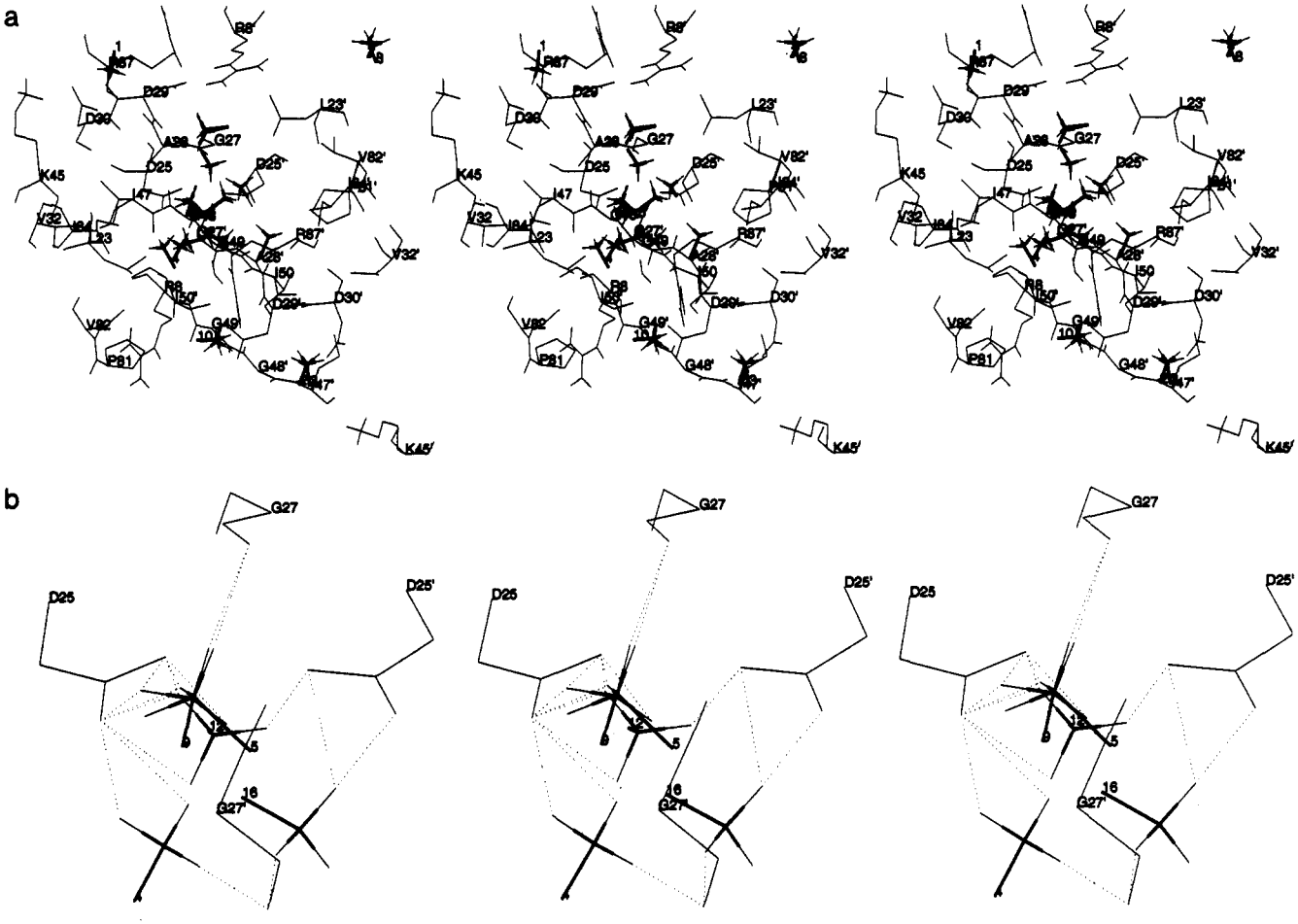


Figure 7. (a) The 20 MCSS-minimized positions of methyl ammonium. C_α atoms of the binding site residues and methyl carbons of the 10 lowest energy methyl ammonium minima are labeled. (b) Methyl ammonium minima 4, 5, 9, 12, and 16, which act as donors with three or four hydrogen bonds (dotted lines) to the catalytic aspartates and/or the backbone CO of Gly 27 or Gly 27′. (The same conventions as in Figure 4 are used.)

Lys 45′ side chain and as donor for the Asp 30′ side chain. The interaction energy of methanol in this position is 8 times the cutoff value (−38.9 kcal/mol), while for NMA it is 5 times the cutoff (−51.8 kcal/mol).

Acetate has relatively few highly favorable positions (Figure 6). Only four acetate minima, two at one open end of the binding site near Arg 8 and Arg 87′, and two at the other end near Arg 8′ (see labeled minima in Figure 6), have an interaction energy lower than the cutoff.

The positively charged functional groups (methyl ammonium and ethyl guanidinium) show a large number of strongly interacting minima. There is a narrower range of energies, relative to the cutoff, than for the polar groups. This is due primarily to the large cutoff energy for the charged groups. The interaction energies are very large in magnitude; the most favorable values are −107.7 kcal/mol for methyl ammonium and −101.5 kcal/mol for ethyl guanidium. The 20 accepted methyl ammonium minima can be grouped into eight clusters, the largest of which contains five minima in the vicinity of the catalytic residues Asp 25 and 25′ (Figure 7, parts a and b). The three lowest minima of methyl ammonium make a salt bridge with the

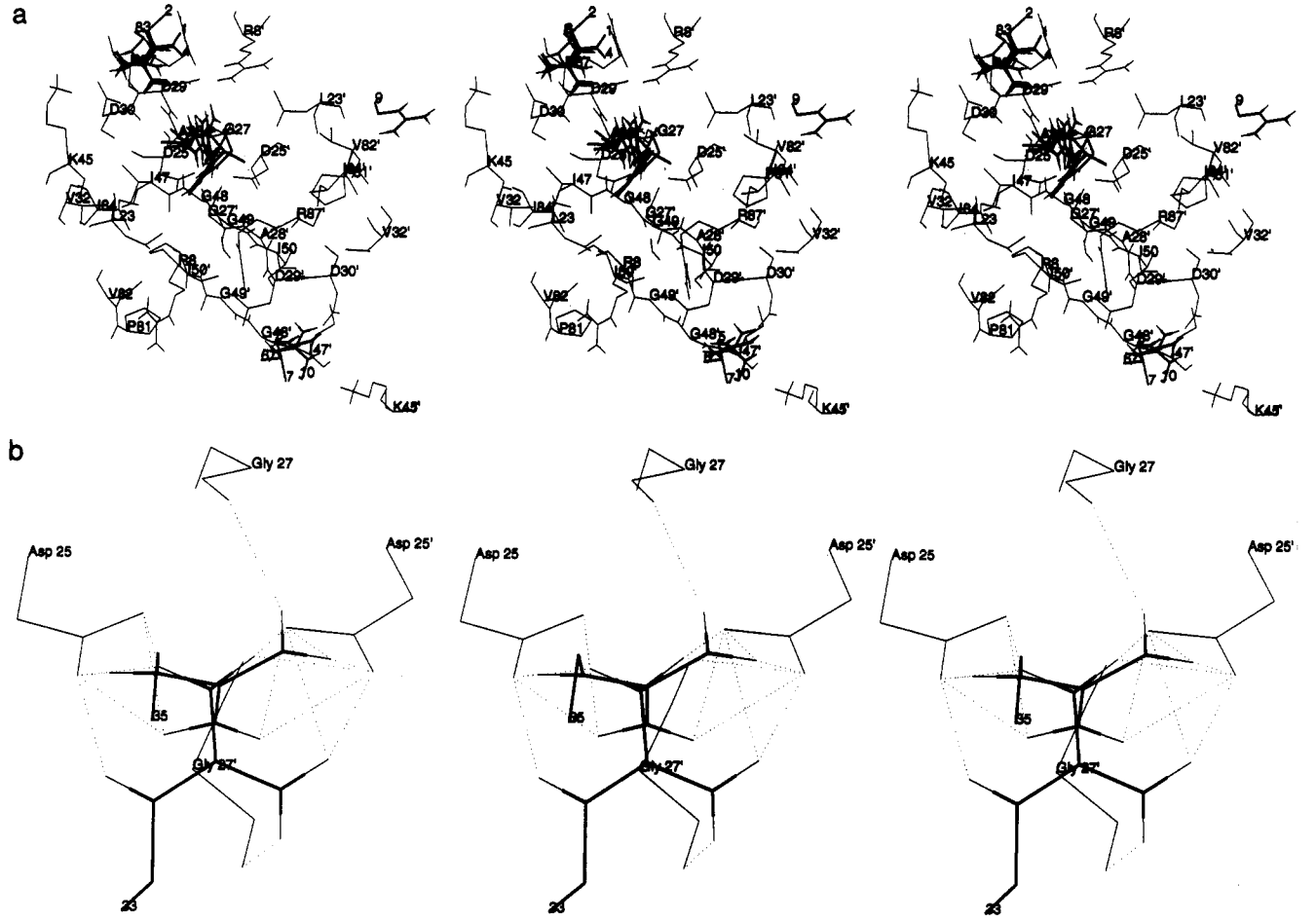


Figure 8. (a) The 10 lowest energy ethyl guanidinum minima and 10 among the 21 minima of the largest cluster. C_α atoms of the binding site residues and methyl carbons of the 10 lowest energy minima are labeled. (b) Ethyl guanidinium minima 33 and 35 each donate seven hydrogen bonds (dotted lines) to the catalytic aspartates and one to the backbone CO of Gly 27 or Gly 27′. (The same conventions as in Figure 4 are used.)

side chain of Asp 30 (site S4, minimum 1) or Asp 30′ (site S4′, minima 2 and 3). None of the eight discarded methyl ammonium positions interact with a charged group.

The largest number of minima (169) were found for the ethyl guanidinium functionality of which only six were discarded by the cutoff criterion. That there are so many minima is due primarily to the several possible orientations of both the ethyl and guanidinium groups in each favorable site (Figure 8, parts a and b). The largest cluster contains 21 minima interacting with the backbone CO group of Gly 27 and Gly 48 and with the side chain of Asp 29 at the S3–S2 site. The interaction energies associated with this cluster range from −54.5 to −39.1 kcal/mol. The second largest cluster is formed by 15 minimized positions interacting with the CO of Gly 27′ and the Asp 29′ side chain; their interaction energy ranges from −76.4 to −57.6 kcal/mol. The third and fourth largest clusters consist of 14 and 13 members, respectively; they interact with the catalytic aspartates. The four lowest minima of ethyl guanidinium participate in a double salt bridge with the Asp 29 and Asp 30 side chains (site S4, Figure 8a). Minima 5–7 are at the opposite end of the binding site (site S4′) where they make a double salt bridge with Asp 29′ and Asp 30′; their interaction energy ranges from −98.6 to −97.9 kcal/mol. Minimum 9 is isolated and has an interaction energy of −93.1 kcal/mol; it makes a salt bridge with the side chain of Glu 34′, which is located on the outside surface of the protein.

Although the three lowest minima of methyl ammonium and 19 among the 20 lowest minima of ethyl guanidinium interact with either Asp 29 and Asp 30 or Asp 29′ and Asp 30′ (at the two open ends of the binding site), several minima of both groups interact with the catalytic aspartates. This suggests that an ammonium group could be introduced into a ligand; e.g. the backbone methylene group in the P1 position could be replaced by a $CHNH_3^+$ or the methylene group could be joined to a $CH_2NH_3^+$ moiety.

3.1.3. Hydrogen Bonds. To analyze the contribution of hydrogen bonds to the interactions, a list of all intermolecular hydrogen bonds was generated with the same criteria as used in section 2.4. Both polar and charged group minima involve positions where several hydrogen bonds are present. There is no simple correlation between the number of hydrogen bonds and the interaction energy (Table VI), although the lowest energy minima of polar and charged groups usually participate in more hydrogen bonds then those with intermediate energy. Most of the discarded minima, with exception of those involving ethyl guanidinium, make one or no hydrogen bond. The hydrogen bonds formed by the lowest energy minima of NMA, methanol and acetate are shown in Figures 4b, 5b, and 6.

Almost all NMAs and acetamides act either as donor or as acceptor or as both with the charged proteinase side chains and with several backbone atoms of the binding site. Table III lists the proteinase side chains that are involved in hydrogen bonds with the NMA minima, while Figure 4b shows the hydrogen bonds formed by the 10

Table VI. Energy and Number of Hydrogen Bonds of the 10 Lowest Energy Minima[a]

minimum[b]	acetate		mamm		eguan		NMA		acetamide		methanol	
1	4	-78.1	2	-107.7	4	-101.5	4 (2)[c]	-51.8	3 (2)	-49.8	2 (1)	-38.9
2	5	-76.3	2	-100.6	4	-101.2	3 (2)	-49.5	2 (1)	-48.1	3 (1)	-29.0
3	3	-71.0	2	-100.2	4	-100.8	2 (1)	-48.5	3 (1)	-39.0	3 (2)	-28.0
4	1	-48.7	3	-71.2	4	-100.8	3 (1)	-42.2	3 (1)	-38.4	2 (0)	-26.9
5	2	-25.5	3	-70.7	5	-98.6	3 (1)	-41.9	2 (2)	-36.8	2 (0)	-26.8
6	4	-25.4	2	-70.4	5	-98.3	3 (1)	-41.2	4 (3)	-36.8	2 (2)	-26.3
7	2	-24.7	2	-69.2	5	-97.9	3 (1)	-40.5	2 (2)	-36.7	1 (0)	-22.2
8	1	-23.9	2	-69.0	4	-95.4	1 (1)	-39.0	2 (1)	-35.4	2 (0)	-22.0
9	2	-22.0	4	-68.9	4	-93.1	2 (2)	-37.4	3 (2)	-34.6	2 (1)	-21.9
10	3	-20.9	2	-68.3	4	-92.5	3 (1)	-37.1	2 (1)	-34.0	2 (1)	-20.6

[a] A cutoff of 3.5 Å for the heavy atoms distance and of 90° for the donor–hydrogen–acceptor angle were used. [b] Sorted by ascending interaction energy. [c] For groups that can act as both donor and acceptor, the total number of hydrogen bonds is listed first and the number of hydrogen bonds in which the group acts as a donor is given in parentheses.

