An Evolutionary Approach for Structure-based Design of Natural and Non-natural Peptidic Ligands

Nicolas Budin, Shaheen Ahmed, Nicolas Majeux and Amedeo Caflisch*

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

Abstract: A new computational approach (PEP) is presented for the structure-based design of linear polymeric ligands consisting of any type of amino acid. Ligands are grown from a seed by iteratively adding amino acids to the growing construct. The search in chemical space is performed by a build-up approach which employs all of the residues 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 binding energy with electrostatic solvation is evaluated to select sequences for further growing. PEP is tested on the peptide substrate binding site of the insulin receptor tyrosine kinase and farnesyltransferase. In both test cases, the peptides designed by PEP correspond to the sequence motifs of known substrates. For tyrosine kinase, tyrosine residues are suggested at position P and P+2. While the tyrosine at P is in agreement with the experimental data, the one at P+2 is a prediction which awaits experimental validation. For farnesyltransferase, it is shown that electrostatic solvation is necessary for the correct design of peptidic inhibitors.

1 INTRODUCTION

The knowledge of gene product sequences generated by the genome projects [1, 2, 3] and the significant advances in experimental protein structure determination and high-throughput homology modelling [4] are providing a large amount of targets for structure-based drug design. Computational approaches that exploit the knowledge of the three-dimensional structure of a protein target have been developed and are used for de novo design [5, 6], improvements of lead compounds, and to help in the selection of monomers to focus combinatorial libraries [7]. Prioritization is done by empirical and knowledge-based scoring functions or force field energy functions [8]. Ligands are built by connecting small molecular fragments or functional groups, often rigid, or even atoms. Atom-based methods usually generate compounds that span a large amount of chemical space [9, 10, 11]. The main disadvantage of compounds generated by atom-based approaches is that they often have complicated structures and are in most cases very difficult to synthesize. Hence, methods that build new compounds by combining predefined fragments are more popular. The number of newly created bonds is small and therefore it is less difficult to control the chemistry, i.e., the synthesizability and the chemical stability of the designed molecules.

Fragment-based ligand design may be achieved in two ways. In the first one, small fragments are docked in the active site [12, 13]. The best positions of each fragment type are retained and connected to generate candidate ligands [14]. Alternatively, a seed fragment is docked in the binding site and the ligand is grown starting from the seed [15]. Both approaches should not be considered as mutually exclusive, but rather as complementary since they are useful to generate candidate ligands with different physico- chemical characteristics and structural properties.

The methods based on the connection of docked fragments have the advantage that the functional groups occupy optimal positions and are oriented such that their interaction with the protein is favorable. On the other hand, the geometry of the bonds connecting the fragments to each other or to a central template is not optimal and has to be accepted initially with a certain tolerance. The mapping of a binding site and fragment assembly into complete ligands can be performed by separated programs [13, 14] or integrated in a single computational tool [16].

The approaches based on the progressive build-up of ligands (called also growing procedures) usually start with a seed fragment placed in an appropriate region of the binding site. New ligands are then grown by sequentially appending building blocks (fragments or atoms). To avoid combinatorial explosion, a large fraction of all building blocks is discarded at every step according to some heuristic scoring. This method has the advantage that the newly formed chemical bonds have a correct geometry and that the intraligand interactions can be taken into account during the design. On the other hand, build-up approaches have difficulties to generate ligands that bind to different pockets if these are separated by gap regions that do not allow specific interactions. Moreover, the success of the growing procedure and therefore the quality of the designed molecules depends dramatically on the position of the seed, since the latter is usually kept fixed. The seed position(s) can be determined from X-ray or NMR structures of ligand-protein complexes. If no structure is available, seeds must be obtained by manual or computer-aided docking. Several programs that implement an automatic build-up strategy have been

*Address correspondence to this author at the Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; Phone: (41 1) 635 55 21; FAX: (41 1) 635 57 12; email: caflisch@bioc.unizh.ch

1386-2073/01 $28.00+.00

described in the literature. The approach was pioneered by Moon and Howe with the program GROW [15], and later used in PRO_LIGAND [17]. In these programs, the library of fragments is however restricted to amino acids and amino acid derivatives. This has the disadvantage that the explored chemical space is relatively small but the designed ligands are synthetically accessible. An important advantage is that the energetics of peptidic ligands can be studied by well parameterized force fields. The conformational flexibility is taken into account by using multiple conformers for each amino acid. The main differences between these programs lie in the scoring functions used to rank the ligands, and in the way the conformation libraries for the amino acids are generated. The scoring function in GROW is based on the AMBER force field [18] supplemented by a solvent accessible surface approximation of solvation [19]. PRO_LIGAND and LUDI [20] use empirical scoring functions combined with a rule based interaction site approach [21, 22]. The GROW and PRO_LIGAND libraries contain low energy conformations whereas LUDI [20] uses amino acid conformations extracted from high-resolution protein structures.

In this paper, we present a new growing procedure (PEP) for docking and design of peptidic ligands consisting of natural and/or non-natural amino acids. PEP uses a genetic algorithm (GA) for conformational optimization of the ligand in a rigid protein. In ligand design the search space is huge because the optimization is performed simultaneously in two different spaces, namely conformational and sequence space. Growing programs usually restrict the search in conformational space to a relatively small number of minima compared to the overall conformational space [15, 20, 17], and therefore may fail to find the correct conformation, especially for amino acids that contain many rotatable bonds [15]. Moreover, since small deviations propagate in the growing procedure, it is necessary to find amino acid conformations which fit nicely to the binding site pocket. This might require a deviation from the minimum conformation of the isolated residue. The GA based conformational optimization used in PEP does not suffer from the limitations due to a finite number of residue conformations. It is unrestricted in conformational space and is able in principle to find the most favorable bound conformation. Although the growing method allows unrestricted chemical space search at each step, further growing must be restricted to a relatively small number of sequences to avoid combinatorial explosion. It is therefore very important to be able to rank correctly the sequences in order to restrict the search to the most favorable ones. The correct ranking of different chemical entities having multiple internal degrees of freedom is not a trivial task. PEP uses an accurate implicit solvation model [23] to effectively rank the designed peptides according to their binding energy in solution. For a large number of protein-small ligand complexes, it was shown that the energies in solution calculated with this implicit solvation model correlate well with the values obtained by finite difference Poisson calculations [13], and in this study it is shown that the model is also appropriate for the ranking of different peptide sequences. The main disadvantage of every growing method is inherent to its sequential approach. The current growing step has no knowledge of the step(s) that will follow and the success of any growing step depends largely on the previous step(s). The orientation selected for the last added amino acid may not correspond to the orientation of the same residue when it is part of a longer sequence, and might therefore not allow further correct growing. In PEP, this problem is partially solved since sequences are ranked according to their ability to allow an additional step of growing besides favorable binding energy in solution.

2 METHODS

The ligand design approach implemented in the program PEP uses a build-up strategy to search for optimal amino acid sequences as well as favorable positions and conformations of the corresponding peptides in the binding site of an enzyme or receptor. At every growing step and for each member of the amino acid library, a genetic algorithm (GA) is used for conformational optimization of the added residue. The implementation of PEP is based on the Molecular Design Classes (MDC), a set of in house modules (written in C++) developed to act as basic layer for structure-based drug design softwares (Fig. (**1**)). The MDC are presented first. Then the growing procedure is described. Finally, the GA and the estimation of the energy are explained.

