Chapter 6

Protein-Protein Interaction for the *De Novo* Design of Cyclin-Dependent Kinase Peptide Inhibitors

Karthiga Arumugasamy, Sunil Kumar Tripathi, Poonam Singh, and Sanjeev Kumar Singh

Abstract

The homology of the inhibitor binding site regions on the surface of cyclin-dependent kinases (CDKs) makes actual CDK inhibitors unable to bind specifically to their molecular targets. Most of them are ATP competitive inhibitors with low specificity that also affect the phosphorylation mechanisms of other non-target kinases giving rise to harmful side effects. So, the search of specific and potent inhibitors able to bind to the desired CDK target is still a pending issue. Structure based drug design minimized the erroneous binding and increased the affinity of the inhibitor interaction. In the case of CDKs their activation and regulation mechanisms mainly depend on protein-protein interactions (PPIs). The design of drugs targeting these PPIs makes feasible and promising towards the discovery of new and specific CDK inhibitors. Development of peptide inhibitors for a target protein is an emerging approach in computer aided drug designing. This chapter describes in detail methodology for use of the VitAL-Viterbi algorithm for de novo peptide design of CDK2 inhibitors.

Key words Molecular docking, Peptide inhibitors, Protein-protein interaction, Structure based drug design, VitAL-Viterbi

1 Introduction

Approximately 20 different chemical classes of CDK inhibitors are available; they are analogues of purine, pyrimidine, and natural metabolites isolated from microbial strains and their derivatives. Flavopiridol, roscovitine, stauroporins, purvalanol, and alsterpaullone are some of the potential drug candidates undergone clinical trials which inhibit the CDK activity. Specificity is the biggest problem in these currently available antagonists. They do not have the specificity to target only one CDK; most of these inhibitors have multiple CDK targets [1]. For example, R-Roscovitine inhibits CDK1, CDK2, CDK5, CDK7, and CDK9 [2]. To design specific kinase inhibitors is essential in order to minimize the

Mar Orzáez et al. (eds.), *Cyclin-Dependent Kinase (CDK) Inhibitors: Methods and Protocols*, Methods in Molecular Biology, vol. 1336, DOI 10.1007/978-1-4939-2926-9_6,

unnecessary side effects of these drugs. CDK inhibitors could be designed through two types of strategies, using either small organic molecules or peptides. Peptide inhibitors promise better specificity of interaction than small molecule ATP competitors.

In CDKs, inhibitor peptides could be derived from cyclin or negative regulators binding regions. Some peptide inhibitors are reported against the CDKs; for example, $p27^{KIP}$ inhibitor acts as model to design a peptide inhibitor of the CDK2/Cyclin A complex activation [3]. In vitro and in vivo studies encourage the potential of a peptide inhibitor derived from the $p21^{Waf1/cip1}$ protein that potently inhibits the CDK2/Cyclin E and CDK4/Cyclin D complexes [4]. Some previously reported complexes of CDK2/Cyclin A/peptide inhibitors are listed in Table 1.

There are different ways to design a specific peptide inhibitor against the active site of a protein. Diverse graphical user interfaces (GUI) and algorithms based software tools help to model the peptide sequence. In this chapter, we describe strategy and VitAL-Viterbi algorithm to design the CDK2 peptide inhibitor based on the protein active site.

Table 1
CDK2/Cyclin A complexed with different peptide chain derived from the natural regulator of cell cycle from PDB (http://www.rcsb.org)

