Journal of Molecular Graphics and Modelling 29 (2010) 546–564

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

The binding modes and binding affinities of epipodophyllotoxin derivatives with human topoisomerase IIα

Pradeep Kumar Naik*, Abhishek Dubey, Komal Soni, Rishay Kumar, Harvinder Singh

Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, Distt. Solan 173215, Himachal Pradesh, India

ARTICLE INFO

Article history:
Received 1 May 2010
Received in revised form 12 October 2010
Accepted 18 October 2010
Available online 26 October 2010

Keywords:
Human topoisomerase IIα
Epipodophyllotoxin
Docking
eMBrAcE

ABSTRACT

Epipodophyllotoxin derivatives have important therapeutic value in the treatment of human cancers. These drugs kill cells by inhibiting the ability of topoisomerase II (TP II) to ligate nucleic acids that it cleaves during the double-stranded DNA passage reaction. The 3D structure of human TP IIα was modeled by homology modeling. A virtual library consisting of 143 epipodophyllotoxin derivatives has been developed. Their molecular interactions and binding affinities with modeled human TP IIα have been studied using the docking and Bimolecular Association with Energetics (eMBrAcE) developed by Schrödinger. Structure activity relationship models were developed between the experimental activity expressed in terms of percentage of intracellular covalent TP II–DNA complexes (log PCPDCF) of these compounds and molecular descriptors like docking score and free energy of binding. For both the cases the r^2 was in the range of 0.624–0.800 indicating good data fit and r^2_{cv} was in the range of 0.606–774 indicating that the predictive capabilities of the models were acceptable. Low levels of root mean square error for the majority of inhibitors establish the docking and eMBrAcE based prediction model as an efficient tool for generating more potent and specific inhibitors of human TP IIα by testing rationally designed lead compounds based on epipodophyllotoxin derivatization.

1. Introduction

Human DNA topoisomerase II (TP II) is a ubiquitous nuclear enzyme involved in the control of DNA topology [1–4]. During the catalytic cycle, the enzyme transiently cleaves dsDNA, passes an intact double helix through the break and reseals it. Due to the requirement for such a DNA strand passage activity in a number of critical nuclear processes, including replication, recombination, and chromosome segregation, TP II is essential for the survival of proliferating eukaryotic cells [5]. Vertebrates contain two isoforms of the enzyme, TP IIα and β [1]. Topoisomerase IIα (TP IIα) levels increase during cell proliferation and this enzyme appears to be the isoform involved in mitosis [2]. To maintain DNA integrity during the strand passage event, the enzyme, a homodimer, forms a covalent phosphotyrosyl adduct between the catalytic Tyr^{804} of each monomer and a strand of the duplex. This covalent enzyme-cleaved DNA complex is referred to as the cleavage complex [1–4]. Because the covalent TP II-cleaved DNA complex (referred to as the *cleavage complex*) is normally a short-lived intermediate in the catalytic cycle of the enzyme, it is tolerated by the cell. However, when present in high concentrations, cleavage complexes become potentially toxic, promoting frameshift mutations, permanent double-stranded DNA breaks, illegitimate recombination, and apoptosis [6,7]. The cytotoxic potential of TP II has been exploited clinically by the development of anticancer drugs that generate high levels of covalent enzyme-DNA cleavage complexes [7]. A number of drugs, such as etoposide (VP16), teniposide (VM26) increase TP II-mediated DNA breakage primarily by inhibiting the ability of the enzyme to religate cleaved nucleic acid molecules [8–12]. As a result, they dramatically increase levels of TP II–DNA cleavage complexes (DNA/TP II/drug ternary complex). The resulting DNA strand breaks initiate multiple recombination/repair pathways and can trigger cell death pathways [13,14].

Despite the wide use of TP II-targeted drugs as antitumour agents, several limitations hamper their benefits. Efforts for improving their clinical efficacy further by overcoming the drug resistance, myelosuppresion and poor bioavailability problems [15] associated with them, were continued to be challenging. Over the years a number of laboratories throughout the world engaged in the synthesis and testing of epipodophyllotoxin derivatives [16–27] to prepare new more potent and less toxic analogues, that is, with better therapeutic indices. The mechanism of action of any drug is very important in drug development. Generally, the drug compound binds with a specific target, a receptor, to mediate its effects. Therefore, suitable drug–receptor interactions are required for high activity. Understanding the nature of these interactions is very significant and theoretical calculations, in particular the molecular docking method, seem to be a proper tool for gaining such

* Corresponding author. Tel.: +91 1792 239227; fax: +91 1792 245362.
E-mail address: pknaik73@rediffmail.com (P.K. Naik).

1093-3263/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2010.10.005

understanding. The docking results obtained will give information on how the chemical structure of the drug should be modified to achieve suitable interactions and for the rapid prediction and virtual prescreening of anti-tumour activity.

The amino acid sequence of human TP IIα is known but the complete three-dimensional structure is not available. This protein comprises of 1531 amino acids. Only part of the human TP-IIα structure (called as ATPase domain) has been available (pdb id: 1ZXN and 1ZXM) covering amino acids from 29 to 426. However, the crystal structure of other part of the human TP IIα has not been available. This fragment (covering 430–1214) of human TP IIα consists of drug binding site which is proved to be the binding site for epipodophyllotoxin analogues [28,2]. In this study, therefore we have constructed the 3D structure of TP IIα-drug binding domain by homology modeling and taken for interaction study between epipodophyllotoxin analogues and TP IIα. Of utmost importance in a structure-based drug design is the reliable filtering of putative hits in terms of their predicted binding affinity (scoring problem) which is based on the *in silico*-generated near native protein–ligand configurations (docking problem). Most of scoring functions used in docking programs are designed to predict binding affinity by evaluating the interaction between a compound and a receptor. However, it should be noted that ligand receptor recognition process is determined not only by enthalpic effects but also by entropic effects. Moreover, the scoring functions have a simplified form for the energy function to facilitate high throughput evaluation of a large number of compounds in a single docking run. These functions may be problematic when used with contemporary docking programs, and can result in a decrease of virtual screening accuracy. To overcome this problem, more precise but time consuming computational methodologies are necessary. Here, we have used and evaluated both docking and molecular mechanics based energy minimization of docking complex for computational modeling of epipodophyllotoxin and its derivatives as potent inhibitor of TP IIα.

2. Computational methods

2.1. Sequence analysis

The protein sequence of human TP IIα was obtained from the protein NCBI database. Sequence similarity search with BLAST in Protein Data Bank (pdb) database gives only one similar protein (33.6% identical), topoisomerase II (pdb ID: 1BJT) from yeast matching with the drug binding domain of human TP IIα. This structure is determined at 2.50 Å resolution [29]. We performed the pairwise alignment of human TP IIα with IBJT as template using the homology module of PRIME [30]. We removed the mismatched sequence part (1–420) from the whole sequence of human TP IIα and then constructed the three-dimensional structure of human TP IIα. The sequence alignment after removing the part of mismatched sequence is shown in Fig. 1.

2.2. Homology model construction

The homology model of the human TP IIα was built using Prime [30] accessible through the Maestro interface (Schrodinger, Inc.). All water molecules were removed. During the homology model building, Prime keeps the backbone rigid for the cases in which the backbone does not need to be reconstructed due to gaps in the alignment. The model was screened for unfavorable steric contacts and remodeled using a rotamer library database of PRIME. Explicit hydrogens were added to the protein and the protein model subjected to energy minimization using the Macromodel (Prime version 1.5) force-field OPLS-2005. Energy minimization and relaxation of the loop regions was performed using 300 iterations in a simple minimization method. The steepest descent energy minimization was carried out until the energy showed stability in the sequential repetition. Model evaluation was performed in PROCHECK v3.4.4 [31] producing plots that were analyzed for the overall and residue-by-residue geometry. Ramachandran plot [32] provided by the program PROCHECK assured very good confidence for the predicted protein. There were only 0.2% residues in the disallowed region and 0.8% residues in generously allowed regions. Nevertheless, PROCHECK assured the reliability of the structure and the protein was subjected to VERIFY3D [33], available from NIH MBI Laboratory Servers.

