OPEN ACCESS Freely available online

PLoS one

VitAL: Viterbi Algorithm for *de novo* Peptide Design

E. Besray Unal, Attila Gursoy, Burak Erman*

Center for Computational Biology and Bioinformatics, Koc University, Istanbul, Turkey

Abstract

Background: Drug design against proteins to cure various diseases has been studied for several years. Numerous design techniques were discovered for small organic molecules for specific protein targets. The specificity, toxicity and selectivity of small molecules are hard problems to solve. The use of peptide drugs enables a partial solution to the toxicity problem. There has been a wide interest in peptide design, but the design techniques of a specific and selective peptide inhibitor against a protein target have not yet been established.

Methodology/Principal Findings: A novel *de novo* peptide design approach is developed to block activities of disease related protein targets. No prior training, based on known peptides, is necessary. The method sequentially generates the peptide by docking its residues pair by pair along a chosen path on a protein. The binding site on the protein is determined via the coarse grained Gaussian Network Model. A binding path is determined. The best fitting peptide is constructed by generating all possible peptide pairs at each point along the path and determining the binding energies between these pairs and the specific location on the protein using AutoDock. The Markov based partition function for all possible choices of the peptides along the path is generated by a matrix multiplication scheme. The best fitting peptide for the given surface is obtained by a Hidden Markov model using Viterbi decoding. The suitability of the conformations of the peptides that result upon binding on the surface are included in the algorithm by considering the intrinsic Ramachandran potentials.

Conclusions/Significance: The model is tested on known protein-peptide inhibitor complexes. The present algorithm predicts peptides that have better binding energies than those of the existing ones. Finally, a heptapeptide is designed for a protein that has excellent binding affinity according to AutoDock results.

Citation: Unal EB, Gursoy A, Erman B (2010) VitAL: Viterbi Algorithm for *de novo* Peptide Design. PLoS ONE 5(6): e10926. doi:10.1371/journal.pone.0010926

Editor: Joel S. Bader, Johns Hopkins University, United States of America

Received March 12, 2010; **Accepted** May 7, 2010; **Published** June 2, 2010

Copyright: © 2010 Unal et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: berman@ku.edu.tr

Introduction

The determination of a specific peptide sequence with affinity to a particular protein surface is a problem of high degree of complexity arising from the fact that each residue of the peptide could be chosen from a pool of twenty natural amino acids. Even for a peptide with three amino acid residues, there exist 8×10^3 possible peptide sequences. Screening of such a large number of molecules is complicated with both experimental and computational techniques. A rational methodology for specific and selective peptide sequence prediction is required. The necessary methodology should be time-efficient and be able to design peptides for given locations on a given set of proteins. A fast and global computational tool is desired. The importance, computational and experimental difficulties and the present state of the art of finding new peptides have recently been discussed and reviewed by Petsalaki et al. [1].

The complete peptide binding problem can be visualized as a three step process: (i) finding a path on the surface of the protein which defines a suitable region for the peptide to bind, (ii) finding the appropriate peptide for this path, and (iii) improving the peptide for a more stable binding required for inhibition. In some cases, the binding surface is known, and a peptide must be designed *de novo*. In other cases, a peptide is given and the best region on the surface that gives the optimum binding conditions is searched. The method of Petsalaki et al., based on a knowledge based bioinformatics approach addressed this problem and could successfully find the binding sites for the peptides. A similar 'indirect' design approach has been adopted by Frenkel et al. [2], using their *de novo* molecular design computational tool Pro_Ligand. Known peptides were docked to unknown locations on given proteins by Hetenyi and Spoel [3] using AutoDock. There have been successful attempts for computational peptide design that use knowledge-based search strategies and use diverse sets of statistical descriptors, different training databases, hydrophobicity scales, motif regularities, etc. [4]. Also, automated peptide binding search techniques [5,6,7] from known epitopes or protein libraries have been successfully used as bioinformatics tools. There have been applications of bioinformatics computational binding tools such as the sequence moment concept, artificial neural networks, fuzzy neural networks and Hidden Markov Model for checking the suitability of inhibitory peptides for binding on MHC class II proteins [4,8,9,10,11,12,13,14,15,16]. Along similar lines, the suitability of a ligand as a drug was tested using Bayesian neural network analysis [17]. Application of genetic algorithms to the design of peptides has been an important line of research, examples of which are: *in silico* peptide screening and application of genetic algorithm to determine inhibitory peptide against Parkinson's-disease-related protein α-Synuclein [18]; peptides as thrombin inhibitors [19,20]; evaluation of energies for peptide

binding to a user-defined protein surface patch via Genetic algorithm for p53, oligopeptidase and DNA gyrase [21]; integer linear programming [22]; design of hexapeptides against stromelysin protein by Singh et al. [23], and a peptide buildup approach together with a genetic algorithm [24,25,26]. We have implemented genetic algorithm and a Markov model for *de novo* heptapeptide design in our recent study [27]. Most of the methods cited in this paragraph use models that depend on structural properties of the peptides and the peptide-protein complex. The prior knowledge of the structural basis of peptide-protein binding strategies is extremely important. The recent paper by London et al. [28], that reviews these strategies is important for understanding the basis of the works cited here. Important points that should be taken into consideration in the sequential generation of ligand molecules concerning the placement of fragments on the surface are discussed by Pegg et al. [29].

Our method offers a novel procedure for *de novo* peptide design that sequentially generates the peptide by docking its residues pair by pair along a chosen path on a protein. We adopt three novel approaches in our design: (i) The first one is the determination of the binding site on a given protein which we determine using the coarse grained Gaussian Network Model (GNM). Recently, we showed that the GNM identifies the surface residues that are suitable binding sites [30,31,32]. Once the binding region is determined, we obtain the binding path on it by docking an arbitrarily chosen peptide using AutoDock [33]. As will be described below, this path is flexible and not very restrictive. (ii) The second novel aspect of the model is the choice of the best fitting peptide to this path. We generate all possible amino acid pairs at each point along the path, calculate the binding energies between these pairs and the specific location on the protein via AutoDock, and form the statistical weight of each pair of amino acids. Once all possible pairs are determined for the full path, we form the Markov based partition function for all possible choices of the peptides using the Ising model matrix multiplication scheme [34]. We evaluate the transition probabilities based on this partition function, and select the best peptide for the given surface employing a Hidden Markov model (HMM) using Viterbi decoding [35]. The types of amino acids are the hidden variables of the algorithm to be obtained as the solution of the problem. (iii) The third novel feature of our approach is the consideration of the suitability of the conformations of the peptides that result upon binding on the surface by including the intrinsic Ramachandran potentials of the $\phi-\psi$ angles [36]. These are the observables of the Viterbi algorithm. As in the choice of the peptides, we assume that the Ramachandran torsion angles obey Markov statistics according to which a given torsion angle depends on the choice of the preceding torsion angle.

The Viterbi algorithm is an efficient way of determining the best state solution of a hidden Markov model based on a given observation sequence [37,38] and is being used widely in the analysis of biological data and in bioinformatics area [39]. Protein structure prediction, where proteins are represented as Markov states, utilizes the Viterbi algorithm. The methodologies are summarized by Bystroff et al. [40]. Analysis of protein, RNA or DNA sequences were also achieved by the Viterbi algorithm as illustrated by a study on gene finding from DNA sequence by Sramek et al. [41]. The prediction of the topology of all-beta membrane proteins combining Viterbi and posterior algorithms was proposed [42] as a modified Viterbi algorithm. Mirabeau et al. determined novel peptide hormones by the Viterbi algorithm by training their algorithm with known receptor protein peptide hormone interactions and testing the model on peptide sequences in databases [43]. A similar Viterbi algorithm was adopted by Sonmez et al., [44]. Noguchi et al. employed the Viterbi algorithm on 3 different studies for peptide design against MHC class II proteins. Training was achieved by non-binding and binding peptides of the target proteins and tests applied to different dataset indicated that the method is able to predict binder peptide sequences [11,14,16].

The plan of the paper is as follows: In the first part, we define the methods needed for the Viterbi algorithm to run properly. Firstly, we define the binding site and path selection procedure. We then explain how the selected binding-path is partitioned into n number of overlapping grids, where n is the desired peptide length. Subsequently, the docking procedure of the 20 amino acids for the first site and 400 dipeptides each of the succeeding n grids on the binding-path are explained. The calculated binding energy scores from the docking procedure are used to obtain the statistical weights for the pair wise dependent choice of amino acids. Knowing the weights, the partition function for a given sequence is calculated using a matrix multiplication scheme and the transition probabilities for each n grid are obtained. We define the various regions on a Ramachandran map as the torsion states of the residues. We then construct the probabilities of the torsion states of a pair of neighboring residues using information from the Protein Data Bank. Inasmuch as we are interested in the denatured conformations of peptides, we construct a coil library from which we obtain the probabilities of the torsion angles [45]. The bound (docked) conformations of each dipeptide on each n grid are used to obtain the torsion angle state of the dipeptide of interest which is used to define the emission probabilities. In the second part of the paper, implementation of the Viterbi algorithm to the peptide binding problem is defined. In the final part, several examples are given.

Methods

The Model

The model consists of five parts: (i) Determining the binding site and the sequence of residues on the surface of the target protein on which the peptide will be docked, (ii) determining the chiral carbon positions of the n residue peptide that will be interacting with the sequence of residues on the target protein surface, (iii) partitioning the path of the n points into a sequence of grids, (iv) docking, by using AutoDock, all 20 peptides to the first grid, and all 400 dipeptides to the succeeding pairs of grids and evaluating their binding affinities, (v) characterizing the $\phi-\psi$ propensities of the dipeptides. We then use the outcome of these five steps for the implementation of the Viterbi algorithm.

Determining the binding site and the sequence of residues on the target protein. The Gaussian network model (GNM) has been shown to predict the protein residues located at specific sites for drug binding [30,31,32]. We employed GNM to predict 'binding site residues' that play major role in peptide binding.

Having identified a site by the GNM, we then need a sequence of residues on the protein that will be in contact with the binding peptide. We choose this sequence of residues using either of the following two approaches: (i) If the protein exists in complex with other proteins, and if the site determined by the GNM lies in an interface in the complex, then the complex is used to determine this sequence of residues on the surface. Recently, two studies indicated that protein-protein interactions, Tuncbag et al. [46] and protein-peptide interactions, Vanhee et al. [47] adopt the same structural motifs as monomeric protein folds. The existence of specific folds at the interaction site increases the stability of the formed protein-peptide complex. The counterpart fold of the

target protein can be safely used as the fold of designed peptide. The protein crystal structure of the target protein is given as input to the web-server HotPoint [48] which predicts the residues of interest on the surface. In Figure 1 an example obtained from the HotPoint server is given. In this specific example the target protein Human Growth Hormone (HGH) is formed by helices and the counterpart protein Human Growth Hormone Receptor (HGHR) is formed by beta-strands. The binding site of HGH, determined by the GNM, is depicted in pink and the residues that interact with the HGH binding-site, determined by HotPoint, are in green. The region shown in pink will contain the residues to which the peptide will bind. The chiral carbons of the residues shown in green are used as grid-centers for the peptide to be designed, which will be explained in the next section. (ii) If an interacting partner of the protein does not exist, a probe peptide made up of all alanine residues, equal in number to the residues of the peptide to be designed, is docked to the binding site using AutoDock. The specific aim of docking an all alanine peptide is to obtain a path which may well be approximated by the backbone contour. Deviations from this path, due to the forces imposed by placing bulky side groups as the peptide grows on the surface for example, are accounted for as described below. The details of AutoDock parameters are given in Appendix S1. Chiral carbon coordinates of the docked peptide are essential for further steps.

Partitioning the Path into Grids. Our aim is to design a peptide of n amino acids. For this purpose, we need n points –one for the center of each amino acid of the peptide- that is close to the sequence of residues on the binding surface. We choose these n points as the spatial coordinates of the chiral carbons of either the docked probe peptide or the interacting protein portion determined by HotPoint. Once this path of n points is defined, n contiguous grid boxes are constructed, each centered around one of the n points. The first grid contains the first amino acid. The first and the second grids along the path contain the first and the second amino acids. The t^{th} and $t+1^{st}$ grids contain the t^{th} and $t+1^{st}$ residues. The n chiral carbon atoms of the path define the centers of the n grid boxes.