PDB ID	Resolution (Å)	Peptide inhibitor derived from natural regulator	Position of the residues	Amino acid sequence	PubMed ID
1H24	2.50	E2F	Residues 87–95	Pro-Val-Lys-Arg-Arg-Leu-Asp-Leu-Glu	12501191
1H25	2.50	Retinoblastoma-associated protein	Residues 869–878	Pro-Lys-Pro-Leu-Lys-Lys-Leu-Arg-Phe-Asp	12501191
1H26	2.24	p53	Residues 376–386	Ser-Arg-His-Lys-Lys-Leu-Met-Phe-Lys	12501191
1H27	2.20	p27	Residues 25–35	Arg-Asn-Leu-Phe-Gly-Pro	12501191
1H28	2.80	p107	Residues 653–663	Gly-Ser-Ala-Lys-Arg-Arg-Leu-Phe-Gly-Glu	12501191
1URC	2.60	Synthetic derivative of p27	Leu-Phe-Gly motif region	Ace-Arg-Lys-Leu-Phe-Gly	15455144

2 Materials

1. Linux cluster system to execute the Viterbi algorithm coding and perform the molecular docking.
2. Protein Data Bank to retrieve the crystal structure of CDK protein associated with inhibitor or substrate proteins (p53, p21, p27, p107, etc.) given as input structure.
3. Hotpoint, web server (http://prism.ccbb.ku.edu.tr/hotpoint) to predict the interest of amino acid residues on the binding surface of protein.
4. AutoDock package for docking analysis and binding free energy calculation.
5. Hyperchem tool for preparation of peptide structure.
6. Coil library, web server (http://www.roselab.jhu.edu/coil) to determine the probabilities of the φ–ψ torsion angles of the peptides.

Prediction of peptide inhibitor through VitAL-Viterbi based algorithm is proved by other target enzymes which compared with their known peptide inhibitors and final peptides shows significant binding free energy [5]. AutoDock has been used to check the reliability of the binding interaction and HyperChem package for amino acids and possible dipeptides [6, 7]. Ramachandran plot have used to characterize the $\varphi\neg\psi$ propensities of the dipeptides. Acetyl-group have added at the N-terminal end of dipeptides for their stability.

3 Methods

The Viterbi algorithm based de novo peptide design sequentially generates the peptide by docking its residues pair by pair along a chosen path on a protein. The prior method needed for the Viterbi algorithm to run properly is given in Fig. 1.

3.1 Prediction of the Active Site Residues of the Protein

In this step, X-ray crystallographic coordinates of CDK2 from Protein Data Bank (PDB) [8] may be given as input file in web server Hotpoint (http://prism.ccbb.ku.edu.tr/hotpoint). Gaussian Network model (GNM) (*see* **Note 1**) can be used to find out the binding site residues of protein by two ways: (1) if the protein exists in complex with other proteins, and (2) if the site determined by the GNM lies in an interface in the complex, then the complex is used to determine sequence of residues on the binding site surface. The regions of hot spot residues present in the CDK2/Cyclin A complex are given in Fig. 2.