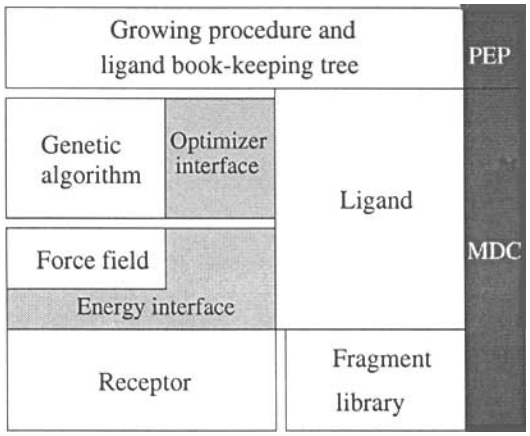

Fig. (1). Layer representation of PEP on top of the MDC. Modules are represented by rectangles. Common rectangle edges symbolize module intercommunication. Laye**61**ark and light gray, respectively.

2.1 Molecular Design Classes (MDC)

The MDC are an implementation of the basic code needed to combine a set of compounds to generate candidate ligands in a receptor active site (Fig. (**1**)). The MDC define two types of molecular descriptors. The first one is used to store both the receptor and the library of molecular fragments that serve as building blocks to create new ligands. The molecule coordinates can be read from files in the pdb or mol2 formats while the parameters are read from a file in the CHARMM parameter format [24]. The second type of molecular descriptor contains the ligand information. Ligands are generated by making chemical bonds between

fragments. The ligand descriptor stores a population of positions and conformations which have rigid body (translation, rotation) and internal (rotatable bonds) degrees of freedom. The position and orientation of the ligands are minimized using a minimizer and an energy function. The MDC define interfaces for the energy function and the minimizer with the advantage that any implementation that respects the appropriate interface can be added directly to the existing code. So far, the minimizer and energy interfaces are implemented by a GA and a force field function, respectively. The different energy contributions of the force field can be selected by the programmer in a modular way. In PEP, the MDC are used to perform an exhaustive search in the chemical space defined by the library of amino acids. The MDC can also be used for structure-based design of combinatorial libraries (Tenette-Souaille et al., in preparation).

2.2 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 which interact most favorably with the functional groups of the receptor binding site. The default number of sequences kept at each growing step is ten. Within the approximation that chemical entity and orientation of a monomer are not affected by the successive monomers, the search is exhaustive, because at every step of growing every amino acid in the library is attached to the actual construct (Fig. (**2**)). Furthermore, the 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. It is computationally prohibitive to compute protein and ligand desolvations during the GA optimization. After all the amino acids have been minimized, the binding energy in solution is calculated 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 growing from 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 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 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 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.

2.3 Template Library and Bond Formation

PEP uses amino acid templates in which the amide can be either primary or secondary (Fig. (**2A**)). This includes L- and D-residues, as well as non-standard 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 [15]. The side chain and backbone rotatable bonds of the last added residue are flexible during the conformational optimization. Moreover, the backbone rotatable bond of the previous residue, which is the closest

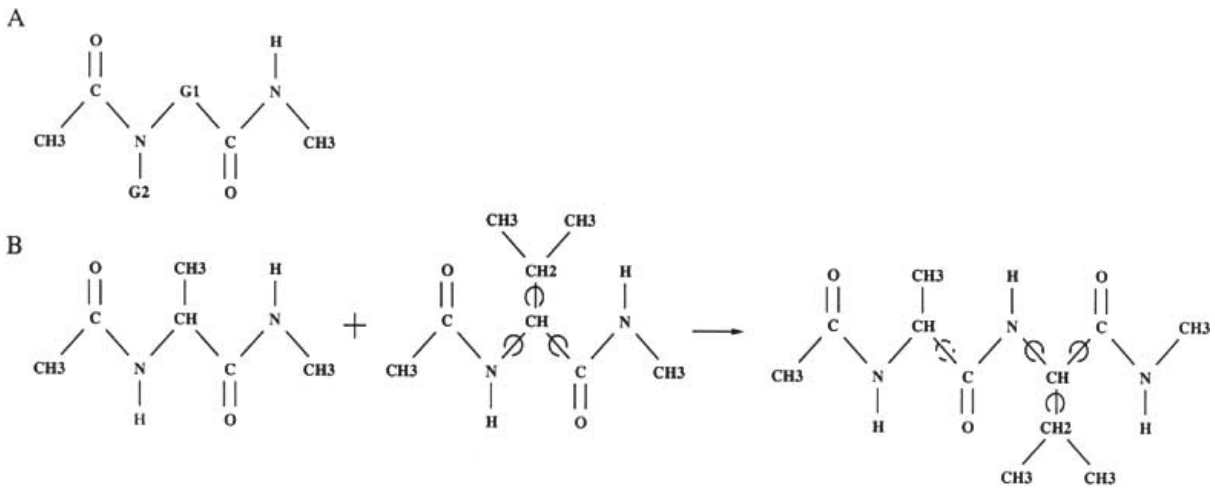

Fig. (2). (A) Amino acid template used by PEP. G2 indicates the substituent position on the template amino group. G1 can be of any type, without size limitation (e.g. -CH2-, standard amino acid side chain, ring etc). (B) Illustration of the flexibility in PEP during the growing. Rotatable bonds are marked with circles. A fully flexible valine is bound to an alanine. Alanine can be either the seed, or a residue positioned during the previous growing cycle. In addition to the valine internal flexibility, the alanine ψ dihedral (dashed circle) is also flexible during the valine conformational optimization.

to the currently minimized amino acid, is also flexible (Fig. (**2B**)). This increases the amount of explored conformational space and prevents the growing direction from being restricted to the optimal orientation of the terminal N-methyl amide group at the previous growing step. This dihedral corresponds to ψ and φ for α-amino acids, when growing in the N to C and C to N direction, respectively.

2.4 Genetic Algorithm

A GA is a stochastic optimization method that mimics the process of natural evolution by manipulating a population of data structures called chromosomes [25, 26]. 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 leads usually 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 last added amino acid and, as mentioned above, the ψ or φ dihedral of the residue closest to the last. A chromosome of N genes therefore encodes the conformation of a molecule with N rotatable bonds. The genes are binary encoded in a string of one byte which gives an integer value between 0 and 255. This integer value is linearly rescaled to a real number between 0 and 2π, which is used as a dihedral angle value for the appropriate rotatable bond. This leads to a theoretical resolution of 1.4 degrees. 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. The details of the GA and the operators will be given elsewhere (Budin et al., in preparation). For each GA conformational optimization, a population of 100 chromosomes was used and 1000 cycles were performed. At each GA cycle, 100 new chromosomes were generated for a total of 10^5 conformations tested during the overall GA optimization. Using the 20 standard amino acids, the growing of a tetrapeptide requires 620 GA runs and 6.2 x 10^7 energy evaluations ([20 + (3 x 200)] x 10^5 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 6.2 x 10^7 energy evaluations for the design of tetrapeptides, 620 include full electrostatic solvation while the remaining ones use the distance dependent dielectric function (see below).

2.5 Energy

The energy terms are implemented in the MDC by modules, each of which calculates a specific energy contribution. The appropriate modules are then combined by the programmer to compute a given energy. In PEP, the modules are combined in three different ways which allow to calculate the binding energy in vacuo (used to check for dead-ends), the total energy in vacuo (used as scoring function for the GA conformational optimization), and the binding energy in solution (used to rank the different sequences after every GA optimization). The different energy contributions are presented first. Then, the three energies used in PEP are explained.

2.5.1 Energy Terms

The energy contributions can be divided in the following categories. First, the terms which represent the intraligand energy (2.5.2). These terms are calculated explicitly for each appropriate set of ligand atoms. Second, the intermolecular energy terms which approximate the interaction energy between the ligand and the protein (2.5.3). Since the receptor is rigid, its van der Waals (vdW) and Coulombic potentials are mapped on look- up tables to improve the efficiency. In these two categories the direct solvation effects, i.e., protein and ligand desolvations, are neglected. The terms that deal with solvation are grouped in the third category (2.5.4). They use a continuum dielectric approximation for the computation of the receptor and ligand desolvations, and the screened ligand- receptor interaction in solution. All of the energy parameters used in the applications presented here are taken from CHARMm22 (MSI Inc.) but any other force field with explicit dihedral, Coulombic, and van der Waals terms could be used.