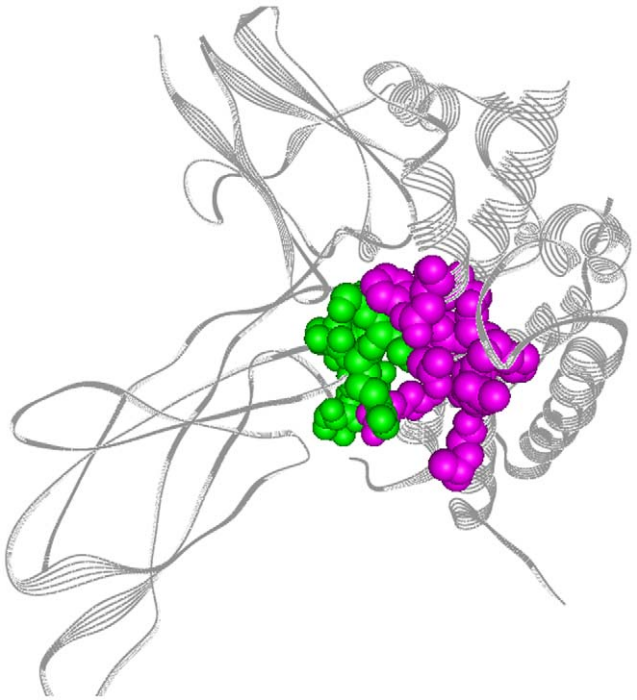


Figure 1. Selection of path by HotPoint server. HGHR (on the left) interacts with HGH (on the right): HotPoint predicts that the regions indicated with green and pink colors interact. The green-colored HGHR residues are selected as binding-sites. The pink-colored HGH residues are selected as binding peptide path.
doi:10.1371/journal.pone.0010926.g001

Docking of amino acids & dipeptides to the binding site. The AutoDock program [33] is used as the docking tool to quantify the binding affinity between the dipeptides and the selected protein surface. Python scripts are written to automate the docking, the probability calculation and the peptide sequence determination procedures. The binding affinity of a given peptide to the surface is determined via AutoDock which gives the affinity in terms of binding energy in kcal/mol.

All 20 amino acids are docked to the 1^{st} grid of the pre-defined path, such that their chiral carbon atoms are forced to coincide with the first grid center. All 400 possible dipeptides are docked to the first and second grid boxes, with the successive chiral carbons located at the grid centers. The pair wise docking of the dipeptides is continued in this way up to the last dipeptide along the path. Pair wise docking rests on the assumption of the Markov property. The amino acids and dipeptides are prepared by the HyperChem software [49]. The dipeptides have Ace-cap on their N-termini.

The grid map is determined by the ADT subprogram of AutoDock. The pre-determined spatial coordinates of the chiral carbon atoms on the path are used as the AutoDock grid box centers. The optimum grid box size is found by trial and error as 2.5 times the length of the distance between the first nitrogen and the last carbon along the backbone of the amino acids. The grids are set such that the chiral carbon atoms of the amino acid coincide with the grid box centers but with a freedom to rotate and translate in the box. This freedom, while keeping the dipeptides to be constrained to the neighborhood of the chosen grids is necessary to account for the side chain differences of the different amino acids. It also decreases the entropy penalty of constraining a peptide to a certain region. The parameters of AutoDock are given in Appendix S2.

Docking the dipeptide on a given grid pair leads to a set of 400 binding affinities, one set for each dipeptide, which are used to determine the probability of binding of each dipeptide to the protein binding site as explained in the following section.

Calculation of transition probabilities. The 20 types of amino acids and the peptide length n determine the number of states in our model; there exists 20n states for the problem. For the Markov model adopted here, one needs the transition probabilities, i.e., the probability of an amino acid occupying the $t+1^{st}$ position, given the amino acid at the t^{th} location. The binding energies of dipeptides obtained by AutoDock are used as the statistical weights for determining the transition probabilities. We adopt the Rotational Isomeric States (RIS) approach of polymer physics which is well suited for this purpose [34,50,51]. For each grid the probabilities are calculated via the formulation given below by *Eqs. 1*, *2*, *3*, and *4*. For the 1^{st} grid 20 binding energies are available, while for the remaining grid pairs 400 binding energies are available; the corresponding statistical weight matrix is given in *Eq.1* and *2*, respectively. The RIS matrix multiplication scheme [34] is used to determine probabilities from the energies.

The statistical weight matrix U_1 for the amino acid bound to the 1^{st} grid box is

$$U_1 = \exp(\beta E_{1;i}) \tag{1}$$

where $\beta = 1/kT$, k being the Boltzmann constant and T the temperature, *i* is any of the 20 amino acids; alanine, cysteine, aspartic

acid, glutamic acid, phenylalanine, glycine, histidine, isoleucine, lysine, leucine, methionine, asparagine, proline, glutamine, arginine, serine, threonine, valine, tryptophan, tyrosine.

The statistical weight matrix U_{t+1} for the dipeptide formed by the t[th] and t+1[st] residues along the peptide is

$$U_{t+1} = \exp\left(\beta E_{t,t+1;i,j}\right) \tag{2}$$

where t,$(1 \le t \le n-1)$, represents the amino acid position number (or the grid box number along the path). The i, j values represent any of the 20 amino acids.

The partition function, Z, of the peptide is obtained according to [34].

$$Z = J^* \prod_{t=2}^{n} U_t \, J \tag{3}$$

where $J^* = U_1$; J = column [1 1 1 . . . 1]. (It is to be noted that in the Flory notation, the J* matrix is given as $J* = \lfloor 1\,0\,0\,\ldots 0 \rfloor$ that would assign alanine to the first residue of peptide. In the present formulation, the choice of the J* matrix allows for the acknowledgement of all of the 20 amino acids to be the first residue).

The probability of having residue i at the t[th] position and residue j at the t+1[st] position is determined by:

$$p_{t,t+1;i,j} = \frac{J^* U_2 \ldots U_t U'_{t+1} \ldots U_n J}{Z} \tag{4}$$

Here, U'_{t+1} is the matrix obtained by equating all elements of U_{t+1} to zero except the ij[th]. The formula given above in *Eq. 4* is used to calculate the probabilities of transition states. Here, we keep the indices t and t+1, but they will be dropped in the application to the Viterbi algorithm for simplicity of representation with the understanding that each pair of sites has its own p_{ij}.

Determining the emission probabilities of the $\phi-\psi$ torsion angles. A fundamental requisite for favorable binding of a peptide to the surface is that the torsion angles of the peptide in the bound conformation should not be forced to have energetically unfavorable $\phi-\psi$ torsion angle values. The apriori probabilities of the torsion angles are needed for this purpose. In this section we explain the formation of these probabilities, called the emission probabilities, and their incorporation into the Viterbi algorithm as the observable variables.

Two sets of probabilities are needed for specifying the conformation of the peptide. The first set gives the probabilities of the torsion states determined by the angles $\phi_t-\psi_t$ of the residue at the t[th] grid. The second set gives the probabilities for the torsion angles $\psi_t-\phi_{t+1}$ of the dipeptide formed by residues at t[th] and t+1[st] grids. The representation of the defined torsion angles are shown in Figure 2. The $\phi_t-\psi_t$ torsion angles of the residue at grid t can select any of eleven regions, as explained below and shown in Figure 3. The torsion angles $\psi_t-\phi_{t+1}$ of the succeeding residue can

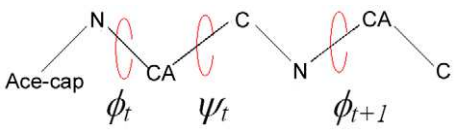


Figure 2. Schematic representation of torsion angles of a dipeptide. Only the backbone atoms of dipeptide are given and the acetyl cap on the N-terminal is shown as ACE, for simplicity.
doi:10.1371/journal.pone.0010926.g002

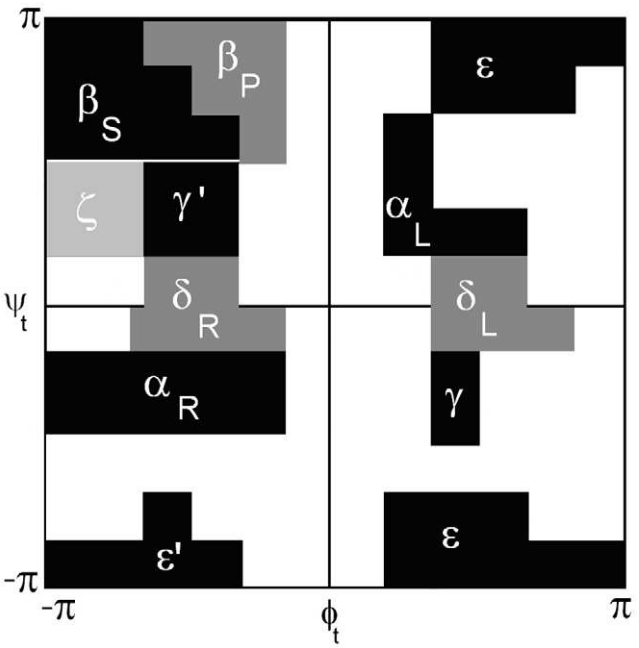


Figure 3. The representation of eleven states on Ramachandran map.
doi:10.1371/journal.pone.0010926.g003

choose the eleven torsion angles as explained below and shown in Figure 4. Each dipeptide is capped at its N-terminus by an acetyl group in order to define the ϕ_t angle.

The ϕ and ψ angles of a residue cannot adopt all values due to backbone intrinsic torsion propensities and attractive and repulsive interactions of atoms that are in close proximity for certain combinations of these two angles. Among the repulsive interactions, steric hindrances resulting from the side groups are the most pronounced. Hydrogen bonds are the most pronounced favorable interactions. Depending on the type of the residue, these angles show preferences for different regions on the Ramachandran map. The frequency of occurrence of these regions for the twenty amino acids can be obtained from the Protein Data Bank, PDB. An examination of the frequency of occurrence of the regions for neighboring units shows that there is strong dependence on residue type [50,51].

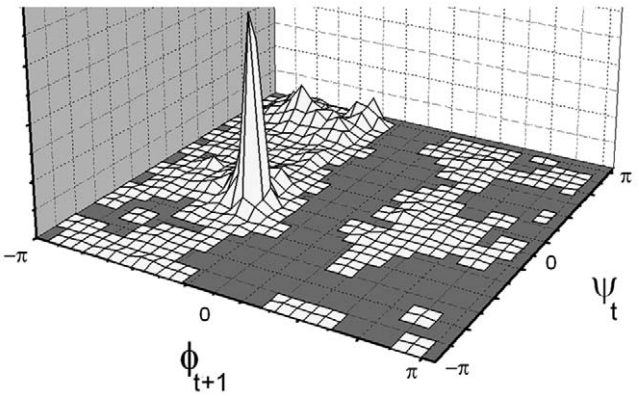


Figure 4. The probability distribution of $\Psi_t-\Phi_{t+1}$ angles on Ramachandran map.
doi:10.1371/journal.pone.0010926.g004

Plotting the $\phi_t-\psi_t$ angles on a Ramachandran map from PDB data, irrespective of residue type shows that there are eleven major isomeric torsion angle states. These are shown in Figure 3. We number the regions from 1 to eleven according to the notation of Reference [52] shown in Table 1.

In obtaining Figure 3, the non-redundant PDB Select database of native proteins was used. There are 197,458 points on the Ramachandran map obtained from the database. 96% of these points fell on the eleven regions shown in Figure 3. The remaining 4% of points were outside of these regions which are known to correspond to strained conformations of the residues resulting from long range perturbations. We ignored this set of 4%.

The probability of occurrence of a residue in any of the eleven regions given above depends on the type of the library used. Since we are interested in the denatured conformations of peptides, we calculated the probabilities over a coil library that we constructed. The coil library was downloaded from the website: http://www.roselab.jhu.edu/coil/. The library contains fragments selected by Dunbrack's PISCES server according to the following criteria: less than 20% sequence identity, better than 1.6 angstrom resolution, and a refinement factor of 0.25 or better [51].