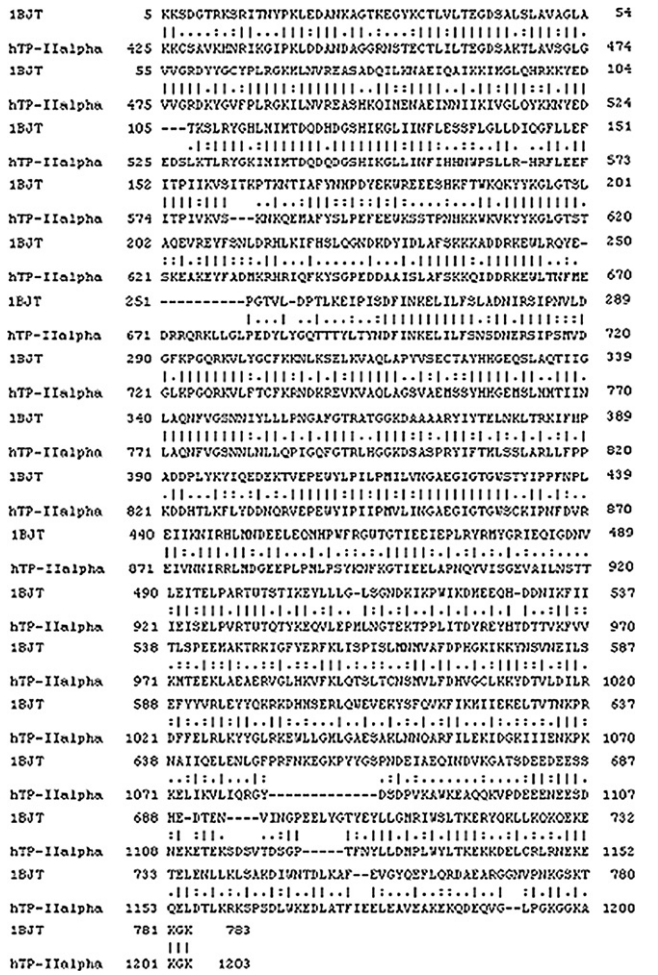


Fig. 1. Alignment of human TP IIα sequence with template (pdb ID: 1BJT).

2.3. Ligand binding site prediction

Several works [34,4,35] revealed that the epipodophyllotoxin drug, etoposide target the catalytic core domain of human TP IIα at the DNA cleavage–ligation site (drug binding site). Further in vitro drug binding assay revealed two binding sites for etoposide on human TP IIα. One is the lower affinity site in the ATPase domain (266 amino acid; N-terminal fragment), while the second one (430–1214 amino acids) binds with higher affinity [28]. Out of the amino acid residues in this catalytic site of TP IIα site directed mutagenesis study has revealed that Tyr^{804} is the critical residue involved in the binding of etoposide with TP IIα [2]. *In silico* prediction of binding site was done for the modeled structure of human TP IIα using SiteMap (Schrodinger package). SiteMap treat entire proteins to locate binding sites whose size, functionality, and extent of solvent exposure meet user specifications. SiteScore, the scor-

ing function used to assess a site's propensity for ligand binding, accurately ranks possible binding sites to eliminate those not likely to be pharmaceutically relevant. It identifies potential ligand binding sites by linking together "site points" that are suitably close to the protein surface and sufficiently well sheltered from the solvent. Given that similar terms dominate the site scoring function, this approach ensures that the search focuses on regions of the protein most likely to produce tight protein–ligand or protein–protein binding. Subsites are merged into larger sites when they are sufficiently close and could be bridged in solvent-exposed regions by ligand atoms. SiteMap evaluates sites using a series of properties. The binding site with highest site score was taken for docking of the epipodophytllotoxin analogues. The algorithm proceeds as follows: the protein is projected onto a 3D grid with a step size of 1.0 Å; grid points are labeled as protein surface, or solvent using certain rules. A grid point is marked as protein if there is at least one atom within 1.6 Å. After the solvent excluded surface is calculated the surface vertices' coordinates are stored. A sequence of grid points, which starts and ends with surface grid points and which has solvent grid points in between, is called a surface–solvent–surface event. If the number of surface–solvent–surface events of a solvent grid exceeds a minimal threshold of 6, then this grid is marked as pocket. Finally, all pocket grid points are clustered according to their spatial proximity. The clusters are ranked by the number of grid points in the cluster. The top three clusters are retained and their centers of mass are used to represent the predicted pocket sites. The binding pocket obtained by *in silico* studies on human TP IIα was consistent with the site directed mutagenesis studies.

2.4. Preparation of the ligands

A total of 143 epipodophyllotoxin analogues (Tables 1 and 2) were used in the study, which were collected from different published articles [16–27]. These compounds were tested for their ability to form intracellular covalent topoisomerase II–DNA complexes. The assay procedures have been described earlier [16]. The activity data are originally expressed as the percentage of cellular protein–DNA complex formed (PCPDCF) and were transformed by taking the logarithm of PCPDCF i.e., $\log_{10}$(PCPDCF). These transformed activities were used in the development of prediction model. To generate statistically robust and most importantly, validated models, all compounds in the original data set were divided randomly into 110 molecules in training set and 33 molecules in test set.

All these epipodophyllotoxin analogues were built from the various scaffold structure (Fig. 2) and substitution of functional groups as mentioned in Tables 1 and 2. We used Maestro-molecular builder for building the scaffold and structural derivatives. LigPrep [36] was used for final preparation of ligands. LigPrep is a utility of Schrödinger software suit that combines tools for generating 3D structures from 1D (Smiles) and 2D (SDF) representation, searching for tatomers and steric isomers and performing a geometry minimization of ligands. The ligands were energy minimized using Macromodel module of Schrödinger with default parameters and applying molecular mechanics force fields (MMFFs). Truncated Newton Conjugate Gradient (TNCG) minimization method was used with 500 iterations and convergence threshold of 0.05 kJ/mol.

2.5. Docking of the ligands

The Glide program [37] was used for docking study. The Glide docking algorithm performs a series of hierarchical searches for locations of possible ligand affinity within the binding site of a receptor. A rough positioning and scoring algorithm is applied during the initial search step, followed by torsional energy optimization on an OPLS-AA non-bonded potential energy grid for enduring candidate poses. The pose conformations of the very best candidates are further refined by using Monte Carlo sampling. Selection of the final docked pose is accomplished using a Glide score, which is a model energy function that combines empirical and force field based terms. The Glide score is a modified and extended version of the ChemScore function [38].

All the ligands were docked to the human TP IIα receptor using Glide 4.0. After ensuring that protein and ligands are in correct form for docking, the receptor-grid files were generated using grid-receptor generation program by selecting the drug binding site, using van der Waals scaling of the receptor at 0.4. The default size was used for the bounding and enclosing boxes. The ligands were docked initially using the "standard precision" method and further refined using "xtra precision" Glide algorithm. For the ligand docking stage, van der Waals scaling of the ligand was set at 0.5. Out of the 50,000 poses that were sampled, 4000 were taken through minimization (conjugate gradients 1000) and the 30 structures having the lowest energy conformations were further evaluated for the favorable Glide docking score. A single best conformation for each ligand was considered for further analysis.

2.6. Molecular mechanics and free energies of binding

After obtaining preferable binding structure from docking simulation, the complex was partially minimized by relaxing ligand and atoms of side chains that are within 7 Å away from the ligand while all other atoms were fixed. Bimolecular Association with Energetics (eMBrAcE) developed by Schrödinger was used for physics based rescoring procedure [39]. The eMBrAcE (*MacroModel v9.1*) program calculates binding energies between ligands and receptors using molecular mechanics energy minimization for docked conformations. eMBrAcE applies multiple minimizations, during which each of the specified pre-positioned ligand is minimized with the receptor. For each ligand, the protein-ligand complex ($E_{\text{lig-prot}}$), the free protein (E_{prot}), and the free ligand (E_{lig}) were all subjected to energy minimization in implicit solvent (generalized Born) [40,41]. It uses traditional molecular mechanics (MM) methods to calculate ligand–receptor interaction energies (G_{ele}, G_{vdW}, G_{solv}), with a Gaussian smooth dielectric constant function method [42] for electrostatic part of solvation energy and solvent-accessible surface for the nonpolar part of solvation energy. A conjugate gradient minimization protocol was used in all minimization. The nonpolar solvent-accessible surface area (SASA) of solvation energy was calculated using Qikprop program. The percentage cellular protein–DNA complex formation is calculated using linear optimized multiple regression as follows:

$$\log(\text{PCPDCF}) = C + \alpha(\Delta G_{\text{vdW}}) + \beta(\Delta G_{\text{ele}}) + \gamma(\Delta G_{\text{solv}}) + \delta(\text{SASA})$$

where α, β, γ and δ are the coefficients for van der Walls, electrostatic, solvation energy terms and SASA, respectively; C is a constant. The approach is simple, fast and straightforward. It benefits the calculation of relative binding affinity needed to evaluate the activity of large set of molecules in rational drug design.