2.5.2 Intraligand Energy

The internal energy contributions consist of the electrostatic $\left(E_{\mathrm{elect}}^{\mathrm{ligand}}\right)$, van der Waals $\left(E_{\mathrm{vdw}}^{\mathrm{ligand}}\right)$, and the strain energy $\left(E_{\mathrm{strain}}^{\mathrm{ligand}}\right)$ of the ligand. The bond lengths and angles are kept constant and have therefore no energy contributions. $E_{\mathrm{vdw}}^{\mathrm{ligand}}$ and $E_{\mathrm{elect}}^{\mathrm{ligand}}$ are sums over the vdW and electrostatic contributions, respectively, calculated explicitly for each pair of ligand atoms *ij*, separated by at least three bonds. A scaling factor of 0.5 is applied to the 1-4 electrostatic interactions (atoms pairs separated by three bonds). The vdW intraligand energy is described as the sum of a steep repulsion and an attractive dispersion term with the 12-6 Lennard-Jones model:

$$E_{vdw}^{ligand} = \sum_{\substack{i,j \\ nonbonding}} \sqrt{\varepsilon_i \varepsilon_j}\left[\left(\frac{R_i^{\mathrm{vdw}} + R_j^{\mathrm{vdw}}}{r_{ij}}\right)^{12} - 2\left(\frac{R_i^{\mathrm{vdw}} + R_j^{\mathrm{vdw}}}{r_{ij}}\right)^{6}\right] \quad (1)$$

where R_i^{vdw} is the van der Waals radius of atom *i* and ε_i is the minimum of the vdW potential between two atoms of type *i* at optimal distance of $2R_i^{\mathrm{vdw}}$. The intraligand electrostatic energy is given by

$$E_{\mathrm{elect}}^{\mathrm{ligand}} = 332.0 \sum_{\substack{i,j \\ nonbonding}} \frac{q_i q_j}{\varepsilon r_{ij}^n} \quad (2)$$

where q_i and q_j are partial charges (in electronic unit), and r_{ij} is the distance in Å between two atoms *i* and *j*. For $n = 1$ and $n = 2$ equation (2) corresponds to the Coulomb law and the distance dependent dielectric model $\left(E_{\mathrm{elect,rdiel}}^{\mathrm{ligand}}\right)$, respectively. There is no sound physical justification in favor of the linear distance-dependent dielectric function even if its agreement with more sophisticated models is remarkable [23, 27]. Nevertheless, it is a simple and useful approximation, since it yields a shorter range interaction than the Coulomb law. Recently, distance-dependent

dielectric models have been used for docking and ligand design [28, 29, 30], as well as in molecular dynamics simulations of peptides [31], protein folding [32] and unfolding [33]. The strain energy is a four-atom term based on the dihedral angle about the axis defined by the middle pair of atoms. $E_{\text{strain}}^{\text{ligand}}$ is the sum of the contribution of all rotatable dihedrals X

$$E_{\text{strain}}^{\text{ligand}} = \sum_{\theta} k_\theta[1 + \cos(n\theta - \delta)] \quad (3)$$

where θ is the dihedral angle, k_θ is the force constant, n is the periodicity, and δ the angle value corresponding to the maximal strain.

2.5.3 Interaction Energy

The electrostatic receptor-ligand interaction energy is calculated by the factorization of the electrostatic potential of the receptor which is kept rigid:

$$E_{elect}^{int} = \sum_{i \varepsilon ligand} q_i \Phi_i = \sum_{i \varepsilon ligand} q_i \left(332.0 \sum_{j \varepsilon receptor} \frac{q_j}{\varepsilon r_{ij}^n} \right) \quad (4)$$

The Coulombic potential of the receptor is computed once over a grid containing the binding site plus a boundary, and stored in a look-up table [13]. The contribution of each ligand atom is computed by multiplying its partial charge q_i with the potential of the receptor interpolated from the surrounding eight points of the grid by the trilinear interpolation method [34].

The vdW interaction between a ligand and the receptor is described with the 12-6 Lennard-Jones model:

$$E_{vdw}^{int} = \sum_{i \varepsilon ligand} \sum_{j \varepsilon receptor} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \quad (5)$$

where r_{ij} is the distance between atoms i and j, A_{ij} and B_{ij} are van der Waals repulsion and attraction parameters. The geometric mean approximation [35, 36, 37, 28] is used to make the ligand and receptor terms factorizable in equation (5): $A_{ij} = \sqrt[3/4]{A_i} \sqrt[3/4]{A_j}$ and $B_{ij} = \sqrt[3/4]{B_i} \sqrt[3/4]{B_j}$, with $A_i = \varepsilon_i (2R_i^{\text{vdw}})^{12}$ and $B_i = 2\varepsilon_i (2R_i^{\text{vdw}})^6$. R_i^{vdw} is the van der Waals radius of atom i and ε_i is the minimum of the van der Waals potential between two atoms of type i at optimal distance of $2R_i^{\text{vdw}}$. When the program starts, for every grid point p the two following "receptor potentials" are calculated and stored in look-up tables which span over the binding site plus a boundary:

$$\varphi^A(p) = \sum_{j \varepsilon receptor} \frac{\sqrt[3/4]{A_j}}{r_{pj}^{12}} \quad \text{and} \quad \varphi^B(p) = \sum_{j \varepsilon receptor} \frac{\sqrt[3/4]{B_j}}{r_{pj}^{6}} \quad (6)$$

where the sums run over the receptor atoms which are within a 10 Å cutoff distance of the grid point. The contribution of ligand atom i is evaluated by multiplying its van der Waals parameters ($\sqrt[3/4]{A_i}$ and $\sqrt[3/4]{B_i}$) with the "receptor potentials" (φ^A and φ^B, respectively) interpolated from the surrounding eight points of the grid by the trilinear interpolation method.

2.5.4 Continuum Electrostatic Energy in Solution

The electrostatic energy in solution of a ligand-receptor complex is evaluated within the continuum electrostatic approximation [23, 27, 38, 39, 40, 41, 42, 43, 44, 45, 46]. The system is partitioned into solvent and solute regions and different dielectric constants are assigned to each region. In this approximation only the intra-solute electrostatic interactions need to be evaluated. This strongly reduces the number of interactions with respect to an explicit treatment of the solvent. Moreover it makes feasible the inclusion of solvent effects in structure-based ligand design where the equilibration of explicit water molecules would be a major difficulty. The electrostatic effects of the solvent are relevant and it has been shown that the continuum dielectric model provides an accurate description of molecules in solution [27, 47]. The difference in electrostatic energy in solution upon binding of a ligand to a receptor can be calculated as the sum of the desolvation of the receptor, screened intermolecular interaction, and desolvation of the ligand [14, 39]. 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 according to

$$\Delta E_{elect,desolv}^{receptor} = \frac{\tau}{8\pi} \sum_{k \varepsilon V_{ligand}} \vec{D}^2(\vec{x}_k) \Delta V_k \quad (7)$$

where $\tau = \frac{1}{\varepsilon_p} - \frac{1}{\varepsilon_w}$ (ε_p and ε_w are the interior and solvent dielectric constants, respectively), V_{ligand} is the volume occupied by the ligand as defined by its molecular surface, and $\vec{D}(\vec{x})$ is the receptor electric displacement. The electric displacement of every partial charge of the receptor is approximated by the Coulomb field and is evaluated over a 3D grid [13].