A given residue may be of type i $(1\leq i\leq 20)$ at the tth grid, and may be in torsion-state m, which is one of the eleven regions shown in Figure 3. The frequency of occurrence of $\phi_t-\psi_t$ in torsion-states m for the i^{th} type of residue was collected in a two dimensional array, $f(i,m)$ which is the normalized frequency that an amino acid is of residue type i in torsion-state m. The singlet probability $b_{i,m}$ that the a residue along the peptide is of type i having the torsion-state m is

$$b_{i,m}=\frac{f(i,m)}{\sum_{m'} f(i,m')} \tag{5}$$

For a given i, $b_{i,m}$ is a column vector of eleven entries. In Figure 5, $b_{i,m}$ values are given for GLY as an example.

The choice of the torsion state of the tth residue places restrictions on the choice of the torsion state of residue t+1 due to the dependence of the torsion angle ϕ_{t+1} on ψ_t. The extent of this dependence for all pairs of residues in the coil library is depicted in Figure 4 where the joint probability distribution of $\psi_t-\phi_{t+1}$ is presented. For uniformity of representation, we keep the same eleven regions of $\psi_t-\phi_{t+1}$. In obtaining Figure 4, the non-redundant PDB Select database of native proteins was used. There are 955,679 points on the Ramachandran map obtained from the database. 91% of these points fell on the eleven regions shown in Figure 3. The remaining 9% of points were outside of these regions. We ignored this set of 9%. The probabilities are calculated over the coil library of http://www.roselab.jhu.edu/coil/.

Table 1. Notation for the eleven states.

State	Notation	Desrcription
1	ε'	Mirror image of the extended region ε
2	ε	The extended regions, $\varphi>0$, $\psi\sim -+180^{\circ}$
3	α_R	Right-handed alpha helix
4	γ	Tight turn region
5	δ_R	The right handed bridge region between two β-strands
6	δ_L	Mirror image of the δ_R region
7	ζ	Region observed mostly in residues preceding PRO
8	γ'	Inverse tight turn region
9	α_L	Mirror image of α_R
10	β_s	Extended beta sheet forming region
11	β_p	Region with extended polyproline-like helices

doi:10.1371/journal.pone.0010926.t001

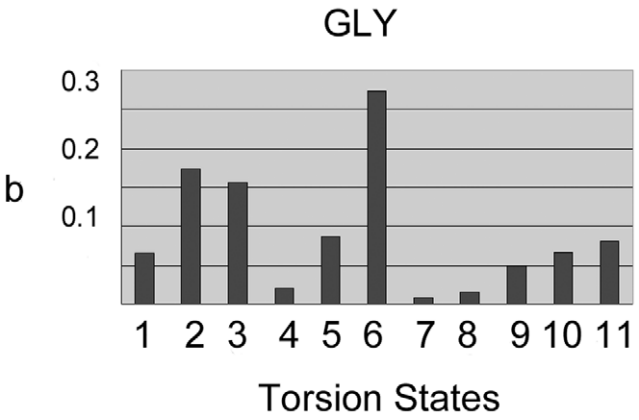


Figure 5. The probability distribution of GLY $\Phi_t-\Psi_t$ angles, derived from coil library. The dipeptide is mostly observed in the 6th state; and the least observation occurs for state 7.
doi:10.1371/journal.pone.0010926.g005

The probability $b_{i,j,m}$ that the residue j is in state m of $\psi_t-\phi_{t+1}$ space when the preceding residue is of type i is

$$b_{i,j,m}=\frac{f(i,j,m)}{\sum_{m'} f(i,j,m')} \tag{6}$$

For a given ij, $b_{i,j,m}$ is a column vector of eleven entries.

For each n grid box, the docked conformation of the dipeptides is determined by AutoDock. Knowing the conformation leads to the $\phi_t-\psi_t$ and $\psi_t-\phi_{t+1}$ angles. Thus, the torsion-state of the angles is determined. Equations 5 and 6 give the apriori probability of observing the torsion-states. The product of the two probabilities, $b_{i,m}$ and $b_{i,j,m}$, is called the emission probability for the torsion state of residue j when the preceding residue i is already prescribed. In the Viterbi algorithm below, this is indicated as $b_{i,j}$. The index m will be removed for simplicity.

Implementation of the Viterbi Algorithm

We follow the notation of Reference [39] in our application of the Viterbi algorithm.

We need the following definitions:

n: Number of residues of the peptide.

t: Grid number, $1\leq t\leq n$

m: Index identifying the torsion state, $1\leq m\leq 11$

$S=\{S_1,S_2,\ldots,S_{20}\}$ The 20 natural amino acids set.

$A=\{a_1,a_2,\ldots,a_{11}\}$: The eleven torsion angle regions.

q_t: The state of the tth grid. For example $q_t=S_i$ means that the state of the tth grid is the amino acid S_i.

a_t: The torsion-state of the amino acid at the tth grid.

$P=(p_{ij})$: The transition probability, $p_{ij}=\Pr(q_{t+1}=S_j|\ q_t=S_i)$.

$b=(b_{i,j})$: The emission probability, $b_{i,j}=\Pr(a_{t+1}=A_j|\ a_t=A_i)$

$\pi=(\pi_i)$: The initial distribution vector, where $\pi_i=\Pr(q_1=S_i)$.

The sets S and A describe the structure of the model, P, b and π describe the parameters.

Our intention is to determine $\arg\max_S \Pr(q_1,q_2,\ldots,q_n|a_1,a_2,\ldots,a_n)$, meaning a peptide sequence made up of preferable torsion angles and with high affinity to protein binding-site. The algorithm is divided into two parts. Firstly, in the forward tracking step, the algorithm finds $\max \Pr(q_1,q_2,\ldots,q_n|a_1,a_2,\ldots,a_n)$. The forward tracking employs emission and transition probabilities. Then the algorithm backtracks to determine an $\alpha_1,\alpha_2,\ldots,\alpha_n$ that realizes this maximum.

For an arbitrary position t and amino acid type i:

$\delta_t(i)$: The maximum probability of all ways to end in state S_i at grid t and have observed the torsion states $a_1,a_2,\ldots,a_t$.

$$\delta_t(i)=\max_{q_1,q_2\cdots q_{t-1}} \Pr(q_1,q_2,\ldots,q_{t-1},q_t=S_i \text{ and } a_1,a_2,\ldots,a_t) \quad (7)$$

where

$$\delta_1(i)=\Pr(q_1=S_1 \text{ and } a_1) \quad (8)$$

then,

$$\max_{q_1,q_2,\ldots,q_n} \Pr(q_1,q_2,\ldots,q_n \text{ and } a_1,a_2,\ldots,a_n) \quad (9)$$

is determined.

Initialization step.

$$\delta_1(i)=\pi_i b_i(\alpha_1),\ 1\le i\le 20 \quad (10)$$

δ_1 is 1×20 array, keeping the binding affinity probabilities of each amino acid for the 1st grid box. π_i is the initial probability of binding of any 20 amino acids to the 1st grid box and is a 1×20 array. All entries of π_i are chosen as unity to give equal chance for selection of any 20 amino acids as the initial residue of the peptide. The binding affinity and the corresponding torsion states are determined by AutoDock. The choice of the first residue will contain the information of its torsion angles from AutoDock, leading to the knowledge of the emission probability $b_i(\alpha_1)$ which is accounted for in *Eq. 10*.

Induction step.

$$\delta_{t+1}(j)=\max_{1\le i\le 20} \delta_t(i)p_{ij}b_{i,j}(a_{t+1}|a_t),\ 1\le t\le n-1, 1\le i,j\le 20 \quad (11)$$

δ_{t+1} is 1×20 array, keeping the binding affinity and torsion angle preference probabilities of each dipeptide for t+1st grid.

Backtracking. For each grid box, the maximum probabilities are kept in $\delta_t(i)$ arrays, as defined in the previous section by *Eq. 11*. Backtracking of those arrays leads to the determination of a peptide sequence with a possible affinity to protein binding-site.

Initially, the last residue of the peptide sequence is determined from the 20 entries of δ_n, the array of the last grid box. The maximum entry of δ_n is selected, which determines the amino acid residue q_n.

We let

$$\Xi_n=\arg\max_{1\le i\le 20} \delta_n(i) \quad (12)$$

and choose $q_n=S_{\Xi_n}$. Thus, q_n is the final state of the last residue of the peptide.

The remaining q_i, that is the amino acid types, for $1\le t\le n-1$ are found recursively by determining:

$$\Xi_t=\arg\max_{1\le i\le 20} \delta_t(i)p_{i\Xi_{t+1}} \quad (13)$$

and then putting $q_t=S_{\Xi_t}$. The backtracking method leads to a peptide sequence with high affinity to the target protein surface.

Quantifying the peptide – Target protein Interaction. The AutoDock software is used for validating the accuracy of the solutions. The tertiary structure of the designed peptide is prepared using HyperChem. The details of the procedures are given in Appendix S2. The binding affinity calculation between the target protein and the designed peptide was carried out by AutoDock.

Test of the Algorithm

Protein-tripeptide case studies. As a proof of concept, the Viterbi algorithm is tested on five known protein-tripeptide complexes. The aim is to demonstrate the feasibility and the reliability of the algorithm.

The binding energy between the target protein and the peptide designed by Viterbi are determined by the AutoDock program. The inhibition constant K_i is also calculated by AutoDock. The visualization of complexes is achieved by Accelerys Discovery Studio 2.5 program [53].

The first complex is HIV-1 protease interacting with the tripeptide Glu-Asp-Leu. The PDB accession number of this complex is **1A30** [54]. This tripeptide is the smallest analogue of HIV-1 transframe octapeptide (TFP) Phe-Leu-Arg-Glu-Asp-Leu-Ala-Phe, which is known to be the most potent inhibitor of the target protein. The inhibition of the protein with this peptide is selective and specific.

The second complex is scytalidocarboxyl peptidase B protein interacting with the tripeptide Ala-Ile-His (**1S2K**) [55]. The protein is pepstatin-insensitive carboxyl peptidase from the organism *Scytalidium lignicolum*. The crystal structure contains the protein and the cleaved angiotensin II peptide: Ala-Ile-His. The tripeptide is bound to the catalytic residues *Gln-53* and *Glu-136* of the protein.

The other complex chosen is the signaling protein from goat mammary gland (SPG-40) and the tripeptide Trp-Pro-Trp (**1ZBW**) [56]. The protein plays role in signaling for reductive remodeling of mammary gland. The remodeling is necessary after cessation of lactation in female mammals. The protein is known to interact with oligosaccharides; the binding enhances protein-protein interactions. The tripeptide Trp-Pro-Trp sits at the active site, to which oligosaccharides bind.

The peptide deformylase (PDF) from *Enterococcus faecium* organism with Met-Ala-Ser tripeptide is another selected complex (**3G6N**), to which the peptide design algorithm is applied [57]. *E. faecium* are found in normal flora of the intestinal track, but the Vancomycin (an antibiotic) resistant types of bacteria cause infection commonly in hospitals and the strain is resistant to all commercially available antibiotics. The PDF protein is essential for bacterial growth, making it a potential drug target. The Met-Ala-Ser motif is recognized by the active site of the protein.

The last test case for tripeptides is concanavalin A (Con A) protein with Tyr-Pro-Tyr peptide (**1HQW**) [58,59]. The peptide is determined to the best binding part out of $\sim1.4\times10^9$ octapeptides. A highly diverse phage library procedure is used to determine this

Table 2. Binding energy; K_i values of HIV-1 protease peptides.

Peptide	Binding Energy (kcal/mol)	K_i
Glu-Asp-Leu	−7.66	2.43 μM
Trp-Tyr-Val	−9.59	94.12 nM

doi:10.1371/journal.pone.0010926.t002

sequence. Con A is a carbohydrate-binding protein; the literature indicates the importance of carbohydrate binding in various biological processes. Consequently, inhibition of carbohydrate-specific proteins is important for novel drug development. The chemical synthesis of oligosaccharides is a sophisticated procedure since protection of sugar monomers is complicated; peptides can be used as ligands for such protein targets. The Tyr-Pro-Tyr peptide is shown to inhibit binding of known monosaccharide ligands to Con A.