The eMBrAcE calculation was performed using the Ligand & Structure-Based Descriptors (LSBD) application of the Schrödinger software package. This calculation was applied to the ligand–receptor complex structures obtained from Glide docking.

The predictive capabilities of the proposed models were determined using leave-one-out cross validation method. The cross validation regression coefficient (q^2_{cv}) was calculated by the following equation:

$$q^2_{cv} = 1 - \frac{\text{PRESS}}{\text{TOTAL}} = 1 - \frac{\sum_{i=1}^{n}(y_{\text{exp}} - y_{pred})^2}{\sum_{i=1}^{n}(y_{\text{exp}} - \bar{y})^2}$$

Table 1
Epipodophyllotoxin analogues (training set) with binding affinity expressed in terms of percentage of cellular protein DNA complex formation (PCPDCF) against the human Topoisomerase II–DNA binding domain.

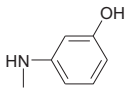
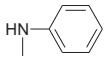
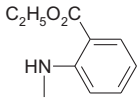
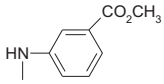
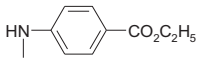
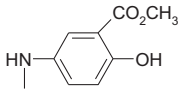
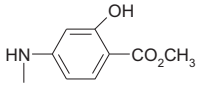
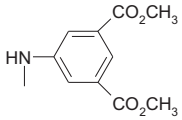
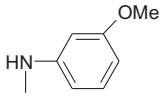
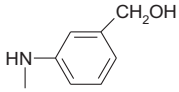
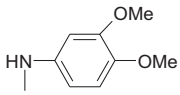
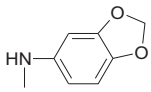
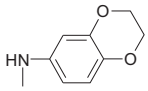
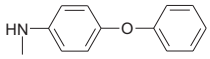
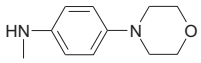
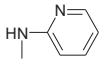
Ligand	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)[a]
1	–OH	1	42.2	1.625
2	$-NHCH_2CH_{20}CH_3$	1	110.8	2.044
3	$-NHCH_2CH{=}CH_2$	1	84.1	1.924
4	$-NHCH_2CH(OH)CH_3$(R)	1	167.2	2.223
5	$-NHCH(CH_3)CH_2OH$(R)	1	161.7	2.208
6		1	290	2.462
7		1	243	2.385
8		1	211	2.324
9		1	4	0.602
10		1	249	2.396
11		1	207	2.316
12		1	83	1.919
13		1	129	2.110
14		1	50	1.699
15		1	104	2.017
16		1	235	2.371
17		1	180	2.255
18		1	164	2.214
19		1	279	2.445
20		1	97	1.986
21		1	140	2.146
22		1	97	1.986

Table 1 (*Continued*)

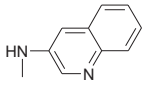
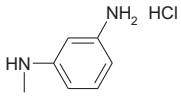
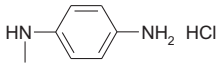
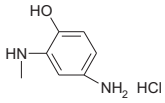
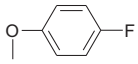
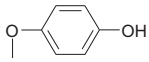
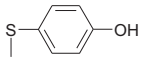
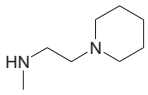
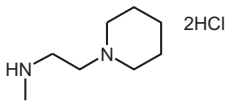
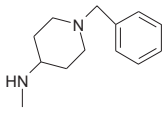
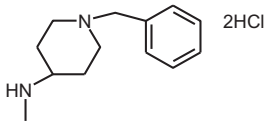
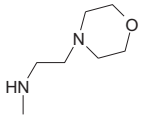
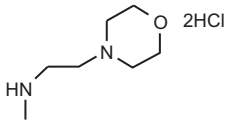
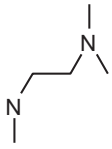
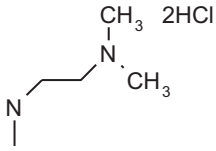
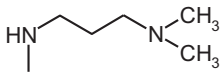
Ligand	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)[a]
23		1	123	2.090
24		1	140	2.146
25		1	330	2.518
26		1	11	1.041
27		1	57	1.756
28		1	34	1.531
29		1	10	1.000
30		1	190	2.278
31		1	183	2.262
32		1	83	1.919
33		1	172	2.235
34		1	77	1.886
35		1	140	2.146
36		1	203	2.307
37		1	183	2.262
38		1	186	2.269

Table 1 (*Continued*)

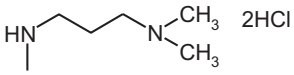
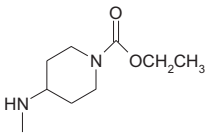
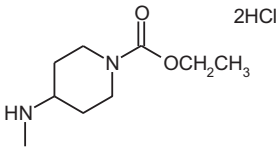
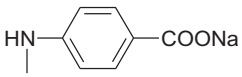
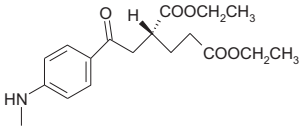
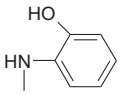
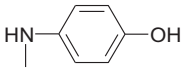
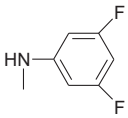
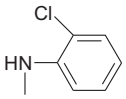
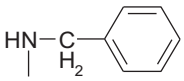
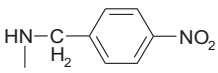
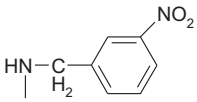
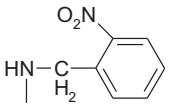
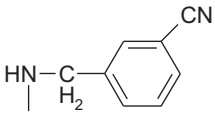
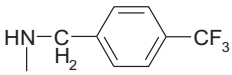
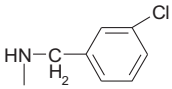
Ligand	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)[a]
39		1	179	2.252
40		1	17	1.230
41		1	138	2.139
42		1	6.9	0.839
43		1	83	1.919
44		1	151	2.179
45		1	211	2.324
46		1	115	2.060
47		1	32	1.505
48		1	181	3.258
49		1	216	2.334
50		1	130	2.114
51		1	144	2.158
52		1	225	2.352
53		1	99	1.995
54		1	159	2.201

Table 1 (*Continued*)

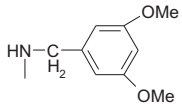
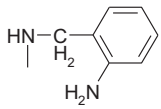
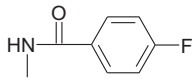
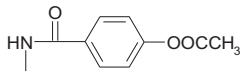
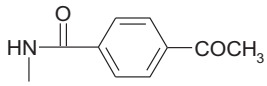
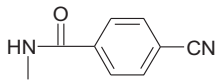
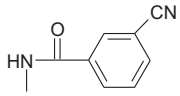
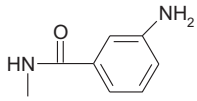
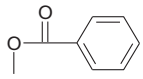
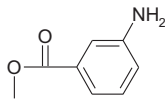
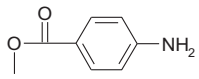
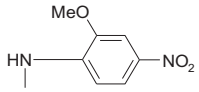
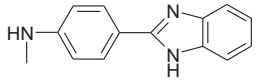
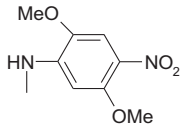
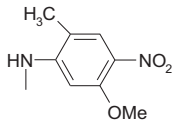
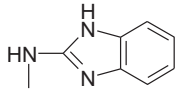
Ligand	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)[a]
55		1	144	2.158
56		1	184	2.264
57		1	117	2.068
58		1	137	2.136
59		1	124	2.093
60		1	159	2.201
61		1	149	2.173
62		1	149	2.173
63		1	94	1.973
64		1	100	2.000
65		1	94	1.973
66		1	83	1.919
67		1	128	2.107
68		1	4.4	0.643
69		1	3.5	0.544
70		1	58	1.763

Table 1 (*Continued*)

