Structure-Based Ligand Design by a Build-Up Approach and Genetic Algorithm Search in Conformational Space

NICOLAS BUDIN, NICOLAS MAJEUX, CATHERINE TENETTE–SOUAILLE, AMEDEO CAFLISCH

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received 13 April 2001; accepted 29 June 2001

ABSTRACT: Program to engineer peptides (PEP) is a build-up approach for ligand docking and design with implicit solvation. It requires the knowledge of a seed from which it iteratively grows polymeric ligands consisting of any type of amino acid, i.e., natural and/or nonnatural from a user-defined library. At every growing step, a genetic algorithm is used for conformational optimization of the last added monomer in the rigid binding site. Pruning is performed at every growing step by selecting sequences according to binding energy with electrostatic solvation. PEP is applied to three members of the caspase family of cysteine proteases using Asp at P_1 as seed. The optimal P_4–P_2 peptide recognition motifs and variants thereof are docked correctly in the active site (backbone root-mean-square deviation < 0.9 Å). Moreover, for each caspase, the P_4–P_2 sequences of potent aldehyde inhibitors are ranked among the 15 hits with the most favorable PEP energy. J Comput Chem 22: 1956–1970, 2001

Keywords: ligand docking; structure-based design; genetic algorithm; implicit solvation; peptidomimetics

Introduction

Computer-aided structure-based ligand design is a multidisciplinary and challenging research topic with broad applications in medicine and biotechnology. It is concerned with the prediction of chemically reasonable compounds that are expected to bind strongly to key regions of biologically relevant molecules (e.g., enzymes and receptors) of known three-dimensional structures so as to inhibit or alter their activity. Despite significant progresses in computational approaches for ligand design and efficient evaluation of binding energy,[1–3] novel procedures for ligand design are required. This is motivated by the genome projects[4–6] and the ever increasing number of protein targets for

Correspondence to: A. Caflisch; e-mail: caflisch@bioc.unizh.ch
Contract/grant sponsor: Novartis Pharma (Basel)

drug design that are being characterized functionally and structurally by major advances in both experimental methods for structure determination and high-throughput homology modelling.[7]

This article describes a computational approach for the structure-based docking and design of peptidic ligands consisting of natural and/or nonnatural amino acids. Ligands are grown from a seed by iteratively adding amino acids to the actual construct. The seed and the last added fragment can be any type of chemical entity, whereas the remaining monomers must have an amino group (primary or secondary) and a carboxyl group. The search in chemical space is performed by a build-up approach that employs all of the residue topologies of a user-defined library. At every growing step, a genetic algorithm is used for conformational optimization of the last added monomer inside the binding site of a rigid target protein. The approach presented in this article makes use of an implicit solvation model to efficiently rank the designed peptides according to their binding energy in solution, and to select sequences for further growing. It is implemented in the program PEP (program to engineer peptides) and applied to three enzymes of the caspase family.

The main disadvantage of any growing procedure is inherent to its sequential approach. The success of any growing step depends largely on the previous step(s), and the current step has no knowledge of the growing step(s) that will follow. In addition, growing must be restricted both at the sequence and conformation levels to avoid combinatorial explosion. Optimal binding modes for each sequence are required to correctly rank different sequences. The correct docking and rank ordering are important because only a limited number of sequences are kept for further growing. Determining the best binding mode for a given peptide is not an easy task because it may not always correspond to the minima of the same peptide when it is part of a longer sequence. PEP performs a dead-end test to check for the feasibility of elongation. Furthermore, during dihedral angle optimization of the nth residue PEP also allows for partial rigid body rotation of the nth residue with respect to the $(n-1)$th residue, which is important for two reasons. First, it provides the growing algorithm with an effective way to do small corrections on peptide backbone orientations that are not optimal for further growing. Second, partial rigid body rotation is used to overcome the geometrical restrictions due to the use of fragments with rigid bond lengths and angles. This allows the algorithm to reproduce X-ray structure binding modes that contain nonideal bond geometries, and that would otherwise be out of reach.

Caspases are cysteine proteases that drive the intracellular pathways leading to apoptosis and pro-inflammatory cytokines activation.[8] As for most proteases, the substrate-binding cleft of caspases can be subdivided in a number of subsites, each consisting of residues important for the recognition of the substrate. These subsites are referred to as $S_n, S_{n-1}, \ldots, S_1, S_1', \ldots, S_m'$ and the catalytic cysteine is placed between the S_1 and S_1' position. The corresponding substrate residues are referred to as P_n to P_m', with the scissile bond between P_1 and P_1'. Caspases have a near absolute specificity for peptides with an aspartic acid at P_1. Using a combinatorial approach with positional scanning of synthetic tetrapeptidyl-aminomethyl coumarin derivatives, Thornerberry and coworkers explored the S_4–S_2 subsite occupancy.[9] Based on S_4 subsite preferences, they subdivided the caspases into three substrate specificity groups. Group I consists of caspase 1, 4, and 5 with a preference for aromatic amino acid residues in the P_4 position; group II contains caspases 2, 3, and 7 with a near absolute specificity for Asp in P_4; and finally, group III includes caspases 6, 8, 9, and 10 with a preference for Leu, Val, Ile, and Asp at P_4. P_4 substrate preferences were confirmed in another study that also investigated P_1 and P_1' specificities.[10] PEP is applied to three caspases (1, 3, and 8) that are representative of the three groups. Using the Asp at P_1 as a seed, PEP finds the correct binding mode for the P_4–P_2 tripeptide sequences of the recognition motifs. Furthermore, for each of the three caspases the P_4–P_2 sequence of one or more known aldehyde inhibitors is reproduced among the 15 better scoring PEP hits. This represents a successful search in the space of the 8000 tripeptides consisting of natural amino acids.

Methods

GROWING PROCEDURE

The aim of PEP is to construct peptides from one or many user-selected starting positions (seeds) by iteratively adding amino acids in conformations that interact most favorably with the residues in the receptor binding site. The default number of sequences kept at each growing step is 10. Within the approximation that chemical entity and orientation of a monomer are not affected by the successive monomers, the search is exhaustive, because at each step of growing every amino acid in the user-defined topology library is attached to the actual

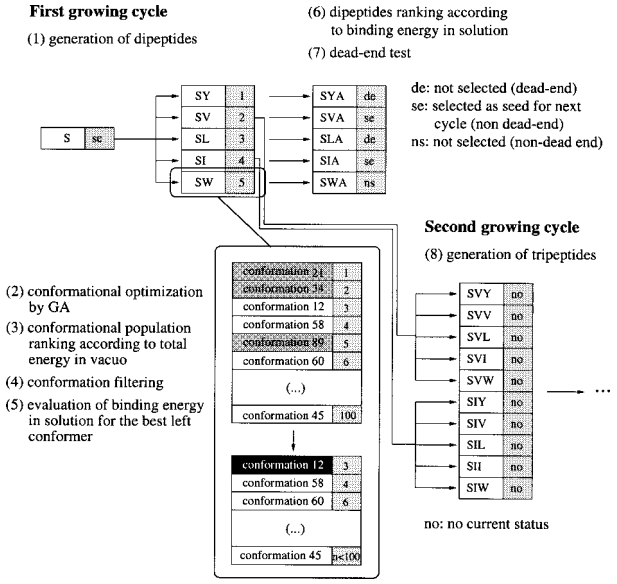

FIGURE 1. Schematic representation of one cycle of growing in PEP. Each ligand is represented by a box containing two fields that indicate the ligand sequence and status (gray background). A number in the status field corresponds to the ligand final ranking according to its binding energy in solution. Serine (S) is used as seed. The fragment library contains five amino acids (Y: tyrosine, V: valine, I: isoleucine, L: leucine, W: tryptophan), and two dipeptide sequences are kept for the next growing cycle. Dead-end testing is illustrated by the rejection of ligand SY, although it is ranked number one. The two best scoring and nondead-end ligands are selected for the next cycle of growing. They correspond to sequences SV and SI, which are ranked two and four, respectively. A detailed illustration of the dipeptide SW conformational population ranking and filtering is shown (bottom left). Each SW structure is represented by a box containing two fields that indicate the conformation number and ranking according to total energy *in vacuo* (light gray background). In the GA, the conformational population size is 100. Conformations that are discarded upon filtering are identified by dark-gray backgrounds. The best remaining SW conformation used for the final ligand ranking is shown with a black background. The small size of the fragment library and the small number of grown peptides are used for the sake of illustration clarity and do not correspond to the actual values used in this study.