The screened ligand-receptor interaction is the intermolecular electrostatic energy in solution ($E_{\text{elect,sol}}^{\text{int}}$). It is calculated, via the GB approximation [23, 27, 42], as the sum of the interaction energies between each ligand atom i and its corresponding list of receptor atoms j [13]:

$$E_{elect,sol}^{int} = \sum_{\substack{i \varepsilon ligand \\ j \varepsilon list_i}} \left(\frac{q_i q_j}{\varepsilon_p r_{ij}} - \frac{q_i q_j \tau}{R_{ij}^{GB}} \right) \quad (8)$$

where

$$R_{ij}^{GB} = \sqrt{r_{ij}^2 + R_i^{eff} R_j^{eff} \exp\left(\frac{-r_{ij}^2}{4 R_i^{eff} R_j^{eff}} \right)} \quad (9)$$

q_i is the value of the partial charge on atom i, while r_{ij} is the distance between atoms i and j. R_i^{eff} is the effective radius of atom i and it is evaluated numerically on a 3D grid covering the solute as described in [23]. It is a quantity depending only on the solute geometry and represents an estimate of the average distance of a charge from the solvent. For a given ligand atom i, list$_i$ contains all the atoms of the receptor residues whose geometrical center lies within a distance of 10 Å from atom i. Additionally, list$_i$ is supplemented with a monopole approximation of the distant

charged residues (charge center outside of the 10 Å sphere), whose treatment reduces the error originating from the use of a cutoff.

The desolvation of the ligand is the electrostatic energy difference upon binding of a charged ligand to an uncharged receptor in solution. The ligand intramolecular energy in solution is calculated by the GB formula as described in [23]:

$$\sum_{i\varepsilon ligand}\left(\frac{q_i^2}{2R_i^{vdw}}\varepsilon_p - \frac{q_i^2\tau}{2R_i^{eff}}\right) + \sum_{i,j\varepsilon ligand}^{i>j}\left(\frac{q_iq_j}{\varepsilon_p r_{ij}} - \frac{q_iq_j\tau}{R_{ij}^{GB}}\right) \quad (10)$$

where R_i^{vdw} is the van der Waals radius of atom i. The first sum is the self-energy term which represents the interaction between the ligand and the solvent. The second term is the screened ligand-ligand interaction. The desolvation energy of a ligand upon binding to an uncharged receptor in solution $\left(\Delta E_{elect,desolv}^{ligand}\right)$ is equal to the difference of the intramolecular energy of the bound and unbound ligand. Both values are calculated using equation (10). For the unbound ligand, the effective radii are calculated as described in [23] and considering as solute the volume enclosed by the molecular surface of the ligand. For the bound ligand, the low dielectric constant (ε_p) is assigned to the volume enclosed by the molecular surface of the receptor-ligand complex.

2.5.5 Total Energy In vacuo and Binding Energy in Solution

During the conformational optimization by the GA, 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:

$$\Delta E_{total} = E_{vdw}^{int} + E_{elect,rdiel}^{int} + E_{vdw}^{ligand} + E_{elect,rdiel}^{ligand} + E_{strain}^{ligand} \quad (11)$$

The last three terms approximate the intraligand energy of the flexible amino acid (Eqs. 1, 2, and 3) while the intermolecular energy is described by the first two terms (Eqs. 4 and 5). The solvent screening effect is approximated by the distance dependent dielectric model. The ligand and receptor desolvation energies are not taken into account in the GA scoring function because their computation would be too CPU intensive since it requires between 5 and 10 seconds CPU time for a single-point calculation. Furthermore, to prevent the sampling of solvent exposed ligand side chains, it is useful to neglect the desolvation penalty during conformational optimization by the GA. The vacuo binding energy used to test dead-ends consists of the first two terms in equation (11).

At each growing step, the best binding modes obtained by the conformational optimization process (usually 200, i.e., 10 actual constructs times 20 residues in the library) are ranked according to the binding energy in solution which includes the following contributions:

$$\Delta E_{binding} = E_{vdw}^{int} + E_{elect,sol}^{int} + \Delta E_{elect,desolv}^{receptor} + \Delta E_{elect,desolv}^{ligand} \quad (12)$$

where it is assumed that the ligand-receptor vdW interaction energy accounts for all the nonelectrostatic contributions to the binding energy [48]. The difference in intraligand energy is neglected because it is difficult to define the energy of an isolated amino acid. Among the non dead-end sequences, the ten with the most favorable $\Delta E_{binding}$ are selected for further growing.

2.6 System Setup

The library of twenty standard L-amino acids was built with the molecular modelling program WITNOTP (A. Widmer, unpublished). Partial charges were assigned with the MPEOE method [49, 50, 51] implemented in WITNOTP which reproduces the all-hydrogen MSI CHARMm22 parameter set [52]. All residues underwent a CHARMM [24] conjugate gradient minimization to a RMS of the energy gradient of 0.02 kcal/mol Å, using the CHARMm22 force field (Molecular Simulations Inc.). This is required for obtaining optimal bond lengths and bond angles values, since these are not modified by the GA.

The 1.9 Å resolution x-ray structure of a 306-residue fragment of the β-chain of the human insulin receptor tyrosine kinase, complexed with a 6-residue peptide substrate and a non-hydrolysable ATP analog [53], was taken from the the Brookhaven PDB database [54] (access code 1IR3). The water molecules, the 6-residue peptide, and the ATP analog were removed. Hydrogen atoms were added with WITNOTP. Partial charges were assigned with the MPEOE method, and hydrogens were minimized with the CHARMM program. The following thirty residues were chosen to define the binding site: His1184, Arg1164, Lys1182, Phe1186, Lys1165, Leu1171, Met1176, Leu1170, Pro1172, Ser1180, Leu1181, Asp1183, Gly1169, Ala1168, Gly1167, Gly1166, Val1185, Val1173, Asn1215, Leu1219, Glu1216, Met1223, Lys1085, Gln1208, Arg1136, Trp1175, Arg1174, Ser1006, Arg1039, Lys1220. Asp(P-1) of the peptidic substrate was used as seed for the growing procedure and its atomic coordinates were supplemented by a methyl amino group connected to the CO by the program WITNOTP.

The 2.5 Å resolution x-ray structure of rat Farnesyltransferase (FTase) complexed with an FPP analog, α-hydroxyfarnesylphosphonic-acid(αHFP), and the peptide CVIM [55] was downloaded from the PDB database (access code 1QBQ). The water molecules and the peptide (CVIM) were removed, whereas αHFP was left inside the FPP binding site. Hydrogen atoms were built with the HBUILD [56] option of the CHARMM program [24]. Partial charges were assigned with the MPEOE method implemented in WITNOTP, and hydrogens were minimized with the CHARMM program. The Zn^{2+} and the following twenty-six residues were chosen to define the binding site: Ala129α, Tyr131α, Lys164α, Asn165α, Tyr166α, Gln167α, His201α, Ala98β, Ser99β, Trp102β, Trp106β, Gly142β, His149β, Ala151β, Pro152β, Met193β, Glu198β, Arg202β, Asp297β, Cys299β, Tyr300β, Trp303β, Asp352β, Tyr361β, His362β, Tyr365β. The first residue of the substrate (Cys) was used as seed and it was prepared for

growing as described above for the Asp(P-1) in the tyrosine kinase.

2.7 Computation Times

For the applications to the insulin receptor tyrosine kinase and farnesyltransferase, one PEP run required 18 and 9 hours on a 550Mhz PentiumIII processor, respectively. Farnesyltransferase required only about half of the time because only three growing steps were performed, while four growing steps were carried out for tyrosine kinase. Multiple PEP runs for each protein were performed in parallel on a cluster of PCs running Linux.