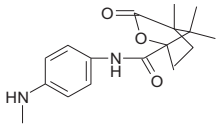
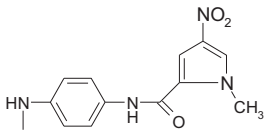
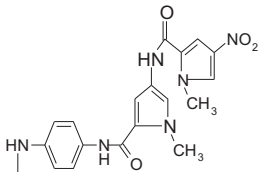
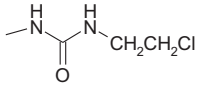
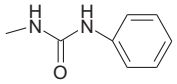
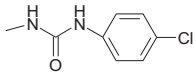
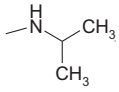
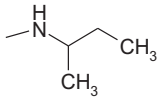
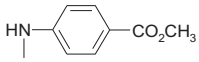
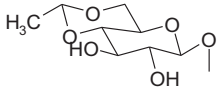
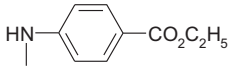
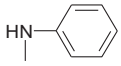
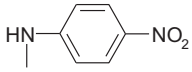
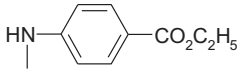
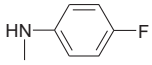
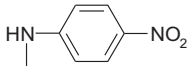
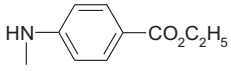
Ligand	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)[a]
71		1	88	1.944
72		1	100	2.000
73		1	26	1.415
74		1	143	2.155
75		1	148	2.170
76		1	125	2.097
77		1	109	2.037
78		1	73	1.863
79		1	207	2.316
80		2	6.1	0.785
81		2	15.6	1.193
82		3	22	1.342
83		4	4	0.602
84		4	99	1.995
85		4	138	2.139
86		4	52	1.716
87		5	75	1.875
88		5	127	2.103

Table 1 (*Continued*)

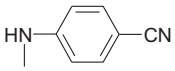
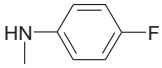
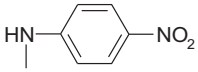
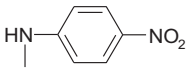
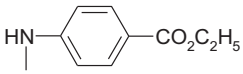
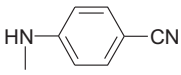
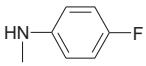
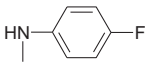
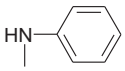
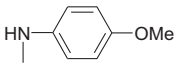
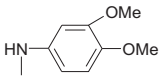
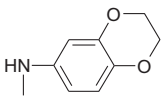
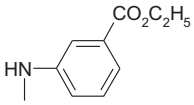
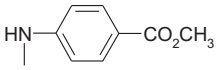
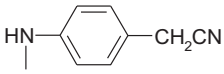
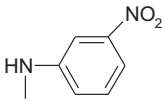
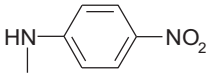
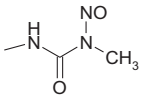
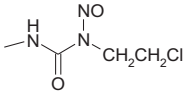
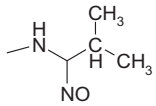
Ligand	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)[a]
89		5	125	2.097
90		5	108	2.033
91		3	23	1.361
92		6	8	0.903
93		6	9	0.954
94		6	12	1.079
95		6	8	0.903
96		7	117	2.068
97		7	105	2.021
98		7	96	1.982
99		7	69	1.839
100		7	119	2.075
101		7	94	1.973
102		7	175	2.243
103		7	146	2.164
104		7	109	2.037
105		7	75	1.875
106		7	200	2.301
107		8	41	1.612
108		8	7	0.845

Table 1 (*Continued*)

Ligand	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)[a]
109		9	1	0.000
110	—NH_2	1	36.4	1.561

[a] CPDCF, percentage of cellular protein–DNA complex formation.

where y_{pred}, y_{exp} and $\bar{y}$ are the predicted, experimental and mean values of experimental activity, respectively. Also the accuracy of the prediction of the developed models were validated by *F*-value, r^2 and r^2_{adj}. A large *F* indicates that the model fit is not a chance occurrence.

3. Results and discussion

The atomic coordinates of human TP IIα were not available in Protein Data Bank, which necessitated developing a protein model. The final model, which we took for further analysis, consisted of 789 amino acid residues. We used both PROCHECK and the VERIFY3D softwares to check the quality of the modeled protein. Ramachandran plot obtained from the program PROCHECK, which checks the stereochemical quality of a protein structures, producing a number of postscript plots, analyzing its overall and residue-by-residue geometry, assured the reliability of the modeled protein with 91.3% residues in most allowed region and 7.8% in additional allowed region. There were only 0.2% residues in disallowed region and 0.8% in generously allowed region. The assessment with VERIFY3D, which derives a "3D–1D" profile based on the local environment of each residue, described by the statistical preferences for: the area of the residue that is buried, the fraction of side-chain area that is covered by polar atoms (oxygen and nitrogen), and the local secondary structure, also substantiated the reliability of the three dimensional structure. The residues that deviated from the standard conformational angles of Ramachandran plot were the members of N

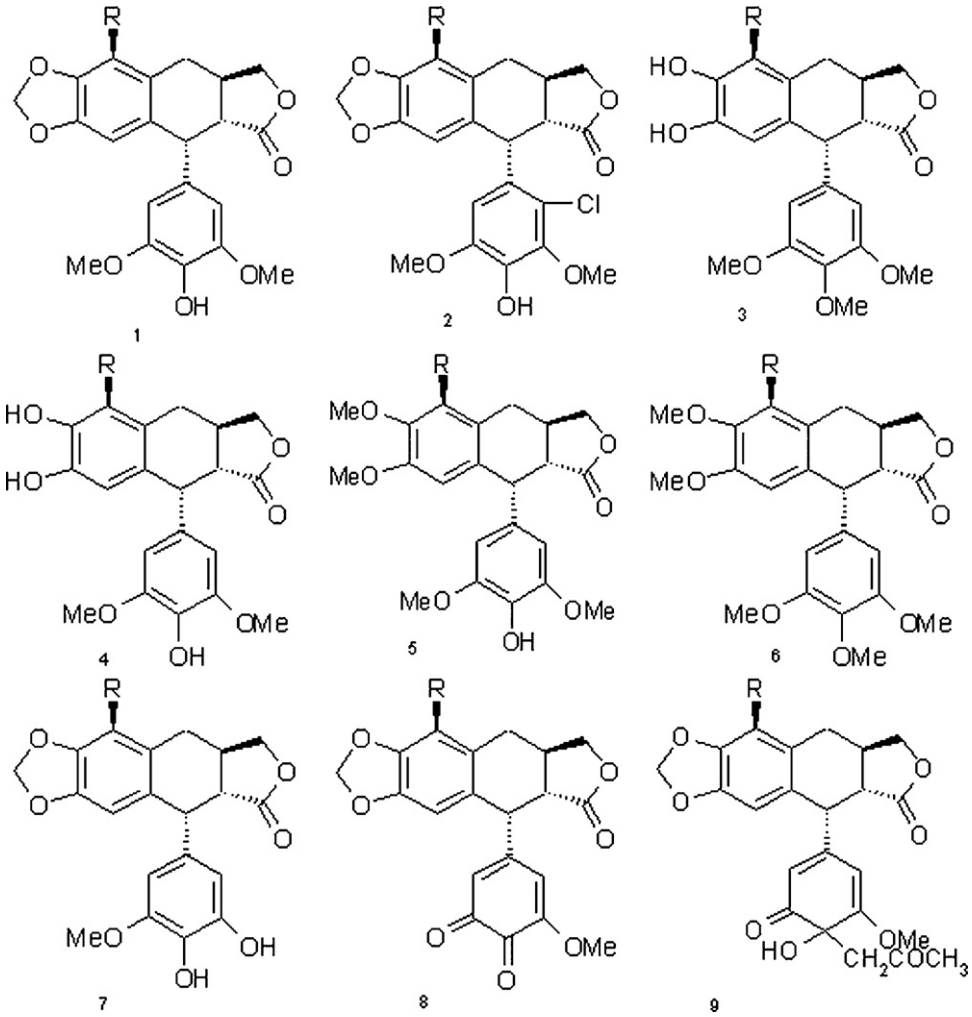


Fig. 2. The various scaffold structures used for building the epipodophyllotoxin analogs.

Table 2
Epipodophyllotoxin analogues (test set) with binding affinity expressed in terms of percentage of cellular protein DNA complex formation (PCPDCF) against the human Topoisomerase II–DNA binding domain.