construct (Fig. 1). The topology library contains the atom types, atomic partial charges, covalent bonds, and a list of rotatable dihedrals for each amino acid. Because it is computationally prohibitive to evaluate protein and ligand desolvations during the GA (genetic algorithm) optimization, the *in vacuo* total energy (intermolecular plus intramolecular) of the last added monomer is optimized by the GA, while most of the already grown ligand is kept rigid. The conformations of the last added monomer are then ranked according to their *in vacuo* energy, and filters are applied (as described below) to discard residues with internal hydrogen bonds, and rotamers that would lead to further growing in a forbidden direction.

After all the amino acids of the topology library have been minimized, the binding energy in solution is calculated for the best rotamer of each residue, and only the highest scoring sequences are retained for the next level of growth. The program then tests if the latter are dead ends, i.e., if there is no space for further growing, or elongating them will lead only to poor interactions with the receptor. The latter case usually happens when the peptide grows away from the receptor surface. To test this, an alanine is attached to the peptide candidate and GA minimized. The corresponding sequence is kept if the *in vacuo* binding energy of the alanine minimum conformation is better than a given energy cutoff (a cutoff value of −10 kcal/mol is used in the applications presented here). The *in vacuo* binding energy is a good indicator of the quality of the interactions between the last added amino acid and the protein: an amino acid conformation has an unfavorable van der Waals energy contribution when it bumps into the protein while its binding energy is very small in absolute value when it grows away from the binding site. This procedure is then repeated on the second growth level; each amino acid in the library is attached to each of the 10 dipeptide sequences retained from the first step, minimized, and then scored. Successive growth levels, therefore, generate peptides that are lengthened by one residue. The procedure terminates when the user-defined peptide length is reached. The output data provided by PEP include residue sequences, energies, and atomic coordinates of the peptide in the PDB format.

DEGREES OF FREEDOM DURING GROWING

PEP uses amino acid templates in which the amine can be either primary or secondary (Fig. 2A). This includes L- and D-residues, as well as nonstandard amino acids and peptoids (N-alkylated peptides). The purpose of the acetyl and amide end groups is twofold: to provide the polar groups for intermolecular hydrogen bonds and to take into account some of the conformational restriction experienced by individual amino acids when they are connected in a polypeptide chain.[11] The seed and

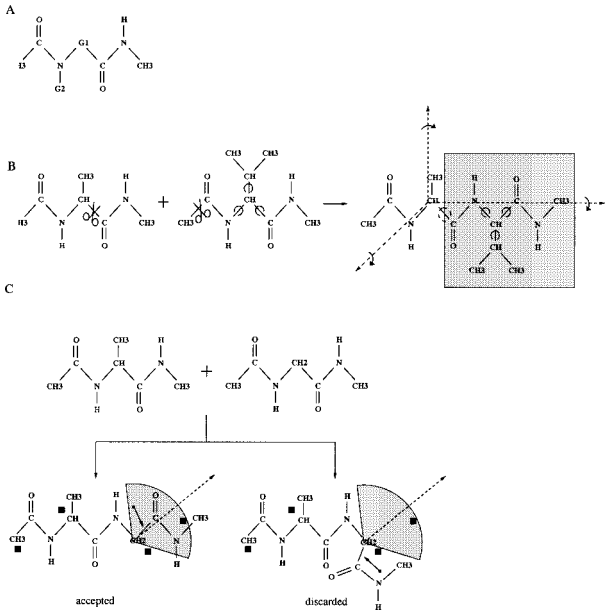

FIGURE 2. (A) Amino acid template used by PEP. G2 indicates the substituent position on the template amino group. G1 can be of any type, without size limitation. (B) Illustration of the flexibility in PEP during the growing. A fully flexible valine is grown from an alanine. Alanine can be either the seed, or a residue positioned during the previous growing cycle. Covalent bonds that are broken upon growing are indicated by a scissor symbol. Rotatable bonds are marked with circular arrows. In addition to the valine internal flexibility, the alanine ψ dihedral (circular dashed arrow) is also flexible during the valine conformational optimization. Cartesian coordinate frame used for partial rigid body rotation are drawn with dashed lines. The peptide region affected by the partial rigid body rotation is delimited by a gray background rectangle. (C) Illustration of the directionality filtering. Two rotamers of the added residue are shown. The CH_2—C bond defines the direction of the rotamer and is indicated by a short arrow. Track points are symbolized by black squares. The range of allowed growing directions is indicated by a gray pie slice that is centered on the vector (dashed arrow) whose orientation is determined by the two closest track points (see Methods section for details).

the last monomer are, however, not restricted to amino acids and can be any molecule. The side chain and backbone rotatable bonds of the last added residue are flexible during the conformational optimization. Additionally, the backbone rotatable bond of the previous residue, which is the closest to the currently minimized amino acid, is also flexible (Fig. 2B). For α-amino acids, this corresponds to ψ and ϕ when growing in the N to C and C to N direction, respectively.

X-ray structures often contain bond lengths/angles that deviate significantly from their ideal values. Minimization is used to obtain a structure that corresponds to a minimum in the force field. This is, however, not possible, because PEP uses residues with optimal bond lengths and angles that are kept constant throughout the growing procedure. The accessible search space is, therefore, limited to ideal covalent geometries. This restriction combined with the ruggedness of the energy function, originating mainly from the van der Waals term, may prevent PEP from finding the correct binding mode. Backbone bond angle deviations are particularly critical because they influence the space that is accessible for growing. To improve sampling, PEP allows partial rigid body rotation of the last added peptide around the three axis of a Cartesian reference frame centered on the C_α of the previous residue (Fig. 2B). Partial rigid body rotation allows to compensate in part for the fixed covalent angles, and provides a mean to access growing directions that otherwise would be out of reach. Moreover, combined with the additional rotatable bond mentioned above, it prevents the growing direction from being restricted to the optimal orientation of the terminal N-methyl amide group at the previous growing step. Although more conformational space could be sampled by allowing for nonideal covalent angles, partial rigid body rotation is much more efficient because it adds only three degrees of freedom. In the present implementation, the rigid body rotation is restricted to ± 8 degrees, and the energy term of the covalent angles around the C_α atom at the center of the Cartesian frame (Fig. 2B) is neglected.

GENETIC ALGORITHM FOR LIGAND CONFORMATIONAL SEARCH

A GA is a stochastic optimization method that mimics the process of natural evolution by manipulating a population of data structures called chromosomes.[12, 13] Starting from an initial randomly generated population of chromosomes, the GA repeatedly applies two mutually exclusive genetic operators, one-point crossover and mutation, which yield new chromosomes (children) that replace appropriate members of the population.