Table VII. van der Waals Contacts of Propane

minima[a]	energy[b]	surface[c]	d	e	pocket[f]	proteinase residues					
1	-7.80	24.4	13	5	S2	Asp 29	Asp 30	Val 32	Gly 48		
2	-7.62	25.2	12	4	S2	Asp 29	Asp 30	Gly 48			
3	-7.48	19.8	8	2	S2	Asp 29	Asp 30	Gly 48			
4	-6.96	29.5	11	5	S2′	Ala 28′	Asp 29′	Asp 30′	Val 32′	Ile 47′	
5	-6.92	22.6	10	4	S2′	Asp 29′	Asp 30′	Val 32′	Ile 47′	Ile 50	
6	-6.87	30.6	13	5	S2′	Ala 28′	Asp 29′	Asp 30′	Val 32′	Ile 47′	Gly 48′
7	-6.51	46.1	18	6	S3	Arg 8′	Leu 23′	Gly 27	Asp 29		
8	-6.16	22.7	13	1	S1–S1′	Asp 25	Asp 25′	Gly 27	Gly 27′	Wat 511	
9	-6.16	26.4	12	1	S1–S1′	Asp 25	Asp 25′	Gly 27	Gly 27′	Ala 28	Wat 511
10	-6.16	49.4	14	4	S3	Arg 8′	Gly 27	Asp 29			
11	-6.12	44.8	15	5	S3′	Arg 8	Leu 23	Gly 27′	Asp 29′		
12	-6.06	38.5	18	6	S3	Arg 8′	Leu 23′	Gly 27	Asp 29		
13	-6.04	24.9	12	0	S1–S1′	Asp 25	Asp 25′	Gly 27	Gly 27′	Wat 511	
14	-6.00	69.8	18	8	Surface	Leu 5	Trp 6	Arg 87′	Asn 88′	Thr 91′	
15	-5.96	52.4	16	5	S3′	Arg 8	Leu 23	Gly 27′	Asp 29′		
16	-5.95	60.8	9	2	Surface	Glu 21′	Glu 34′	Val 82′	Asn 83′		
17	-5.85	23.1	9	1	S1′	Asp 25	Asp 25′	Gly 27′	Ile 84	Wat 511	
18	-5.85	23.6	7	1	S1′	Asp 25	Gly 27′	Ile 50′	Wat 511		
19	-5.84	23.8	11	3	S1′	Asp 25	Asp 25′	Gly 27	Ile 84	Wat 511	
20	-5.84	24.7	11	2	S1–S1′	Asp 25	Asp 25′	Gly 27	Gly 27′	Wat 511	

[a] The 20 lowest energy propane and isobutane minima are listed. [b] Interaction energy in kcal/mol. [c] Accessible surface area in Å^2. [d] Number of intermolecular heavy atom contacts (distance value smaller than 4.2 Å). [e] Number of intermolecular contacts with protein carbon atoms (distance value smaller than 4.2 Å). [f] The binding pocket is defined as the residues of the protein that have at least one atom within 4.2 Å of any atom of the ligand.

lowest energy NMA minima. For eight of these minima, the NMA group acts as both donor and acceptor and forms between 4 (lowest minimum) and 1 (third highest minimum) hydrogen bonds (minima 8 and 9 act only as donors); in the majority of cases there are more acceptor hydrogen bonds to the carbonyl than donor hydrogen bonds from the NH group. This is in accord with simulation studies, which have shown that C=O's tend to act as acceptors for two hydrogen bonds, while NH's act as donors of one hydrogen bond.[29]

In the methanol minima, the OH hydrogen often interacts with backbone carbonyl oxygens and with charged side chains, such as those of Glu 21, Asp 25, Asp 29, Asp 30, and the corresponding residues of the other chain of the proteinase (see Table III). The 10 lowest energy minima of NMA (Figure 4b), acetamide, and methanol (Figure 5b) form hydrogen bonds with the side chains of Arg 8, Lys 45′, Arg 87′ (donors) and Asp 29′, Asp 30′ (acceptors).

Among the 20 accepted methyl ammonium minima, 2, 4, 13, and 1 donate 4, 3, 2, and 1 hydrogens, respectively. Ten of these minima participate in at least one hydrogen bond with the catalytic residues Asp 25 and Asp 25′ (Figure 7a).

Most of the 163 accepted ethyl guanidinium minima make three or more hydrogen bonds; among these, 42 minima are involved in 1–7 hydrogen bonds with the catalytic residues Asp 25 and Asp 25′ (Figure 8b).

3.1.4. Apolar Groups. Apolar groups tend to bind in cavities where they have a large number of favorable van der Waals contacts with the protein. For propane and isobutane the results are very similar so we describe only the former. There is a distribution of interaction energies which peaks strongly at a value 2–3 times the cutoff value. However, the absolute range of interaction energies is relatively small and, in particular, the 20 lowest minima are very close in energy (see Table VII). The propane and isobutane groups yield 15 and 16 clusters, respectively. The two largest clusters of both apolar functionalities are at the S1 and S1′ subsites (Figure 9). This is in agreement with the requirement of hydrophobic residues at these subsites in most of the sequences of HIV-1 proteinase inhibitors.[30,31] Among the 20 lowest propane minima, two or more bind in each subsite of the proteinase binding pocket; the same is true for isobutane. Table VII lists the binding pockets of the best 20 propane minima; the proteinase residues involved correspond to the residues which contact the side chains of inhibitors MVT-101 and A-74704 (cf. Table 2 of ref 7 and Table 1 of ref 21). In many cases, polar and charged proteinase residues are involved and some of the contacts are to the aliphatic portions of these sidechains. (See Table VII, column 5.) The six lowest energy minima of propane and the four lowest minima of isobutane involve the S2 or S2′ subsites; S2 is the pocket of the Ile P2 side chain of the inhibitor.

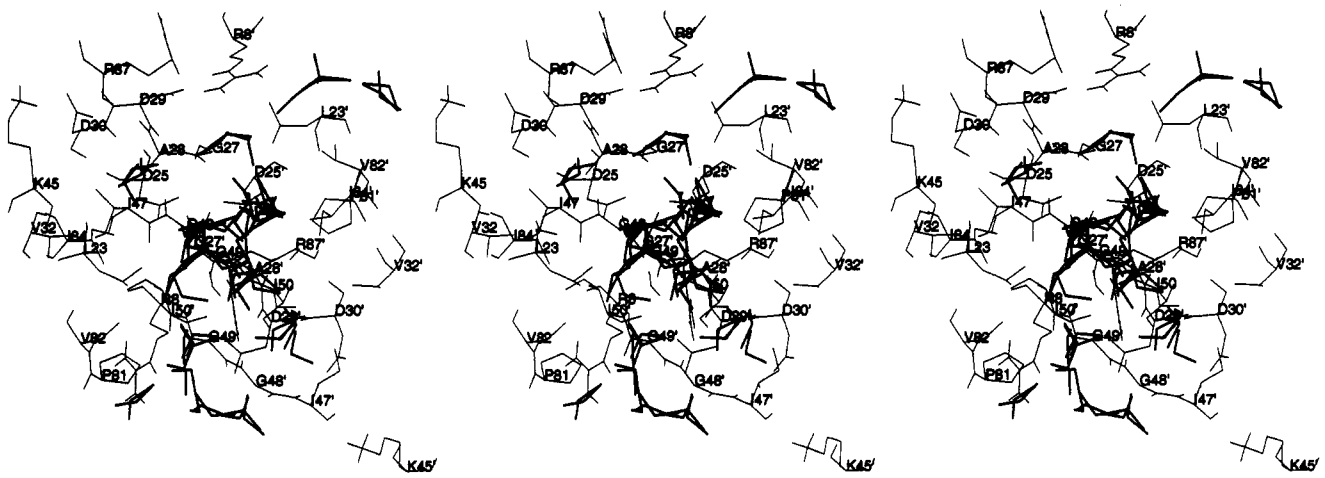


Figure 9. The 79 propane minima. (The same conventions as in Figure 4 are used.)

Table VIII

a. NMA Minima Found by MCSS Near the Inhibitor Backbone

AA site	Ace-Thr P4–P3	Thr-Ile P3–P2	Ile-Nle P2–P1	Nle-Nle P1–P1′	Nle-Gln P1′–P2′	Gln-Arg P2′–P3′	Arg·amide P3′–P4′
rank	30	55	40	11	20	78	49
energy[a]	−32.5	−22.0	−31.0	−39.9	−34.2	−16.1	−27.2
RMS[b]	2.1	2.0	1.9	2.4	0.9	1.2	1.8
c	2	4	4	1	6	3	1
d	2	4	0	0	0	3	1
e	3	2	2	6	1	3	0
E_0[f]	−28.1	−8.0	−23.7		−30.6	−9.2	−23.1
E_{min}[g]	−37.6	−22.0	−34.0		−34.2	−16.1	−27.2
rank[h]	15 (2.2)	55	22 (3.3)		20	78	49

b. Minima Found by MCSS Near the Inhibitor Side Chains

AA site	Thr P3	Ile P2	Nle P1	Nle P1′	Gln P2′	Arg P3′
group	methanol	propane	propane	propane	acetamide	ethyl guanidinium
rank	68	3	47	25	89	41
energy[a]	−9.3	−7.5	−5.1	−5.7	−25.1	−59.5
RMS[b]	1.8	1.8	0.9	2.0	2.1	2.4
E_0[f]	−7.7	2.3	−3.3	2.6	−13.1	−29.0
E_{min}[g]	−18.8	−7.5	−5.1	−5.9	−25.6	−59.5
rank[i]	20 (2.7)	3	47	18 (2.1)	85 (2.2)	41

[a] Interaction energy with the HIV-1 aspartic proteinase binding site. All values are in kcal/mol. [b] All-atom RMS deviations in Å between a NMA group found by MCSS and the nearest corresponding group in the experimental structure of the inhibitor. [c] Number of NMA minima within 2.6 Å of the experimental structure. [d] Same as in *c* for the complexed proteinase without Wat 511. [e] Same as in *c* for the native proteinase. [f] Interaction energy of the corresponding group in the inhibitor X-ray structure. [g] Interaction energy of the corresponding group in the inhibitor X-ray structure after conjugate gradient minimization. [h] Gives rank of minimum to which the NMA converged from the peptide group in the inhibitor; the RMS deviation of the different minima are given in parentheses (in Å). [i] Gives rank of minimum to which the group converged from the corresponding side chain group in the inhibitor; the RMS deviation of the different minima are given in parentheses (in Å).

The algorithm of Lee and Richards[32] was used in the CHARMM program to determine the accessible surface area of for the 10 lowest minima of propane and isobutane; a probe radius of 1.4 Å was chosen. The accessible surface area of isolated propane (isobutane) measures about 221 Å^2 (245 Å^2). Upon binding, the average value of the accessible surface area was 30 and 38 Å^2 for the 10 lowest energy minima of propane and isobutane, respectively; i.e., a significant reduction in accessible surface area takes place. The accessible surface area of the 20 lowest propane minima are given in Table VII.

3.1.5. Comparison of MCSS Minima with the MVT-101 Inhibitor. As a first step in evaluating the results, we compare the MCSS minima with the positions of corresponding moieties in the MVT-101 inhibitor. All of the backbone peptide groups of the inhibitor correspond to one or more NMA minima in the MCSS functionality map. (See Table VIII and Figure 10.) The intermolecular backbone hydrogen bonds of the experimental structure are present in several NMA minima. In particular, the NMA minima listed in Table VIIIa form the backbone donor and acceptor hydrogen bonds observed in the crystal structure of the inhibitor complex (Figure 10). Also, all inhibitor side-chain positions were found by at least one appropriate functional group with an RMS deviation of 2.4 Å or less. (See Table VIIIb and Figure 10.) It is likely that some of the deviations are due to the fact that the binding of the complete inhibitor is not ideal in terms of individual functional group interactions. The largest RMS deviation between inhibitor peptide group and nearest NMA minimum is at position P1–P1′ (2.4 Å, Table VIIIa); this discrepancy is due to the absence of the oxygen atom in the reduced inhibitor.

A conjugate gradient minimization of NMA replacing each peptide group of the inhibitor (except the reduced one) was performed in the static proteinase; a convergence criterion for the energy gradient of 10^{-4} kcal/(mol Å) was used. The initial interaction energies for the six groups are listed in Table VIIIa. The minimization of the inhibitor NMA positions yielded six among the 83 NMA minima

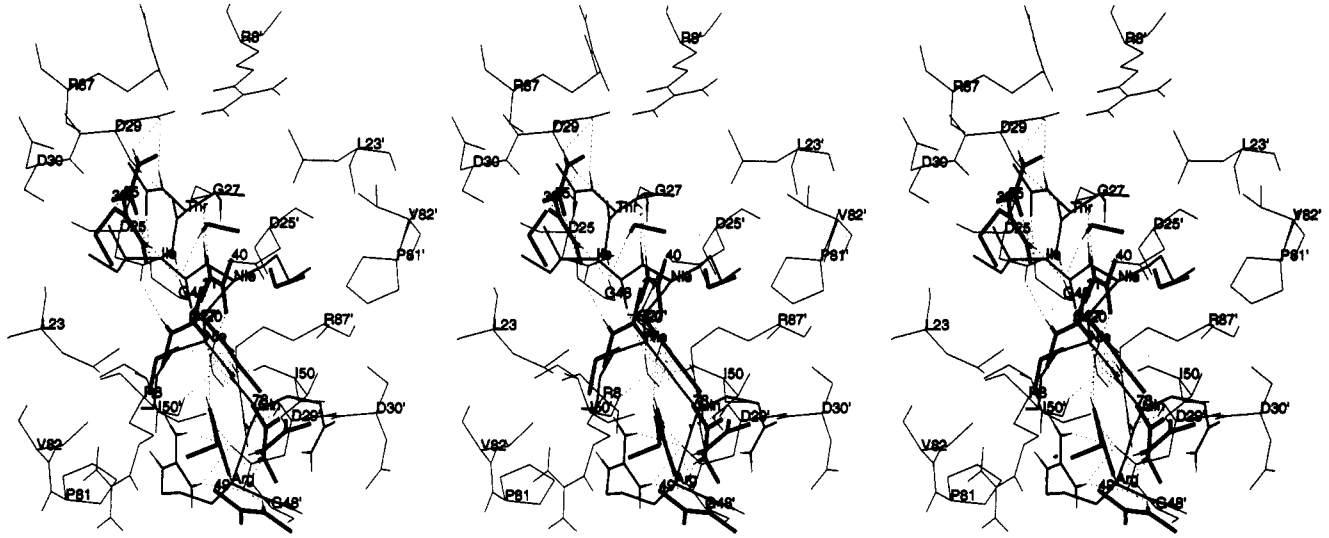


Figure 10. Stereoview of MCSS minima (thick lines) nearest to the corresponding groups of the crystal structure of MVT-101 (medium lines); binding site residues participating in hydrogen bonds and/or hydrophobic contacts with the inhibitor are shown (thin lines) together with the hydrogen-bonded water molecule (Wat 511). Intermolecular backbone hydrogen bonds are dotted. C_α atoms of the complexed proteinase (one letter code) and of the inhibitor (three letter code) are labeled, together with the α-carbon of the NMA-minimized positions.