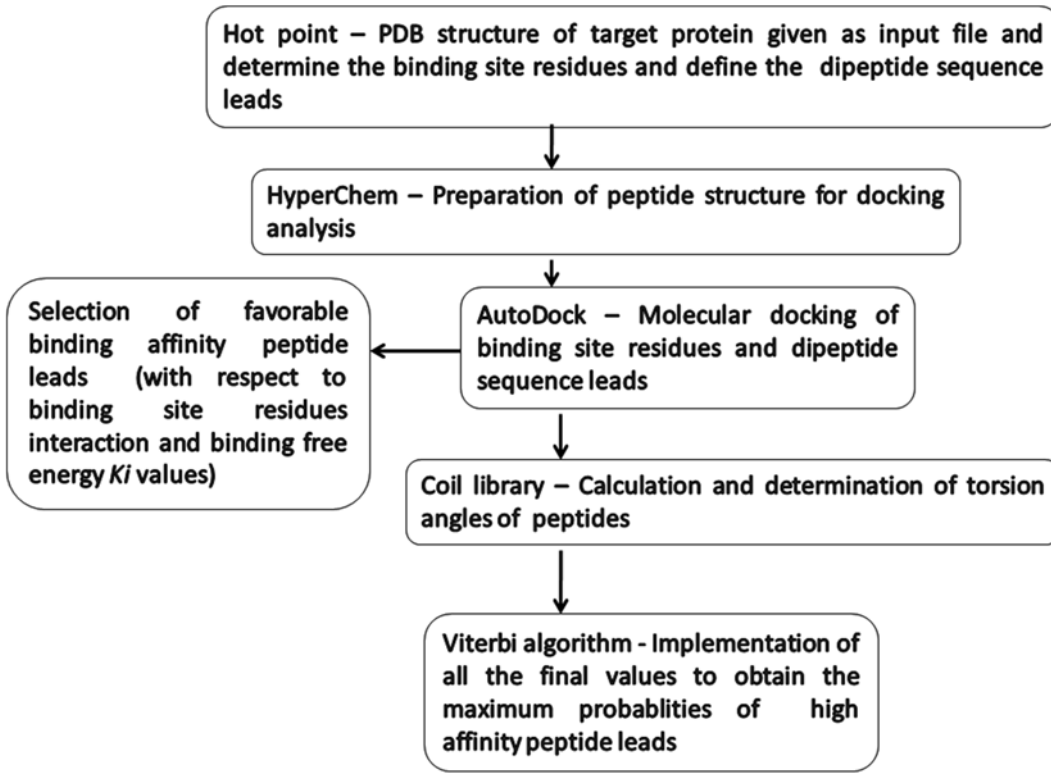

Fig. 1 The representation of steps involved in the Viterbi algorithm based de novo peptide design

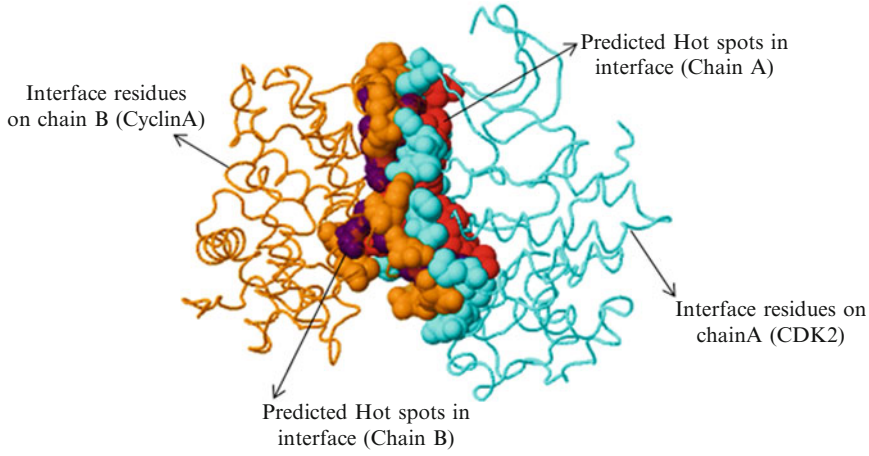

Fig. 2 Hot spot server aid to predict the CDK2/Cyclin A complex (PDB ID: 1FIN) binding site residues. The space fill model denotes the active site amino acids of CDK2 and Cyclin A (predicted hot spots in interface of CDK2 in chain A are shown as *red* and predicted hot spots in interface of Cyclin A in chain B are shown as *purple*)

3.2 Determination of "n" Amino Acid in the Grid Box

To accomplish this step, the grid box is generated based on the binding site residue prediction (probe peptide or the interacting protein portion) in CDK2 protein. The first and the second grids along the path contain the first and the second amino acids. The tth and $t+1$st grids contain the tth and $t+1$st residues. The "n" chiral carbon atoms of the path define the centers of the "n" grid boxes. Grid box is determined (possible region of interaction between ligand and receptor molecule) by AutoDock software tool [7]. The GOLD and GLIDE module can also be used for grid generation and docking process [9].

3.3 Amino Acid and Dipeptide Preparation

HyperChem tool has to be used to prepare the structure of amino acids and dipeptides for molecular docking (*see* **Note 2**) [6]. At the N-terminal end of dipeptide, acetyl group has to be added for their stability.