Data Structure in Chromosomes

Amino acids can have many rotatable bonds. It, therefore, takes too long to perform an exhaustive conformational search, unless a large increment angle is used. This, however, usually leads to poor results because of the ruggedness of the energy

landscape due to the van der Waals term. In the GA used in PEP, each chromosome contains so-called genes that encode the values of the angles of rotation around the rotatable bonds of the added residue, the ψ or ϕ dihedral angle of the preceding residue, and three angles that define the rigid body orientation of the added residue. A chromosome of $N + 4$ genes, therefore, encodes the orientation and the conformation of a residue with N rotatable bonds. The rotatable bond genes are binary encoded in a string of six bits that describes an integer value between 0 and 64. This integer value is linearly rescaled to a real number between 0 and 2π with a theoretical resolution of 5.6 degrees. The genes that encode partial rigid body rotation are binary encoded in a string of four bits that corresponds to an integer value between 0 and 16. This integer value is rescaled to a real number that ranges from -8 to $+8$ degrees, and that corresponds to the angle difference between the current and the initial residue orientation around the corresponding axis. Small rigid body rotations up to 8 degrees are, therefore, performed in both directions around the three rigid body axis of rotation. The theoretical rigid body orientation accuracy is 1 degree.

Fitness Scaling and Parent Selection

Both genetic operators (see below) are applied to parent chromosomes randomly selected from the existing population with a bias toward the fittest. This selection is analogous to spinning a roulette wheel with each member of the population having a slice of the wheel that is proportional to its fitness. The emphasis on the survival of the fittest introduces an evolutionary pressure into the algorithm, and ensures that over time the population should move toward the minimum conformation(s). The chromosomes are first ordered by decreasing energies. To avoid premature convergence, linear normalization is used for the chromosome fitness values. A constant value is assigned to the last chromosome in the list and the remaining fitness values are increased linearly. In the present application, 100 chromosomes were used, the worst chromosome was assigned a fitness value of 500, and the increment was 10. This corresponds to a selection pressure of 1.5 (the selection pressure represents the relative probability that the best individual will be chosen as a parent compared with the average individual). Because, as mentioned above, the probability of selection is proportional to the fitness, the chromosome with the most favorable energy has a three-time (1500/500) larger probability to be selected as a parent than the chromosome with the poorest energy, irrespective of the absolute values of the energy. The parameters used in fitness scaling were chosen after preliminary test runs but have not been systematically optimized.

Genetic Operators

One-point crossover is a binary operator that creates two new chromosomes by swapping two segments of two parent chromosomes after a randomly selected gene. Mutation is a unary operator that leads to a new chromosome by randomly flipping the bits of selected genes of the parent chromosome. The crossing over and mutation operators are mutually exclusive, meaning that either one or the other of these operators can be applied during the generation of a new chromosome, but not both. At each reproduction event, the operator is chosen using the roulette wheel method so that mutation and crossing over are selected with a chance of 80 and 20%, respectively. These parameters are also used in a GA method for flexible ligand docking developed by others.[14] Once the mutation operator has been selected for a given chromosome, each of its genes has three chances out of 10 to be mutated.

Evolution of the Population

The selection of the members of the population that should be replaced by new chromosomes is a crucial step. To avoid premature convergence it is very important to keep structural diversity. Hence, the nicheing method[13] is modified such that both the energy difference and the conformational similarity are taken into account to determine if a given member of the population should be replaced by a new chromosome (Fig. 3). Each new chromosome is tested for similarity with the energy sorted population, starting from the best ranking member, until a similar chromosome is found. Two chromosomes are considered similar if all their genes have a dissimilarity score that is smaller than the dissimilarity cutoff. For two given genes, the dissimilarity score is obtained by dividing the difference of their encoded values by the difference of their extrema values (360° for a rotatable bond). The dissimilarity score therefore ranges from 0 to 1, 0 being the score obtained by two identical genes. If a similar chromosome is found in the population, it is replaced by the new chromosome only if the energy of the new one is more favorable. Otherwise, the new chromosome is discarded. If no similar chromosome is found in the population, the dissimilarity cutoff

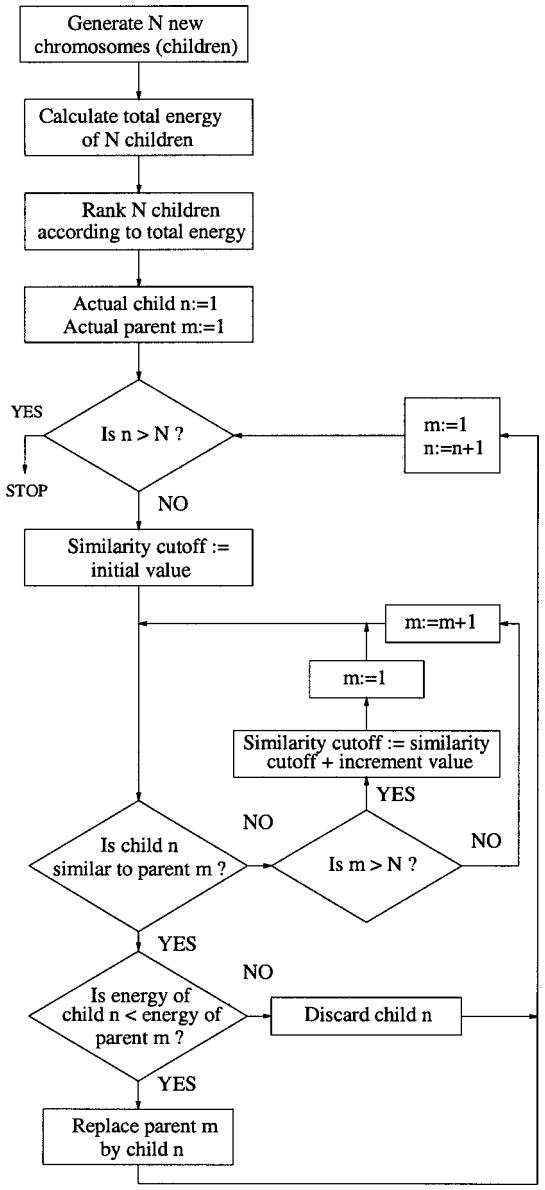

FIGURE 3. Schematic representation of one cycle of conformational optimization by the GA with emphasis on the chromosome replacement algorithm. Both parent and children populations have a size of *N*. Variable assignment are symbolized by ":=". Conditional statements are enclosed by diamonds.

is linearly increased until a similar chromosome is found. The whole procedure is repeated until all the new chromosomes have been tested. The initial dissimilarity cutoff and increment values were 0.001 (0.36 degrees) and 0.05 (18 degrees), respectively. The very small initial value is necessary to discard identical conformations originating from the mating event.

GA Parameters

For each GA conformational optimization, a population of 100 chromosomes was used and 1500 cycles were performed. At each GA cycle, 100 new chromosomes were generated for a total of 1.5×10^5 conformations tested during the overall GA optimization. These parameters yielded good convergence after preliminary docking runs on HIV-1 aspartic protease, tyrosine kinase and the three caspases. Using the 20 standard amino acids, the growing of a tetrapeptide requires 620 GA runs and 9.3×10^7 energy evaluations ($[20+(3\times200)]\times1.5\times10^5$ where the number in brackets is the sum of the GA optimizations performed at the first and three subsequent growing steps, and the 200 originates from 10 kept sequences times 20 residues). Of the 9.3×10^7 energy evaluations required by the design of tetrapeptides, 620 include full electrostatic solvation while the remaining ones use the distance dependent dielectric function (see below).