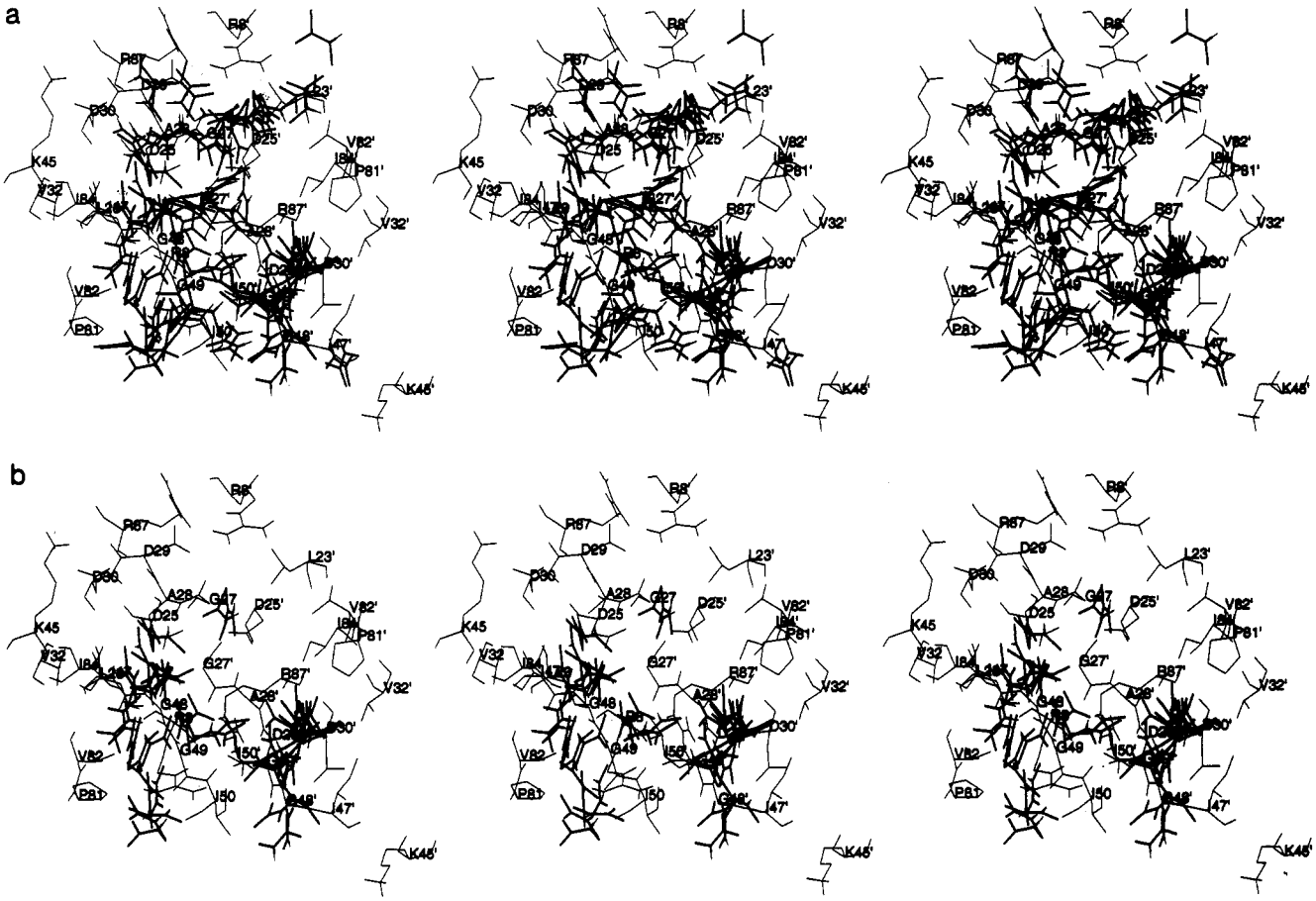


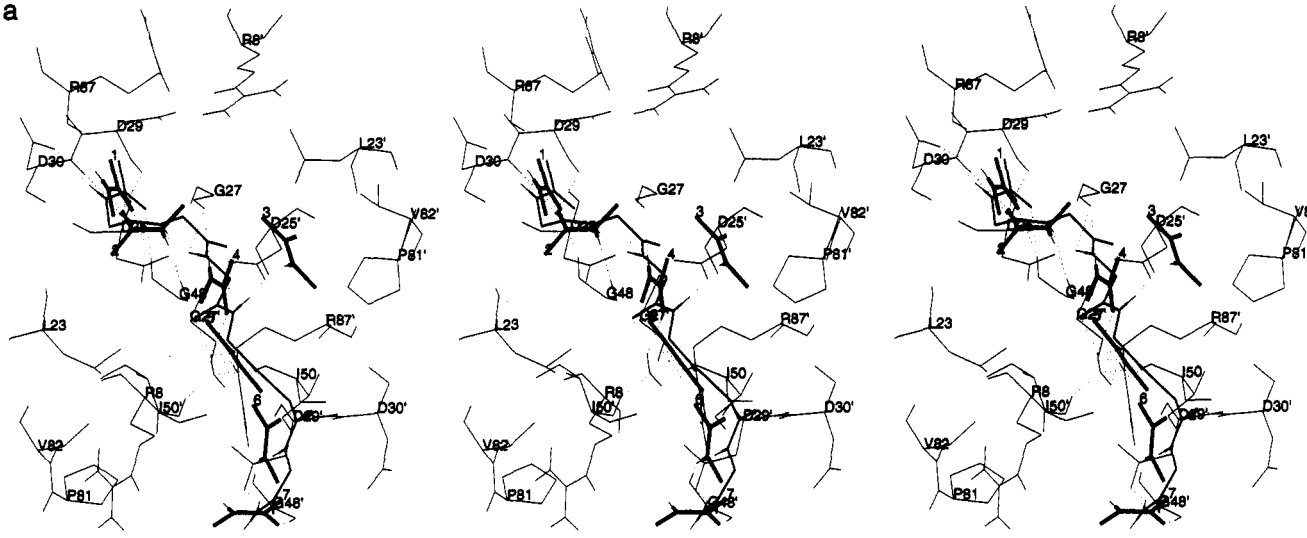
Figure 11. (a) Stereoview of the 104 accepted NMA minima (thick lines) in the binding site of the native proteinase (thin lines). (b) Stereoview of 49 NMA minima (thick lines) in the native proteinase (thin lines), which were not present in the complexed proteinase structure (within 2.6-Å RMS deviation).

found by MCSS (Table VIIIa); four of the minima are the same and two are different from the closest MCSS minimum. For the six side chains, three of the positions converged to the closest MCSS minimum on minimization and three did not (see Table VIIIb). Both for the main chain and the side chains, the different minima have better interaction energies with the protein and are only slightly further from the corresponding inhibitor position than the closest MCSS minimum.

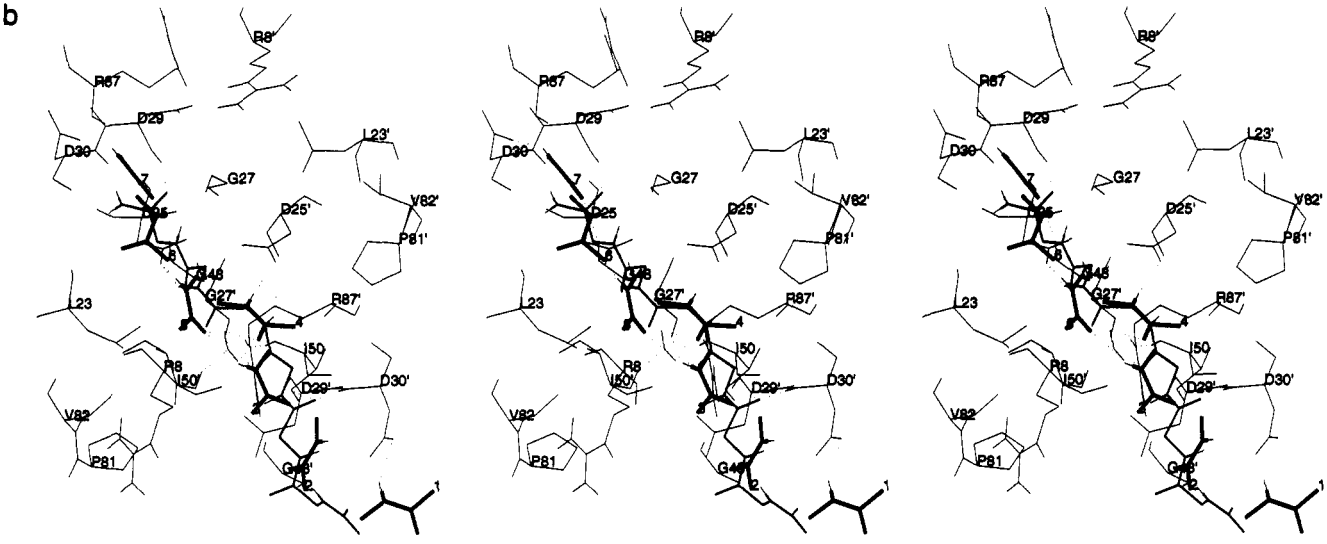
The best backbone and side-chain MCSS minima in terms of closeness to the hexapeptide inhibitor position are generally not the lowest energy minima for the functional groups; e.g. from Table VIIIa, the NMA minima rank between 11 and 78 relative to the total number of 83 minima within the cutoff. This makes it clear that all of the MCSS minima have to be considered in the construction of oligopeptides (section 3).

Four methanol minima donate a hydrogen bond to Asp

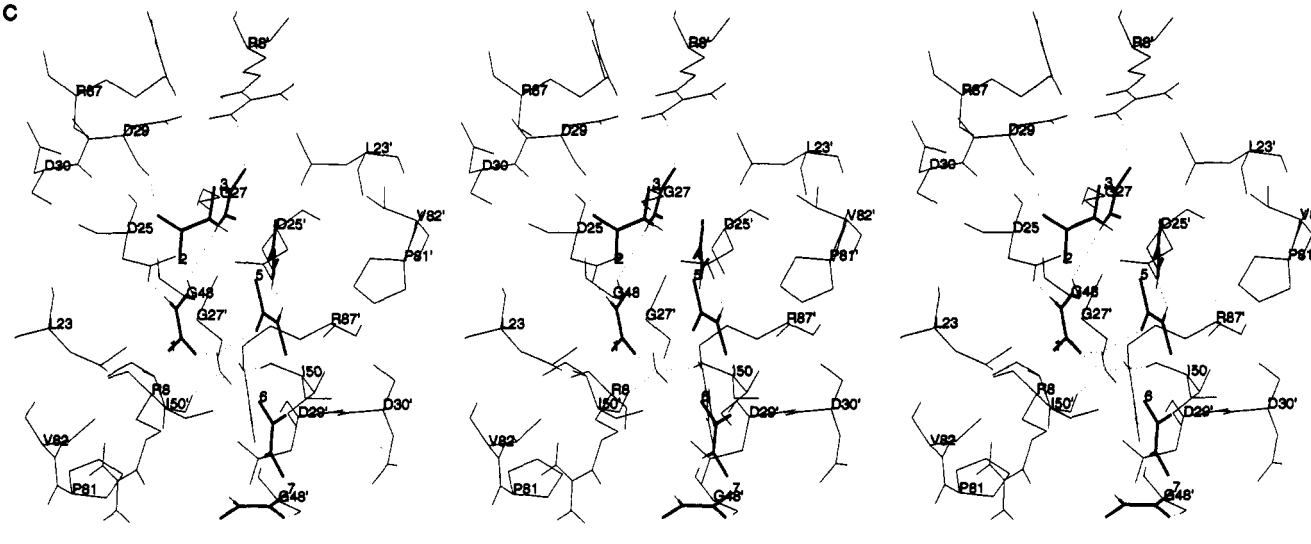
a


R8'
R67
D29
D30
L23'
G27
D25'
V82'
P81'
G48
L23
R87'
I50
R8
I50'
D29'
D30'
V82
P81
G48'

b


R8'
R67
D29
D30
L23'
G27
D25'
V82'
P81'
G48
G27'
L23
R87'
I50
R8
I50'
D29'
D30'
V82
P81
G48'

c


R8'
R67
D29
D30
L23'
G27
D25
D25'
V82'
P81'
G48
G27'
L23
R87'
I50
R8
I50'
D29'
D30'
V82
P81
G48'