3 RESULTS AND DISCUSSION

3.1 Tyrosine Kinase

Most kinase inhibitor projects aim at blocking the binding of ATP [57]. This approach is however difficult because the intracellular ATP concentration is high, and the ATP binding sites are very similar among different kinases, which leads to selectivity problems [58]. Inhibitors directed at the substrate binding site do not suffer of these limitations, but they usually have low binding affinities due to the shallowness of the substrate site. The insulin receptor is an $\alpha_2\beta_2$ transmembrane glycoprotein with intrinsic kinase activity [59] that mediates the physiological effect of insulin [60]. It has been shown that tyrosine residues in the YMXM motif, where X can be any standard amino acid, are efficient substrates of the insulin receptor in vitro and in vivo [61]. The eighteen residues peptide KKKLPATGD**YMNM**SPVGDis an insulin receptor substrate with a reported K_m of 24 μM [61]. Its x-ray structure in complex with the activated insulin receptor kinase is available (code 1IR3) [53]. Of the eighteen residues, supporting electron density is seen for only six, from the P-2 to the P+3 residue (GDYMNM, P-residue is the acceptor tyrosine). The shallow substrate binding site of tyrosine kinase represents a challenging test case for PEP and the solvation model.

As a first test, a single sequence was docked into the substrate binding site. Starting from the Asp(P-1) residue, twenty growing runs restricted to the sequence YMYM were carried out, each one with a different initial random number value for the GA. The aim was twofold: to check that the program was able to find the right binding mode of the peptide, and verify that the results were independent of the seed values used. Since in the x-ray structure the side chain of Asn(P+2) does not interact with the receptor, a tyrosine at P+2 was used for the docking test. Out of the twenty conformations of YMYM generated by PEP, nineteen have the correct backbone and side chain orientations (apart from one conformation with a different rotamer for the χ_2 of Met at P+1) (Fig. (**3A**)). The remaining one has the worst energy and a different binding mode. All the binding features previously described [53] are reproduced in the nineteen correctly placed structures. The four backbone hydrogen bonds are reproduced: two between the backbone polar groups of Met(P+1) and Leu1171, and two between Met(P+3) and Gly1169. The largest deviation is found in the hydrogen bond between the CO of Met(P+3) and the NH of Gly1169. For this hydrogen bond, the distance between the heavy atoms ranges from 3.4 Å to 4 Å in the nineteen conformations docked by PEP, while it is 3.0 Å in the x-ray structure.

The hydroxyl group of Tyr(P) is hydrogen bonded both as donor and acceptor to the side chain of Asp1132 and Arg1136, respectively (Fig. (**3A**)). Finally the methionine side chains of the peptide fit into two adjacent hydrophobic pockets (Fig. (**3B**)). Most of the binding interactions are still present in the conformation with the worst energy and different binding mode. The hydrogen-bonding of the Tyr(P) side chain and Met(P+1) backbone are conserved. However, the methionine side chains are not placed in the hydrophobic pockets and the Met(P+3) backbone hydrogen bond is lost. The average backbone RMSD of the twenty peptides compared to the x-ray structure is 0.67 Å. It is remarkable that the structure with the incorrect binding mode and the one with the misplaced Met(P+1) side chain have the worst and the second worst energy values, respectively. Moreover, the incorrect structure is separated by an energy gap of about 4 kcal/mol from the nineteen best conformations. These results show that the GA approach is very reproducible (95% in this case), and that the binding energy in solution is able to rank correctly different binding modes of the same peptide generated by different GA runs.

A more stringent test of PEP is the design of tetrapeptide sequences. Starting from the same amino acid seed as for the docking test (Asp (P-1)), fifteen PEP runs unrestricted in sequence space were performed with different initial random number values for the GA. A library containing the twenty naturally occurring amino acids was used at each growing step. Since the grown sequence is part of a longer peptide, the dead-end selection procedure was performed even for the fourth growing step in order to discard sequences that would not allow further growing. To analyze the results one has to consider that the GA performs a stochastic search; hence, the most favorable sequences are those which are generated in many runs and with a good binding energy in solution. A given sequence can be generated by PEP n times (with up to n different conformations) out of m growing runs with $n \leq m$. 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. Table 1 contains sequences that were generated most often in the GA runs, and only the five with the most favorable average binding energy in solution are listed for each set of sequences having the same occurrence value. The backbones are oriented correctly with the four aforementioned hydrogen bonds formed, and the largest backbone RMS deviation is 0.54 Å. The most frequent peptides contain Tyr at P and P+2 (Fig. (**3B**-**D**)). The former is in agreement with experimental data, while the Tyr at P+2 is a prediction. It would be interesting to test this prediction by the synthesis and test of the YMYM or a longer peptide. Ala and Met are found at P+1 and hydrophobic residues are clearly favored at P+3. Tyr and Phe fit very well in the hydrophobic pocket occupied by Met(P+3) in the x-ray structure, and the hydroxyl group of the Tyr donates a hydrogen bond to the CO of Leu1181. The sequence with the best binding energy (YMYM) corresponds to the YMXM motif.