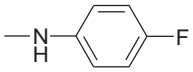
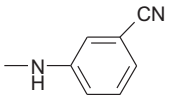
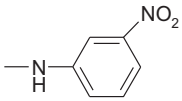
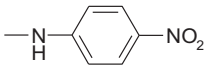
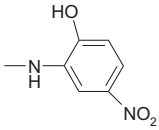
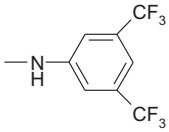
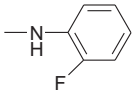
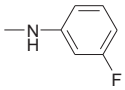
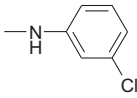
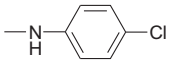
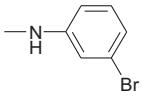
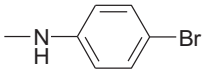
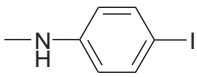
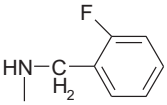
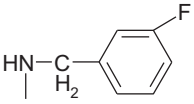
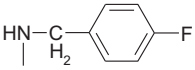
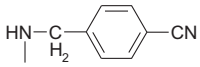
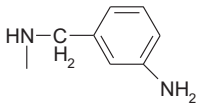
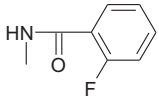
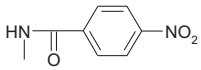
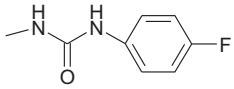
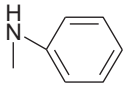
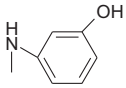
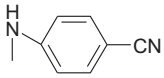
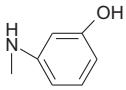
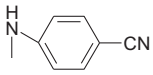
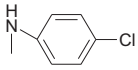
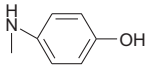
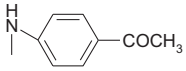
Ligands	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)
1	$-NHCH_2CH_2OH$	1	121.4	2.084
2	$-NHCH_2CH_2CH_3$	1	69.7	1.843
3	$-NHCH_2CH_2CH_2OH$	1	89.2	1.950
4		1	213	2.328
5		1	137	2.136
6		1	230	2.361
7		1	323	2.509
8		1	15	1.176
9		1	21	1.322
10		1	121	2.082
11		1	158	2.198
12		1	51	1.707
13		1	99	1.995
14		1	62	1.792
15		1	179	2.252
16		1	64	1.806
17		1	126	2.100
18		1	216	2.334
19		1	169	2.227

Table 2 (*Continued*)

Ligands	R	Structure type	Cellular protein–DNA complex formation (%)	Log(PCPDCF)
20		1	284	2.453
21		1	191	2.281
22		1	128	2.107
23		1	160	2.204
24		1	118	2.072
25		3	9	0.954
26		3	4	0.602
27		4	62	1.792
28		4	18	1.255
29		3	33	1.518
30		7	128	2.107
31		7	77	1.886
32		7	83	1.919
33		7	147	2.167

PCPDCF, percentage of cellular protein–DNA complex formation.

terminal domain of the protein. This was an ignorable condition since the N-terminal end was not critical in our study. The distance of these residues to the active site residues also were found to be more than 10 Å, which suggested that those residues would interfere little with the binding of ligands in the drug binding site region of TP IIα. The structural comparison of template protein and human TP IIα model showed significant similarity in overall structure and binding site residues (Fig. 3). Active site was identified considering the amino acids which are essential for binding of etoposide with TP IIα from experimental study. The output from the Sitemap program (Fig. 4) showed coherent active sites for the target protein as reported in site directed mutagenesis study [24].

One of the key challenges in computer-aided drug discovery is to maximize the capabilities of the method in use for predicting and rank-ordering the binding affinities of compounds for a given target protein. The efficiency of a prediction method is predominantly determined by these capabilities. Various descriptors extracted from the structural information on ligand–receptor complex may provide an advantageous solution to creating a reliable binding-affinity-prediction model. Here, we combined the results obtained from a standard docking protocol with data from structure based calculations of free energy of binding (eMBrAcE) and then investigated the utility of both the methods on the virtual screening efficiency for epipodophyllotoxin derivatives.

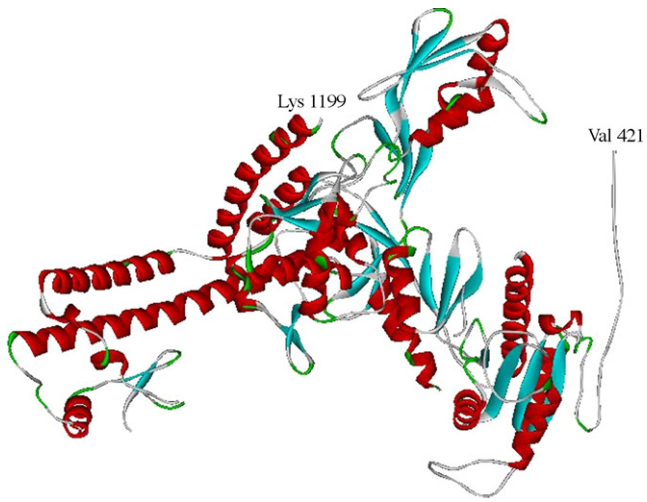


Fig. 3. The structural comparison of template (pdb ID:1BJT) and modeled structure of human TP-IIα.

Docking simulation of epipodophyllotoxin derivatives to the homology modeled TP IIα was performed using the Glide program (Schrodinger package). All the 143 epipodophyllotoxin ligands with known binding affinity expressed in terms of percentage of cellular protein–DNA complex formations (PCPDCF) were docked into the defined binding site. The binding mode of a cognate ligand within the binding site is represented in Fig. 5. In this figure we can observe that the molecule well fitted to the defined binding pocket. All the 143 epipodophyllotoxin analogues were also found to be good binder with TP IIα. For each ligand in the virtual library, the pose with the lowest Glide score was rescored using eMBrAcE. These approaches predict the binding free energy for set of ligands to receptor.

3.1. Building models for prediction of binding affinity using Glide score

Prediction model for prediction of binding of epipodophyllotoxin with TP IIα were built by considering the Glide score (GScore) as a descriptor. The Eq. (1) of the model and the corresponding

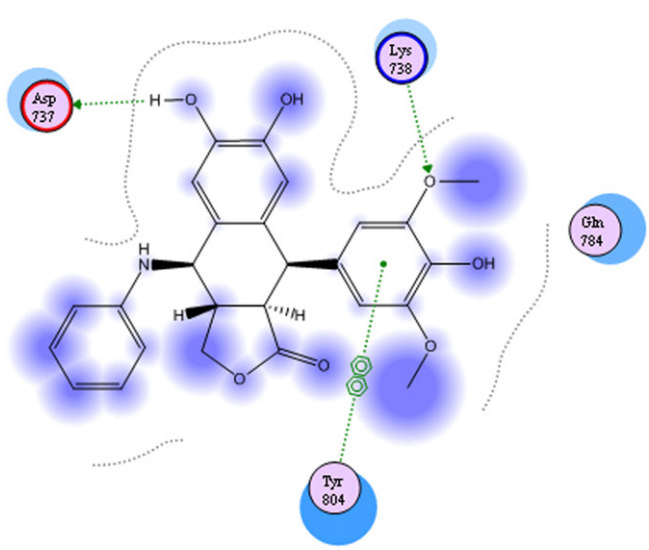


Fig. 4. Ligplot of human TP IIα-epipodophyllotoxin binding site.