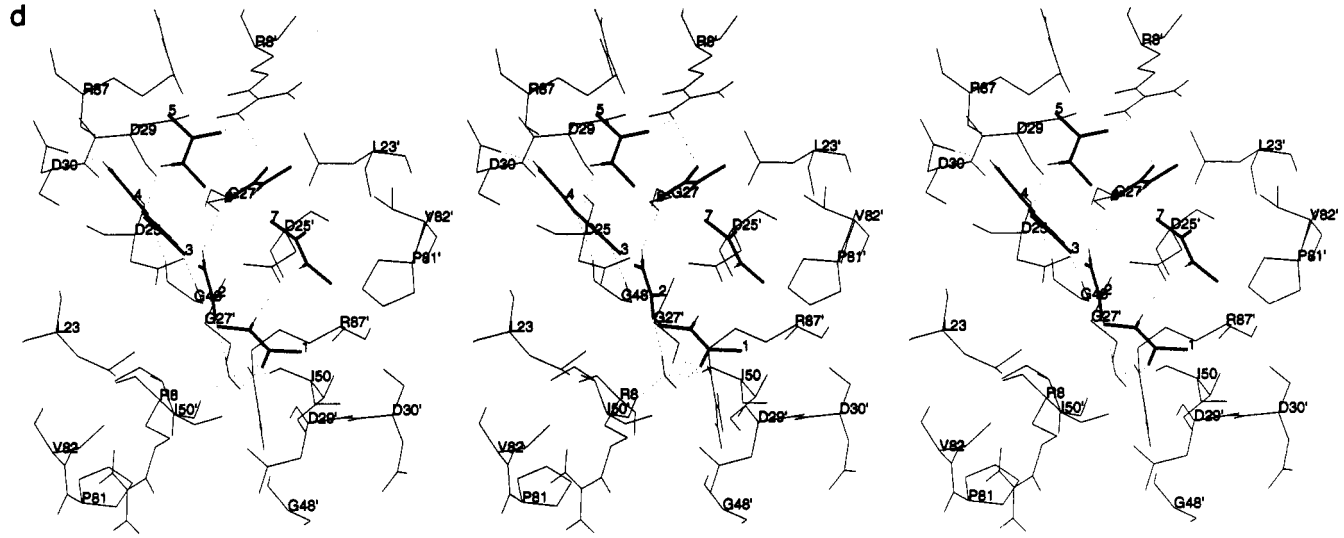


Figure 12. (a–d) Four sequences of 7 NMA minima (thick lines) in the complexed proteinase (thin lines) found by the connecting algorithm: (a) extended, same direction as in crystal structure; before (thick lines) and after (medium lines) minimization; (b) extended, opposite direction from the crystal structure; before (thick lines) and after (medium lines) minimization; (c) turn at residues 1–5; (d) turn at residues 3–6.

25 (Figure 5a). Two of them have a heavy-atoms RMS deviation less than 1.8 Å from the corresponding atoms in the hydroxyethylamine moiety of JG365 and the hydroxyethylene moiety of U-85548, two substrate-based inhibitors with known crystal structures that bind 10^3 times more tightly than MVT-101 (see also section 3.3).[22,28] The distance between the oxygen atom of methanol minimum 48 and the hydroxyl oxygen of JG365 (U-85548) is 1.1 Å (1.5 Å).

3.1.6. Role of Wat 511. To investigate the importance of Wat 511, which is present in all crystallographic complexes of HIV-1 aspartic proteinase with inhibitors.[7,21,22,28,33,34] further MCSS runs were performed for NMA, methanol, acetate, and methyl ammonium with the HIV-1 structure from the complex,[7] but without Wat 511. The MCSS results with or without Wat 511 are essentially the same for the charged and polar groups that were examined, except in the vicinity of the S2 to S2′ subsites (see Tables VIIIa, row with footnote d). Approximately the same number of minima were found and the same positions and orientations were present. For methyl ammonium and acetate, the RMS deviation between corresponding minima is less than 0.1 Å in most cases. For methanol, two additional minima were found; they occupy the position of Wat 511 acting as acceptor for the amino groups of the flap residues Ile 50 and Ile 50′. The NMA minima in the P4–P3, P3–P2, P2′–P3′, and P3′–P4′ sites have the same positions and orientations as in the run with Wat 511. The 11 NMA minima between P2 and P2′ (i.e., in subsites P2–P1, P1–P1′, P1′–P2′) characterized by a RMS deviation from the corresponding peptide group in the experimental inhibitor structure that is smaller than 2.6 Å, were not present in the run without Wat 511. (See Table VIIIa.) This finding supports the essential role of this water molecule in mediating hydrogen bonds between the proteinase flaps and the inhibitor backbone. This is a major difference between the structures of complexes between HIV-1 proteinase and inhibitors,[7,21,22,28,33,34] and those of cellular aspartic proteinases and inhibitors,[35–41] which do not contain this fully buried water molecule. Hence, the functional replacement of this water molecule should yield improved specificity.[21,22] The presence of two additional methanol minima at the water position confirms that it should be possible to design an HIV-1 proteinase selective inhibitor, which contains an oxygen atom in the appropriate position. One possibility is to covalently cross-link the inhibitor main-chain P2 and P1′ CO groups with a CH_2OCH_2 moiety; cross-links of peptide main chain and side chains have been suggested to be of general utility in inhibitor design.[1b] In addition to mimicking the interactions of the water molecule, such an inhibitor might bind better because the entropy loss involved in confining the water molecule and in selecting a specific main-chain conformation would not be present.

3.1.7. MCSS Results for the Native Proteinase. The MCSS methodology was also utilized for NMA in the free structure of HIV-1 aspartic proteinase.[9] The crystal structure of the unliganded enzyme does not contain a water molecule corresponding to Wat 511; accordingly no such molecule was included in the model. With the same methodology, 104 minima were accepted and 31 were discarded. Figure 11a shows the accepted NMA minima in the free proteinase structure, and Figure 11b shows those that are bound in significantly different regions from the results for the complexed proteinase. The additional minima are near the flaps and near residues 80–82 and 80′–82′. These are the regions that are most affected by the binding of the inhibitor. The lowest energy minima (1, 2, 3, 4, 6, and 8) are at the same open end of the binding site as in the liganded structure, while minimized positions 5, 7, 9, and 10 are at the other open end.

Table V, shows the clusters of NMA minima in the free proteinase together with their interaction energy ranges; they can be compared with the results for the complexed structure in IV. From the RMS deviation between NMA minima in the free and complexed proteinase, only four among the 40 lowest energy NMA minima in the free proteinase were not present in the complexed structure (within 2.6-Å RMS deviation), whereas 14 among the 20 highest energy minima were not present; conversely, only four among the 40 lowest energy NMA minima in the complexed proteinase were not found in the native structure, whereas 11 among the 20 highest were not found. This implies that most of the additional NMA minima found in the free structure are characterized by a relatively small interaction energy. It is clear that differences between the complexed and native structure affect the distribution of functional group positions in the binding

Table IX. All-Atom RMS Deviation between Monte Carlo-Minimized MVT-101 Structures

	U2	U3	U4	U5	U6	U7		D2	D3	D4	D5	D6	D7	D8
U1	8.0	7.4	4.0	4.5	4.9	4.2	D1	2.4	2.5	3.1	2.8	4.7	3.3	5.9
U2		2.3	8.0	6.6	5.8	7.0	D2		3.4	3.9	3.6	5.0	3.8	6.0
U3			7.5	6.0	5.4	6.4	D3			2.1	1.8	5.3	3.4	5.8
U4				5.2	5.6	5.0	D4				2.1	5.7	3.2	5.1
U5					2.1	3.4	D5					5.3	3.5	5.6
U6						4.1	D6						4.3	8.1
							D7							6.3

[a] RMS deviations (in Å) after MCM minimization.

Table X. RMS Deviation and Energy of the 15 Lowest Energy Complexes[a]

structure	RMS[b] C_α–C_β vs X [Å]	RMS[c] backbone vs X [Å]	RMS[d] all vs X [Å]	RMS[e] all vs Xm [Å]	E_{tot}[f] conjug gradient [kcal/mol]	E_{tot}[g] final [kcal/mol]	$E_{vdW\ int}$[h] final [kcal/mol]	$E_{ele\ int}$[i] final [kcal/mol]	$E_{internal}$[j] final [kcal/mol]
Xm[k]	0.9	1.2	2.6	0.0	−310.9	−326.4	−40.0	−160.2	−126.2
U1	3.1	3.0	4.2	3.5	−271.5	−324.9	−49.4	−163.5	−112.0
U2	6.5	6.2	7.1	6.7	−292.0	−310.1	−36.8	−192.3	−81.0
U3	5.9	5.7	6.7	6.1	−283.1	−303.3	−40.1	−197.4	−65.8
U4	1.7	2.3	2.1	3.0	−270.4	−301.5	−48.7	−146.7	−106.1
U5	3.6	1.9	5.1	4.5	−258.1	−297.0	−36.5	−190.0	−70.5
U6	4.0	2.2	5.3	4.6	−275.9	−291.4	−49.5	−176.1	−65.8
U7	3.6	3.0	4.9	4.3	−270.1	−287.3	−34.5	−176.2	−76.6
D1	2.1				−272.7	−334.4	−44.6	−170.2	−119.6
D2	2.3				−306.1	−330.6	−45.8	−187.6	−97.2
D3	0.9				−299.3	−330.2	−48.0	−181.2	−101.0
D4	2.0				−299.6	−322.6	−48.0	−190.3	−84.3
D5	1.1				−278.2	−312.2	−45.1	−161.1	−106.0
D6	4.6				−272.2	−289.9	−37.1	−190.3	−62.5
D7	3.1				−278.8	−283.9	−33.0	−206.8	−44.1
D8	5.4				−260.5	−280.9	−38.9	−175.4	−66.5

[a] Results for the 15 lowest energy conformations selected by conventional minimization (see text). Structures U have the same chain direction as in the crystal structure of the complex, whereas the D conformations have the opposite direction. Both are sorted by ascending total energy. [b] RMS deviation of the C_α and C_β atoms between Monte Carlo-minimized structures (300 cycles) and X-ray inhibitor (X). This comparison criterion was introduced since it does not depend on the chain direction. [c] Backbone-atom (N, C_α, C, O) RMS deviation between Monte Carlo-minimized structures (300 cycles) and X. [d] All-atom RMS deviation between Monte Carlo-minimized structures (300 cycles) and X. [e] All-atom RMS deviation between Monte Carlo-minimized structures (300 cycles) and the structure obtained after Monte Carlo minimization of the X-ray inhibitor (Xm). [f] Total energy inter- plus intramolecular after conventional minimization to 0.01 kcal/(mol Å). [g] Total energy after 300 cycles of MCM. [h] van der Waals interaction energy between inhibitor and proteinase after MCM. [i] Electrostatic interaction energy between inhibitor and proteinase after MCM. [j] Internal energy of the inhibitor after MCM. [k] X-ray structure after 300 cycles of MCM.

site. Since these differences are predominantly (although not exclusively) among the higher energy minima, it is likely that common peptides would be found in calculations for the two protein structures. Thus, studies of native proteinase could have been employed to obtain useful information, if the complexed structure were not available.

Ten NMA minima accept one hydrogen bond from the backbone NH group of flap residues Gly 48 or Gly 48′; further, minima 41 and 42 act as acceptors for both the Ile 50 and Ile 50′ NH groups (only four NMA minima in the complexed proteinase act as acceptors for NH groups of flap residues).

The RMS deviation between NMA minima in the unliganded proteinase and peptide groups in the inhibitor were computed. For this purpose, the position of the inhibitor inside the uncomplexed binding site was obtained by rigidly superimposing the crystal structures of the free and complexed proteinase. It was found that all peptide group positions, except the P3′–P4′ position, were within 2.6 Å of at least one MCSS minimum (see Table VIIIa). Further, all intermolecular backbone hydrogen bonds are present in several NMA minima. Failure to obtain the P3′–P4′ position is due to the lack of accessible space between the Arg 8 and Asp 29′ side chains, which make a salt bridge in the native proteinase and move to an alternative position in the complexed system. Results for hexapeptide main chains obtained by connecting these NMA minima are described briefly at the end of section 3.2.1.

3.2. Hexapeptide Results. 3.2.1. Construction of Hexapeptides. All possible ways of connecting NMA minima to form hexapeptide backbones were evaluated. A stepwise procedure was used with the criterion for adding an NMA group being that the pseudoenergy of the partial chain have a value smaller than 50 kcal/mol. This yielded 8034 different sequences of seven NMA groups (i.e., including the blocking groups). The lowest energy structure has a pseudoenergy of E_{ps} = 20.1 kcal/mol; of this, 12.2 kcal/mol was in the six bonds (E_b) and 7.9 kcal/mol in nonbonded contacts (E_{nb}). The sequences were clustered, and a representative of each cluster was selected with the algorithm described in section 2.2. This yielded 104 clusters and one representative of each was selected for further study. Most of the resulting backbone structures had an extended form in the binding site (see Figure 12, parts a and b) with the terminal NMA minima positioned at or near the S4 and S4′ site, respectively. Further, the central NMA position was located at the S2–S2′ sites in 97 of 104 sequences. Only two main chains showed a turn in the middle; the chain consisting of NMA minima 14–74–44–73–21–78–54 had a turn at residues 1–5 (Figure 12c), while the chain consisting of NMA minima 29–27–64–17–37–45–75 had a turn at residues 3–6 (Figure 12d). Among the 104 structures, 27 had the same (parallel) chain direction as in the crystal structure of MVT-101, and the remaining 77 had the opposite (antiparallel) direction. The latter are favored by better electrostatic interactions. The main chain consisting of NMA minima

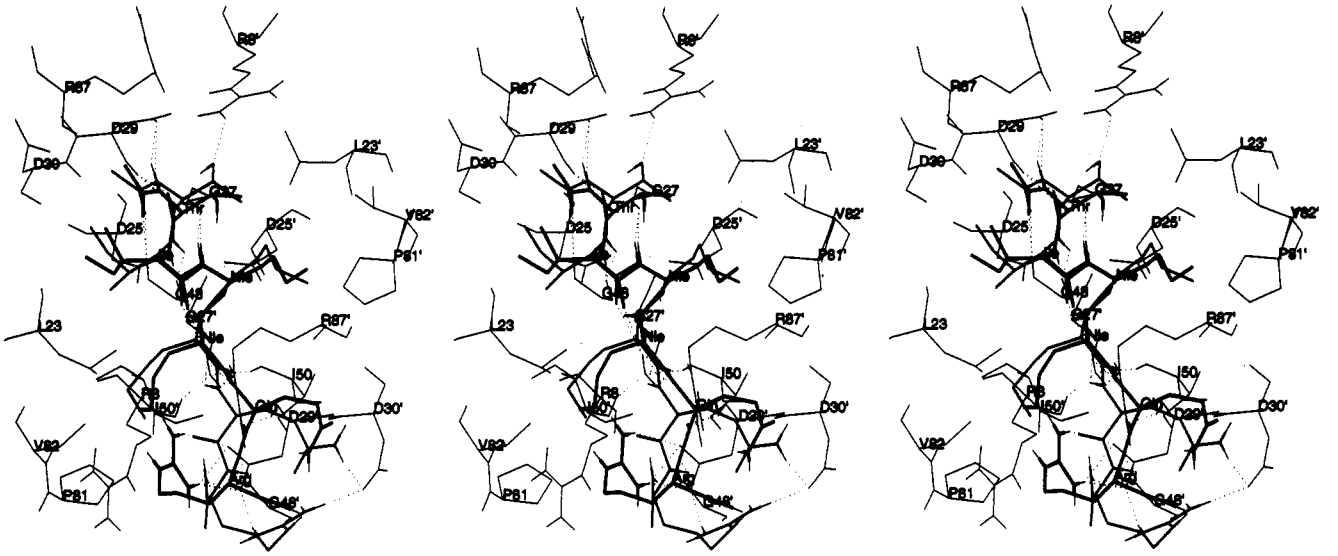


Figure 13. Stereoview of Monte Carlo-minimized structure of MVT-101 (Xm, medium lines) superimposed on the crystal structure (X, thick lines). Intermolecular backbone hydrogen bonds are dotted. C_α atoms of the proteinase (one letter code) and of the X structure of the inhibitor (three letter code) are labeled.