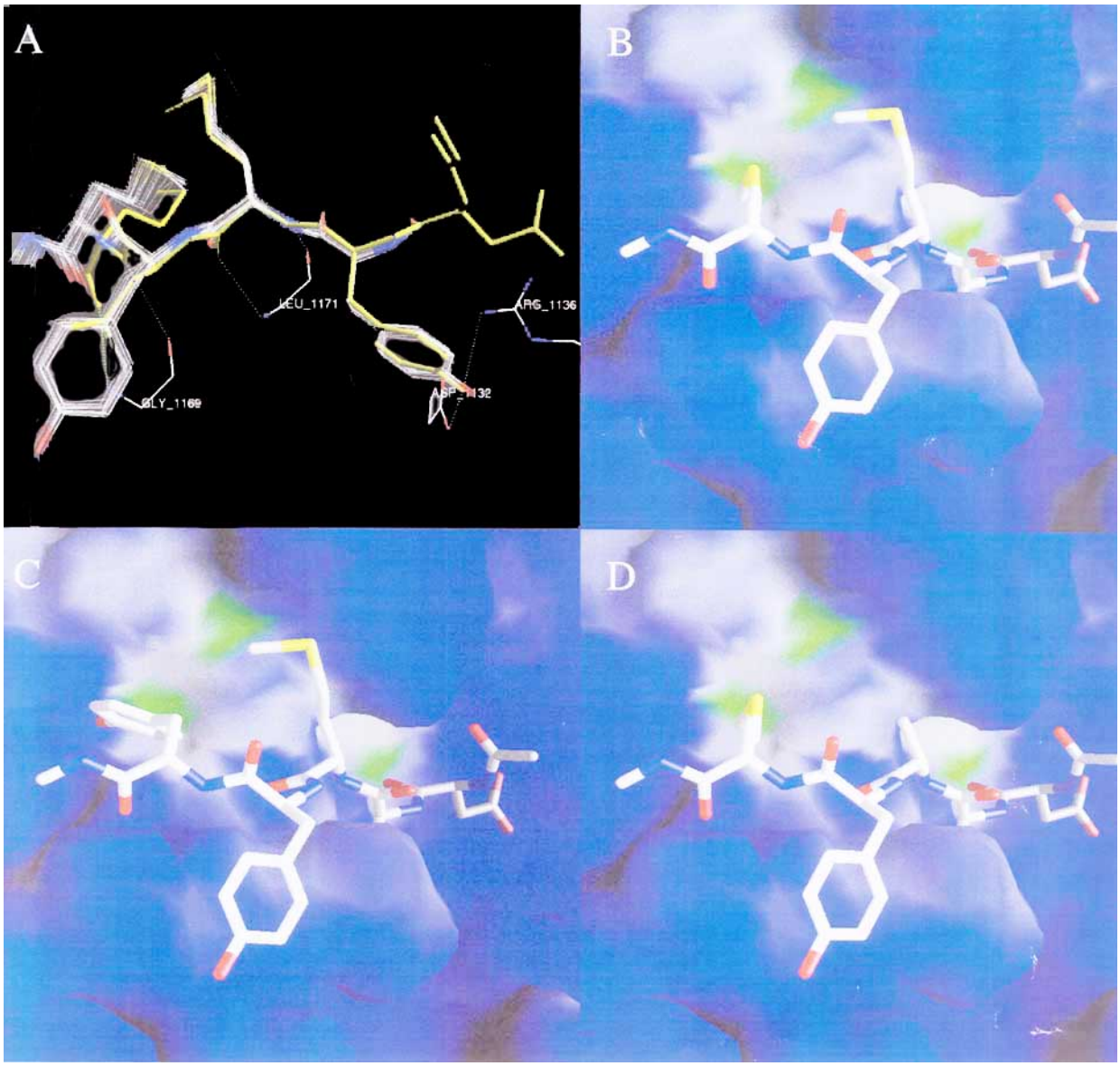

Fig. (3). (A) Nineteen YMYM tetrapeptide conformations docked by PEP in the active site of the insulin receptor tyrosine kinase. The substrate x-ray structure (YMNM) is shown in yellow, and hydrogen bonds are indicated by green dotted lines. (B) The molecular surface of the tyrosine kinase active site is displayed together with the YMYM tetrapeptide grown by PEP. The side chain of Tyr(P) is not visible because it is buried in a deep pocket. Hydrophobic regions are displayed in green and hydrophilic in blue [48]. Figure made with GRASP [73]. (C) Same as (B), with the YMYY tetrapeptide designed by PEP. (D) Same as (B), with the YAYM tetrapeptide designed by PEP.

3.2 Farnesyltransferase

Ras proteins play a critical role in signal transduction pathways that control cell growth and differentiation. Mutants of three human Ras proteins (Ha-Ras, Ki-Ras and N-Ras) are found in 20-30% of all human cancers [62, 63, 64], which makes them an attractive target for antitumoral drug design. A promising approach for interfering in the Ras function involves inhibition of the enzyme farnesyltransferase (FTase). This enzyme covalently links the isoprenoid moiety of farnesyl pyrophosphate (FPP) to the C-terminal part of Ras as well as to other membrane associated proteins [65, 66, 67]. This, among other fast posttranslational modifications, is required for their attachment to the plasma membrane, which is essential for their biological activity [68, 69, 70]. Ras processing and membrane association are signaled by a carboxyterminal tetrapeptide sequence present on all Ras proteins. This sequence is normally referred to as the CaaX motif where 'C' stands for a cysteine, 'a' is generally an aliphatic amino acid, and 'X' typically is a methionine and less frequently a Ser, Ala, Phe or Leu [71].

Starting from the Cys residue that coordinates the zinc atom, fifteen unrestricted growing were performed, with

Table 1. Insulin Receptor Tyrosine Kinase Inhibitors Generated by PEP

Sequence				Occurrences[a]	Relative binding energy[b]	Backbone RMS deviation[c]
P	P+1	P+2	P+3	[%]	[kcal/mol]	[Å]
Y	A	Y	M[d]	87	1.4	0.37
Y	A	Y	Y	87	1.9	0.35
Y	A	Y	A	87	2.5	0.36
Y	A	Y	G	87	2.7	0.38
Y	A	Y	I	87	2.7	0.37
Y	**M**	**Y**	**M**[e]	73	0.0	0.54
Y	M	Y	Y[f]	73	1.2	0.46
Y	M	Y	F	73	1.6	0.47
Y	M	Y	A	73	1.7	0.47
Y	M	Y	V	73	1.9	0.48

The sequence in bold corresponds to the YMXM motif; [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 average binding 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 3D; [e]Peptide shown in Figure 3B. [f]Peptide shown in Figure 3C.

different initial random number values for the GA. PEP generated tripeptides that belong to the consensus motif of FTase peptidic inhibitors (Table 2). It is striking that the most frequent tripeptides contain Val and Ile at a1 and a2, respectively. Furthermore, among the sequences with the highest occurrence (12 of 15 runs) the one with the second best energy (VIM) corresponds to the tripeptide sequence in the x-ray structure [55]. The conformation of the PEP hits overlap the x-ray structure (Fig. (**4A**)). In particular, the backbone RMS deviation of the first five sequences in their minimum energy conformation is smaller than 1.0 Å. The deviations of the backbone are larger than for tyrosine kinase because only two intermolecular hydrogen bonds are formed between the peptide substrate and FTase. The heavy atom RMSD between the VIM tripeptide generated by PEP and the x-ray structure is 1.23 Å. The two hydrogen bonds between the backbone of the CVIM peptide and FTase are reproduced. These are the hydrogen bonds between the CO of a2 and the guanidinium of Arg202β and between the C-terminal carboxy group and the NH_2 of Gln167α (Fig. (**4A**)). The second most frequent set of sequences contain a isoleucine and glutamine at positions a1 and a2, respectively. Glutamine at position a2 was found recently by combinatorial tetrapeptide libraries in very efficient substrates [72].

Another series of fifteen PEP runs were carried out using the binding energy in vacuo at the selection step. The aim was to determine if solvation is required to get the correct peptide sequence. Although the peptides generated by PEP still have the right backbone conformation (Fig. (**4B**)), large polar/charged side chains are now clearly favored (Table 3). The most frequent tripeptides contain glutamine and lysine at position a1 and a2, respectively. Polar residues are also predominant for the C-terminal residue. These results indicate that the desolvation penalty of the ligand can not be neglected. Moreover, the percentage of occurrence is much lower compared to the results obtained with solvation

Table 2. Farnesyltransferase Inhibitors Generated by PEP

Sequence			Occurrences[a] [%]	Relative binding energy[b] [kcal/mol]	Backbone RMS deviation[c] [Å]
V	I	L	80	2.5	0.59
V	**I**	**M**[d]	80	3.3	0.60
V	I	H	80	3.6	0.59
V	I	T	80	3.6	0.54
V	I	G	80	3.9	0.91
I	Q	M	60	0.0	1.06
I	Q	Q	60	1.2	1.28
I	Q	L	60	1.5	0.72
I	Q	H	60	2.3	1.03
I	Q	S	60	2.6	1.19

The sequence in bold corresponds to the peptide in the x-ray structure; [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 average binding 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 4A.