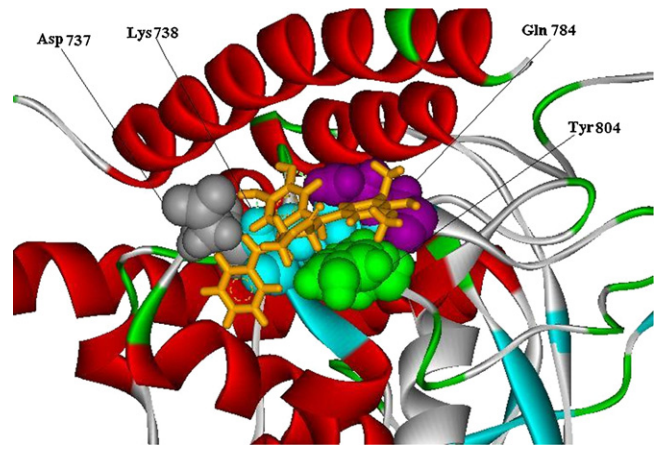


Fig. 5. Binding mode of epipodophyllotoxin derivative (5) within the binding site of human TP IIα.

statistics are shown below:

$$\log(\text{PCPDCF}) = 0.183(\pm 0.131) - 0.642(\pm 0.048) \times \text{GScore}$$
$$(N = 110;\ r^2 = 0.624;\ s = 0.321;\ F = 179.0;\ r^2_{cv} = 0.606;\ \text{PRESS} = 11.65) \quad (1)$$

The root mean square error (RMSE) between the experimental PCPDCF and the predicted PCPDCF obtained by the regression model was 0.271, which is an indicator of the robustness of the fit and suggested that the calculated PCPDCF based on Glide score is reliable. The quality of the fit can also be judged by the value of the squared correlation coefficient (r^2), which was 0.624 for the data set. Fig. 6 graphically shows the quality of fit. The statistical significance of the prediction model is evaluated by the correlation coefficient r^2, standard error, F-test value, leave-one-out cross-validation coefficient r^2_{cv} and predictive error sum of squares PRESS. The regression model developed in this study is statistically ($r^2_{cv} = 0.606$, $r^2 = 0.624$, $F = 179.0$) best fitted and consequently used for prediction of formation of complexes with TP IIα (ln PCPDCF) of the epipodophyllotoxin analogues as reported in Table 3.

3.2. Linear optimization of energy parameters vs. binding affinity

One docking structure from each molecule docking result was picked up as final docked structure in TP IIα and was imported into eMBrAcE for further calculations. As the Glide treats a receptor rigidly during docking simulation, an energy minimization was performed to the docked complex. A vdW energy and electrostatic

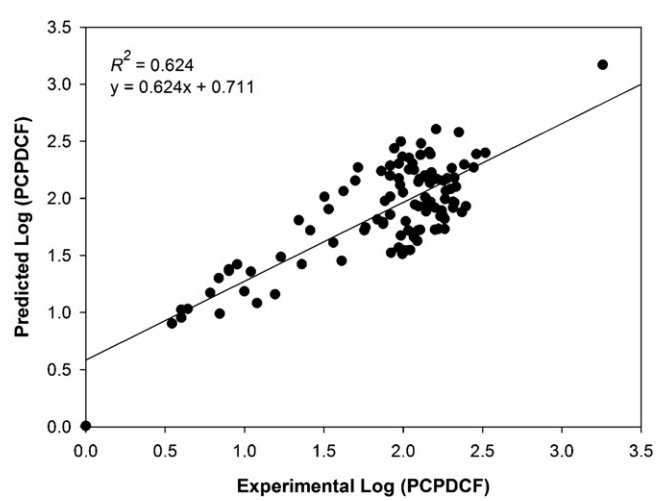


Fig. 6. Models for predicting binding affinity (log PCPDCF) of the epipodophyllotoxin derivatives based on Glide score for the training set.

Table 3
Predicted log(PCPDCF) of epipodophyllotoxin analogues using Glide score (XP) as a descriptor for training set compounds based on Eq. (1).

Ligand	GScore	Experimental log(PCPDCF)	Predicted log(PCPDCF)	Ligand	GScore	Experimental log(PCPDCF)	Predicted log(PCPDCF)
1	−2.927	1.625	2.062	56	−2.819	2.264	1.993
2	−2.124	2.044	1.547	57	−2.300	2.068	1.660
3	−2.087	1.924	1.523	58	−3.140	2.136	2.199
4	−2.413	2.223	1.732	59	−2.390	2.093	1.717
5	−3.771	2.208	2.604	60	−2.700	2.201	1.916
6	−3.432	2.462	2.386	61	−2.780	2.173	1.968
7	−3.290	2.385	2.295	62	−3.430	2.173	2.385
8	−3.110	2.324	2.180	63	−3.100	1.973	2.173
9	−1.310	0.602	1.024	64	−2.120	2.000	1.544
10	−2.720	2.396	1.929	65	−3.300	1.973	2.302
11	−2.700	2.316	1.916	66	−2.605	1.919	1.856
12	−2.852	1.919	2.014	67	−2.400	2.107	1.724
13	−3.421	2.110	2.379	68	−1.320	0.643	1.030
14	−3.070	1.699	2.154	69	−1.122	0.544	0.903
15	−2.516	2.017	1.798	70	−2.431	1.763	1.744
16	−2.640	2.371	1.878	71	−3.510	1.944	2.437
17	−3.070	2.255	2.154	72	−2.911	2.000	2.052
18	−3.090	2.214	2.167	73	−2.392	1.415	1.718
19	−3.250	2.445	2.270	74	−2.752	2.155	1.950
20	−2.320	1.986	1.672	75	−3.040	2.170	2.135
21	−2.650	2.146	1.884	76	−2.720	2.097	1.929
22	−3.604	1.986	2.497	77	−3.381	2.037	2.354
23	−2.248	2.090	1.626	78	−3.200	1.863	2.238
24	−3.114	2.146	2.182	79	−2.783	2.316	1.970
25	−3.450	2.518	2.398	80	−1.540	0.785	1.172
26	−1.830	1.041	1.358	81	−1.520	1.193	1.159
27	−2.394	1.756	1.720	82	−2.531	1.342	1.808
28	−2.681	1.531	1.904	83	−1.200	0.602	0.953
29	−1.560	1.000	1.184	84	−3.396	1.995	2.363
30	−3.100	2.278	2.173	85	−3.120	2.139	2.186
31	−2.410	2.262	1.730	86	−3.251	1.716	2.270
32	−3.138	1.919	2.198	87	−2.502	1.875	1.789
33	−2.585	2.235	1.843	88	−3.093	2.103	2.169
34	−2.793	1.886	1.976	89	−3.055	2.097	2.144
35	−2.737	2.146	1.940	90	−2.389	2.033	1.717
36	−3.240	2.307	2.263	91	−1.930	1.361	1.422
37	−2.549	2.262	1.820	92	−1.861	0.903	1.378
38	−2.930	2.269	2.064	93	−1.929	0.954	1.421
39	−2.587	2.252	1.844	94	−1.401	1.079	1.082
40	−2.030	1.230	1.486	95	−1.839	0.903	1.364
41	−2.843	2.139	2.008	96	−3.217	2.068	2.248
42	−1.741	0.839	1.301	97	−2.119	2.021	1.544
43	−3.274	1.919	2.285	98	−3.010	1.982	2.116
44	−3.182	2.179	2.226	99	−2.541	1.839	1.814
45	−2.770	2.324	1.961	100	−2.742	2.075	1.944
46	−3.303	2.060	2.304	101	−2.155	1.973	1.567
47	−2.850	1.505	2.013	102	−2.662	2.243	1.892
48	−3.900	3.258	2.687	103	−3.460	2.164	2.405
49	−2.987	2.334	2.101	104	−3.224	2.037	2.253
50	−3.577	2.114	2.480	105	−2.481	1.875	1.776
51	−3.140	2.158	2.199	106	−2.958	2.301	2.082
52	−3.730	2.352	2.578	107	−1.977	1.612	1.452
53	−2.070	1.995	1.512	108	−1.256	0.845	0.989
54	−2.401	2.201	1.725	109	−0.800	0.000	0.007
55	−3.110	2.158	2.180	110	−2.224	1.561	1.611

PCPDCF, percentage of cellular protein–DNA complex formation.

energy between ligand and receptor as well as solvation energy were calculated for each minimized complex. Also solvent accessible surface area (SASA) change was calculated using Qikprop. All these energies are listed in Table 4. A scheme similar to Linear Response was developed to predict the protein–DNA complex formation based on these energies. The predicted value of log(PCPDCF) of these analogues is listed in Table 4. The predicted activity has good correlation to the actual activity. The predicted activity based on linear optimization of different energy components used in the calculation represented the actual activity well. Several papers have been reported, in which a reasonable correlation between calculated activity and experimental activity for a small set of ligands have been obtained. Although these energy components are added directly together in most of these applications, it is still a challenge to apply these methods into large set of ligands. Normally, these different energy components (vdW, electrostatic, solvation) were calculated using more than one method. To same set of structure, different force field or different methods will produce different values of energy. This suggests that these energy components need to be scaled before an equation is obtained to get a better expression for these energy components. A set of weights can be used to scale these energies to get free energy expression by linearly combining these energies. Some scoring functions [24] used this strategy, which were optimized using a test set of molecules. In the work, a linear combination strategy was used to express the protein complex formation in presence of epipodophyllotoxin analogues by four energy components calculated from different methods. The equation of the model and the corresponding statistics are shown

Table 4
Calculated energy components and predicted log(PCPDCF) of epipodophyllotoxin analogues for the Training set compounds based on Eq. (2).