15–67–76–40–20–78–54 was the closest to the inhibitor backbone (2.5-Å RMS deviation). Furthermore, it yielded the peptide structure closest to the experimental MVT-101 structure (conformation U4, see section 3.2.2). No main chain could be constructed from the closest NMA minima listed in row 1 of Table VIIIa (30–55–40–11–20–78–49) because the pseudoenergy was 161 kcal/mol. Similarly, the set of minima resulting from the inhibitor when it was separated into NMA that were minimized (row 9 of Table VIIIa) also did not yield a hexapeptide because they had a pseudoenergy of 219 kcal/mol. This suggests that MVT-101 does not have an optimized backbone position. Among the 104 backbones there were sequences, such as 15–55–40–20–78–51–5 and 73–47–55–40–20–78–51, that used four out of the seven of the closest NMA minima.

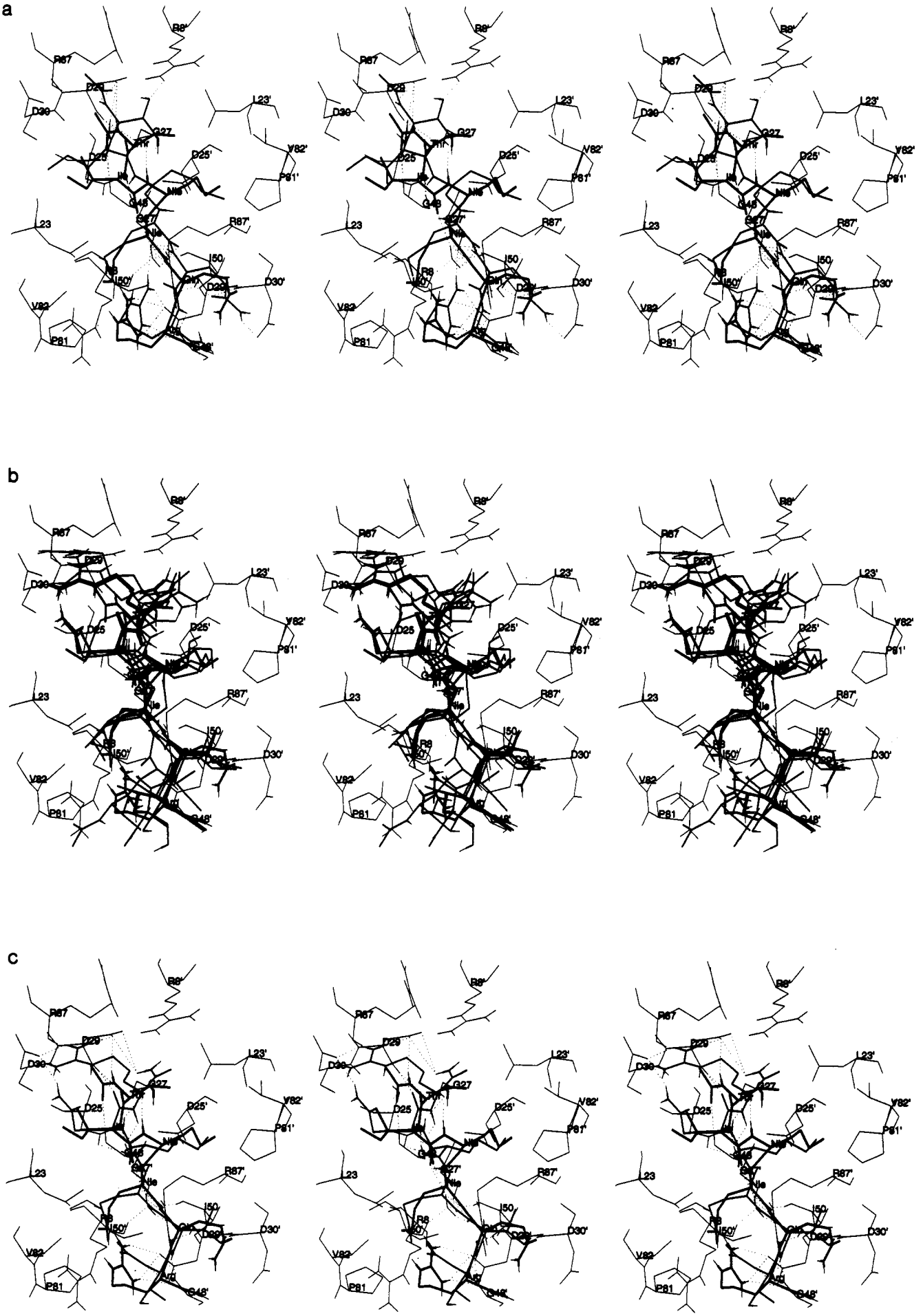
Given the 104 backbone representatives, all possible ways of connecting the appropriate side chain to each C_α atom were evaluated. For each of the six C_α atom in the 104 backbones, the MCSS-minimized position characterized by the lowest value of the pseudoenergy (eq 1) was chosen. The entire inhibitor structure was accepted if each candidate side chain had a pseudoenergy value smaller than 50 kcal/mol. This yielded a total of 63 inhibitor conformations. Among the 63 conformations, 24 pointed in the same direction as the experimental inhibitor complex and 39 pointed in the opposite direction. Only three inhibitors of the 27 backbone structures oriented in the same direction as in the crystal structure were discarded, while 38 oriented in the opposite direction were rejected because of the lack of appropriate MCSS side-chain minima in the vicinity of all of the C_α atoms. As an example, the sequence of NMA minima 4–9–65–13–74–44–73 was not accepted because of the lack of appropriate side-chain minima.

Application of the connecting algorithm to the NMA minima found in the native proteinase (section 3.1.7) yielded 4430 different sequences of seven NMA groups, corresponding to the hexapeptide with blocking groups. These sequences were clustered into 115 clusters and one representative of each was selected. It was found that three (nine) representative sequences had RMS deviations of 3.0 Å (3.5 Å) or less, from the corresponding backbone atoms of the X-ray inhibitor structure. The binding modes are generally similar to those found in the liganded proteinase structures. Most of the 115 sequences are extended along the binding site in both directions. Two extended sequences, whose N-terminal part or C-terminal part makes an angle of about 60° with the axis of the experimental structure of the inhibitor, represent possible alternative binding modes in the native HIV-1 aspartic proteinase structure.

3.2.2. Minimization Results for Hexapeptides. The 63 inhibitor conformations were minimized by the conjugate gradient algorithm[25] in the presence of the rigid protein; a convergence criterion of 0.01 kcal/(mol Å) was chosen. Minimization of the backbone structures prior to the attachment of the side chains was also attempted but abandoned because the procedure for side chain attachment yielded only 47 inhibitor structures. It was found that some parts of the inhibitor backbone had moved into cavities required for the side chains.

After minimization there were 15 inhibitor structures with a total potential energy (intermolecular plus intramolecular) lower than the cutoff value of −250 kcal/mol. Among these, seven had the same chain direction as in the inhibitor crystal structure (called the "up" direction U), while eight were oriented in the opposite direction ("down", D). The 15 structures were further minimized by the MCM method with a temperature of 310 K used in the Metropolis criterion (see section 2.3). Each complex was first submitted to 100 MCM cycles involving random changes in all dihedral angle variables (40 in the present inhibitor). Then 200 MCM cycles were performed that involved random changes in only the side-chain torsion angles (21 in the present inhibitor). Average values of the acceptance ratio were 19% for the first 100 cycles and 27% for the last 200 cycles, the latter value being larger because of the restriction to sidechain alterations. In the Monte Carlo minimization the average RMS displacement was 1.7 Å (minimum value of 0.8 Å and maximum value of 3.1 Å). Before the MCM procedure, only U5 and U6 were similar to each other (all-atom RMS deviation smaller than 2.6 Å). After MCM, U3 was similar to U2, U6 to U5, both D3 and D2 to D1, D5 to D4, and both D5 and D4 to D3 (cf. Table IX).

The decrease in energy after the 300 MCM cycles was −27.4 kcal/mol on average, with a maximum of −61.7 kcal/


a
b
c
R8'
R87
D29
L23'
D30
G27
Thr
D25
D25'
V82'
P81'
Nle
G48
G27'
L23
R87'
I50
I50'
R8
D29'
D30'
V82
P81
Arg
G48'

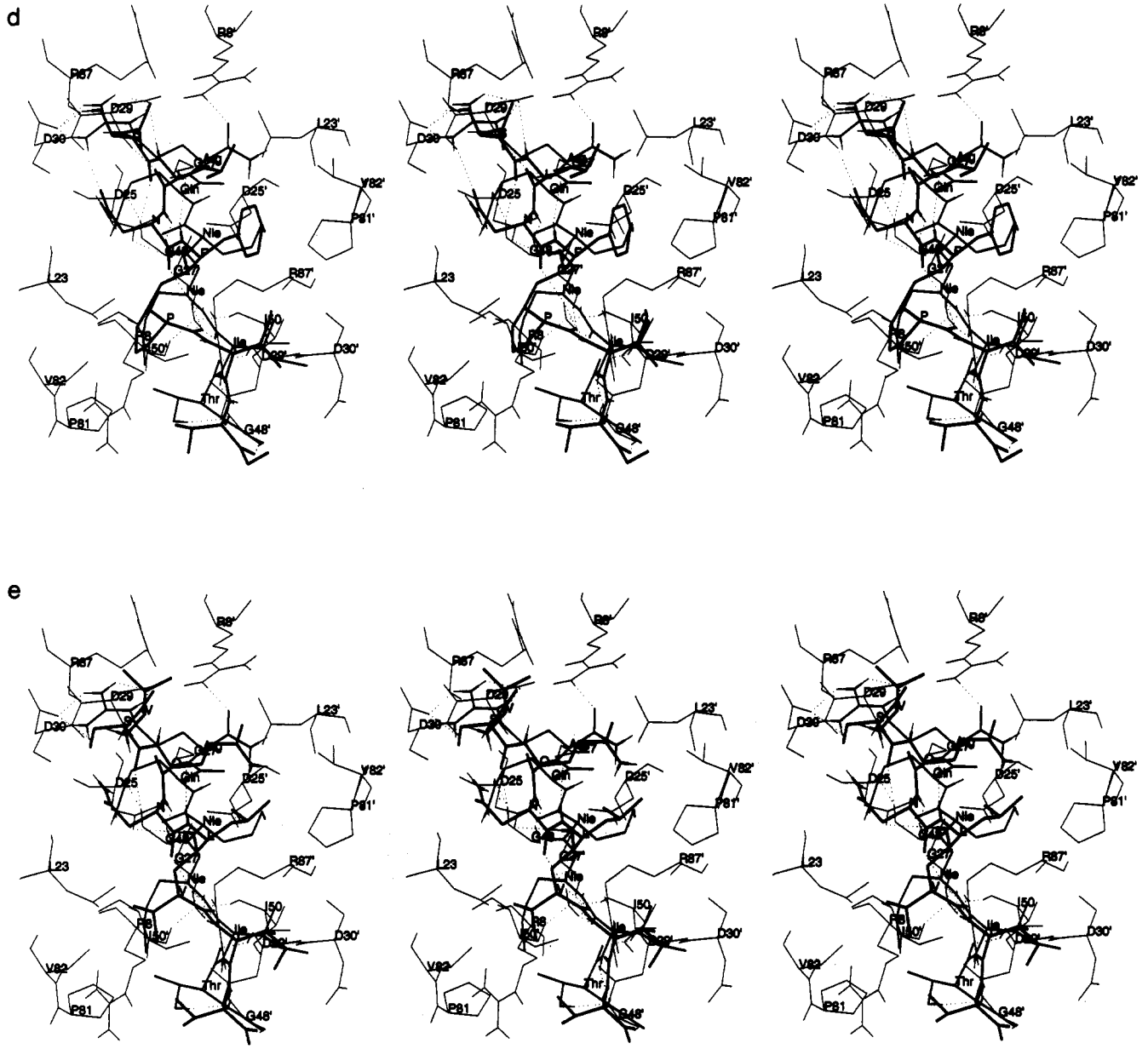


Figure 14. (a) Stereoview of computed structure U4 (medium lines) superimposed on X (thick). Proteinase structure is shown in thin lines. (b) Computed structures D1–D5 (medium lines) superimposed on X (thick). (c) Computed structure D3 (medium lines) superimposed on X (thick). (d) Computed structure D1 (medium lines) superimposed to the structure of inhibitor JG-365 (thick). The position of inhibitor JG-365 was determined by superimposing the proteinase structure. (e) Computed structure D1 (medium lines) superimposed on the structure of inhibitor U-85548E (thick). The position of inhibitor U-85548E was determined by superimposing the proteinase structures. C_α atoms of the proteinase (one letter code) and of the computed structure of the inhibitor (three letter code) are labeled.

mol and a minimum of −5.1 kcal/mol. Table X lists the final energy values for the seven up conformations (U1–U7) and the eight down conformations (D1–D8), together with their RMS deviation from the crystal structure of the inhibitor (X). The total energy, as well as the van der Waals and electrostatic contributions to the interaction energy are given; the electrostatic term is always greater in magnitude and has considerably more variability than the van der Waals term. The inhibitor internal energy also is significant and has a large variation in the different structures; it ranges from −44.1 (D7) to −119.6 kcal/mol (D1) for the D conformers and from −65.8 (U3 and U6) to −112.0 kcal/mol (U1) for the U conformers. When the same MCM procedure was applied to the X-ray structure of the inhibitor, it yielded an RMS deviation of 2.6 Å from X and the energy was lowered by −15.5 kcal/mol from an initial value of −310.9 kcal/mol; the internal energy was equal to −126.2 kcal/mol. (This structure is referred to as Xm.) The major alteration was an interchange of the positions of the Arg P3′ side chain and the amide C-terminal group. The Arg P3′ side chain donates 2 and 1 hydrogen bonds to Asp 29′ and Asp 30′, respectively (Figure 13) in the minimized structure, Xm, whereas in X, Arg P3′ donates 3 hydrogen bonds to Asp 29′.