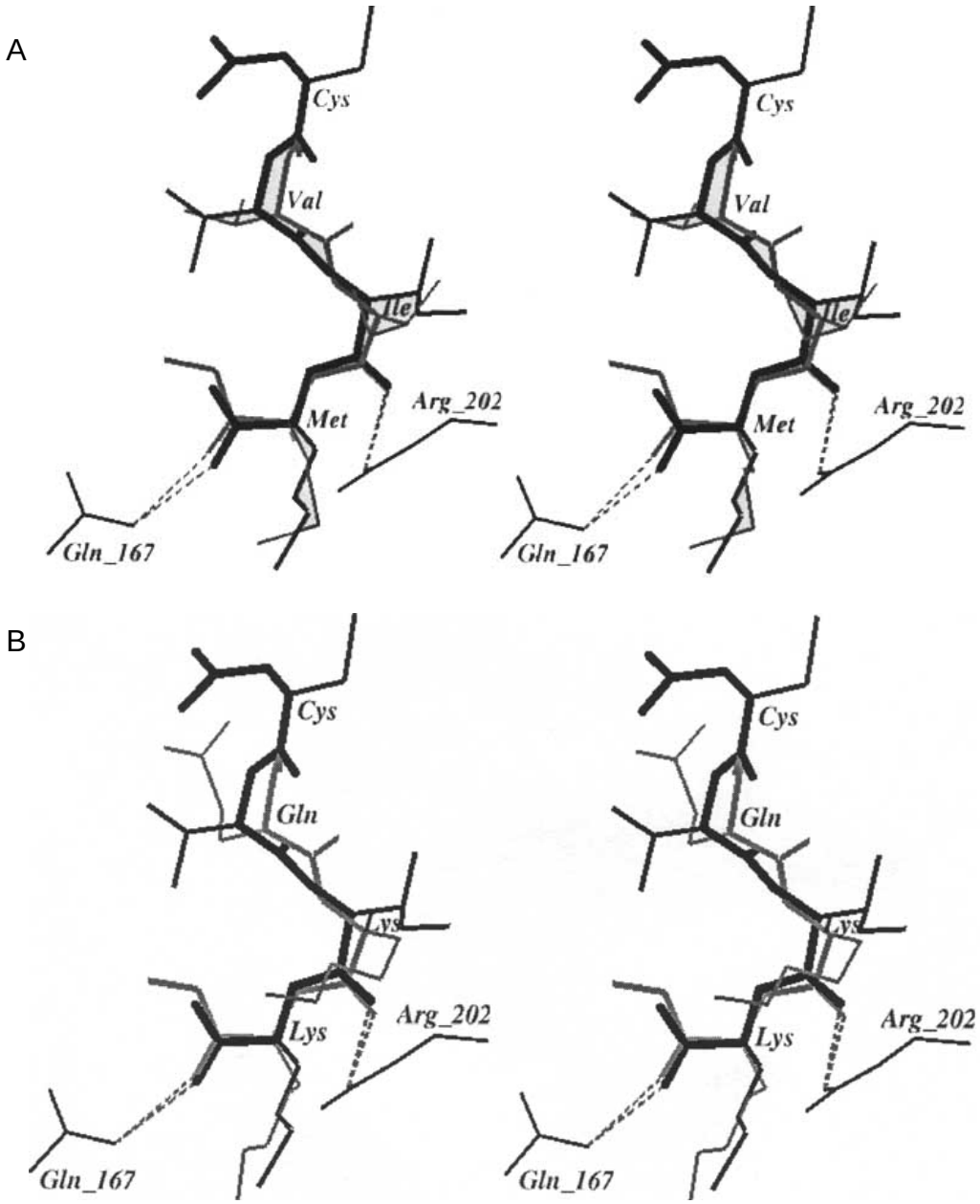

Fig. (4). Stereo-view of PEP results on FTase. (A) The x-ray structure of the CVIM substrate is shown with thick black lines and the CVIM sequence grown by PEP in thin grey lines. Conserved hydrogen bonds with two FTase side chains are indicated by dashed lines. (B) Same as (A), with the CQKK sequence (thin lines) grown by PEP using the vacuo binding energy.

(compare second column of tables 2 and 3). Hence, neglecting solvation it is more difficult to obtain convergence in chemical space.

4 CONCLUSIONS

PEP is a ligand build-up approach that uses at each growing step a GA for conformational optimization of the last added monomer. The ligands are linear combinations of monomers connected by amide bonds and can consist of peptides, peptoids (N-substituted amino acids), and/or every kind of organic fragment with an amino and carboxy group. The main advantage of this type of compounds is that they can be synthesized by combinatorial or parallel approaches provided that the monomers are available. PEP uses an accurate evaluation of binding energy in solution to drive the search in chemical space. For the two test cases presented in this study, the peptidic ligands designed by PEP have both sequence and binding mode in agreement with experimental results.

The main limitation of PEP is the requirement of a correctly oriented seed fragment. The placement of the seed

Table 3. Farnesyltransferase Inhibitors Generated Using the Vacuo Binding Energy

Sequence			Occurrences[a] [%]	Relative binding energy[b] [kcal/mol]	Backbone RMS deviation[c] [Å]
Q	K	K[d]	34	0.0	0.61
Q	K	R	34	2.8	0.58
Q	K	T	34	13	0.66
Q	K	Y	34	13.1	0.76
Q	K	Q	34	14.4	0.71

[a]Percentage of PEP runs (out of 15) that generated a given sequence; [b]Vacuo binding energy averaged over all conformations of a given sequence. The average binding 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 vacuo energy was used to calculate the RMS deviation; [d]Peptide shown in Figure 4B.

fragment, although separated from the growing method itself, has a great influence on the outcome of the procedure. A poorly positioned seed can prevent ligands from reaching important interaction sites in the receptor. We are currently implementing in PEP some limited flexibility in the orientation of the seed fragment (Budin et al. in preparation). A number of methods are available for choosing reasonable seed positions. If an x-ray structure with a bound peptidic ligand is available, one amino acid of the ligand can be used as seed position. Otherwise, it is possible with most modelling systems to manually dock a seed fragment into the receptor site. However, identifying the optimal placement of the seed is not always a trivial problem. Another option is to use an automatic fragment docking program like SEED, an in house developed program for exhaustive docking of small ligands with electrostatic solvation [13].

A second limitation is that in the current implementation of PEP the protein target is kept rigid. The easiest way to overcome this problem is to run PEP on different conformations of the same target sampled by molecular dynamics or, if available, from crystal structures of complexes with different inhibitors. This is not a genuine solution of the problem, but it is easy to realize because of the very favorable price/performance ratio of PC clusters. The large amount of output data has to be further processed (clustering in sequence and conformational space).

Another limitation is the neglect of the solute entropic penalty upon binding. Simple approximations based on counting the number of frozen rotatable bonds could be implemented. The electrostatic contribution to differences in solvent entropy is taken into account in the continuum dielectric approach, whereas the nonpolar contribution could be approximated by the change in solvent accessible surface.

ACKNOWLEDGMENTS

We thank C. Ehrhardt (Novartis Pharma Inc., Basel) for helpful comments and for suggesting the insulin receptor tyrosine kinase as test case. We also thank C. Tenette-Souaille and J. Apostolakis for helpful discussions, and A. Widmer (Novartis Pharma Inc., Basel) for the molecular modelling program WITNOTP. This work was supported by the Swiss National Science Foundation (Nationalfonds, grant nr. 31-53604.98) and Novartis Pharma Inc. 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 (email: caflisch@bioc.unizh.ch).

REFERENCES

[1] Hattori, M. *Nature* **2000**, *405*, 311-319.

[2] Dunham, I. *Nature* **1999**, *402*, 489-495.

[3] Adams, M. D. *Science* **2000**, *287*(5461), 2185-2195.

[4] Sánchez, R.; Sali, A. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 13597-13602.

[5] Caflisch, A.; Wälchli, R.; Ehrhardt, C. *News Physiol. Sci.* **1998**, *13*, 182-189.

[6] Kubinyi, H. *Curr. Opin. Drug Design Discov.* **1998**, *1*, 4-15.

[7] Weber, L.; Wallbaum, S.; Broger, C.; Gubernator, K. *Angew. Chem. Int. Ed.* **1995**, *34*, 2280-2282.

[8] Apostolakis, J.; Caflisch, A. *Comb. Chem. High Throughput Screen.* **1999**, *2*, 91-104.

[9] Pearlman, D.; Murcko, M. *J. Comput. Chem.* **1993**, *14*, 1184-1193.

[10] Gehlhaar, D. K.; Moerder, K. E.; Zichi, D.; Sherman, C. J.; Ogden, R. C.; Freer, S. T. *J. Med. Chem.* **1995**, *38*, 466-472.