Protein-heptapeptide case studies. In the second step of our test, we applied the method to the design of heptapeptides against Proteinase K, and HLA-B*2705 proteins. Both of these proteins have known peptide ligands in the literature.

Proteinase K is a serine protease with broad specificity. The PDB accession number of the inhibitor peptide Pro-Ala-Pro-Phe-Ala-Ala-Ala in complex with the protein is **1P7V** from the organism *Engyodontium album*. The article about the crystal structure and interaction details of this protein-peptide complex has not yet been published.

HLA-B*2705 is a disease-associated human MHC class I allele HLA-B27 subtype protein. The protein is the target for nonapeptides. The self peptide sequence of this protein is Arg-Arg-Lys-Trp-Arg-Arg-Trp-His-Leu. The copy number of this peptide in ankylosing spondylitis patients is observed to increase [60]. Viral peptide Arg-Arg-Arg-Trp-Arg-Arg-Leu-Thr-Val derived from Epstein-Barr virus membrane has shown to have affinity to HLA-B*2705 [61]. The glucagon receptor-derived peptide Arg-Arg-Arg-Trp-His-Arg-Trp-Arg-Leu has also proven to interact with HLA-B*2705 [62]. The PDB accession codes of those 3 peptides in complex with HLA-B*2705 are **1OGT**, **1UXS**, **2A83**.

Predicting a peptide for a protein with no known peptide ligand. The Human Growth Hormone (HGH) is responsible for linear growth in vertebrates via stimulation of skeletal and visceral growth. The protein also plays a role in carbohydrate metabolism and fat mobilization from tissues [63]. The PDB accession code of the protein is **1HGU** [64]. A crystal structure of the protein-peptide complex is not available in PDB. Consequently, the most possible binding site of the protein is determined by GNM.

Results

HIV-1 protease peptide

The Viterbi program designed the Trp-Tyr-Val tripeptide with high binding affinity for the HIV-1 protease. The binding affinity was calculated as −9.59 kcal/mol by AutoDock; the K_i value was 94.12 nanomolar. The binding affinity of the known inhibitor Glu-Asp-Leu was determined by AutoDock as −7.66 kcal/mol; the K_i value was 2.43 micromolar. The affinity terms are summarized in the Table 2. The binding region of the peptides is shown in Figure 6. Figure 6A indicates the complex formed by the HIV-1 protease with Glu-Asp-Leu, while Figure 6B indicates the complex formed by the HIV-1 protease with Trp-Tyr-Val. As the figures imply, both of the peptides bind to the same active-site on the protein.

The HIV-1 protease is formed by 2 identical chains: chain A and chain B. Glu-Asp-Leu interacts with the chain A residues Asp-25, Gly-27 Ala-28, Asp-29, Asp-30, Met-46, Gly-48; and with the chain B residues Arg-8, Asp-25, Val-82. The peptide makes five Hydrogen bonds with Gly-27, Asp-29, Asp-30 and Gly-48 shown in Figure 7A. A salt bridge is present between Asp-29 of the chain A and Glu residue of the peptide. The sulfide atom of Met-46 from the chain A makes a bond with the oxygen atom of Asp.

Trp-Tyr-Val interacts with the chain A residues Asp-25, Gly-27, Ala-28, Asp-29, Asp-30, Gly-48, Pro-81, Val-82, Ile-84; and with the chain B residues Asp-25, Gly-27, Ala-28, Ile-50, Thr-80, Pro-81, Val-82, Ile-84. The peptide makes nine Hydrogen bonds with Asp-25 of the both chains, Ile-50, Thr-80 and Val-82 as shown in Figure 7B. There are 2 salt bridges between Asp-25 of the chain A and the Trp residue of the peptide. Another 2 salt bridges are present between Asp-25 of the chain B and Trp residue of the peptide.

Scytalidocarboxyl peptidase B peptide

The Viterbi program designed the tripeptide Arg-Arg-Arg as potential binder peptide for the scytalidocarboxyl peptidase B protein. The binding affinity of this peptide was calculated as −12.96 kcal/mol; the K_i value was 315.78 picomolar. The known inhibitor Ala-Ile-His binding affinity for the target protein was determined as −5.33 kcal/mol; the K_i value was 124.38

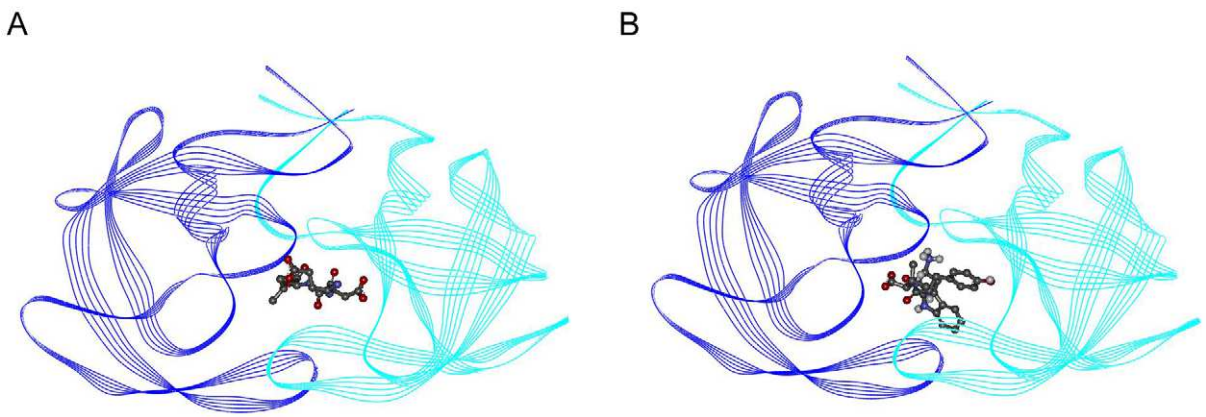


Figure 6. HIV-1 protease peptide complexes. (A) Glu-Asp-Leu and HIV-1 protease. (B) Trp-Tyr-Val and HIV-1 protease.
doi:10.1371/journal.pone.0010926.g006

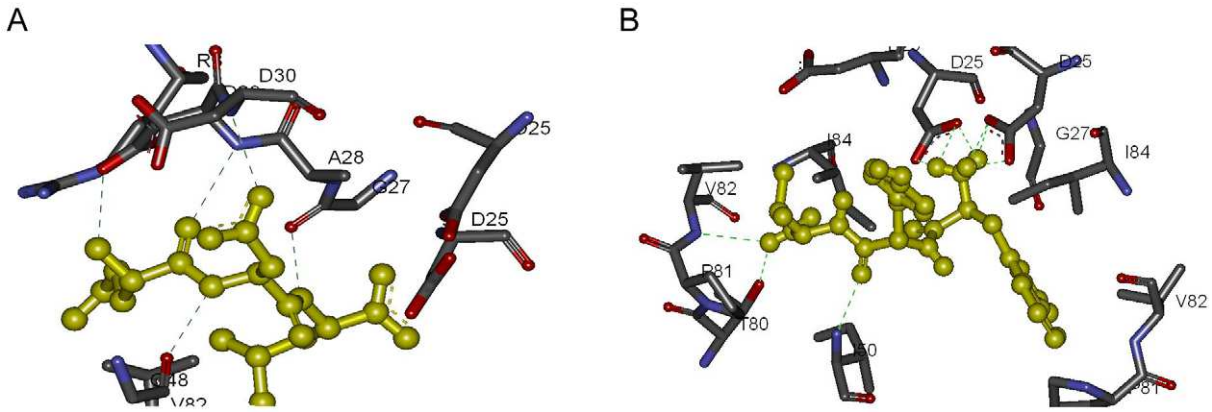


Figure 7. Detailed analysis of HIV-protease peptide complexes. Hydrogen bonds are indicated with green lines. (A) Glu-Asp-Leu and HIV-1 protease. (B) Trp-Tyr-Val and HIV-1 protease.
doi:10.1371/journal.pone.0010926.g007

micromolar. The affinity terms are summarized in Table 3. The binding region of the peptides is shown in Figure 8; Figure 8A indicates the complex formed by the scytalidocarboxyl peptidase B with Ala-Ile-His, while Figure 8B indicates the complex formed by the scytalidocarboxyl peptidase B with Arg-Arg-Arg. As the figures imply, both of the peptides bind to the same active-site.

Ala-Ile-His interacts with the residues Gln-53, Asp-57, Trp-67, Glu-136, Phe-138, Glu-139, Glu-140, Cys-141, and Cys-148. The peptide makes two Hydrogen bonds with Glu-139 as shown in Figure 9A. $\pi-\pi$ stacking between Trp-67 and His residues of the peptide is present. There is a salt bridge between Glu-139 and Ala residue of the peptide. Sulfide atom and aromatic ring interaction is present between residues Cys-141 and His; Cys-148 and His. S-O bonding is present between the residues: Cys-141 and Ala; Cys-148 and Ala.

Arg-Arg-Arg interacts with the residues Trp-6, Trp-39, Gln-53, Asp-57, Tyr-59, Asp-65, Trp-67, Glu-69, Glu-73, Glu-136, Phe-138, Glu-139, Glu-140, and Cys-141. The peptide makes fifteen Hydrogen bonds with the residues Asp-57, Tyr-59, Glu-69, Glu-73, Glu-136, Glu-139, and Glu-140 as indicated in Figure 9B. π-cation interactions are observed between *P*he-138 and Arg-1 residue of the peptide; Trp-39 and Arg-3 residues. There are salt bridges Glu-69 and Arg-1; Glu-140 and Arg-1; Glu-136 and Arg-3; Asp-65 and Arg-3; also an internal salt bridge between Arg-1 and Arg-3 residues of the peptide. Arg-1 interacts with Cys-141 through S-O bonding.

SPG-40 peptide

The Viterbi program designed the tripeptide Trp-Tyr-Tyr as the sequence with possible binding affinity to SPG-40. The binding affinity of this peptide was calculated as -10.97 kcal/mol; the K_i value was 9.04 nanomolar. The known inhibitor Trp-Pro-Trp binding affinity for the target protein was determined by AutoDock as -9.10 kcal/mol; the K_i value was 215.27 nanomolar. The affinity terms are summarized in Table 4. The binding region of the peptides is shown in Figure 10. Figure 10A indicates the complex formed by SPG-40 with Trp-Pro-Trp, while Figure 10B indicates the complex formed by SPG-40 with Trp-Tyr-Tyr. As the figures imply, both of the peptides bind to the active-site of the protein. Trp-Pro-Trp interacts with the residues Trp-10, Arg-14, Asn-79, Thr-267, and Glu-269. The peptide makes two Hydrogen bonds with *Asn-79* shown in Figure 11A. There is a salt bridge between Glu-139 and Ala residue of the peptide.

Trp-Tyr-Tyr interacts with the residues Trp-10, Arg-14, Cys-20, Phe-37, Trp-78, Asn-79, Asp-186, Arg-242, Glu-269, Ile-272, Trp-331, Asp-334, and Leu-335. The peptide makes six Hydrogen bonds with the residues Arg-14, Arg-242, and Glu-269 shown in Figure 11B. $\pi-\pi$ stacking is observed between Trp-10 and Trp-1 residue of the peptide and Phe-37 and Trp-1 residues. π- cation interactions are observed between Trp-78 and Tyr-2; Trp-10 and Trp-1; also between Tyr-2 and Trp-1 residues of the peptides; Tyr-2 and Tyr-3 residues of the peptides. There are salt bridges between Glu-269 and Trp-1. There exists aromatic ring and sulfide interaction between Cys-20 and Tyr-2.

Table 3. Binding energy; K_i values of scytalidocarboxyl peptidase B peptides.

Peptide	Binding Energy (kcal/mol)	K_i
Ala-Ile-His	−5.33	124.38 μM
Arg-Arg-Arg	−12.96	315.78 pM

doi:10.1371/journal.pone.0010926.t003

PDF peptide

The Viterbi program designed Val-Trp-Trp as peptide with possible binding affinity to PDF. The binding affinity of this peptide was calculated as -9.52 kcal/mol and the K_i value was 105.71 nanomolar. The known inhibitor Met-Ala-Ser binding affinity for the target protein was determined as -8.07 kcal/mol. The K_i value was 1.21 micromolar. The affinity terms are summarized in *Table 5*. The binding region of the peptides is shown in Figure 12. Figure 12A indicates the complex formed by the PDF with Met-Ala-Ser, while Figure 12B indicates the complex formed by the PDF with Val-Trp-Trp. As the figures imply, both peptides bind to the active-site on the protein. The Fe^{+2} ion in the active site is shown with CPK representation.