The backbone conformation of U4, U5, and U6 are similar to X; their RMS deviations are 2.3, 1.9, and 2.2 Å, respectively (Table X). After MCM minimization U4 converged toward X (RMS deviation of 2.1 Å). The major differences are (see Figure 14a) in the position of the acetyl terminus, which is rotated by about 60°, in the Thr P3 side chain, which makes a hydrogen bond with the Arg 8′ side chain (Thr P3 is the only side chain lying partially outside of density in the X-ray structure[7]), and in the Gln P2′ acetamide group, which is rotated of about 90° and acts as an acceptor for the NH group of both Asp 29′ and Asp 30′ and donates to one oxygen of the Asp 30′ side

Table XI. Hydrogen Bonds[a]

X-ray		U4		U5		D3[b]	
Donor	Acceptor	Donor	Acceptor	Donor	Acceptor	Donor	Acceptor
Thr P3 NH	$O_{\delta 2}$ Asp 29	+	+	+	+	+	+
Ile P2 NH	CO Gly 48	+	$O_{\delta 2}$ Asp 29	+	+	+	+
Nle P1 NH	CO Gly 27	+	CO Gly 48	+	+	+	+
Gln P2′ NH	CO Gly 27′	+	+	+	+	+	+
Arg P3′ NH	CO Gly 48′	+	+	+	+	+	+
Ami P4′ NH	$O_{\delta 2}$ Asp 29′	–	–	+	CO Gly 48′	+	+
–	–	–	–	–	–	Thr P3 NH	$O_{\delta 1}$ Asp 29′
–	–	–	–	–	–	Nle P1′ NH	$O_{\delta 2}$ Asp 25
–	–	–	–	–	–	Ami P4′ NH	$O_{\delta 2}$ Asp 29
Gly 48 NH	CO Ace P4	+	CO Thr P3	+	+	–	–
Asp 29 NH	CO Thr P3	+	CO Ile P2	+	+	+	+
Wat 511 H1	CO Ile P2	+	CO Nle P1′	+	+	+	+
Asp 29′ NH	CO Gln P2′	–	–	–	–	+	+
–	–	Gly 48′ NH	CO Arg P3′	Asp 29′ NH	CO Arg P3′	–	–
Arg P3′ $N_{\eta 1}$	$O_{\delta 1}$ Asp 29′	+	+	–	–	–	–
Arg P3′ $N_{\eta 2}$	$O_{\delta 1}$ Asp 29′	+	+	–	–	–	–
Arg P3′ $N_{\eta 2}$	$O_{\delta 2}$ Asp 29′	+	+	+	+	+	+
–	–	Gln P2′ $N_{\epsilon 2}$	$O_{\delta 1}$ Asp 30′	Gln P2′ $N_{\epsilon 2}$	$O_{\delta 1}$ Asp 29′	Gln P2′ $N_{\epsilon 2}$	$O_{\delta 2}$ Asp 30
–	–	Arg P3′ $N_{\eta 1}$	CO Gly 27′	Arg P3′ $N_{\eta 1}$	$O_{\delta 1}$ Asp 30′	Arg P3′ $N_{\eta 1}$	$O_{\delta 2}$ Asp 30
–	–	–	–	Arg P3′ N_{ϵ}	$O_{\delta 1}$ Asp 30′	–	–
–	–	–	–	Arg P3′ $N_{\eta 1}$	$O_{\delta 2}$ Asp 30′	Arg P3′ $N_{\eta 2}$	$O_{\delta 2}$ Asp 30
–	–	Arg 8′ $N_{\eta 1}$	O_{γ} Thr P3	Arg 8 $N_{\eta 1}$	$O_{\epsilon 1}$ Gln P2′	–	–
–	–	Asp 29′ NH	$O_{\epsilon 1}$ Gln P2′	–	–	–	–
–	–	Asp 30′ NH	$O_{\epsilon 1}$ Gln P2′	–	–	–	–
Thr P3 NH	CO Thr P3	–	–	Nle P1 NH	+	–	–
Nle P1′ NH	CO Ile P2	+	CO Nle P1′	–	–	–	–
Gln P2′ NH	CO Gln P2′	+	+	–	–	+	+
–	–	Thr P3 OH	CO Ace P4	Ile P2 NH	CO Ace P4	–	–
–	–	Arg P3′ NH	CO Arg P3′	–	–	–	–
–	–	Arg P3′ N_{ϵ}	CO Gln P2′	–	–	–	–
–	–	Arg P3′ $N_{\eta 1}$	CO Gln P2′	–	–	–	–
–	–	Ami P4′ NH	$O_{\epsilon 1}$ Gln P2′	–	–	Ami P4′ NH	CO Gln P2′

[a] A cutoff of 3.5 Å for the heavy-atoms distance and of 90° degrees for the donor–hydrogen–acceptor angle were used. A plus sign means some polar group as in X-ray structure, whereas a pair of minus signs means that the hydrogen bond was not present. The bottom of the table contains the inhibitor intramolecular hydrogen bonds. [b] This structure has direction opposite to the X-ray inhibitor conformation; accordingly, for the proteinase residue listed in the first two columns one has to consider Gly 27′ as Gly 27, Asp 29 as Asp 29′, Gly 48 as Gly 48′ and so on.

chain (see Table XI and Figure 14a). Of the 11 experimental inhibitor backbone-proteinase (or Wat 511) hydrogen bonds, 9, 5, and 4 are present in U5, U6, and U4, respectively (Figure 14a and Table XI). Furthermore, in U4 the CO groups of Thr P3, Ile P2, and Arg P3′ accept from the NH's of Gly 48, Asp 29, and Gly 48′, respectively, while the NH group of Ile P2 and Nle P1 donate to the Oδ of Asp 29 and the CO group of Gly 48, respectively. The Arg P3′ side chain of U4 makes the same salt bridge with the Asp 29′ side chain as in the X-ray structure of the complex (Table XI). In both U5 and U6 the side chain of residue n points into the pocket occupied by the residue $n + 1$ in the X-ray structure. Hence, the all-atom RMS deviation from the X-ray structure is larger than that of U4, although the backbone RMS deviation is slightly smaller. Before MCM, the U4 structure had the 12th lowest energy (among the 63) and a RMS deviation from X of 2.4 Å; after MCM, its energy value is 9th and its RMS deviation from X is only 2.1 Å. The RMS deviation between U4 and Xm is larger than that between U4 and X because of the differences in the position of Arg P3′ and the C-terminal amide group.

Structures U2 and U3 are translated two residues along the binding cleft relative to U4 and X; their side chain n lies in the groove of side chain $n + 2$. This suggests that an octapeptide inhibitor could be constructed. This position allows the arginine side chain of the inhibitor to make a salt bridge with Asp 30′ instead of Asp 29′. The arginine side chain of U1 also makes a salt bridge with Asp 30′, which lies at the end of the HIV-1 proteinase binding site. Since the distance of the Asp 30′ side chain from Arg 8 is larger than that from Asp 29′, the Arg P3′ side chain of the inhibitor points toward the carboxyl group of the former aspartate and avoids the electrostatic repulsion of the Arg 8 side chain. The lack of a solvation correction leads to unrealistically low electrostatic energies. All other MCM inhibitor structures are characterized by lower electrostatic interaction energies than U4 (Table X).

The down conformations, D1–D5 (i.e., 5 of the 6 lowest energy conformations) have backbone positions similar to the X-ray result, although their chain direction is inverted. It is interesting that the C_{α}–C_{β} RMS deviation (see caption of Table X) between D3 and X is only 0.9 Å and the one between D5 and X is only 1.1 A; D1, D2, and D4 have deviations smaller than 2.3 Å. In D1–D5 the sidechain of residue Pi′ points into the Si site (Figure 14b). Ten among the 11 inhibitor backbone-proteinase (or Wat 511) hydrogen bonds of the crystal structure are present in the D3 conformation (Figure 14c and Table XI); only the acetyl P4 CO-NH Gly 48′ is missing. One among the 3 Arg P3′ side chain H bonds with Asp 29 is present in D3; furthermore, Arg P3′ of D3 makes two hydrogen bonds with Asp 30. In most of the D conformations, the Ile P2 side chain occupies the S2′ site of the proteinase and makes hydrophobic contacts with Ile 47′, Gly 48′, and Ile 50. In D1–D5 the residues Nle P1 and Nle P1′ point in the same direction and make the same hydrophobic contacts (with Leu 23, Ile 50, Pro 81, Val 82, and Ile 84 and the corresponding residues on the second chain of the proteinase) as Nle P1′ and Nle P1, respectively, in the crystal

structure (see Figure 14b). Also, Gln P2′ points into the S2 pocket, Arg P3′ is involved in a salt bridge with the Asp 30 side chain, while the C-terminal amide occupies several positions but is mainly directed toward the broad P3 site. The D1–D5 structures are well behaved in terms of the internal, van der Waals, and electrostatic energy contributions (see Table X).

3.2.3. Addendum. After the present study was completed, a copy of the paper was sent to Dr. A. Wlodawer, who transmitted it to Dr. Maria Miller. She informed us that a refined crystal structure based on 2-Å data for the complex between HIV-1 aspartic proteinase and MVT-101 showed that the inhibitor binds in two orientations, which correspond to D and U of the calculations. Moreover, the D orientation is the predominant conformation (70%) and that the U conformation, described in the original paper, is a minor conformer (30%). This is in agreement with the simulation results. Dr. Maria Miller kindly provided the new coordinates prior to publication (M. Miller *et al.*, to be published). Comparison of the predominant (D) conformer in the new crystal structure with D_1, D_3, D_4, and D_5 from the calculations demonstrates that the main-chain position, side-chain orientations, and hydrogen bonds for positions P2 through P2′ are very similar. The sidechains of Ile P2, Nle P1, Nle P1′, and Gln P2′ in the computed structures D1, D3, and D5 occupy the same subpockets as the corresponding side chains in the refined MVT-101 crystal structure and participate in the same interactions with the proteinase. For P2 to P2′, the all-atom RMS deviations of MVT-101 including main-chain and side-chain atoms are 1.76, 3.06, 1.09, 1.75, and 1.21 Å relative to D1, D2, D3, D4, and D5, respectively. Both the S3 and S3′ pockets, which involve surface residues, are less well defined in the crystal structure than S1,S1′ and S2,S2′ and the position of the inhibitor P3 and P3′ residues appears to be affected by both the solvent and crystal contacts. In the crystal, the side chain of Thr P3 at the N terminus is partially exposed to solvent and there is a water molecule that couples the hydroxyl hydrogen of the P3 to the carboxylates of Asp 29′ and Asp 30′. In the computed structures no water was included and the Thr P3 hydroxyl group donates a hydrogen bond directly to the Asp 29′ side chain. Hence, the Thr P3 side chain in the simulation has a slightly different position from that in the crystal structure. The C terminal Arg P3′ side chain appears to be buried in a hydrophobic pocket between the side chains of Phe 53 and Pro 81′ in the crystal structure. It is surprising to have such a charged side chain with limited polar interactions; further, the Arg P3′ side chain is not in electron density in the structure. In the simulation the Arg P3′ side chain makes salt bridges with the side chains of Asp 30 and Asp 29.

The agreement between the modeled and experimental structures for the P2–P2′ region is gratifying, particularly because the present study was performed without any knowledge of the presence of a second orientation for the bound MVT-101 inhibitor.

3.3. Comparison with Other Proteinase-Inhibitor Complexes. A crystal structure of a complex between HIV-1 proteinase and acetyl pepstatin, Ace-Val-Val-Sta-Ala-Sta, exists at 2.0-Å resolution with an R factor of 0.176.[33] Since statine is only one carbon–carbon bond shorter than a dipeptide, the main-chain length of this inhibitor is approximatively the same as that of MVT-101. The inhibitor acetyl pepstatin is bound in two approximately symmetric orientations, U and D; reflecting differences in the flap positions, the atoms in the two orientations of the inhibitor adopt very similar but not identical positions. Its peptidyl backbone is extended, and the intermolecular hydrogen bonding network is similar in both orientations to that of the MVT-101 complex. That both U and D orientations are observed is in accord with the two main-chain orientations of the MVT-101 inhibitor generated in the present work. Also, the essential correspondence of the main-chain H bonds in the U and D conformations is confirmed. In the computed MVT-101 conformations and the X-ray structure of the acetyl pepstatin proteinase complex,[33] the side-chain position of inhibitor residue Pi in the U orientation points into the site occupied by the sidechain Pi′ in the D orientation and *vice versa* (compare Figure 14b of this paper with Figure 7 of ref 33). It is of interest that an inhibitor with as few polar group interactions as acetyl pepstatin binds so well (K_i = 20 nM). This supports the importance of hydrophobic interactions in determining binding constants, even when the specificity is due to polar or charged groups.