[11] Todorov, N. P.; Dean, P. M. *J. Comput.-Aided Mol. Design* **1998**, *12*, 335-349.

[12] Miranker, A.; Karplus, M. *Proteins: Structure, Function and Genetics* **1991**, *11*, 29-34.

[13] Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhard, C.; Caflisch, A. *Proteins: Structure, Function and Genetics* **1999**, *37*, 88-105.

[14] Caflisch, A. *J. Comput.-Aided Mol. Design* **1996**, *10*, 372-396.

[15] Moon, J. B.; Howe, W. J. *Proteins: Structure, Function and Genetics* **1991**, *11*, 314-328.

[16] Böhm, H. J. *J. Comput.-Aided Mol. Design* **1992**, *6*, 61-78.

[17] Frenkel, D.; Clark, D. E.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. *J. Comput.-Aided Mol. Design* **1995**, *9*, 213-25.

[18] Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, Jr., S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765-784.

[19] Ooi, T.; Oobatake, M.; N'emethy, M.; Scheraga, H. A. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 3086-3090.

[20] Böhm, H. J. *J. Comput.-Aided Mol. Design* **1996**, *10*, 265-272.

[21] Böhm, H. J. *J. Comput.-Aided Mol. Design* **1994a**, *8*, 243-256.

[22] Clark, K. P.; Ajay, *J. Comput. Chem.* **1995**, *16*(10), 1210-1226.

[23] Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. A* **1997**, *101*, 8098-8106.

[24] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187-217.

[25] Davis, L. Handbook of Genetic Algorithms Van Nostrand Reinhold, New York NY, **1991**.

[26] Goldberg, D. E. Genetic Algorithms in Search Optimization and Machine Learning Addison-Wesley, Reading MA, **1989**.

[27] Scarsi, M.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **1998**, *102*, 3637-3641.

[28] Luty, B.A.; Wasserman, Z.R.; Stouten, P.F.W.; Hodge, C.N.; Zacharias, M.; McCammon, J.A. *J. Comput. Chem.* **1995**, *16*, 454-464.

[29] Caflisch, A.; Ehrhardt, C. *Structure-based combinatorial ligand design*. In Veerapandian, P., editor, Structure-based drug design, pages 541-558, **1997**.

[30] Horvath, D. *J. Med. Chem.* **1997**, *40*, 2412-2423.

[31] Ferrara, P.; Apostolakis, J.; Caflisch, A. *J. Phys. Chem. B* **2000**, *104*, 5000-5010.

[32] Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Structure, Function and Genetics* **2000**, *39*, 252-260.

[33] Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928-1931.

[34] Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. Numerical Recipes in Fortran Cambridge University Press, **1992**.

[35] Hagler, A. T.; Huler, E.; Lifson, S. *J. Am. Chem. Soc.* **1974**, *96*, 5319-5327.

[36] Pattabiraman, N.; Levitt, M.; Ferrin, T. E.; Langridge, R. *J. Comput. Chem.* **1985**, *6*, 432-436.

[37] Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. *J. Comput. Chem.* **1992**, *13*, 505-524.

[38] Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, 157, 671-679.

[39] Gilson, M. K.; Honig, B. H. *Proteins: Structure, Function and Genetics* **1988**, *4*, 7-18.

[40] Bashford, D.; Karplus, M. *Biochemistry* **1990**, *29*, 10219-10225.

[41] Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comp. Phys. Comm.* **1991**, *62*, 187-197.

[42] Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127-6129.

[43] Hawkins, G. D.; Cramer, C. J.; Trulhar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122-129.

[44] Hawkins, G. D.; Cramer, C. J.; Trulhar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824- 19839.

[45] Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578-1599.

[46] Qiu, Di ; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005-3014.

[47] Marrone, T. J.; Gilson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996**, *100*, 1439-1441.

[48] Scarsi, M.; Majeux, N.; Caflisch, A. *Proteins: Structure, Function and Genetics* **1999**, *37*, 565-575.

[49] Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219-3288.

[50] No, K.T.; Grant, J.A.; Scheraga, H.A. *J. Phys. Chem.* **1990**, *94*, 4732-4739.

[51] No, K.T.; Grant, J.A.; Jhon, M.S.; Scheraga, H.A. *J. Phys. Chem.* **1990**, *94*, 4740-4746.

[52] Momany, F.A.; Klimkowski, V.J.; Sch=E4fer, L. *J. Comput. Chem.* **1990**, *11*, 654-662.

[53] Hubbard, S. R. *EMBO J.* **1997**, *16*(18), 5573-5581.

[54] Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, Jr., E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535-542.

[55] Strickland, C. L.; Windsor, W. T.; Syto, R.; Wang, L.; Bond, R.; Wu, Z.; Schwartz, J.; Le, H. V.; Beese, L. S.; Weber, P. C. *Biochemistry* **1998**, *37*, 16601- 16611.

[56] Brünger, A.; Karplus, M. *Proteins: Structure, Function and Genetics* **1988**, *4*, 148-156.

[57] Al-Obeidi, F.A.; Wu, J.J.; Lam, K.S. *Biopolymers* **1998**, *47*, 197-223.

[58] Toledo, L.M.; Lydon, N.B.; Elbaum, D. *Curr. Med. Chem.* **1999**, *6*(9), 775-805.

[59] Ullrich, A.; Bell, J.R.; Chen, E.Y.; Herrera, R.; Petruzzelli, L.M.; Dull, T.J.; Gray, A.; Liao, Y.C.; Tsubokawa, M. *Nature* **1985**, *313*, 756-761.

[60] Ebina, Y.; Ellis, L.; Jarnagin, K.; Edery, M.; Graf, L.; Clauser, E.; Ou, J.H.; masiarz, F.; Kan, Y.W.; Goldfine, I.D. *Cell* **1985**, *40*, 747-758.

[61] Shoelson, S. E.; Chatterjee, S.; Chaudhuri, M.; White, M. F. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2027-2031.

[62] Barbacid, M. *Annual Review of Biochemistry* **1987**, *56*, 779-827.

[63] Bos, J. L. *Cancer Research* **1989**, *49*, 4682-4689.

[64] Keely, P.; Parise, L.; Juliano, R. *Trends in Cell Biology* **1998**, *8*, 101-106.

[65] Farnsworth, C. C.; Wolda, S. L.; Gelb, M. H.; Glomset, J. A. *J. Biol. Chem.* **1989**, *264*, 20422-20429.

[66] Casey, P. J.; Solski, P. A.; Der, C. J.; Buss, J. E. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 8323-8327.

[67] Maltese, W. A.; Robishaw, J. D. *J. Biol. Chem.* **1990**, *265*, 18071-18074.

[68] Maltese, W. A. *FASEB Journal* **1990**, *4*, 3319-3328.

[69] Cox, A. D.; Der, C. J. *Current Opinion in Cell Biology* **1992**, *4*, 1008-1016.

[70] Magee, A. I.; Newman, C. M.; Giannakouros, T.; Hancock, J. F.; Fawell, E.; Armstrong, J. *Biochemical Society Transactions* **1992**, *20*, 497-499.

[71] Reiss, S. J.; Stradley, L. M.; Gierasch, M. S.; Brown, M. S.; Golddstein, J. L. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 732-736.

[72] Boutin, J. A.; Marande, W.; Petit, L.; Lyonel, A.; Desmet, C.; Canet, E.; Fauchere, J. L. *Cellular Signalling* **1999**, *11*, 59-69.

[73] Nicholls, A.; Sharp, K.A.; B., Honig *Proteins: Structure, Function and Genetics* **1991**, *11*, 281-296.

Received: 1 August 2000 **Accepted: 23 November 2000**