TOTAL ENERGY *IN VACUO* AND BINDING ENERGY IN SOLUTION

During the conformational optimization by the GA [point (2) in Fig. 1], the sum of the intraligand and intermolecular energies is calculated for each new conformation of the flexible amino acid. The total energy consists of the following contributions:

$$E_{\text{total}} = E_{\text{vdW}}^{\text{interm}} + E_{\text{elect, rdiel}}^{\text{interm}} + E_{\text{vdW}}^{\text{ligand}} + E_{\text{elect, rdiel}}^{\text{ligand}} + E_{\text{strain}}^{\text{ligand}} \quad (1)$$

The last three terms approximate the intraligand energy of the flexible amino acid, which consists of the van der Waals ($E_{\text{vdW}}^{\text{ligand}}$), electrostatic ($E_{\text{elect, rdiel}}^{\text{ligand}}$), and the strain energy ($E_{\text{strain}}^{\text{ligand}}$). The latter contains only the dihedral terms, because the bond lengths and angles are kept constant. These energy contributions are calculated explicitly for each appropriate set of ligand atoms. The receptor–ligand interaction energy is described by the first two terms, where $E_{\text{vdW}}^{\text{interm}}$ and $E_{\text{elect, rdiel}}^{\text{interm}}$ are the intermolecular van der Waals and Coulombic energy, respectively. Because the receptor is rigid, both energy terms are mapped on look-up tables to improve the efficiency.[15] The solvent screening effect is approximated by the distance dependent dielectric model. The desolvation penalty is neglected during conformational optimization by the GA because of computational efficiency. All of the energy parameters used in the applications presented here are taken from CHARMm22 (MSI Inc., San Diego) but

any other force field with explicit dihedral, Coulombic, and van der Waals terms could be used.

At each growing step, the best binding modes obtained by the conformational optimization of each sequence are ranked according to their binding energy in solution [point (5) in Fig. 1], where the electrostatic energy is evaluated within the continuum electrostatic approximation:[16–18]

$$\Delta G_{\text{binding}} = \Delta G_{\text{vdW}}^{\text{interm}} + \Delta G_{\text{elect, sol}}^{\text{interm}} + \Delta G_{\text{elect, desolv}}^{\text{receptor}} + k\Delta G_{\text{elect, desolv}}^{\text{ligand}} + \Delta G_{\text{entropy}}^{\text{ligand}} \quad (2)$$

It is assumed that the ligand–receptor vdW interaction energy ($\Delta G_{\text{vdW}}^{\text{interm}}$) accounts for all the nonelectrostatic contributions to the binding energy.[19] The sum of the receptor desolvation ($\Delta G_{\text{elect, desolv}}^{\text{receptor}}$), screened intermolecular interaction ($\Delta G_{\text{elect, sol}}^{\text{interm}}$), and ligand desolvation ($\Delta G_{\text{elect, desolv}}^{\text{ligand}}$) represents the difference in electrostatic energy in solution upon binding of a ligand to a receptor. The desolvation of the receptor is the electrostatic energy difference upon binding of an uncharged ligand to a charged receptor in solution. It is calculated by numerical integration of the energy density of the electric field. For this integration, the electric displacement of every partial charge of the receptor is approximated by the Coulomb field and the energy density is discretized over a 3D grid.[18] The screened ligand–receptor interaction and the desolvation of the ligand are evaluated using the generalized Born approximation.[16, 18, 20] A scaling factor of $k = 0.6$ is applied to the ligand desolvation term to take into account the fact that the desolvation is smaller for a residue that is part of a larger ligand.[21]

$\Delta G_{\text{entropy}}^{\text{ligand}}$ is a penalty term that represents the loss of entropy when ligand side chain rotatable bonds are frozen upon binding to the receptor

$$\Delta G_{\text{entropy}}^{\text{ligand}} = c \sum_i \frac{6 - n_i}{4} \quad (3)$$

where $c = 1.0$ kcal/mol and n is the number of heavy atoms covalently bound to the two atoms at the center of the quartet, for example, $n = 2$ in butane and $n = 4$ in 2,3-dimethylbutane. The index i runs over all side chain rotatable bonds of the ligand.

$\Delta G_{\text{binding}}$ does not contain the dihedral term because it is not possible to directly compare the strain for different sequences.

CONFORMATIONAL FILTERING

The genetic algorithm is used to minimize the vacuo energy of the conformation of the last added residue. It is, however, designed to keep population heterogeneity, and therefore produces clusters of conformations that correspond to local minima. In cases where multiple binding modes are found, it is very difficult to predict which one should be used for the next growing step, especially when they have similar energies. Moreover, the global minimum conformation of the last added residue may not correspond to the lowest energy in the context of a longer peptide. This is a clear limitation of the growing approach. This problem, however, arises mostly with residues that do not make strong interaction with the binding site surface, and is well illustrated by the S_3 pocket of the three caspases investigated in this study. S_3 contains an Arg, which is usually involved in a hydrogen bond or salt bridge with the side chain of the ligand that fills the pocket. Additionally, the ligand backbone is hydrogen bonded to the protein. PEP finds two binding modes for apolar residues in P_3. The first one corresponds to the backbone binding mode found in the X-ray structure, whereas in the second mode, the peptide backbone carbonyl is hydrogen bonded to the Arg. The second mode has the most favorable energy *in vacuo* and binding energy in solution, and is, therefore, selected by PEP. A similar problem arises when polar residues are placed in hydrophobic pockets. The binding mode that contains an intraligand hydrogen bond has a more favorable energy than its counterpart, which interacts only weakly with the pocket surface. The obvious solution would be to keep the best energy conformation for each residue binding mode, but this would lead to an intractable number of sequences/conformations to be grown.

Two filters are, therefore, applied on the final conformations found by the GA to find out the most suitable conformational minima for the following growing step. The first filter discards residue conformations that contain one or more intraresidue hydrogen bonds. These have favorable *in vacuo* energies mainly because of the internal energy contributions. The second filter is a two-step procedure used to screen rotamers according to their ability to be grown in a specific direction. First, in a preprocessing step PEP is used to dock a known peptidic ligand in the protein binding site. The resulting C_α atom coordinates define a path that is used to filter peptide rotamers during ligand design. For every designed sequence, the vector that defines the growing direction is given by the two atoms that are the closest to the flexible part in the next growing cycle (Fig. 2C). For α-amino acids, this corresponds to the C_α and carbonyl carbon atoms

when growing in the N to C direction, and the C_α and nitrogen atoms when growing from C to N. Their two closest path points make up a direction vector that defines the peptide growing direction. The angle between the two vectors is calculated, and the rotamer is discarded if the angle absolute value is above a given angle cutoff. In the present study, a cutoff value of 60 degrees was used. It should be clear that this procedure does not filter out rotamers according to their position in the binding site, but restricts only the direction of the growing. A large part of the binding site is, therefore, still accessible for design. Hence, the direction vectors and dead-end test are complementary and not redundant because the former influences the overall orientation of the peptide backbone whereas the latter is used to forbid specific cul-de-sac regions in the binding site or growing completely out of it. Moreover, no bias is introduced because the directions are generated automatically by the program.