After the modeling work reported here was completed, we received the X-ray coordinates of synthetic HIV-1 proteinase complexed with the inhibitors JG-365[22] and U-85548E.[28] The complex with inhibitor JG-365, which has the sequence Ace-Ser-Leu-Asn-Phe-ψ[CH(OH)CH_2N]-Pro-Ile-Val-OMe and a K_i of 0.24 nM,[30] was refined to an R factor of 0.146 at 2.4-Å resolution. The complex with inhibitor U-85548E, which has the sequence H-Val-Ser-Gln-Asn-Leu-ψ[CH(OH)CH_2]-Val-Ile-Val-OH (K_i < 1.0 nM),[31] was refined to an R factor of 0.138 at 2.5-Å resolution.

The main-chain position and side-chain orientations of JG-365 and U-85548E in the binding site are similar to that of MVT-101, even though JG-365 and U-85548E are one and two amino acid longer, respectively, at the N-terminus; i.e., they occupy the corresponding sites with the $CHOHCH_2$ group in the same site as the CH_2NH group in MVT-101. The Ser P4 side chain makes a hydrogen bond with the side chain of Asp 30 in both JG-365 and U-85548E, whereas the Val P5 of U-85548E appears to be accessible to solvent. It is of interest that there are displaced positions of computed MVT-101, as described earlier, in which residue n points into the pocket corresponding to the side-chain n + 1 (structures U5, U6, and D6) or n + 2 (U2, U3, and D8); the symmetry of the site would suggest that inhibitors in n − 1 and n − 2 would also occur. The intermolecular backbone hydrogen bonding scheme is the same in all three complexes; they all have the same 11 backbone hydrogen bonds. In both JG-365 and U-85548E the backbone hydroxyl group interacts favorably with the catalytic Asp 25 and Asp 25′. This hydroxyl group appears to be important for binding, since both JG-365 and U-85548E bind to proteinase >10^3–10^4-fold more tightly than closely related inhibitors that lack such a group.[30,31] The hydroxyl group is not present in inhibitor MVT-101, which possesses a reduced peptide bond at position P1–P1′; its binding constant is 780 nM. The importance of the hydroxyl group is in accord with the methanol MCSS minima found in a corresponding position (see section 3.1.5).

The two computed structures characterized by the lowest value of the total energy (D1 and D2) possess the same position and orientation as JG-365 and U-85548E (see Figure 14, parts d and e). The D1 and D2 side chains occupy the same subsites of the JG-365 and U-85548E

Table XII. Docking of Dipeptides to One Open End (Near S4') of the Binding Site

name	sequence	E^b backbone	E^c (S1)	E^d (S2)	E^e MCM	$E_{tot}{}^f$ MCM	% $E_{int}{}^g$
T1	Ala-Gln	-120.1	-2.1	-14.9	-120.5	-164.9	88
T2	Leu-Thr	-120.2	-4.0	-28.0	-114.7	-139.0	75
T3	Nle-Arg	-122.1	-3.1	-77.6	-172.5	-179.3	85
T4	Arg-Thr	-118.0	-77.6	-28.2	-168.5	-195.8	75
T5	Ile-Ser	-116.7	-6.0	-29.0	-87.4	-129.2	58

[a] The sequences were generated as explained in the text. [b] Interaction energy of the three NMA groups. All energy values are in kcal/mol. [c] Interaction energy of the MCSS group selected for the side chain of residue 1. [d] Interaction energy of the MCSS group selected for the side chain of residue 2. [e] Interaction energy after MCM. [f] Total energy after MCM. [g] Percent of the interaction energy of the MCSS minima conserved in the dipeptide after MCM.

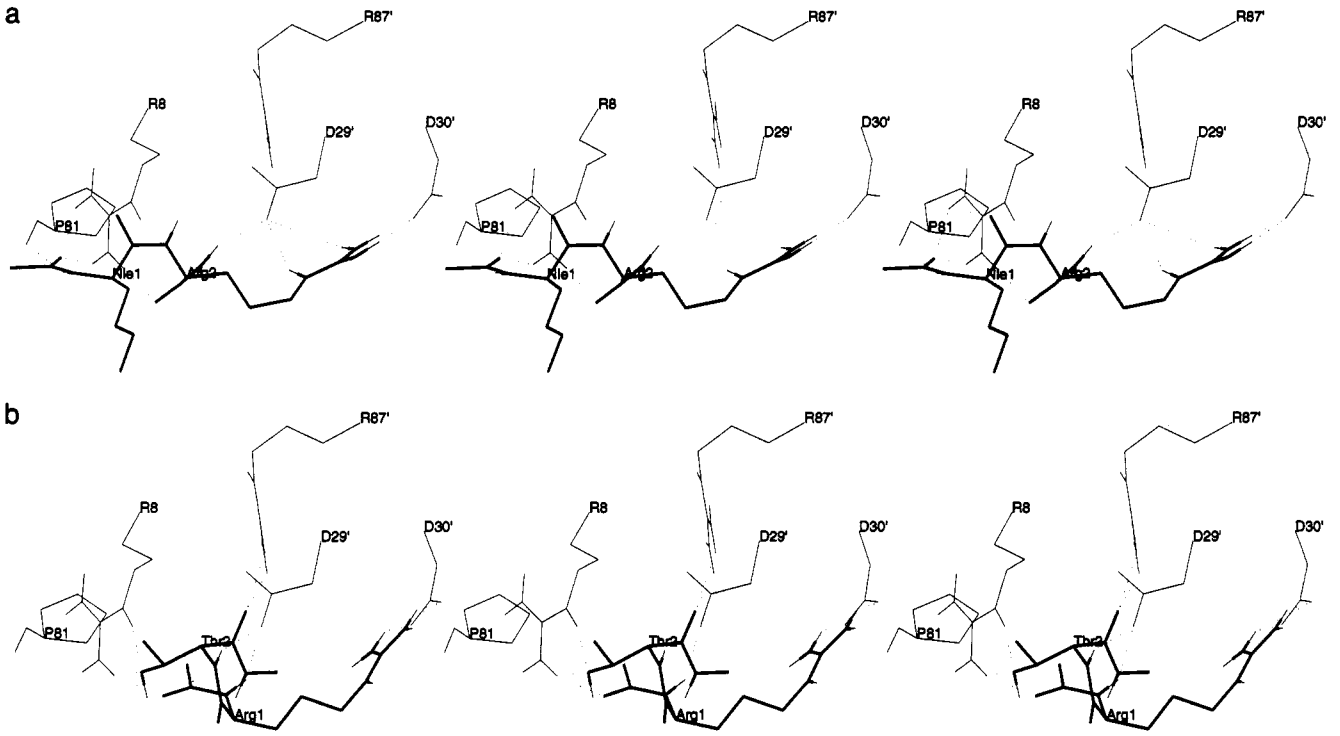


Figure 15. Stereoview of two terminal blocked dipeptides (CH_3CO-X-X-$NHCH_3$) docked at one open end of the binding site (thin lines): (a) Nle-Arg (T3) and (b) Arg-Thr (T4). Intermolecular hydrogen bonds are shown as dotted lines.

inhibitors; the carbonyl group at the N-terminus in D1 and D2 is located at the same positions as the corresponding moiety at the C-terminus P4'; the Thr P3 side chain is slightly shifted with respect to the Val P3' side chain of both JG-365 and U-85548E; the Ile P2, Nle P1, and Nle P1' residues make the same hydrophobic contacts as the Ile P2', Pro P1' (Val P1'), and Phe P1 (Leu P1) side chains in JG-365 (U-85548E). Furthermore, the Gln P2' side chain of D1 points into the same pocket as the Asn P2 of both JG-365 and U-85548E, while the salt bridge between the Arg P3' side chain and Asp 30 corresponds to the hydrogen bond between the Ser P4 side chain of both JG-365 and U-85548E and the proteinase Asp 30 side chain. Last, the C-terminal amide of D1 and D2 is involved in the same polar interactions (mainly with the Arg 8' side chain) as the Gln P3 side chain of U-85548E.

3.4. Construction of Dipeptide Ligands at Open End of Binding Site. Once the MCSS methodology has been applied to a sufficient number of functional groups, the connection and minimization procedure can be utilized to design new oligopeptides. As described in (section 3.1), several of the low-energy minima of polar and charged groups bind to the region near residues Arg 8, Asp 29', Asp 30', and Arg 87' at one open end of the HIV-1 proteinase binding site. Although these may not be the best binding sites because of their significant solvent exposure, we use this region to illustrate the possibilities for peptide design. It is also of interest because such peptides could be connected to longer peptides in the binding site to provide specificity via their interaction with the charged amino acids at the open ends that are present only in HIV-1 proteinase and not in other proteinases. As a simple example, we construct terminal-blocked dipeptides (CH_3-CO-X-X-$NHCH_3$) by connecting sets of three of the 10 lowest energy NMA minima at the same open end of the binding site (Figure 4b). They could represent the terminal part of a peptide consisting of more than six residues or be part of a smaller peptide. A search through all 720 combinations of three different NMA minima yielded five sequences with a pseudoenergy E_{ps} (see eq 1) smaller than 50 kcal/mol. The lowest value of E_{ps} was found for minima 8, 5, and 10 (see Figure 3b); the E_{ps} value was 19.5 kcal/mol, which derived from the two stretched bonds. None of the five dipeptide sequences had significant bad contacts.

Side chains were built onto each of the five dipeptide backbones by evaluating the pseudoenergy E_{ps} (eq 1) without moving either the backbone or the side-chain minima. All the accepted MCSS minima of acetamide, methanol, acetate, methyl ammonium, ethyl guanidinium, propane, and isobutane described in section 3.1 were considered for each C_α atom. This represents the main difference with respect to the calculations for the MVT-101 inhibitor, where only the minima corresponding to the known side chain were included. The minimized position characterized by the lowest value of E_{ps} was

selected for each pair of sidechains. The resulting dipeptide sequences are listed in Table XII, together with the interaction energies of the constitutive MCSS minima. In a full search, more than one minimum would be tried for each side chain.

The five dipeptides T1–T5 were then submitted to 1000 MCM cycles with all variable dihedrals included in the random perturbation. The resulting conformations bind in different ways but all of them make several hydrogen bonds with the charged residues Arg 8, Asp 29′, Asp 30′, and Arg 87′. Dipeptides T3 and T4 are characterized by an energy of interaction with the protein that is about 50 kcal/mol greater in absolute value than that of the three remaining dipeptide ligands. The stronger interaction is due primarily to the presence of the Arg side chain, which like the minimum energy hexapeptide structures D1–D3 makes a salt bridge with Asp 30′ in both cases. The Arg is position two in T3 and position one in T4. This is possible because the peptides have opposite orientations. Both T3 and T4 participate in 10 intermolecular hydrogen bonds (Figure 15, parts a and b). In all cases the interaction energy of the dipeptides increases as a result of the MCM procedure. This is due to the fact that the total energy is minimized in MCM and improvements in the internal geometry are made that inhibit ideal intermolecular contacts. In structures T1–T4, the interaction energy after MCM is more than 75% of the sum of the interaction energy of their constitutive MCSS minima (Table XII). This suggests that these peptides have found a geometry that corresponds to a near-optimum binding mode.

The D1–D3 hexapeptides do not overlap the T3 and T4 dipeptides, and it is possible to connect them to each other. For example, the Thr P3 hydroxyl (in D1 or D2) could be linked to the Arg 2 backbone nitrogen of T3 or the N-terminal CH_3 of T4.

4. Conclusions

The procedure described here for rational ligand design divides a very complex problem into a series of solvable tasks. The present paper is concerned with the search for all optimal positions and orientations of a set of functional groups and the connection of the resulting functional group minima to generate peptide ligands. As a test of the method, it was applied to the construction of hexapeptide ligands with the sequence of inhibitor MVT-101 in the HIV-1 proteinase binding site. In addition, alternative functional group interactions were considered and some possible dipeptide ligands were constructed.

The search of the binding site for functional group minima was made by the MCSS method. At least one MCSS minimum was found within 2.4 Å of the experimental result for all positions of the groups that correspond to those present in MVT-101. Further, all of the inhibitor–enzyme hydrogen bonds and most of the hydrophobic interactions are reproduced by the MCSS minima.

The MCSS positions of the side-chain and main-chain functional groups of MVT-101 differ somewhat from the positions seen in the cocrystal structure. This arises in part from imprecision in the MCSS calculation, but also from the constraints associated with linking these functional groups into a single peptide. This implies that the design of peptide ligands, which could include all natural as well as easily obtained nonnatural side chains, may yield peptides with superior electrostatic and steric contacts with the protein.

An algorithm for evaluating the possible connections between MCSS-minimized positions was developed to build peptide ligand conformations from the MCSS functionality maps. The method is very fast; e.g., a systematic search of all possible ways of forming hexapeptide main chains from the 83 NMA minima was made. The combination of a method for an exhaustive search for functional group minima in the binding site with a highly efficient method for constructing molecules from them provides a novel and effective approach to the theoretical design of candidate peptide ligands.

Sixty-three replicas of the MVT-101 hexapeptide, *N*-acetyl-Thr-Ile-Nle-ψ[CH_2-NH]-Nle-Gln-Arg·amide, were generated at different positions inside the binding site of the HIV-1 proteinase. The 15 lowest energy minima were divided into two groups; seven had the same (parallel) orientation as MVT-101 and eight had the opposite (antiparallel) orientation; the latter tended to be lower in energy. [*Addendum*: This is in agreement with recent high-resolution crystallographic data, which have shown that there is a minor and a major conformer in the binding site. (M. Miller *et al.* Private communication.) Moreover, the P_2–P_2' segment of the computed structure characterized by the lowest (third lowest) total energy has an all-atom RMS deviation of 1.76 Å (1.09 Å) from the corresponding segment of the MVT-101 inhibitor in its major orientation.]