Ligand	Experimental log(PCPDCF)	ΔG_{vdW} (kcal/mol)	ΔG_{ele} (kcal/mol)	ΔG_{solv} (kcal/mol)	SASA	Predicted log(PCPDCF)
1	1.625	−120.4	−191.1	301.4	561.7	1.750
2	2.044	−91.1	445.7	−444.1	639.8	2.129
3	1.924	−80.4	300.7	−210.3	643.9	1.943
4	2.223	−51.7	324.9	−297.8	236.4	2.494
5	2.208	−68.3	275.0	−244.9	340.0	2.346
6	2.462	−108.1	−26.4	139.7	184.3	2.254
7	2.385	−143.1	45.6	91.5	400.7	2.013
8	2.324	−129.0	−18.1	162.7	396.3	1.979
9	0.602	−192.8	249.9	796.3	732.6	0.616
10	2.396	−151.1	−48.1	153.0	275.7	2.142
11	2.316	−134.2	−142.6	275.5	199.2	2.148
12	1.919	−84.0	−40.9	106.1	728.3	1.704
13	2.110	−125.4	57.8	36.5	349.5	2.133
14	1.699	−106.5	−65.9	150.3	853.5	1.523
15	2.017	−62.7	−257.1	280.7	317.0	2.108
16	2.371	−143.9	65.8	55.3	416.0	2.026
17	2.255	−41.2	−528.6	583.6	65.5	2.215
18	2.214	−107.8	56.4	95.1	281.8	2.141
19	2.445	−154.8	86.3	71.0	235.5	2.191
20	1.986	−172.1	189.5	−5.3	106.2	2.351
21	2.146	−116.1	243.9	−602.7	774.6	2.314
22	1.986	−36.3	336.2	−314.0	668.1	2.027
23	2.090	−139.9	641.2	−531.5	737.0	1.977
24	2.146	−57.5	−388.0	469.8	291.4	1.999
25	2.518	−82.8	957.8	−841.1	392.7	2.527
26	1.041	−138.3	298.2	577.5	692.0	0.908
27	1.756	−102.6	190.5	68.5	347.4	2.005
28	1.531	−104.3	−167.0	272.1	665.0	1.657
29	1.000	−76.0	−55.5	738.7	715.8	0.950
30	2.278	−53.2	949.4	−867.3	724.7	2.201
31	2.262	−131.4	1381.9	−1186.5	767.6	2.219
32	1.919	−87.8	733.1	−708.1	722.1	2.156
33	2.235	−93.1	966.1	−837.8	830.4	2.026
34	1.886	−104.4	612.6	−484.1	659.7	2.033
35	2.146	−67.2	344.5	−260.1	698.1	1.915
36	2.307	−57.4	−15.3	76.1	651.0	1.814
37	2.262	−100.7	−245.1	270.0	366.1	2.050
38	2.269	−73.9	951.0	−934.1	694.1	2.312
39	2.252	−128.5	1359.6	−1200.9	707.7	2.320
40	1.230	−64.5	460.7	374.1	771.1	0.968
41	2.139	−102.5	578.8	−461.2	798.9	1.874
42	0.839	−68.6	−1185.2	949.2	722.6	1.499
43	1.919	−82.1	−346.4	419.7	315.2	1.999
44	2.179	−144.7	−303.5	423.6	116.8	2.172
45	2.324	−35.2	−162.8	211.5	312.2	2.136
46	2.060	−137.7	13.4	149.0	287.7	2.093
47	1.505	−126.5	153.5	−4.9	638.0	1.793
48	3.258	−69.6	136.2	−797.9	203.0	3.280
49	2.334	−92.3	436.7	−317.1	322.7	2.333
50	2.114	−76.3	186.4	−159.0	306.0	2.340
51	2.158	−86.7	601.9	−403.1	241.1	2.412
52	2.352	−94.2	507.7	−411.7	231.2	2.500
53	1.995	−75.3	516.6	−416.9	765.3	1.908
54	2.201	−108.2	499.7	−341.8	330.5	2.306
55	2.158	−119.9	323.0	−243.1	309.4	2.332
56	2.264	−85.3	536.1	−424.7	473.2	2.227
57	2.068	−89.4	185.7	−17.4	334.7	2.131
58	2.136	−182.8	−58.7	241.7	335.2	1.968
59	2.093	−110.5	−104.7	173.9	393.9	2.034
60	2.201	−77.5	−57.9	120.2	404.9	2.061
61	2.173	−108.0	−113.8	251.2	258.5	2.097
62	2.173	−145.3	74.1	105.5	231.2	2.164
63	1.973	−56.2	−174.0	224.3	274.1	2.167
64	2.000	−102.5	−272.1	359.0	321.3	2.009
65	1.973	−88.7	−212.1	287.1	233.3	2.156
66	1.919	−49.8	−62.8	119.7	241.3	2.254
67	2.107	−133.3	263.5	−110.9	219.4	2.310
68	0.643	−216.2	238.6	884.6	595.9	0.663
69	0.544	−251.8	−326.9	906.4	761.0	0.852
70	1.763	−105.1	−173.1	243.0	726.2	1.629
71	1.944	−113.2	−14.7	116.9	447.2	1.979
72	2.000	−181.5	43.9	79.2	318.1	2.114
73	1.415	−164.4	−21.8	139.5	886.2	1.456
74	2.155	−77.0	−131.3	246.0	300.4	2.076
75	2.170	−109.4	46.3	57.2	246.5	2.234

Table 4 (*Continued*)

Ligand	Experimental log(PCPDCF)	ΔG_{vdW} (kcal/mol)	ΔG_{ele} (kcal/mol)	ΔG_{solv} (kcal/mol)	SASA	Predicted log(PCPDCF)
76	2.097	−87.3	−114.1	224.0	244.6	2.151
77	2.037	−62.7	242.1	−214.6	319.5	2.356
78	1.863	−84.0	354.4	−309.8	592.9	2.083
79	2.316	−91.8	−78.4	195.1	257.4	2.145
80	0.785	−124.9	−59.5	690.8	937.7	0.754
81	1.193	−79.7	−285.4	270.9	576.4	1.848
82	1.342	−113.3	89.2	−1.2	871.3	1.577
83	0.602	−236.8	285.2	1278.0	270.6	0.503
84	1.995	−104.7	−98.9	157.4	738.6	1.667
85	2.139	−56.2	−167.1	207.5	477.8	1.955
86	1.716	−83.5	3.9	69.7	712.2	1.735
87	1.875	−86.0	6.5	91.5	795.7	1.612
88	2.103	−74.8	−103.0	165.7	449.8	1.988
89	2.097	−131.5	46.5	99.2	782.7	1.579
90	2.033	−147.8	−49.3	208.9	407.4	1.928
91	1.361	−85.7	−318.2	303.0	752.1	1.634
92	0.903	−183.6	−205.8	362.5	996.8	1.186
93	0.954	−78.2	3.0	72.8	970.3	1.445
94	1.079	−98.5	−91.3	180.6	894.9	1.460
95	0.903	−149.6	−19.3	170.0	980.8	1.314
96	2.068	−36.3	−184.5	209.8	597.0	1.836
97	2.021	−72.9	−19.4	77.4	666.0	1.796
98	1.982	−64.1	−316.0	317.3	691.2	1.688
99	1.839	−155.2	5.7	104.7	723.4	1.663
100	2.075	−131.7	51.8	65.7	684.2	1.726
101	1.973	−105.3	−148.8	231.3	744.6	1.605
102	2.243	−91.5	−25.3	103.3	230.9	2.250
103	2.164	−124.1	−71.3	160.8	693.9	1.689
104	2.037	−105.2	−376.6	349.0	318.4	2.100
105	1.875	−67.7	−352.3	386.9	681.8	1.638
106	2.301	−191.6	−142.0	345.6	25.1	2.244
107	1.612	−122.8	−7.3	103.8	666.2	1.744
108	0.845	−50.3	−329.5	928.8	668.3	0.972
109	0.000	−149.9	1062.9	372.4	728.6	0.566
110	1.561	−143.3	1045.0	−138.7	572.7	1.385