Met-Ala-Ser interacts with the residues Gly-57, Val-59, Gly-60, His-76, Gly-113, Leu-115, Tyr-150, His-157, His-161, and Phe-167. The peptide makes no Hydrogen bonds with the protein, Figure 13A. A sulfide atom and aromatic ring interaction occurs between the residues His-76 and Met; Tyr-150 and Met; His-157 and Met; His-161 and Met; Phe-167 and Met.

Val-Trp-Trp interacts with the residues Met-4, Gln-45, Gly-57, Gly-58, Gly-60, Leu-108, Glu-110, Gly-111, Glu-112, Gly-113,

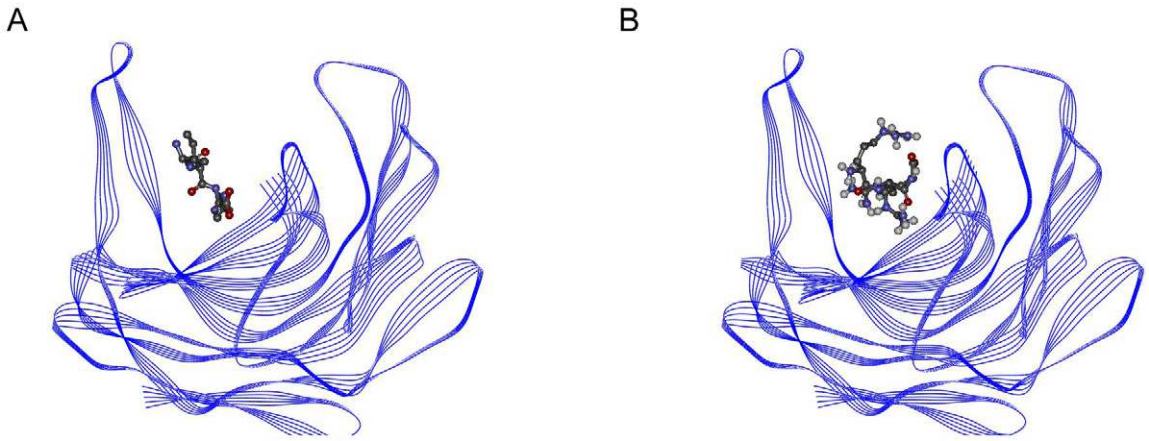


Figure 8. Scytalidocarboxyl peptidase B peptide complexes. (A) Ala-Ile-His and scytalidocarboxyl peptidase B. (B) Arg-Arg-Arg and scytalidocarboxyl peptidase B.
doi:10.1371/journal.pone.0010926.g008

Cys-114, Leu-115, Tyr-150, His-157, Glu-158, Met-166, and Glu-187. The peptide makes four Hydrogen bonds with Gly-111, Gly-113 and Tyr-150 as shown in Figure 13B. There are internal salt bridges between Val-1 and Trp-3 residues of the peptide. There is an aromatic ring and sulfide atom interaction between the residue pairs Cys-114 and Trp-3, Met-4 and Trp-3, Met-166 and Trp-3.

Con A peptide

The Viterbi program designed the tripeptide Gly-Ala-Tyr for Con A. The binding affinity of this peptide was calculated as −5.70 kcal/mol; the K_i value was 65.83 micromolar. The known inhibitor Tyr-Pro-Tyr binding affinity for the target protein was determined as −5.70 kcal/mol. The K_i value was 144.38 micromolar. The affinity terms are summarized in *Table 6*. The binding region of the peptides is shown in *Figure 14*; *Figure 14a* indicates the complex formed by the Con A with Tyr-Pro-Tyr, while *Figure 14b* indicates the complex formed by the Con A with Gly-Ala-Tyr. As the figures imply, both peptides bind to the active-site of the protein. Tyr-Pro-Tyr interacts with the residues Thr-15, Ser-21. The peptide makes one Hydrogen bond with Thr-15.

Gly-Ala-Tyr interacts with the residues Thr-11, Tyr-12, Pro-13, Thr-15, Asp-16, His-205, Pro-206, and Arg-228. The peptide makes seven Hydrogen bonds with Thr-11, Tyr-12, Asp-16, Pro-206 and Arg-228. $\pi-\pi$ stacking is probable between Tyr-12 and Tyr residues.

Proteinase K peptide

The Viterbi program designed the Trp-Tyr-Tyr-Tyr-Tyr-Tyr-Tyr heptapeptide with possible binding affinity to Proteinase K. The binding affinity of this peptide was calculated as −11.59 kcal/mol. The K_i value was 3.22 nanomolar. The known inhibitor Pro-Ala-Pro-Phe-Ala-Ala-Ala binding affinity for the target protein was determined as −8.52 kcal/mol; the K_i value was 569.86 nanomolar. The affinity terms are summarized in Table 7. Both peptides bind to the active-site on the protein.

Pro-Ala-Pro-Phe-Ala-Ala-Ala interacts with the residues Asn-67, His-69, Asn-99, Gly-100, Tyr-104, Leu-133, Gly-134, Gly-135, Gly-136, Ala-158, Gly-160, Asn-161, Asn-162, Trp-212, Ile-220, Ser-221, Thr-223, Ser-224, and Met-225. The peptide makes one Hydrogen bond with Gly-102. There exist a pi-cation interaction between Tyr-104 and Pro-1. A sulfide atom and aromatic ring interaction is observed between Cys-73, Met-225 residues and Phe-4. The sulfide atom oxygen interaction is present between Cys-73 and Ala-7. The stacking of ring structures is observed for His-69 and Phe-4.

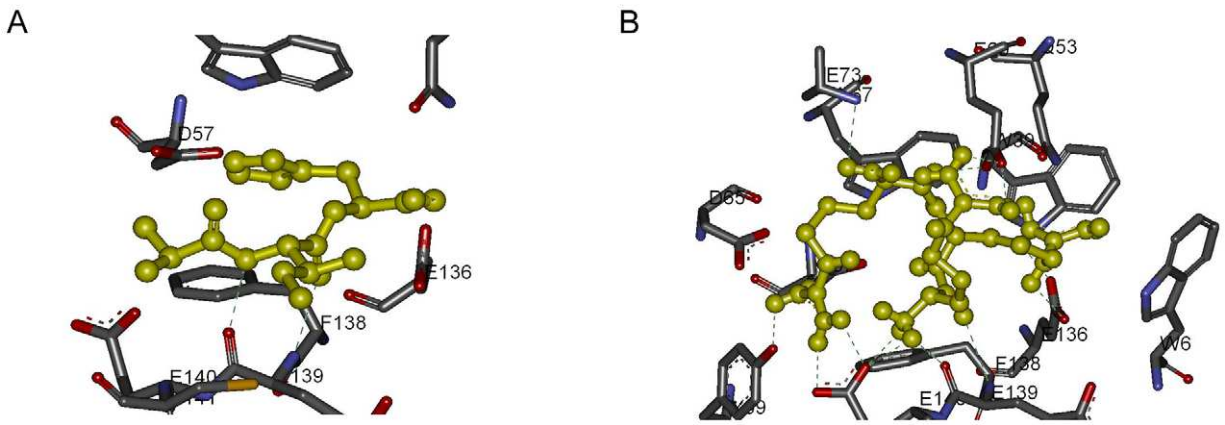


Figure 9. Detailed analysis of scytalidocarboxyl peptidase B peptide complexes. Hydrogen bonds are indicated with green lines. (A) Ala-Ile-His and scytalidocarboxyl peptidase B. (B) Arg-Arg-Arg and scytalidocarboxyl peptidase B.
doi:10.1371/journal.pone.0010926.g009

Table 4. Binding energy; K_i values of SPG-40 peptides.

Peptide	Binding Energy (kcal/mol)	K_i
Trp-Pro-Trp	−9.10	215.27 nM
Trp-Tyr-Tyr	−10.97	9.04 nM

doi:10.1371/journal.pone.0010926.t004

Trp-Tyr-Tyr-Tyr-Tyr-Tyr-Tyr interacts with the residues Asn-67, His-69, Asn-99, Ser-101, Gly-102, Gln-103, Tyr-104, Leu-133, Gly-134, Ala-158, Gly-160, Asn-161, Asn-162, Tyr-169, Ser-170, Ala-172, Trp-212, Ile-220, and Ser-224. The peptide makes three Hydrogen bonds with Gln-103, Ser-170 and Asn-161. A sulfide atom and aromatic ring interaction is observed between Cys-73, Met-225 residues and Trp-1, Tyr-2, Tyr3 residues of the peptide. $\pi-\pi$ stacking is observed between His-69, Trp-212 and Trp-1; His-69 and Tyr-2; Tyr-169 and Tyr-3; Phe-192 and Tyr-4; Tyr-104 and Tyr-6; also between Tyr-5 and Tyr-6 of the peptide. A pi-cation interaction is present between His-69 and Trp-1.

HLA-B*2705 peptide

The Viterbi program designed the Trp-Arg-Trp-Trp-Lys-Tyr-Tyr heptapeptide for HLA-B*2705. The binding affinity of this peptide was calculated as −8.97 kcal/mol; the K_i value was 265.82 nanomolar. The known inhibitors Lys-Trp-Arg-Arg-Trp-His-Leu, Arg-Trp-His-Arg-Trp-Arg-Leu, Arg-Trp-Arg-Arg-Leu-Thr-Val binding affinity for the target protein were determined by AutoDock as −6.84, −7.21, and −8.52 kcal/mol, respectively. The affinity terms are summarized in Table 8. Both peptides bind to the active-site on the protein.

Lys-Trp-Arg-Arg-Trp-His-Leu interacts with residues Arg-62, Ile-66, Lys-70, Thr-73, Asp-77, Tyr-99, His-114, Lys-146, Trp-147, Val-152, Gln-155, Lue-156, and Tyr-159. The peptide makes nine Hydrogen bonds with Ile-66, Lys-70, Asp-77, Tyr-84, Tyr-99, Thr-143, Trp-147, and Gln-155. There exist pi-cation interactions between Lys-146 and His-6; Lys-1 and Trp-2; Tyr-159 and Lys-1; and Tyr-99 and Lys-1. A salt-bridge is present between Asp-77 and Arg-3. A sulfide atom and aromatic ring interaction is observed between Cys-164, Met-5 residues and Trp-2. Stacking of ring structures is observed for Tyr-99 and Trp-2; Tyr-159 and Trp-2; Trp-2 and Trp- 7 in the peptide; Trp-147 and Trp-5; Trp-133 and Trp-5; Trp-147 and His-6.

Trp-Arg-Trp-Trp-Lys-Tyr-Tyr interacts with the residues Ala-69, Thr-73, Glu-76, Thr-80, Arg-83, His-114, Lys-146, Trp-147, Ala-150, Val-152, Gln-155 and Leu-156. The peptide makes three Hydrogen bonds with Thr-80, Arg-83 and Gln-155. A sulfide atom and aromatic ring interaction is observed between Cys-67 and Trp-4 residue of the peptide. $\pi-\pi$ stacking is observed between Trp-147 and Trp-1; His-114 and Trp-3; Trp-147 and Trp-3; Tyr-99 and Trp-3; Trp-133 and Trp-3; Trp-147 and Tyr-6.

HGH peptide

The Viterbi program designed Trp-Glu-Leu-Met-Phe-Phe-Tyr heptapeptide for HGH. The binding affinity of this peptide was calculated as −8.05 kcal/mol; the K_i value was 1.25 micromolar. The affinity terms are summarized in Table 9.