Terminal-blocked dipeptides, which are characterized by optimal interactions (*in vacuo*) with one open end of the binding site, were designed. It would be of interest to determine if such peptides bound effectively and, if so, whether by themselves they inhibited the proteinase. Alternatively, the dipeptides generated here could be added to known ligands to make a tighter binding inhibitor. Since retroviral proteinases have a special arrangement of charged amino acids at the binding site open ends, the longer inhibitor should be characterized by a high degree of specificity.

In addition to the validation of the MCSS procedure and the demonstration of the utility of the pseudoenergy function approach for constructing peptide ligands, a number of results of the present analysis are of interest. First, there are many strongly binding functional-group minima in the active site that are not used by any known inhibitor. This includes both minima of the mainchain (NMA) and sidechain type. Particularly suggestive are the open end (S4 and S4′) regions with many charged side chains that contribute to the dipeptide interaction and several methyl ammonium and ethyl guanidinium minima that interact with the catalytic aspartates. Modifications of known inhibitors to include such positively charged functionalities between the P1 and P1′ sites could lead to improved inhibition. [A reviewer pointed out that the use of an aminostatine has been tried in renin inhibitors and that the potency of these inhibitors is similar to or greater (depending on the stereochemistry) than that of their counterparts; further, the aminostanine inhibitors are characterized by a better solubility.[51]] The methanol functionality map suggests that the role of the hydrogen-bonded water molecule interacting with the flaps of the proteinase and the inhibitor can be replaced by an oxygen atom covalently bonded to the ligand main chain; an aliphatic alcohol group and a cross-link containing an ether group are suggested. Since this buried water molecule is not present in the mammalian aspartic proteinases, the design of such a ligand could aid in developing a selective inhibitor that is specific for retroviral aspartic proteinases. More generally, the present results raise the question whether inhibitors that mimick substrates or transition

states, although easy to design and often successful, necessarily represent the most strongly binding ligands in the region of the active site.

Second, a comparison of the results without the active site water and for the inhibitor-free structure with those of the complexed structure including the water suggests that application of the present methodology could yield useful results even if the complexed proteinase structure were not available.

Third, none of the hexapeptide inhibitors constructed in the binding site by the present procedure are identical to the X-ray structure and when the latter is minimized with the same force field it moves away from the X-ray structure in the rigid protein. This suggests that improvements in the energy function may be required. The energy contribution that is likely to have the largest error is the electrostatic term, which is obtained from an *in vacuo* Coulomb calculation. Also, no direct term for the hydrophobic interaction is present in the potential function. A variety of approaches to overcome these difficulties are being examined.[42-48]

Acknowledgment. This work was supported in part by a grant from the National Institutes of Health. A. C. was supported by the Swiss National Science Foundation. We thank Dr. L. Caves and E. Evensen for helpful discussions and program development. Versions of CHARMM developed at Harvard by R. Elber and L. Caves that contain the replica formalism were used in the MCSS calculations. We thank Dr. J. Erickson and Dr. A. Wlodawer for crystallographic coordinates of HIV-1 aspartic proteinase-inhibitor complexes. We are indebted to Dr. M. Miller for making her manuscript available before publication. We thank a referee for detailed and helpful comments. The calculations were done on a SGI 340 GTX, an IBM RISC/6000 550, and a CONVEX C220. Inquiries concerning the CHARMM program and the MCSS method should be addressed to the corresponding author (email: marci@tammy.harvard.edu).

References

(1) (a) Ripka, W. C.; Blaney, J. M. Computer graphics and molecular modeling in the analysis of synthetic targets. *Top. Stereochem.* **1991**, *20*, 1–85. (b) Hruby, V. J.; Al-Obeidi, F.; Kazmierski, W. Emerging approaches in the molecular design of receptor-selective peptide ligands: conformational, topographical and dynamic considerations. *Biochem. J.* **1990**, *268*, 249–262.

(2) Jolles, G.; Woolridge, K. R. H. *Drug Design: Fact or Fantasy?* Academic Press: London, 1984.

(3) Dean, P. M. *Molecular foundations of drug-receptor interactions*; Cambridge: University Press, 1987.

(4) Noble, M. E. M.; Verlinde, C. L. M. J.; Groendijk, H.; Kalk, K. H.; Wierenga, R. K.; Hol, W. G. J. Crystallographic and molecular modelling studies on trypanosomal triosephosphate isomerase: a critical assessment of the predicted and observed structures of the complex with 2-phosphoglycerate. *J. Med. Chem.* **1991**, *34*, 2709–2718.

(5) Miranker, A.; Karplus, M. Functionality maps of binding sites: a multiple copy simultaneous search method. *Proteins: Struct., Funct., and Genet.* **1991**, *11*, 29–34.

(6) Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078–1082.

(7) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. *Science* **1989**, *246*, 1149–1152.

(8) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F., Jr.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.

(9) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. H. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **1989**, *245*, 616–621.

(10) Li, Z.; Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 6611–6615.

(11) Li, Z.; Scheraga, H. A. Structure and free energy of complex thermodynamic systems. *J. Mol. Struct. (Theochem.)* **1988**, *179*, 333–352.

(12) Nayeem, A.; Vila, J.; Scheraga, H. A. A comparative study of the simulated-annealing and Monte Carlo-with-minimization approaches to the minimum-energy structures of polypeptides: [Met]-Enkephalin. *J. Comput. Chem.* **1991**, *12*, 594–605.

(13) Caflisch, A.; Niederer, P.; Anliker, M. Monte Carlo-docking of oligopeptides to proteins. *Proteins: Struct., Funct., and Genet.* **1992**, *13*, 223–230.

(14) Caflisch, A.; Niederer, P.; Anliker, M. Monte Carlo-minimization with thermalisation for global optimization of polypeptide conformations in cartesian coordinate space. *Proteins: Struct., Funct., Genet.* **1992**, *14*, 102–109.

(15) (a) Goodford, P. J. A computational procedure for determining energetically binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857. (b) Bobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **1989**, *32*, 1083–1094. (c) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding sites on Molecules of Known Structure. 1. Ligand Probe Groups with the Ability to Form two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36*, 140–147. (d) Wade, R. C.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding sites on Molecules of Known Structure. 2. Ligand Probe Groups with the Ability to Form More Than Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36*, 148–156.

(16) Moon, J. B.; Howe, W. J. Computer design of bioactive molecules: A method for receptor-base de novo ligand design. *Proteins: Struct., Funct., and Genet.* **1991**, *11*, 314–328.

(17) Böhm, H. J. The Computer Program LUDI: A new method for De Novo design of enzyme inhibitors. *J. Comput. Aid. Mol. Des.* **1992**, *6*, 61–78.

(18) Böhm, H. J., LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads, *J. Comput. Aid. Mol. Des.* **1992**, *6*, 593–606.

(19) Elber, R.; Karplus, M. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* **1990**, *112*, 9161–9175.

(20) Brooks, R. R.; Bruccoleri, B. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(21) Erickson, J.; Neidhart, D. J.; VanDrie, J.; Kempf, D. J.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, J.; Wideburg, N.; Kohlbrenner, W. E.; Simmer, R.; Helfrich, R.; Paul, D. A.; Knigge, M. Design, activity, and 2.8 Å crystal structure of a C symmetric inhibitor complexed to HIV-1 protease. *Science* **1990**, *249*, 527–533.

(22) Swain, A. L.; Miller, M. M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. X-ray crystallographic structure of a complex between a synthetic protease of human immunodeficiency virus 1 and a substrate-based hydroxyethylamine inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.

(23) Roux, B.; Yu, H.-A.; Karplus, M. Molecular basis of the Born model of ion solvation. *J. Phys. Chem.* **1990**, *94*, 4683–4688.

(24) Schulz, G. E.; Schirmer, R. H. Principles of Protein Structure. *Principles of Protein Structure*; Springer-Verlag: New York, 1979.

(25) Fletcher, R.; Reeves, C. M. Function minimization by conjugate gradients. *Comput. J.* **1964**, *7*, 149–154.

(26) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

(27) Brünger, A. T.; Karplus, M. Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 148–156.

(28) Jaskolski, M.; Tomasselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrikson,, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. Structure at 2.5 Å resolution of chemically synthesized human immunodeficiency virus type 1 protease complexed with a hydroxyethylene-based inhibitor. *Biochemistry* **1991**, *30*, 1600–1609.

(29) Rossky, P. J.; Karplus, M. Solvation. A molecular dynamics study of a dipeptide in water. *J. Am. Chem. Soc.* **1979**, *101*, 1913–1937.

(30) Rich, D. H.; Green, J.; Toth, M. V.; Marshall, G. R.; Kent, S. B. H. Hydroxyethylamine analogues of the p17/p24 substrate cleavage site are tight-binding inhibitors of HIV protease. *J. Med. Chem.* **1990**, *33*, 1285–1288.

(31) Tomasselli, A. G.; Hui, J. O.; Sawyer, T. K.; Staples, D. J.; Bannow, C.; Reardon, I. M.; Howe, W. J.; DeCamp, D. L.; Craik, C. S.; Heinrikson, R. L. Specificity and inhibition of proteases from human immunodeficiency viruses 1 and 2. *J. Biol. Chem.* **1990**, *265*, 14675–14683.

(32) Lee, B.; Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **1971**, *55*, 379–400.

(33) Fitzgerald, P. M. D.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C.; Herber, W. K.; Dixon, R. A. F.; Darke, P. L. Crystallographic analysis of a complex between human immunodeficiency virus type 1 and acetyl-pepstatin at 2.0 Å resolution. *J. Biol. Chem.* **1990**, *265*, 14209–14219.
(34) Bone, R.; Vacca, J. P.; Anderson, P. S.; Holloway, M. K. X-ray crystal structure of the HIV protease complex with L-700,417, an inhibitor with pseudo C_2 symmetry. *J. Am. Chem. Soc.* **1991**, *113*, 9382–9384.
(35) Suguna, K.; Padlan, E. A.; Smith, C. W.; Carlson, W. D.; Davies, D. R. Binding of a reduced peptide inhibitor to the aspartic proteinase from *Rhizopus chinensis*: Implications for a mechanism of action. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 7009–7013.
(36) James, M. N. G.; Sielecki, A. R. *Biological macromolecules and Assemblies*. F. A. Jurnak, F. A., McPherson, A., Ed., 1987; Vol. 3, pp 413–482.
(37) Blundell, T. L.; Cooper, J.; Foundling, S. I.; Jones, D. M.; Atrash, B.; Szelke, M. On the rational design of renin inhibitors: X-ray studies of aspartic proteinases complexed with transition state analogues. *Biochemistry* **1987**, *26*, 5585–5590.
(38) Šali, A.; Veerapandian, B.; Cooper, J. B.; Foundling, S. I.; Hoover, D. J.; Blundell, T. L. High-resolution X-ray diffraction study of the complex between endothiapepsin and an oligopeptide inhibitor: the analysis of the inhibitor binding and description of the rigid body shift in the enzyme. *EMBO J.* **1989**, *8*, 2179–2188.
(39) Veerapandian, B.; Cooper, J. B.; Šali, A.; Blundell, T. L. X-ray analyses of aspartic proteinases. III. Three-dimensional structure of endothiapepsin complexed with a transition-state isostere inhibitor of renin at 1.6 Å resolution. *J. Mol. Biol.* **1990**, *216*, 1017–1029.
(40) Suguna, K.; Padlan, E. A.; Bott, R.; Boger, J.; Parris, K. D.; Davies, D. R. Structures of complexes of rhizopuspepsin with pepstatin and other statine-containing inhibitors. *Proteins: Struct., Funct., Genet.* **1992**, *13*, 195–205.
(41) Dhanaraj, V.; Dealwis, C. G.; Frazao, C.; Badasso, M.; Sibanda, B. L.; Tickle, I. J.; Cooper, J. B.; Driessen, H. P. C.; Newman, M.; Aguilar, C.; Wood, S. P.; Blundell, T. L.; Hobart, P. M.; Geoghegan, K. F.; Ammirati, M. J.; Danley, D. E.; O'Connor, B. A.; Hoover, D. J. X-ray analyses of peptide-inhibitor complexes define the structural basis of specificity for human and mouse renins. *Nature* **1992**, *357*, 466–472.
(42) Northrup, S. H.; Pear, M. R.; Morgan, J. D.; McCammon, J. A.; Karplus, M. Molecular dynamics of ferrocytochrome c. *J. Mol. Biol.* **1981**, *153*, 1087–1109.
(43) Warwicker, J.; Watson, H. C. Calculation of electric potential in the active site cleft due to α-helix dipoles. *J. Mol. Biol.* **1982**, *157*, 671–679.
(44) Gilson, M. K.; Honig, B. H. Energetics of charge-charge interactions in proteins. *Proteins: Struct., Funct., Genet.* **1988**, *3*, 32–52.
(45) Bashford, D.; Karplus, M. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **1990**, *29*, 10219–10225.
(46) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. Electrostatics and diffusion of molecules in solution: simulations with the university of Houston Brownian dynamics program. *Comp. Phys. Commun.* **1991**, *62*, 187–197.
(47) Wesson, L.; Eisenberg, D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1992**, *1*, 227–235.
(48) Pettitt, B. M.; Karplus, M.; Rossky, P. J. Integral equation model for aqueous solvation of polyatomic solutes: application to the determination of the free energy surface for the internal motion of biomolecules. *J. Phys. Chem.* **1986**, *90*, 6335–6345.
(49) Ben-Naim, A.; Marcus, Y. Solvation thermodynamics of nonionic solutes. *J. Chem. Phys.* **1984**, *81*, 2016–2027.
(50) Gao, J. Theoretical studies of condensed phase chemistry. Ph.D. thesis, Purdue University, 1987.
(51) Arrowsmith, R. J.; Carter, K.; Dann, J. G.; Davies, D. E.; Harris, C. J.; Morton, J. A.; Lister, P.; Robinson, J. A.; Williams, D. J., Novel renin inhibitors: Synthesis of aminostatine and comparison with statine-containing analogues. *J. Chem. Soc., Chem. Commun.* **1986**, 755–757.