3.4 Docking and Binding Energy Quantification

The AutoDock program has to be used as the docking tool to quantify the binding affinity between the dipeptides and the selected protein surface. In the initial part of docking, first grid box path docked with 20 amino acids, then 400 possible dipeptides (20×20) are docked with first and second grid box, thus their chiral carbon atoms forced to overlap with first grid center 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. AutoDock program gives the bound conformation of protein-peptide, the binding energy and Ki value.

3.5 Principle to Find the Probabilities of Dipeptide Binding

Rotational Isomeric States (RIS) approach (*see* **Note 3**) has to be used to determine the type and length of the amino acid. Statistical analysis of the binding energies of dipeptides determines the significant and possible interaction and transition probabilities, and derivation of equation elaborately given by Unal et al. [5].

3.6 Selection of Favorable Dipeptide Probability by Coil Library

1. Energetically favorable φ, ψ torsion angle peptides used and other conformations are excluded. Two sets of probabilities are needed for specifying the conformation of the peptide. $\Phi t - \psi t$ (grid box 1) and $\Phi t - \psi t_{+1}$ (grid box 2) set of torsion angles selected for the probability analysis.
2. Depends on the chemical nature of the amino acids the repulsive, attraction, steric hindrances effects are formed. Hydrogen bond formation is consider as most favorable interaction between protein-peptide (*see* **Note 4**).
3. Web server, coil library (http://www.roselab.jhu.edu/coil) that contains the precalculated torsion angle for the fragments as well as crystal structure of the peptides available in the PDB [8]. Number of peptide sequence structure is retrieved from the coil library data and selected by various criteria (*see* **Note 5**).

Table 2
Backbone torsion angle calculation of known CDK2 complex inhibitor p27 derived peptide sequence (PDB ID: 1H27) using coil library tool

Amino acid	φ	Ψ	ω	Secondary structure
LEU	−103.16	151.07	−170.20	E
ILE	−129.93	144.76	179.59	E
ASN	−112.51	−178.10	178.73	C
THR	−76.15	−11.78	−173.20	T
GLU	−98.21	9.80	−175.27	T
GLY	88.18	3.65	176.72	C
ALA	−77.99	146.21	173.54	E
ILE	−125.12	150.21	165.44	E

T—β-turn; C—Coil; E—β-strand, secondary structure conformation formed by the amino acids (http://www.roselab.jhu.edu/coil)

The p27 peptide sequence (PDB ID: 1H27) taken as example and their torsion angles are given in Table 2.

4. Description of variables and equations to obtain the possible peptide conformation was given by Unal et al. [5].

3.7 Accuracy of the Protein-Peptide Interaction

After the prediction of possible and reliable peptide structure can be docked with receptor CDK2 or CDK2/Cyclin complexes. In case of CDK2, some previously reported peptide inhibitors structure and information are available [3, 10] so the binding energy and interactions of the newly designed peptide inhibitors can be compared with the parent inhibitors.

3.8 Implementation of Viterbi Algorithm

All the previous steps final values are employed in Viterbi algorithm equations. This process is divided into two steps, namely forward tracking and backward tracking method. Both of the steps are elaborated by Unal et al. [5]. Each step leads to increase the probability of the peptide sequence and to determine a peptide sequence with possible affinity to protein binding site.

Different algorithms are available for successful design of peptide inhibitors. The predicted peptide inhibitors should go for experimental validation to examine the potential against applicable target. Thus, in terms of accuracy the prediction must be accurate and reliable for such novel and potent peptide inhibitors. In addition to VitAL-Viterbi, Rosetta is one of the important algorithm based tool (http://www.rosettacommons.org/) which gives the accurate and reliable results for such potent inhibitors [11, 12]. CDKs have

high sequence similarity and many natural inhibitors for regulation of their mechanism. It will improve the search of inhibitors based on the binding pocket of protein to find out the potent and reliable specific inhibitors against CDKs. Development of more computational algorithms evokes the emerging of novel peptide inhibitors for CDKs. In general, the algorithm should be simple and give the accurate output form of peptide antagonist or ligand.