Trp-Glu-Leu-Met-Phe-Phe-Tyr interacts with the residues Met-14, His-21, Gln-22, Phe-25, Arg-64, Glu-65, Gln-66, Thr-175, Arg-178, Cys-182, and Cys-189. The peptide makes five Hydrogen bonds with Arg-64 and Arg-178. There exist a pi-cation interaction between Arg-178 and Tyr-7. A sulfide atom and aromatic ring interaction is observed between Met-170 and Trp-1; His-18 and Met-4; Met-14, Cys-182, Cys-189 and Phe-5; Met-14, Cys-182, Cys-189 and Phe-6; Cys-182, Cys-189, Met-14 and Tyr-7. Sulfur – oxygen interactions exist between Ser-188 and Met-4; Cys-189 and Met-4; Cys-182, Cys-189 and Phe5; Cys-182, Cys-189 and Phe-6; Cys-182 and Tyr-7. Stacking of ring structures is observed for His-21 and Trp1; Phe-25 and Trp-1; His-18 and Trp-1; His-18 and Phe-5.

Discussion

The Viterbi algorithm is successful in predicting tripeptides to the five proteins. The binding affinities of the designed tripeptides are all superior to the binding affinities of their known tripeptide ligands. The comparison is made by using the affinities given by the AutoDock. The method is also successful in predicting heptapeptides to the two proteins, proteinase K and HLA*B2705. The method was able to determine a better potential binding peptide for these two proteins.

The ability of our algorithm for *de novo* peptide design is proven for the HGH protein case-study. The binding affinity of the peptide is comparably better than some known peptide inhibitor affinities for their own target proteins.

The binding surfaces of all target proteins, except Con A, have both hydrophilic and hydrophobic residues. The Con A binding

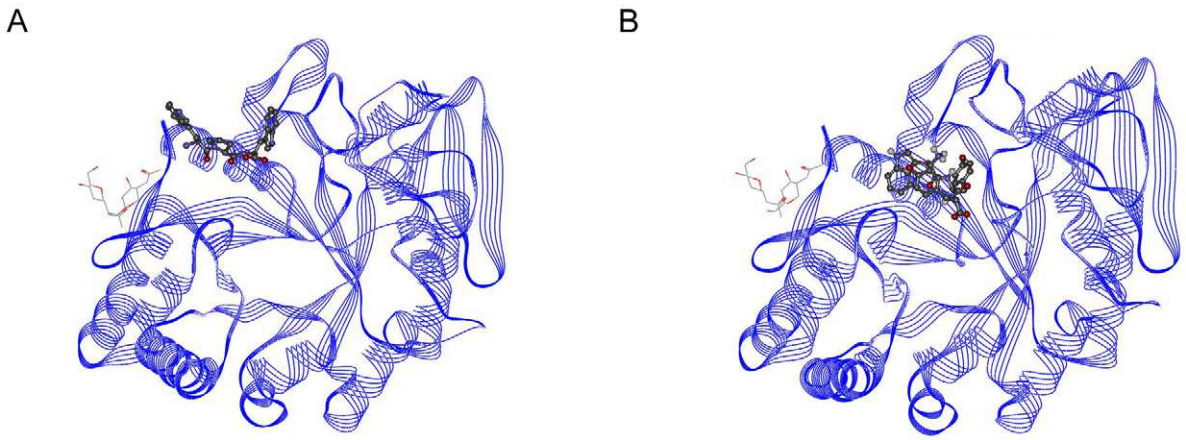


Figure 10. SPG-40 peptide complexes. (A) Trp-Pro-Trp and SPG-40. (B) Trp-Tyr-Tyr and SPG-40.
doi:10.1371/journal.pone.0010926.g010

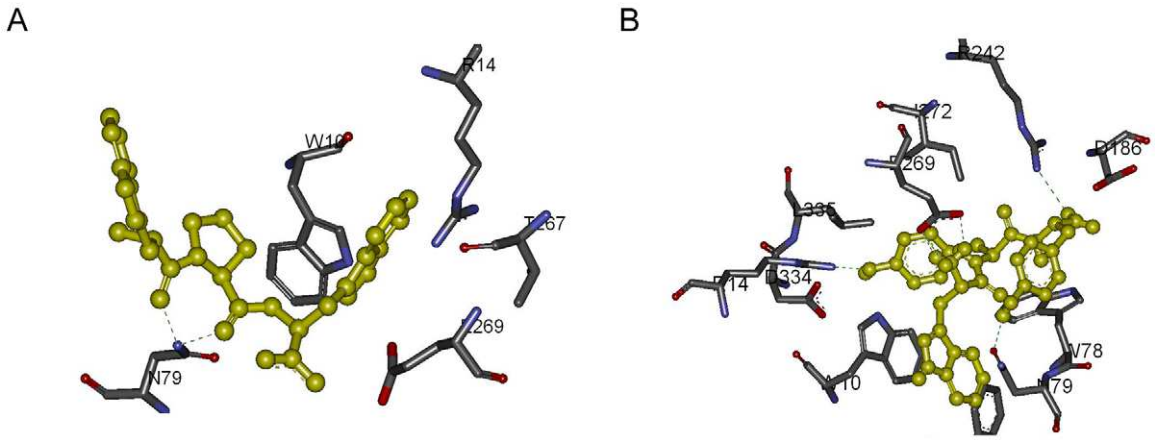


Figure 11. Detailed analysis of SPG-40 and peptide complexes. Hydrogen bonds are indicated with green lines. (A) Trp-Pro-Trp and SPG-40. (B) Trp-Tyr-Tyr and SPG-40.
doi:10.1371/journal.pone.0010926.g011

surface is made up of only hydrophilic residues. The results imply that a binding surface with both hydrophilic and hydrophobic residues could lead to a potential binder peptide design by our method. A surface exposed to water, with all hydrophilic residues, may not lead to very potent peptide designed by the Viterbi algorithm. This may be due to low number of residues in each grid box, since there is no cavity on the Con A surface. All other protein targets described in this paper have specific cavities as the binding surfaces. Consequently, the number of protein residues interacting with amino acids/dipeptides is low for the binding surfaces that are highly exposed to solvent. Balanced number of hydrophilic-hydrophobic residues in a grid box leads to more specific interactions. The specific interactions lead to design of a specific peptide with affinity to the selected protein surface.

The HIV-1 protease binding surface has −4 net charges, 50% of residues that form the surface is hydrophobic; the Trp-Tyr-Val peptide designed for this surface has two hydrophobic and one hydrophilic residues. The hydrophobic residue number is in harmony with the number of hydrophobic residues of the protein (chain A: Ile-50, Pro-81, Val-82, Ile-84; chain B: Pro-81, Val-82, Ile-84). The hydrophilic residue Tyr, which sits in the middle of the peptide, is in close proximity to the polar and basic residues of the protein surface. The known peptide of this protein has one hydrophobic and two basic residues; the only hydrophobic residue is close to Val-82 of chain B. The known peptide Glu-Asp-Leu makes interactions with only the hydrophilic and the charged residues; while the designed peptide also interacts with the hydrophobic residues of the binding pocket. Although the number of the intermolecular hydrogen bonds and the salt-bridges are the same for the protein-known peptide and the protein-designed peptide; the electrostatic compatibility between the protein and the designed peptide is more appropriate.

The Scytalidocarboxyl peptidase B binding surface has −4 net charges, 25% of residues is hydrophobic; the designed peptide Arg-Arg-Arg has +3 net charges. There are seven basic residues in the binding pocket; so the designed peptide stabilizes itself by electrostatic complementarity and the salt bridge formation. The known peptide Ala-Ile-His has two hydrophobic and one acidic residues; this peptide can interact with only 50% of the basic residues of the binding region. The designed peptide size is larger than that of the Ala-Ile-His peptide and the designed peptide also contains more positive charge. The tripeptide Arg-Arg-Arg makes a fine interaction with protein surface. The designed peptide makes three times more hydrogen bonds with the protein, when compared to the known peptide-protein interaction. The π interactions and the sulfur atom bonds are observed in both the protein-known peptide and the protein-designed peptide systems. The electrostatic compatibility between the protein and the designed peptide is more appropriate.

Table 5. Binding energy; K_i values of PDF peptides.

Peptide	Binding Energy (kcal/mol)	K_i
Met-Ala-Ser	−8.07	1.21 μM
Val-Trp-Trp	−9.52	105.71 nM

doi:10.1371/journal.pone.0010926.t005

The SPG-40 protein binding region has no net charge, since it contains one acidic and one basic residues. 20% of residues is hydrophobic; the designed Trp-Tyr-Tyr peptide has two hydrophilic and one hydrophobic residues. The only hydrophobic residue Trp is surrounded by six hydrophobic residues; four of those hydrophobic amino acids have aromatic ring in their side chains enabling the π interactions. Our method designed a peptide with two Tyr residues, which are similar to saccharide monomers. Also Trp residue of the designed peptide has aromatic ring structure that is also observed in saccharides. The similarity is important since the binding molecule of the target protein is oligosaccharides. The known peptide is made up of all hydrophobic residues. The designed peptide makes more hydrogen bonds with the protein, compared to the known peptide. The π interactions and the sulfur atom bonds are observed only in the protein-designed peptide.

The peptide deformylase binding surface has no charge, 33% of residues is hydrophobic; the designed Val-Trp-Trp peptide is made up of all hydrophobic residues. The known peptide consists of two hydrophobic and one hydrophilic residues. The known peptide makes five sulfide – aromatic ring interactions; while the designed peptide makes five sulfide – aromatic ring interactions and four Hydrogen bonds.

The Con A binding surface is exposed to water with no charge; the peptide Gly-Ala-Tyr has two hydrophilic and one small hydrophobic residues. The known peptide is also made up of two hydrophilic and one hydrophobic residues. Both of the peptides have their hydrophilic residue as the 2nd amino acid. The designed

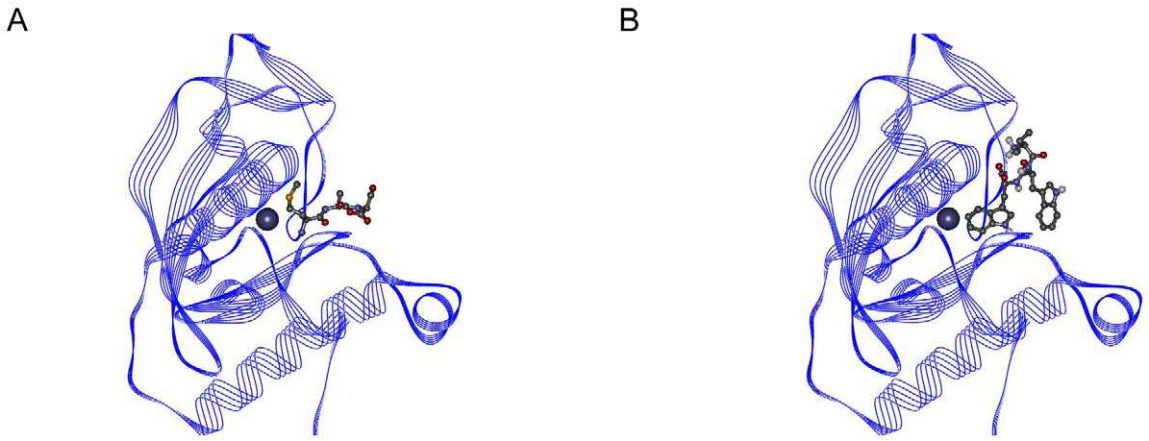


Figure 12. PDF peptide complexes. (A) Met-Ala-Ser and PDF. (B) Val-Trp-Trp and PDF.
doi:10.1371/journal.pone.0010926.g012

peptide perfectly covers the binding surface, while the known peptide is more distant to the surface. The Ala and Gly residues give flexibility to the peptide with their small side-chains. Our method kept the Tyr residue, which is similar to saccharide monomers. As stated in the case of SPG-40, the similarity to saccharides is important since the binding molecule of this target protein is also oligosaccharides.

The proteinase K binding surface has +1 net charge, 20% of residues that form the surface is hydrophobic; the 14% of residues of the designed peptide Trp-Tyr-Tyr-Tyr-Tyr-Tyr-Tyr is hydrophobic. The Asn-161 residue of protein plays role in formation of hydrogen bond with both the known and the designed peptides. The residues Cys-73, Met-225 form stable sulfide aromatic ring interactions with both the known and the designed peptides; but the number of interactions formed are higher for the Viterbi designed peptide. The aromatic ring stacking number is superior for the designed peptide, since it has more residues with side-chains containing ring structures when compared to the known peptide sequence Pro-Ala-Pro-Phe-Ala-Ala-Ala.