SYSTEM SETUP

The entries 1IBC, 1CP3, and 1QDU were downloaded from the PDB database.[22] 1IBC is the 2.73-Å resolution X-ray structure of the human cysteine protease interleukin-1β converting enzyme (caspase 1), complexed with the tetrapeptide inhibitor acetyl-Trp-Glu-His-Asp-aldehyde (Ac-WEHD-CHO).[23] 1CP3 is the 2.3-Å resolution X-ray structure of the human cysteine protease apopain (caspase 3), complexed with the tetrapeptide inhibitor acetyl-Asp-Val-Ala-Asp fluoromethyl ketone (Ac-DVAD-fmk).[24] 1QDU is the 2.80-Å resolution X-ray structure of the human cysteine protease caspase 8, complexed with the peptidic inhibitor benzyloxycarbonyl-Glu-Val-Asp-dichloromethylketone (Z-EVD-dcbmk).[25] For all structures water molecules and inhibitor were removed. Hydrogen atoms were added to the three systems and minimized with the CHARMM program.[26]

COMPUTATION TIMES

Growing tripeptides using the 20 natural amino acids requires about 20 h on a 550-MHz PentiumIII processor, while docking a single sequence takes less than 10 min. For the caspase applications, multiple PEP runs were performed in parallel on a cluster of PCs running the Linux operating system.

Results and Discussion

The enzymes of the caspase family represent an interesting test case for PEP, because comprehensive substrate specificity data are available[9] and inhibition constants (K_i) of a number of reversible inhibitors (mainly tetrapeptide aldehydes, ketones, and nitriles) have been measured.[27–30] Furthermore, several X-ray structures of caspase/inhibitor complexes have been determined.[23–25, 30, 31] The aim of the present study was to investigate the PEP ability to find the relevant binding modes of high-affinity inhibitors and to determine the preferred P_4–P_2 recognition motifs of caspases. The latter is a particularly challenging test case for any energy-based ligand design approach because both the caspase binding site and preferred peptide recognition motifs contain several charged side chains. For groups with formal charges, the continuum electrostatic approximation used by PEP for sequence ranking shows the largest deviation from finite-difference Poisson calculation.[18]

First, docking runs were performed with the optimal recognition motifs, i.e., the sequences Ac-WEHD, Ac-DEVD, and Ac-LETD for caspases 1, 3, and 8, respectively. The Asp residue at P_1 from the X-ray structure was used as starting seed and the P_4–P_2 sequences were grown in the C to N direction. Growing in the opposite direction (N to C) using the P_4 residue as seed (design of P_3–P_1) is also possible but was not attempted, as the backbone of residue P_1 would bump into the protein because of the covalent bond with the catalytic cysteine. In preliminary docking runs without the partial rigid body rotation, PEP was not able to reproduce the binding mode of Ac-WEHD in caspase 1. For each sequence, 20 docking runs with partial rigid body rotation were then performed with different initial random number values for the GAs. Binding modes were clustered according to backbone atom positions. For the three enzymes, the backbone conformation found by PEP with the highest frequency overlaps the inhibitor backbone in the X-ray structure.

For the search in sequence space (design), the C_α coordinates of the highest scoring docked conformation of the aforementioned peptide recognition motifs were used to define the direction of growing. The same seed as for the docking runs [Asp(P_1)] was used. On each of the three enzymes 15 PEP runs unrestricted in sequence space were performed with different initial random number values for the GA. A library containing the topologies of the 20

naturally occurring amino acids was used at each growing step. To analyze the results one has to consider that the GA performs a stochastic search; hence, the most favorable sequences are those that are generated in many runs and with a good binding energy in solution. If a given sequence is generated by PEP n times, it can assume up to n different conformations, when m growing runs are performed ($n \leq m$). Usually, most of the conformations of a given sequence are very similar among each other, which indicates that the GA search in conformational space reaches convergence. The sequences generated by PEP are first sorted according to the highest occurrence and then by the binding energy in solution averaged over the n conformations. A few of the sequences with low occurrence can have a very favorable relative binding energy. Yet, the fact that these sequences are not grown completely in most runs indicates that they have a poor average binding energy. It is, therefore, necessary not only to perform multiple PEP runs in parallel but also to rank the sequences according first to occurrence and then average binding energy. Tables I, II, and III contain sequences that were generated most often in the GA runs for caspases 1, 3, and 8, respectively. Docking and growing runs performed on the three enzymes are discussed in detail in the following sections.

CASPASE 1

Docking of the Optimal Peptide Recognition Motif (WEHD)

Eighteen of the 20 conformations of Ac-WEHD generated by PEP have the correct backbone orientation (Fig. 4A). The 18 backbone conformations of Ac-WEH (Asp at P_1 is neglected because it was used as seed) have an average RMSD from the X-ray structure of 0.88 Å. The four hydrogen bonds observed in the X-ray structure between the peptide backbone and caspase 1 are reproduced. These involve the NH of Asp(P_1) and the CO of Ser 339, the backbone polar groups of Glu(P_3) and Arg341, and the acetyl oxygen and the imidazole of His342. The binding mode of the His(P_2) side chain differs significantly from its X-ray structure counterpart. The all atom RMSD from the X-ray structure averaged over the 18 conformations is 1.99 ± 0.25 Å and 1.07 ± 0.12 Å with and without the His(P_2) side chain, respectively. PEP places the imidazole side chain in a cleft formed by the side chains of Pro177, His237, and Arg341. The Coulombic interaction between the

TABLE I.
P_4–P_2 Sequences Designed by PEP in Caspase 1.

Sequence			Occurrence[a]	Relative Binding Energy[b]	Backbone RMS Deviation[c]
P_4	P_3	P_2	(%)	(kcal/mol)	(Å)
W	**D**	**T**[d]	80	0.0	0.67
W	**D**	L	80	1.8	0.57
N	**D**	**T**	80	2.7	0.89
D	**D**	**T**	80	3.6	0.80
H	**D**	**T**	80	3.6	1.00
F	**D**	**T**	80	3.7	0.82
N	**D**	L	80	4.0	0.85
F	**D**	L	80	4.0	0.89
G	**D**	**T**	80	4.2	0.73
H	**D**	L	80	4.4	0.89

Among the 200 final sequences in each of the 15 PEP runs, 31, 33, and 18 sequences were found in 12, 11, and 10 runs, respectively. Residues in optimal substrate sequences are indicated in bold.[9] The P_4–P_2 optimal recognition motif of caspase 1 is WEH[9] and the aldehyde inhibitor Ac-WEHD-CHO has a K_i of 56 pM.[29]

[a] Percentage of PEP runs (out of 15) that generated a given sequence.

[b] Binding energy averaged over all conformations of a given sequence. The energy values are relative to the one of the most favorable sequence.

[c] Backbone RMS deviation from the X-ray structure. For each sequence, the conformation with the best energy was used to calculate the RMS deviation.

[d] Peptide shown in Figure 4B.

TABLE II.
P_4–P_2 Sequences Designed by PEP in Caspase 3.

Sequence			Occurrence[a]	Relative Binding Energy[a]	Backbone RMS Deviation[a]
P_4	P_3	P_2	(%)	(kcal/mol)	(Å)
W	**E**	**V**[b]	93	0.0	0.58
F	**E**	H	93	0.5	0.57
W	**E**	H[c]	93	1.3	0.70
F	**E**	**V**	93	1.6	0.57
H	**E**	**V**	93	1.9	0.53
H	**E**	H	93	2.0	0.43
Y	**E**	H	93	2.1	0.77
Y	**E**	**V**	93	2.5	0.52
N	**E**	**V**	93	3.2	0.53
Q	**E**	H	93	4.9	0.86

Among the 200 final sequences in each of the 15 PEP runs, 38, 40, and 20 sequences were found in 14, 13, and 12 runs, respectively. The P_4–P_2 optimal recognition motif of caspase 3 is DEV,[9] and the aldehyde inhibitor Ac-DEVD-CHO has a K_i of 230 pM.[29]
[a] See caption of Table I.
[b] Peptide shown in Figure 4D.
[c] The aldehyde inhibitor Ac-WEHD-CHO has a K_i of 2 μM for caspase 3.[29]

imidazole nitrogen and Arg341 favors this binding mode over the interaction with the hydrophobic S_2 pocket observed in the X-ray structure. This is supported by a docking run performed with a higher dielectric constant ($4r_{ij}$, where r_{ij} is the interatomic distance), which yielded the correct orientation of the His(P_2) side chain. The Glu(P_3) side chain is positioned correctly, and its salt bridge with Arg341 is reproduced. The Trp(P_4) side chain is placed in the S_4 pocket and overlaps its X-ray counterpart. It is interesting to note that PEP finds the correct binding mode of Trp(P_4) and Glu(P_3), despite the misplaced side chain of His(P_2) at the first growing step.