4 Notes

1. Based on the statistical thermodynamic model, GNM has been proposed the potential of residue-residue interactions. Thus offers a model for determining structurally and functionally important residues in relation to ligand-protein interactions. Although, provides the protein transferring information from one point to the other [13].
2. HyperChem tool has been used to get the amino acid structure, and other tools like Chimera, Sybyl, and Schrodinger can also be used to get the amino acid structure [9].
3. The α-helical and β-strand amino acid shows significant energy differences. RIS polymer physics facilitates to determine that energy difference either due to favorable or unfavorable interactions with the peptide sequence. Torsion angles adjacent to the peptide bond (Ramachandran map) should be considered to check the reliable conformation of the protein. Higher order interdependences between bond dihedral angles can be ignored [14].
4. Different levels of correlations among the φ ψ angles are already identified and studied for native as well as denatured proteins. Ramachandran map gives the correlations among the φ and ψ and angles of a residue resulting from exclusion of steric overlaps that hold both for denatured and native proteins.
5. The selection criteria may be chosen by the user. For example, less than 20 % sequence identity, 1.6 Å resolution, and 0.25 refinement factor was followed to select the peptides derived from CDK2/p27 complex.

Acknowledgments

The authors thank the DST New Delhi (SR/FT/CS-66/2010) for financial support. AK and SKT gratefully acknowledge DST (New Delhi) for JRF and CSIR (New Delhi) for SRF respectively. One of the authors thanks CSIR-CDRI for support.

References

1. Gadek JW, Maurer M, Zulehner N et al (2011) Whether to target single or multiple CDKs for therapy? That is the question. J Cell Physiol 226:341–349
2. Senderowicz AM (2000) Small molecule modulators of cyclin-dependent kinases for cancer therapy. Oncogene 19:6600–6606
3. Andrews MJI, McInnes C, Kontopidis G et al (2004) Design, synthesis, biological activity and structural analysis of cyclic peptide inhibitors targeting the substrate recruitment site of cyclin dependent kinase complexes. Org Biomol Chem 2:2735–2741
4. Mutoh M, Lung FDT, Long YQ et al (1999) $p21^{Waf/cip1}$ carboxyl-terminal peptide exhibited cyclin-dependent kinase inhibitory activity and cytotoxicity when introduced into human cells. Cancer Res 59:3480–3488
5. Unal EB, Gursoy A, Erman B (2010) VitAL: Viterbi algorithm for *de novo* peptide design. PLoS One 5(6), e10926
6. HyperChem (TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA
7. Morris GM, Goodsell DS, Halliday RS et al (1998) Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J Comput Chem 19:1639–1662
8. Berman HM, Westbrook J, Feng Z et al (2000) The Protein Data Bank. Nucleic Acids Res 28(1):235–242
9. Liao C, Sitzmann M, Pugliese A et al (2011) Software and resources for computational medicinal chemistry. Future Med Chem 3(8): 1057–1085
10. Lowe ED, Tews I, Cheng KY et al (2002) Specificity determinants of recruitment peptides bound to phospho-CDK2/cyclin A. Biochemistry 41:5625
11. Kuhlman B, Dantas G, Ireton GC et al (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302: 1364–1368
12. Sievers SN, Karanicolas J, Chang HW et al (2011) Structure-based design of non-natural amino-acid inhibitors of amyloid fibril formation. Nature 475:96–100
13. Tuncbag N, Kar G, Gursoy A et al (2009) Towards inferring time dimensionality in protein-protein interaction networks by integrating structures: the p53 example. Mol Biosyst 5:1770–1778
14. Keskin O, Yuret D, Gursoy A et al (2004) Relationships between amino acid sequence and backbone torsion angle preferences in proteins. Proteins 55:992–998