The HLA-B*2705 protein binding surface is made up of 44% by hydrophobic residues. Both the designed and the self-peptide have 3 hydrophobic residues, which are able to make interactions with the hydrophobic residues of binding surface. The self-peptide net charge is +3; the designed peptide charge is +2. The algorithm is successful to keep Lys and Arg residues, which are known to play major role in HLA-B*2705 binding. The residue Gln-55 is observed to play a role in hydrogen bonding for both the known and the designed peptide. The designed peptide makes three net hydrogen bonds, while the known peptide makes only two bonds.

The HGH binding region is made up of hydrophobic residues by 18%. The net charge of the surface is +2 and the net charge of the designed peptide is −1. Consequently, there exist an electrostatic complementarity between the target protein and the designed peptide. There exist two Cys and one Met residues on protein binding site, which are potential stabilizing residues of peptide binding. The designed peptide also has a Met residue, which makes interactions with the sulfur atom on its side-chain. Numerous sulfur - aromatic ring, sulfur – oxygen interactions are observed due to the Cys and Met residues of both the peptide and the protein. The Trp and Phe residues of the peptide enable stacking of aromatic rings.

The detailed analysis of the designed peptide interactions and comparison of those peptides with the known peptides indicate that our method is able to detect the requirements of binding surface; such as hydrophilicity, electrostatic compatibility, aromatic interactions, sulfur atom and its interactions.

The residues Trp, Tyr and Val are highly favored on the designed peptides. The binding energy of dipeptides containing

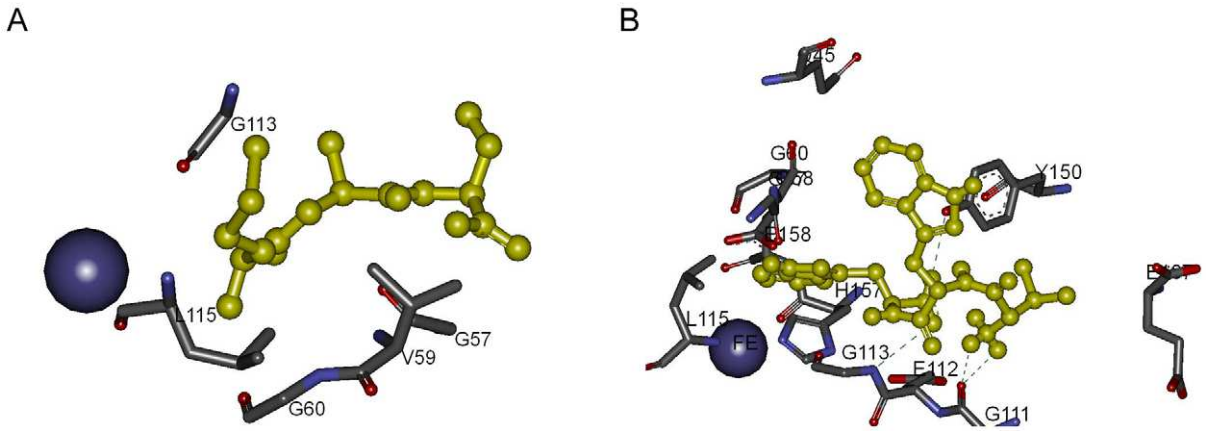


Figure 13. Detailed analysis of PDF and peptide complexes. Hydrogen bonds are indicated with green lines. (A) Met-Ala-Ser and PDF. (B) Val-Trp-Trp and PDF.
doi:10.1371/journal.pone.0010926.g013

Table 6. Binding energy; K_i values of Con A peptides.

Peptide	Binding Energy (kcal/mol)	K_i
Tyr-Pro-Tyr	−5.24	144.38 μM
Gly-Ala-Tyr	−5.70	65.83 μM

doi:10.1371/journal.pone.0010926.t006

Table 7. Binding energy; K_i values of Proteinase K peptides.

Peptide	Binding Energy (kcal/mol)	K_i
Pro-Ala-Pro-Phe-Ala-Ala-Ala	−8.52	569.86 nM
Trp-Tyr-Tyr-Tyr-Tyr-Tyr-Tyr	−11.59	3.22 nM

doi:10.1371/journal.pone.0010926.t007

Trp and Tyr are observed to be superior to other dipeptides. Consequently the transition probabilities of those dipeptides are favored. This may be the reason for high occurrence of those residues in the designed peptide sequences. The conservation of Trp may also be due to the formation of stacking and consequent stabilization. When the protein – designed peptide complexes are analyzed in detail, it is obvious that Trp residues form high number of stacking due to its aromatic ring; this stabilizes the formed complex. The same observation applies to Val for hydrophobic regions; since it has a smaller side-chain compared to other hydrophobic residues - except Ala -; the amino acid is able to interact with small hydrophobic cavities. The HIV-1 protease, SPG-40, HLA-B*2705 and Con A proteins have Trp and Tyr residues on their own inhibitory peptides; and proteinase K protein binding surface contains Trp, Phe and His residues on its binding surface. Nature seems to protect interactions containing amino acids with aromatic side-chains; our methodology also conserves those residues. A recent study of London et al. [28] reveals that few hot-spots on the peptide are responsible of the protein binding. The hot-spot residues are enriched in Phe, Leu, Trp, Tyr and Ile. Out of 36 amino acids of the peptides determined by VitAL, the 21 are Trp and Tyr residues. Also the peptide for the HGH protein has two Phe and one Leu residues.

Our algorithm designed a positively charged peptide for the negatively charged scytalidocarboxyl peptidase B protein surface. The method also conserved the Lys and Arg residues for HLA-B*2705 protein, which are known the play major role in peptide binding to this surface. This proves that our method not only favors the aromatic and the hydrophobic interactions, but also the electrostatic complementarity.

The only case that the method failed to determine an outstanding binder peptide is for the Con A protein case. The surface of this protein is highly exposed to water, as stated before. Consequently, we may state that our algorithm works best for cavities on protein surfaces, with both hydrophilic and hydrophobic residues present.

The study of 103 protein-peptide complexes by London et al. [28] showed that most peptides do not alter the target protein conformation minimizing the entropic cost of binding. This statement is supportive for our studies, since the protein conformation is kept rigid and the peptide is relaxed free to change conformation in AutoDock runs. London et al. also indicates that peptides of length 6–11 are observed to have coiled conformation generally. The peptides designed by VitAL are observed to have coiled conformation when bound to their target protein.

The peptide design based on a pre-determined binding surface is shown to be successful on the case-studies. Peptides are modeled as sequences of Markov chains where the states defined for each residue are dependent on the states of the neighboring residues along the chain. This assumption allows for the application of the RIS formalism to calculate the binding probability and conformational properties of the peptides. Here, we used a knowledge-based approach to determine the statistical weights of the torsion angle states of the 20 amino acids and the dependences of the statistical weights on the neighboring residues. The partition function for a given peptide is determined using the RIS multiplication of the statistical weight matrices. The Viterbi algorithm is implemented to our method in order to determine a potential binding peptide using the probability values from RIS multiplication scheme. For the Viterbi Algorithm, the binding probabilities are set as the transition state probabilities, while the torsion state probabilities are set as the emission probabilities. The peptide design; the binding affinity of the designed peptide; the peptide – protein interactions are analyzed in detail. The importance of the binding surface selection is highlighted; a binding path with no cavity and made up of all hydrophilic residues was shown to be not very suitable for determination of a

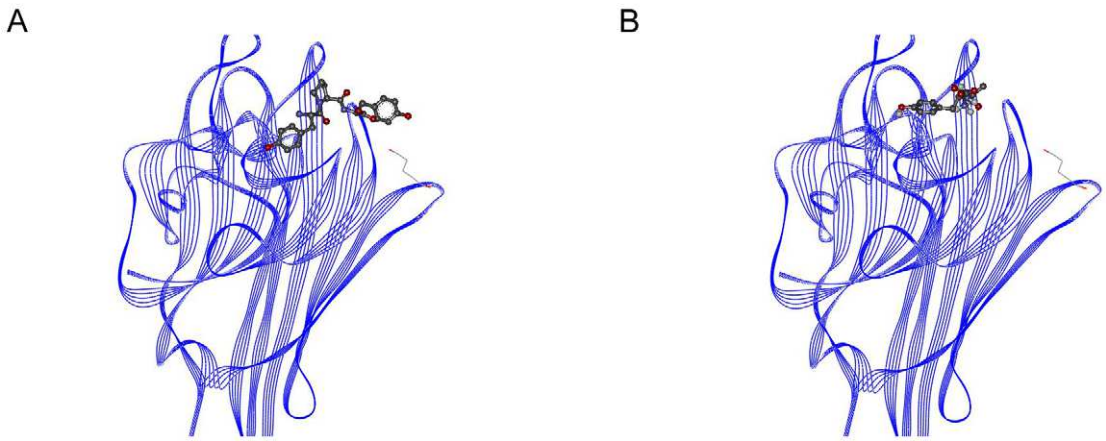


Figure 14. Con A peptide complexes. (A) Tyr-Pro-Tyr and Con A. (B) Gly-Ala-Tyr and Con A.
doi:10.1371/journal.pone.0010926.g014

Table 8. Binding energy; K_i values of HLA-B*2705 peptides.

Peptide	Binding Energy (kcal/mol)	K_i
Lys-Trp-Arg-Arg-Trp-His-Leu	−6.84	9.73 μM
Arg-Trp-His-Arg-Trp-Arg-Leu	−7.21	9.73 μM
Arg-Trp-Arg-Arg-Leu-Thr-Val	−8.52	9.73 μM
Trp-Arg-Trp-Trp-Lys-Tyr-Tyr	−8.97	265.82 nM

doi:10.1371/journal.pone.0010926.t008

candidate binder peptide. The method is shown to be successful to determine peptide according to the specific properties of the binding surface.

Haack et al. [65] indicated that the introduction of D-amino acids can significantly increase resistance to proteases and thus improve the potential use of peptides as therapeutic agents. For further improvements on VitAL we aim to add D-forms of amino acids into our library.

The algorithm requires O(mn) memory and $O(mn^2)$ time to run; where n is the peptide length and m is 20 - the number of states- [39]. The program details are given in Appendix S3. The time-consuming part for our methodology is the docking process. The Viterbi algorithm works efficiently to determine a single peptide is superior to other possible peptides, i.e. the most possible peptide sequence. The 1-best and posterior algorithms may also be employed to determine de novo peptide sequences, which have the same occurrence probability [42].

Table 9. Binding energy; K_i values of HGH peptide.

Peptide	Binding Energy (kcal/mol)	K_i
Trp-Glu-Leu-Met-Phe-Phe-Tyr	−8.05	1.25 μM

doi:10.1371/journal.pone.0010926.t009

Supporting Information

Appendix S1 AutoDock parameters, docking procedures.
Found at: doi:10.1371/journal.pone.0010926.s001 (0.03 MB DOC)

Appendix S2 Quantifying the peptide - target protein interaction via AutoDock.
Found at: doi:10.1371/journal.pone.0010926.s002 (0.02 MB DOC)

Appendix S3 Program details.
Found at: doi:10.1371/journal.pone.0010926.s003 (0.02 MB DOC)

Author Contributions

Conceived and designed the experiments: EBU AG BE. Performed the experiments: EBU. Analyzed the data: EBU. Contributed reagents/materials/analysis tools: EBU AG. Wrote the paper: BE.