Design of Tetrapeptide Inhibitors

The best PEP hit is shown in Fig. 4B and the 10 P_4–P_2 sequences with the highest occurrence and

TABLE III.
P_4–P_2 Sequences Designed by PEP in Caspase 8.

Sequence			Occurrence[a]	Relative Binding Energy[a]	Backbone RMS Deviation[a]
P_4	P_3	P_2	(%)	(kcal/mol)	(Å)
R	**E**	**T**	100	0.0	0.85
V	**E**	**T**[b]	100	0.5	0.75
T	**E**	**T**	100	0.8	0.74
Q	**E**	**T**	100	1.2	0.79
V	**E**	**I**	100	1.4	0.64
L	**E**	**T**	100	1.6	0.84
G	**E**	**T**	100	1.6	0.87
I	**E**	**T**[c]	100	2.0	0.79
V	**E**	**V**	100	2.0	0.68
N	**E**	**I**	100	2.1	0.60

Among the 200 final sequences in each of the 15 PEP runs, 55, 33, and 80 sequences were found in 15, 14, and 13 runs, respectively. The P_4–P_2 optimal recognition motif of caspase 8 is LET.[9]
[a] See caption of Table I.
[b] Peptide shown in Figure 4F.
[c] The aldehyde inhibitor Boc-IETD-CHO has a K_i of 1 nM for caspase 8.[29, 30]

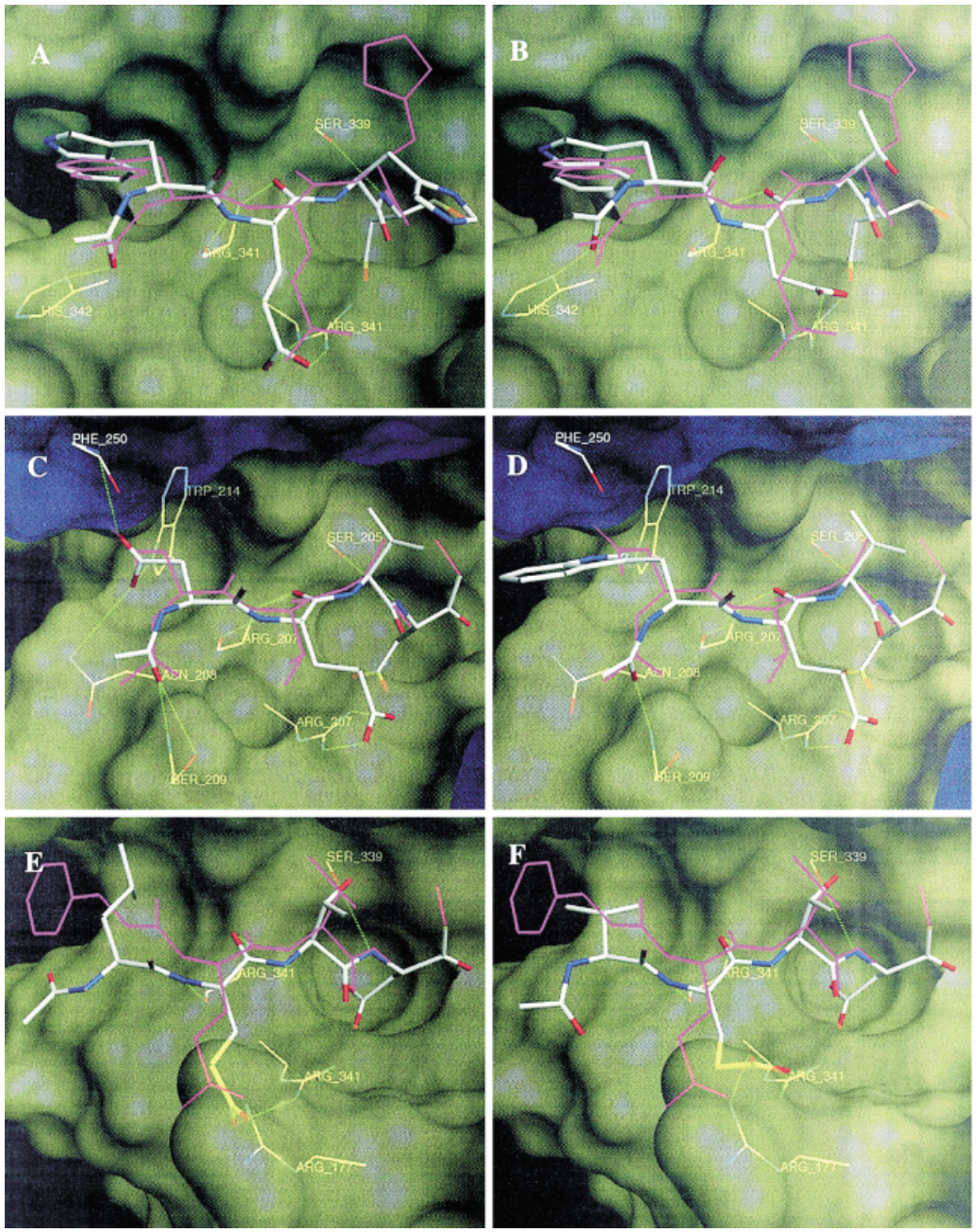

FIGURE 4. (A) Ac-WEHD tetrapeptide docked by PEP in the active site of caspase 1 (thick cylinders; atoms of carbon, oxygen, and nitrogen are colored in white, red, and blue, respectively). The substrate X-ray structure (Ac-WEHD-CHO) is shown in magenta, and hydrogen bonds are indicated by green dotted lines. (B) same as (A), with the Ac-WDTD tetrapeptide designed by PEP. (C) Ac-DEVD tetrapeptide docked by PEP in the active site of caspase 3. The substrate X-ray structure (Ac-DVAD-fmk) is shown in magenta. (D) Same as (C), with the Ac-WEVD sequence designed by PEP. (E) Ac-LETD tetrapeptide docked by PEP in the active site of caspase 8. The substrate X-ray structure (Z-EVD-dcbmk) is shown in magenta. (F) Same as (E), with the Ac-VETD sequence designed by PEP. (A–F) The solvent accessible surface of the enzyme is shown in yellow, and caspase residues involved in hydrogen bonds are shown in thin lines colored by atom type.

most favorable energy are listed in Table I. They contain Thr and Leu at P_2, Asp at P_3, and mainly residues with aromatic side chains at P_4, with a clear preference for Trp. It has been shown experimentally that Thr at P_2 belongs to the recognition motif, whereas peptides with Leu(P_2) show relatively weak substrate activity.[9] No experimental results for caspase 1 inhibitors with Leu at P_2 are available. Visual inspection reveals that the S_2 pocket is less pronounced compared to its counterpart in caspases 3 and 8. The P_2 side chain is, therefore, more exposed to solvent, which penalizes large hydrophobic residues. The PEP solvation model used for ligand ranking does not include a penalty for exposed hydrophobic residues unless they desolvate polar regions of the protein, and this might explain the occurrence of Leu at P_2 in the PEP sequences. Asp(P_3) is found in active sequences, although Glu is slightly better at this position. Aromatic residues are clearly favored at P_4 in optimal tetrapeptide recognition motifs.[9] The PEP sequences 1 (WDT), and 6 (FDT) correspond to optimal peptide motifs for caspase 1.