References

1. Petsalaki E, Stark A, Garcia-Urdiales E, Russell RB (2009) Accurate Prediction of Peptide Binding Sites on Protein Surfaces. Plos Computational Biology 5: -.
2. Frenkel D, Clark DE, Li J, Murray CW, Robson B, et al. (1995) Pro-Ligand - an Approach to De-Novo Molecular Design.4. Application to the Design of Peptides. Journal of Computer-Aided Molecular Design 9: 213–225.
3. Hetenyi C, van der Spoel D (2002) Efficient docking of peptides to proteins without prior knowledge of the binding site. Protein Science 11: 1729–1737.
4. Juretic D, Vukicevic D, Ilic N, Antcheva N, Tossi A (2009) Computational Design of Highly Selective Antimicrobial Peptides. Journal of Chemical Information and Modeling 49: 2873–2882.
5. Mayrose I, Penn O, Erez E, Rubinstein ND, Shlomi T, et al. (2007) Pepitope: epitope mapping from affinity-selected peptides. Bioinformatics 23: 3244–3246.
6. Moreau V, Fleury C, Piquer D, Nguyen C, Novali N, et al. (2008) PEPOP: Computational design of immunogenic peptides. Bmc Bioinformatics 9: -.
7. Stein A, Aloy P (2008) A molecular interpretation of genetic interactions in yeast. Febs Letters 582: 1245–1250.
8. Brusic V, Rudy G, Honeyman M, Hammer J, Harrison L (1998) Prediction of MHC class II-binding peptides using an evolutionary algorithm and artificial neural network. Bioinformatics 14: 121–130.
9. Hammer J, Bono E, Gallazzi F, Belunis C, Nagy Z, et al. (1994) Precise Prediction of Major Histocompatibility Complex Class-Ii Peptide Interaction Based on Peptide Side-Chain Scanning. Journal of Experimental Medicine 180: 2353–2358.
10. Hammer J, Gallazzi F, Bono E, Karr RW, Guenot J, et al. (1995) Peptide Binding-Specificity of Hla-Dr4 Molecules - Correlation with Rheumatoid-Arthritis Association. Journal of Experimental Medicine 181: 1847–1855.
11. Hanai T, Noguchi H, Matsubara Y, Takeda K, Honda H, et al. (2000) Computational Design of Proteinous Drug Employing Hidden Markov Model. Genome Informatics 11: 394–395.
12. Harrison LC, Tait BD, Honeyman MC, Brusic V (1993) Major Histocompatibility Complex (Mhc) Molecules in Insulin-Dependent Diabetes. Journal of Leukocyte Biology. pp 105–105.
13. Honeyman MC, Brusic V, Stone NL, Harrison LC (1998) Neural network-based prediction of candidate T-cell epitopes. Nature Biotechnology 16: 966–969.
14. Kato R, Noguchi H, Honda H, Kobayashi T (2003) Hidden Markov model-based approach as the first screening of binding peptides that interact with MHC class II molecules. Enzyme and Microbial Technology 33: 472–481.
15. Noguchi H, Hanai T, Honda H, Harrison LC, Kobayashi T (2001) Fuzzy neural network-based prediction of the motif for MHC class II binding peptides. Journal of Bioscience and Bioengineering 92: 227–231.
16. Noguchi H, Kato R, Hanai T, Matsubara Y, Honda H, et al. (2002) Hidden Markov Model-Based Prediction of Antigenic Peptides That Interact with MHC Class II Molecules. Journal of Bioscience and Bioengineering 94: 264–270.
17. Ajay A, Walters WP, Murcko MA (1998) Can we learn to distinguish between "drug-like" and "nondrug-like" molecules? Journal of Medicinal Chemistry 41: 3314–3324.
18. Abe K, Kobayashi N, Sode K, Ikebukuro K (2007) Peptide ligand screening of a-synuclein aggregation modulators by *in silico* panning. BMC Bioinformatics 8: 451.
19. Kamphausen S, Holtge N, Wirsching F, Morys-Wortmann C, Riester D, et al. (2002) Genetic algorithm for the design of molecules with desired properties. Journal of Computer-Aided Molecular Design 16: 551–567.
20. Riester D, Wirsching F, Salinas G, Keller M, Gebinoga M, et al. (2005) Thrombin inhibitors identified by computer-assisted multiparameter design. Proceedings of the National Academy of Sciences of the United States of America 102: 8597–8602.
21. Belda I, Madurga S, Llora X, Martinell M, Tarrago T, et al. (2005) ENPDA: an evolutionary structure-based *de novo* peptide design algorithm. Journal of Computer-Aided Molecular Design 19: 585–601.
22. Klepeis JL, Floudas CA, Morikis D, Tsokos CG, Lambris JD (2004) Design of peptide analogues with improves activity using a novel *de novo* protein design approach. Industrial & Engineering Chemistry Research 43: 3817–3826.
23. Singh J, Ator MA, Jaeger EP, Allen MP, Whipple DA, et al. (1996) Application of Genetic Algorithms to Combinatorial Synthesis: A Computational Approach to Lead Identification and Lead Optimization. J Am Chem Soc 118: 1669–1676.
24. Budin N, Ahmed S, Majeux N, Caflisch A (2001) An evolutionary approach for structure-based design of natural and nonnatural peptidic ligands. Combinatorial Chemistry & High Throughput Screening 4: 661–673.
25. Budin N, Majeux N, Caflisch A (2001) Fragment-based flexible ligand docking by evolutionary optimization. Biological Chemistry 382: 1365–1372.
26. Budin N, Majeux N, Tenette-Souaille C, Caflisch A (2001) Structure-based ligand design by a build-up approach and genetic algorithm search in conformational space. Journal of Computational Chemistry 22: 1956–1970.
27. Unal EB, Gursoy A, Erman B (2010) Inhibitor Peptide Design for NF-kappaB: Markov Model and Genetic Algorithm Approaches. 5th International Symposium on Health Informatics and Bioinformatics (HIBIT'10). Antalya, Turkey: IEEE Xplore.
28. London N, Movshovitz-Attias D, Schueler-Furman O (2010) TheStructural Basis of Peptide-Protein Binding Strategies. Structure 18: 188–199.
29. Pegg SCH, Haresco JJ, Kuntz ID (2001) A genetic algorithm for structure-based *de novo* design. Journal of Computer-Aided Molecular Design 15: 911–933.

30. Haliloglu T, Erman B (2009) Analysis of Correlations between Energy and Residue Fluctuations in Native Proteins and Determnation of Specific Sites for Binding. Physical Review Letters 102: 088103.
31. Haliloglu T, Gul A, Erman B (2010) Predicting the Important Residues and Interaction Pathways in Proteins Using the Gaussian Network Model. Application to Binding and Stability of HLA Proteins. Submitted to PLOS computational Biology.
32. Haliloglu T, Seyrek E, Erman B (2008) Prediction of Binding Sites in Receptor-Ligand Complexes with the Gaussian Network Model. Physical Review Letters 100: 228102.
33. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and and Empirical Binding Free Energy Function. Journal of Computational Chemistry 19: 1639–1662.
34. Flory PJ (1969) Statistical Mechanics of Chain Molecules. New York: Wiley.
35. Rabiner LR (1989) A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of The IEEE 77: 257–286.
36. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. J Mol Biol 7: 95–99.
37. Forney GDJ (1973) The Viterbi Algorithm. Proceedings of IEEE 61: 268–278.
38. Viterbi AJ (1967) Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm. IEEE Transactions on Information Theory 13: 260–269.
39. Ewens WJ, Grant GR (2001) Statistical methods in bioinformatics: an introduction. New York: Springer.
40. Bystroff C, Krogh A (2008) Hidden Markov Models for Prediction of Protein Features. Methods in Molecular Biology 413: 173–198.
41. Sramek R, Brejova B, Vinar T (2009) On-line Viterbi Algorithm for Analysis of Long Biological Sequences. In: Giancarlo R, Hannenhalli S, eds. Lecture Notes in Computer Science. Philadelphia, PA, USA: Springer.
42. Fariselli P, Martelli PL, Casadio R (2005) A new decoding algorithm for hidden Markov models improves the prediction of the topology of all-beta membrane proteins. BMC Bioinformatics 6.
43. Mirabeau O, Perlas E, Severini C, Audero E, Gascuel O, et al. (2007) Identification of novel peptide hormones in the human proteome by Hidden Markov model screening. Genome Resarch 17: 320–327.
44. Sonmez K, Zaveri NT, Kerman IA, Burke S, Neal CR, et al. (2009) Evolutionary Sequence Modeling for Discovery of Peptide Hormones. Plos Computational Biology 5: -.
45. Fitzkee NC, Fleming PJ, Rose GD (2005) The Protein Coil Library: a structural database of nonhelix, nonstrand fragments derived from the PDB. Proteins 58: 852–854.
46. Tuncbag N, Kar G, Gursoy A, Keskin O, Nussinov R (2009) Towards inferring time dimensionality in protein-protein interaction networks by integrating structures: the p53 example. Molecular Biosystems 5: 1770–1778.
47. Vanhee P, Stricher F, Baeten L, Verschueren E, Lenaerts T, et al. (2009) Protein-Peptide Interactions Adopt the Same Structural Motifs as Monomeric Protein Folds. Structure 17: 1128–1136.
48. Tuncbag N, Gursoy A, Keskin O (2009) Identification of Computational Hot Spots in Protein Interfaces: Combining Solvent Accessibility and Inter-Residue Potentials Improves the Accuracy. Bioinformatics 25: 1513–1520.
49. HyperChem(TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
50. Keskin O, Yuret D, Gursoy A, Turkay M, Erman B (2004) Relationships between amino acid sequence and backbone torsion angle preferences in proteins. Proteins-Structure Function and Bioinformatics 55: 992–998.
51. Unal EB, Gursoy A, Erman B (2009) Conformational energies and entropies of peptides, and the peptide-protein binding problem. Phys Biol 6: 36014.
52. Karplus PA (1996) Experimentally observed conformation-dependent geometry and hidden strain in proteins. Protein Science. pp 1406–1420.
53. DS Modeling 2.5, Accelrys Inc.
54. Louis JM, Dyda FD, Nashed NT, Kimmel AR, Davies DR (1998) Hydrophilic Peptides Derived from the Transframe Region of Gag-Pol Inhibit the HIV-1 Protease. Biochemistry 37: 2105–2110.
55. Fujinaga M, Cherney MM, Oyama H, Oda K, James MNG (2004) The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. PNAS 101: 3364–3369.
56. Kumar J, Ethayathulla AS, Srivastava DB, Singh N, Sharma S, et al. (2009) Crystal structure of the complex formed between signalling protein from goat mammary gland (SPG-40) and a tripeptide Trp-Pro-Trp at 2.8A resolution.
57. Nam KH, Kim K, Kim EEK, Hwang KY (2009) Crystal structure of an *Ef*PDF complex with Met-Ala-Ser based on crystallographic packing. Biochemical and Biophysical Research Communications 381: 630–633.
58. Oldenburg KR, Loganathan D, Goldstein IJ, Schultz PG, Gallop MA (1992) Peptide ligands for a sugar-binding protein isolated from a random peptide library. PNAS 89: 5393–5397.
59. Zhang Z, Qian M, Huang Q, Jia Y, Tang Y, et al. (2001) Crystal Structure of the Complex of Concanavalin A and Tripeptide. Journal of Protein Chemistry 20: 59–65.
60. Hulsmeyer M, Fiorillo M, Bettosini F, Sorrentino R, Saenger W, et al. (2004) Dual, HLA-B27 Subtype-dependent Conformation of a Self-peptide. J Exp Med 199: 271–281.
61. Fiorillo M, Ruckert C, Hulsmeyer M, Sorrentino R, Saenger W, et al. (2005) Allele-dependent Similarity between Viral and Self-peptide Presentation by HLA-B27 Subtypes. J Biol Chem 280: 2962–2971.
62. Rucker C, Fiorillo M, Loll B, Moretti R, Biesiadka J, et al. (2006) Conformational Dimorphism of Self-peptides and Molecular Mimicry in a Disease-associated HLA-B27 Subtype. J Biol Chem 281: 2306–2316.
63. Lin-Su K, Wajnrajch M (2002) Growth Hormone Releasing Hormone (GHRH) and the GHRH Receptor. Reviews in Endocrine & Metabolic Disorders 3: 313–323.
64. Chantalat L, Jones N, Korber F, Navaza J, Pavlovsky A (1995) The Crystal-Structure of Wild-Type Growth-Hormone at 2.5 Angstrom Resolution. Protein Pept Lett 2: 333.
65. Haack T, Gonzalez MJ, Sanchez Y, Giralt E (1997) D-amino acids in protein de novo design. II. Protein-diastereomerism versus protein-enantiomerism. Letters in Peptide Science 4: 377–386.
66. Humphrey W, Dalke A, Schulten F (1996) VMD - Visual Molecular Dynamics. J Molec Graphics 14: 33–38.