CASPASE 3

Docking of the Optimal Peptide Recognition Motif (DEVD)

Nineteen of the 20 conformations of Ac-DEVD generated by PEP have a backbone binding mode that overlaps the Ac-DVAD inhibitor X-ray structure (Fig. 4C). The 19 backbone conformations of Ac-DEVD have an average RMSD from the X-ray structure of 0.48 Å. The five inhibitor backbone hydrogen bonds observed in the X-ray structure are reproduced: one between the NH of Asp(P_1) and the CO of Ser205, two between the backbone polar groups of Glu(P_3) and Arg207, and two between the oxygen of the acetyl group at the N terminus and Ser209. The hydrophobic pocket S_2 is filled by the Val(P_2) side chain, and the Glu(P_3) makes a salt bridge with Arg207. Although the orientation of the Glu(P_3) side chain cannot be confirmed by the inhibitor X-ray structure, which contains a Val at P_3, it is most probably correct because a similar salt bridge is observed for Glu(P_3) in both the caspase 1 and caspase 8 X-ray structures. Finally, Asp(P_4) reproduces the binding mode of its counterpart in the X-ray structure, and accepts hydrogen bonds from the side chains of Asn208 and Trp214, and the backbone NH of Phe250. It is important to note that PEP correctly docks the tetrapeptide Ac-DEVD in a structure of caspase 3 from a complex with a different inhibitor (Ac-DVAD).

Design of Tetrapeptide Inhibitors

The 10 P_4–P_2 sequences designed by PEP with the highest occurrence and most favorable energy are listed in Table II, and the best is shown in Figure 4D. They contain Val and His at P_2, Glu at P_3, and aromatic side chains at P_4. This is consistent with the experimentally determined recognition motif at P_2 and P_3. On the other hand, substrate specificity data indicate that caspase 3 is highly specific for Asp at P_4.[9] This discrepancy is probably due to the approximations inherent to the continuum electrostatics model used for sequence ranking. The difference in electrostatic energy between Asp(P_4) and Trp(P_4) is only -0.4 kcal/mol ($\Delta\Delta G^{\text{interm}}_{\text{elect, sol}} = -15.2$ kcal/mol, $\Delta\Delta G^{\text{ligand}}_{\text{elect, desolv}} = 17.8$ kcal/mol, $\Delta\Delta G^{\text{receptor}}_{\text{elect, desolv}} = -3$ kcal/mol), although Asp at P_4 is hydrogen bonded to Asn208, Trp214, and Phe250. This does not balance the van der Waals interaction, which is less favorable ($\Delta\Delta G^{\text{interm}}_{\text{vdW}} = 4$ kcal/mol) for Asp(P_4) than Trp(P_4). It is important to note that the Asp(P_1) position used as seed for the growing procedure is taken from the X-ray structure of a tetrapeptide that shows only low substrate activity. Nevertheless, starting from this nonideal seed, PEP ranks the optimal recognition motif (DEVD) as number 14, and proposes a binding mode that is in agreement with the available X-ray structure.[23, 25]

CASPASE 8

Docking of the Optimal Peptide Recognition Motif (LETD)

The 20 conformations of Ac-LETD generated by PEP have the correct backbone orientation (Fig. 4E). The average value of the backbone RMSD from the X-ray structure is 0.89 Å. As observed in the X-ray structure, the tetrapeptide Ac-LETD docked by PEP forms an antiparallel β-sheet interaction that involves Thr (P_2) and Glu(P_3), and the backbone polar groups of residues Ser339 and Arg341, respectively. PEP places the Thr(P_2) side chain in the S_2 pocket, which is occupied by the Val side chain in the Z-EVD inhibitor in the X-ray structure.[25] The Glu(P_3) side chain reproduces the double salt bridge (with Arg341 and Arg177) observed in the X-ray structure, and the Leu(P_4) side chain fills the rather hydrophobic S_4 pocket. As for caspase 3, it is important to underline that PEP finds the correct binding mode of the recognition motif Ac-LETD using a conformation of caspase 8 from a complex with Z-EVD, a rather different inhibitor.

Design of Tetrapeptide Inhibitors

The 10 designed P_4–P_2 sequences with highest occurrence and best energy are listed in Table III. They contain Thr, Ile, and Val at P_2, Glu at P_3, and different residue types at P_4. These results are in agreement with the known optimal recognition motifs of caspase 8, which have Thr, Leu, and Ile at P_2, strict specificity for Glu at P_3, and a rather promiscuous S_4 pocket.[9, 29] It is striking that the optimal recognition motif is reproduced by 2 of the 10 designed sequences, namely Ac-VETD (Fig. 4F) and Ac-LETD, which rank number 2 and 6, respectively. Moreover, apart from the blocking group the PEP sequence 8 corresponds to the known aldehyde inhibitor Boc-IETD-CHO, which has a K_i of 1 nM for caspase 8.[29, 30]

Conclusions

The build-up approach implemented in PEP allows searching in ligand conformational space (docking) and chemical space (design). Docking is performed by a genetic algorithm that is able to find ligand binding modes that are very similar to their counterparts observed in X-ray structures. The search in chemical space is exhaustive within a library of amino acids specified by the user. The ability of PEP to correctly dock (within 0.9 Å backbone RMSD) the peptide recognition motifs was shown on three members of the caspase family of enzymes. The design of P_4–P_2 sequences yielded known caspase inhibitors among the top 15 PEP hits.

There are three advantages of PEP with respect to previous build-up procedures.[11, 32, 33] First, PEP uses a library of residue topologies without any predefined structural information. Hence, the conformational optimization in PEP is not restricted to the discrete space of a library of low energy conformations. A predefined library of conformers might not contain the correct conformation of monomers with many internal degrees of freedom.[11] Second, sequences are ranked by PEP according to both their binding energy and their ability to be further grown. The latter is checked by a procedure that consists in elongating the current sequence with an alanine, and discarding it if the minimized alanine binding energy is worse than a given cutoff. This is important, because a partially grown sequence, especially if optimally buried, may not allow further growing. Third, partial rigid body rotation of the last added monomer is used in PEP at each growing cycle to increase the accessible space. This is useful, because the growing direction defined by a suboptimal orientation of monomer $n-1$ may prevent monomer n from optimally interacting with the protein binding site.

Although PEP allows an exhaustive, though discrete, search in chemical space at each growing step, further growing is restricted to a relatively small number of sequences to avoid combinatorial explosion. The correct ranking of the sequences is, therefore, of utmost importance. In PEP sequences are ranked according to a binding energy that includes solvent effects in a continuum approximation.[16] Test runs performed on caspases and other enzymes using the *in vacuo* energy without solvation yielded a large and unrealistic number of charged residues in the top ranking PEP hits (Budin and Caflisch, unpublished results). Hence, the solvation energy used in PEP is useful for the ranking of different peptide sequences. Although the results are much better than using an *in vacuo* energy function, the approximations used in the implicit solvation model (Coulomb field approximation, generalized Born formula) yield an error that can be rather large for functional groups with formal charges. This results in the relatively poor ranking of Asp(P_4) compared to Trp(P_4) in the S_4 pocket of caspase 3. Another source of error in the PEP energy is the crude approximation of the energy of the free ligand. The contribution of the free receptor cancels out. On the other hand, taking into account the ligand reference state, which is important for design but not for docking, is currently beyond the limits of routine calculations, because it includes finding the most probable conformations in solvent and averaging with the correct Boltzmann weights. Structure-based ligand design is essentially the extension of the docking problem into chemical space. In docking, part of the systematic error due to the approximations inherent in the force field cancels out. This is not so when evaluating and comparing the binding affinity of different molecules (design). When the search space is very small, for example, docking a mainly rigid fragment into a rigid protein, a simple empirical scoring function might suffice. This is not the case for docking into a flexible protein or for ligand design. In general, the larger the space to be sampled, conformational and eventually also chemical space, the more accurate has to be the energy function even for a qualitative ranking of candidate ligands.[3]

The main limitation of PEP is the necessity of an anchor fragment correctly placed in the target binding site. The seed is usually obtained from the X-ray structure of a known inhibitor complexed with the

target. It can be used to design novel sequences as demonstrated on caspase 3 where PEP is able to predict sequences of potent inhibitors starting from the Asp(P_1) binding mode of a weak inhibitor. To reduce the influence of the seed position, we are currently investigating a new procedure where partial rigid body motion and internal flexibility of the seed is allowed during the conformational optimization of the first added monomer (first growing cycle). Preliminary results on the insulin receptor tyrosine kinase show that PEP finds the correct binding mode of the ligand when rigid body translations up to 0.5 Å and rotations up to 20 degrees are allowed to the seed. Although this procedure does not completely solve the issue of the seed, it greatly reduces the influence of its initial placement on the success of the growing procedure.

The protein rigidity is another limitation in the current version of PEP. There is a large amount of experimental evidence, mainly crystal structures, which show that the uncomplexed and ligand bound conformation of the binding site might differ substantially.[34, 35] Typical rearrangements due to ligand binding include rotation and displacement of side chains[36] but also the relative motion of entire domains constituting the binding site (e.g., p38 MAP kinase[37]). It is clear that some protein side chain flexibility should be allowed during conformational optimization of the ligand, at least for the hydroxyl groups of the serine, threonine, and tyrosine side chains, because their conformation is usually not specified by the X-ray structure and must, therefore, be assigned when hydrogens are added. Further versions of PEP will try to address this issue.

In principle, PEP can be expanded to grow any type of linear compounds consisting of monomers provided that their chemical structure (atom types, partial charges, covalent bonds, and rotatable bonds) can be defined in a topology library. We are currently trying to extend PEP into a more general combinatorial ligand design program.

Acknowledgments

We thank Dr. C. Ehrhardt and S. Ahmed for useful discussions. We also thank A. Widmer for providing the molecular modelling program WITNOTP (unpublished) which was used to make Figure 4. The program PEP (for SGI or PC running the Linux operating system), as well as the library of amino acids, are available for not-for-profit institutions from the last author.

References

1. Ajay, A.; Walters, W. P.; Murcko, M. A. J Med Chem 1998, 41, 3314.
2. Zou, X.; Sun, Y.; Kuntz, I. D. J Am Chem Soc 1999, 121, 8033.
3. Apostolakis, J.; Caflisch, A. Comb Chem High Throughput Screen 1999, 2, 91.
4. Hattori, M. Nature 2000, 405, 311.
5. Dunham, I. Nature 1999, 402, 489.
6. Adams, M. D. Science 2000, 287, 2185.
7. Sánchez, R.; Šali, A. Proc Natl Acad Sci USA 1998, 95, 13597.
8. Wolf, B. B.; Green, D. R. J Biol Chem 1999, 274, 20049.
9. Thornberry, N. A.; et al. J Biol Chem 1997, 272, 17907.
10. Stennicke, H. S.; Renatus, M.; Meldal, M.; Salvesen, G. S. Biochem J 2000, 350, 563.
11. Moon, J. B.; Howe, W. J. Proteins Struct Funct Genet 1991, 11, 314.
12. Davis, L. Handbook of Genetic Algorithms; Van Nostrand Reinhold: New York, 1991.
13. Goldberg, D. E. Genetic Algorithms in Search Optimization and Machine Learning; Addison-Wesley: Reading, MA, 1989.
14. Jones, G.; Willett, P.; Glen, R. C. J Mol Biol 1995, 245, 43.
15. Majeux, N.; Scarsi, M.; Caflisch, A. Proteins Struct Funct Genet 2001, 42, 256.
16. Scarsi, M.; Apostolakis, J.; Caflisch, A. J Phys Chem A 1997, 101, 8098.
17. Scarsi, M.; Apostolakis, J.; Caflisch, A. J Phys Chem B 1998, 102, 3637.
18. Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Proteins Struct Funct Genet 1999, 37, 88.
19. Scarsi, M.; Majeux, N.; Caflisch, A. Proteins Struct Funct Genet 1999, 37, 565.
20. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J Am Chem Soc 1990, 112, 6127.
21. Caflisch, A. J Comput-Aided Mol Design 1996, 10, 372.
22. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Nucleic Acids Res 2000, 28, 235.
23. Rano, T. A.; Timkey, T.; Peterson, E. P.; Rotonda, J.; Nicholson, D. W.; Becker, J. W.; Chapman, K. T.; Thornberry, N. A. Chem Biol 1997, 4, 149.
24. Mittl, P. R. E.; Di Marco, S.; Krebs, J. F.; Bai, X.; Karanewsky, D. S.; Priestle, J. P.; Tomaselli, K. J.; Grütter, M. G. J Biol Chem 1997, 272, 6539.
25. Blanchard, H.; Kodandapani, L.; Mittl, P. R. E.; Di Marco, S.; Krebs, J. F.; Wu, J. C.; Tomaselli, K. Grütter, M. G. Structure 1999, 7, 1125.
26. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
27. Thornberry, N. A.; et al. Nature 1992, 356, 768.
28. Nicholson, D. W.; et al. Nature 1995, 376, 37.
29. Garcia-Calvo, M.; Peterson, E. P.; Leiting, B.; Ruel, R.; Nicholson, D. W.; Thornberry, N. J Biol Chem 1998, 273, 32608.

30. Blanchard, H.; Donepudi, M.; Tschopp, M.; Kodandapani, L.; Wu, J. C. G.; Grütter, M. J Mol Biol 2000, 302, 9.
31. Wilson, K. P.; Black, J. F.; Thomson, J. A.; Kim, E. E.; Griffith, J. P.; Navia, M. A.; Murcko, M. A.; Chambers, S. P. A.; Aldape, R.; Raybuck, S. A. J.; Livingston, D. Nature 1994, 370, 270.
32. Böhm, H. J. J Comput-Aided Mol Design 1996, 10, 265.
33. Frenkel, D.; Clark, D. E.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. J Comput-Aided Mol Design 1995, 9, 213.
34. Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Science 1989, 246, 1149.
35. Rini, J. M.; Schulze-Gahmen, U.; Wilson, I. A. Science 1992, 255, 959.
36. Arevalo, J. H.; Stura, E. A.; Taussig, M. J.; Wilson, I. A. J Mol Biol 1993, 231, 103.
37. Wang, Z.; Canagarajah, B. J.; Boehm, J. C.; Kassisa, S.; Cobb, M. H.; Young, P. R.; Abdel-Meguid, S.; Adams, J. L.; Goldsmith, E. J. Structure 1998, 6, 1117.