PCPDCF, percentage of cellular protein–DNA complex formation.

below:

$$\log(\text{PCPDCF}) = 2.636 + 0.00021\,\Delta G_{vdW} - 0.00072\,\Delta G_{ele} - 0.00123\,\Delta G_{solv} - 0.00112\,\text{SASA} \quad (N = 110,\ r^2 = 0.800,\ s = 0.237,\ F = 105.1,\ r^2_{cv} = 0.774,\ \text{PRESS} = 6.672) \tag{2}$$

The statistical significance of the prediction model is evaluated by the correlation coefficient r^2, standard errors, F-test value, leave-one-out cross-validation coefficient r^2_{cv} and predictive error sum of squares PRESS. The regression model developed based on linear combination of energy components in this study is statistically ($r^2_{cv} = 0.774$; $r^2 = 0.800$ and $F = 105.1$) best fitted (Fig. 7) and consequently used for prediction of PCPDCF of the epipodophyllotoxin analogues as reported in Table 4. The root mean square error (RMSE) between the experimental PCPDCF values and the predicted PCPDCF values obtained by the regression model was also very less (0.182), which is an indicator of the robustness of the fit and suggested that the calculated PCPDCF based on above structure based approach is reliable.

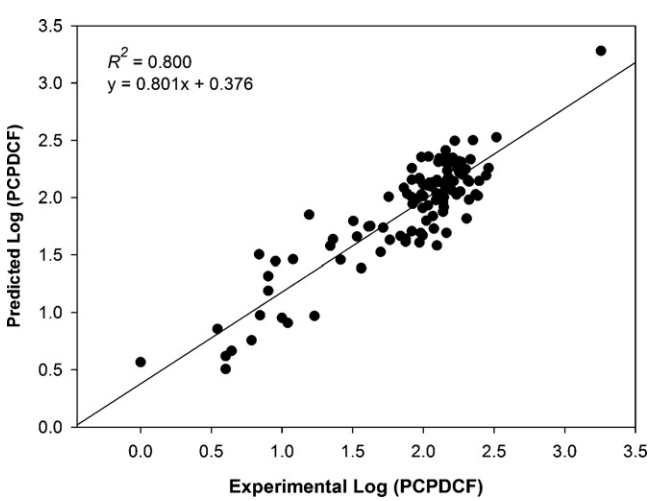


Fig. 7. Models for predicting binding affinity (log PCPDCF) of the epipodophyllotoxin derivatives based on eMBrAcE for the training set.

To judge the accuracy of the prediction models developed, we have taken a separate data set called as test set consisting of 33 compounds (Table 2). Their potencies and chemical structures were obtained from literature [23]. Experimentally determined biological activity of the drugs based on in vitro study is also provided in order to evaluate the accuracy of predictions. For all compounds, both the prediction models (Eqs. (1) and (2)) produce exactly the same trend for relative potencies, even though the exact magnitudes of these values do not match very well (Tables 5 and 6). The overall RMSE between the experimental and predicted PCPDCF value was in the range of 0.229–0.236, which means that both the structure based modeling were able to predict the activity of 33 epipodophyllotoxin analogues more reliably. Figs. 8 and 9 graphically shows the quality of fit for the test set. These results indicate that methodologies with a better prediction precision in binding affinities such as Glide XP and eMBrAcE, though more time-consuming, in comparison to Glide SP, can provide a significant advantage in prioritizing candidate compounds with high biological activity (low micromolar or nanomolar activity).

We have presented herein a FEB calculation on the binding affinity of 143 epipodophyllotoxin derivatives with TP IIα. The

Table 5
Predicted log PCPDCF of epipodophyllotoxin analogues using Glide score (XP) as a descriptor for test set compounds based on Eq. (1).

Ligand	GScore	Experimental log(PCPDCF)	Predicted log(PCPDCF)	Ligand	GScore	Experimental log(PCPDCF)	Predicted log(PCPDCF)
1	−3.531	2.084	2.450	18	−3.106	2.334	2.177
2	−2.470	1.843	1.769	19	−3.350	2.227	2.334
3	−3.442	1.950	2.393	20	−3.590	2.453	2.488
4	−3.069	2.328	2.153	21	−3.705	2.281	2.562
5	−2.552	2.136	1.821	22	−2.596	2.107	1.849
6	−3.561	2.361	2.470	23	−2.372	2.204	1.706
7	−3.750	2.509	2.591	24	−3.050	2.072	2.141
8	−1.600	1.176	1.210	25	−1.955	0.954	1.438
9	−2.360	1.322	1.698	26	−1.350	0.602	1.050
10	−2.653	2.082	1.887	27	−2.501	1.792	1.789
11	−2.830	2.198	2.000	28	−2.330	1.255	1.679
12	−2.255	1.707	1.631	29	−1.947	1.518	1.433
13	−2.594	1.995	1.849	30	−2.288	2.107	1.652
14	−2.660	1.792	1.891	31	−2.517	1.886	1.799
15	−2.541	2.252	1.814	32	−2.197	1.919	1.594
16	−2.826	1.806	1.997	33	−2.377	2.167	1.709
17	−2.902	2.100	2.046				

PCPDCF, percentage of cellular protein–DNA complex formation.

magnitude of the binding affinity can be a key factor that decides the activeness of an individual inhibitor. An energetic evaluation of the binding affinity will provide a way to estimate the activity of inhibitors. In any binding energy calculation, the correct binding structure of each ligand has to be determined first prior to binding energy estimation. No experimental structure of epipodophyllotoxin with TP IIα is available. We used flexible docking to determine the binding structure of the epipodophyllotoxin analogues with TP IIα. Very similar binding structures were obtained for a set of analogues. This makes a credible prediction model of the biological activity (PCPDCF) calculation possible. The calculated docking scores and binding free energy value of a set of structural analogues demonstrates excellent linear correlation to the experimental activity. These models could be useful to predict the range of activities for new epipodophyllotoxin analogues. The information that we have expressed in this study may lead to the designing (synthesis) of more potent epipodophyllotoxin derivatives for inhibition of TP IIα. Although the current study does not involve a large number of receptors and test sets of compounds, our evaluation data should add valuable information that may enhance the practice of computerized drug discovery.

Table 6
Calculated energy components and predicted log(PCPDCF) of epipodophyllotoxin analogues for the test set compounds based on Eq. (2).

Ligand	Experimental log(PCPDCF)	ΔG_{vdW} (kcal/mol)	ΔG_{ele} (kcal/mol)	ΔG_{solv} (kcal/mol)	SASA	Predicted log(PCPDCF)
1	2.084	−31.1	257.1	−167.3	605.8	1.972
2	1.843	−96.6	215.4	−165.6	652.6	1.934
3	1.950	−125.8	457.4	−325.5	624.3	1.982
4	2.328	−54.8	−99.1	134.6	270.3	2.228
5	2.136	−28.4	−50.2	102.5	682.7	1.776
6	2.361	−115.7	−241.4	311.8	326.7	2.037
7	2.509	−82.1	−177.4	213.2	236.2	2.220
8	1.176	−57.8	17.3	158.8	725.0	1.728
9	1.322	−103.8	−56.8	187.4	773.0	1.560
10	2.082	−164.9	−58.0	162.7	691.6	1.669
11	2.198	−122.0	−110.3	216.4	357.9	2.023
12	1.707	−154.7	−159.9	344.3	709.2	1.502
13	1.995	−129.1	152.2	14.6	692.8	1.706
14	1.792	−121.0	121.6	5.1	694.1	1.740
15	2.252	−96.2	70.9	−6.1	592.1	1.798
16	1.806	−83.4	−246.2	264.1	712.7	1.673
17	2.100	−127.6	621.7	−492.8	695.9	1.989
18	2.334	−121.4	402.5	−276.7	603.1	1.874
19	2.227	−21.7	192.3	−221.2	719.2	1.960
20	2.453	−138.3	530.4	−404.6	430.0	2.242
21	2.281	−125.7	431.8	−335.2	695.0	1.933
22	2.107	−117.0	104.2	24.1	716.2	1.705
23	2.204	−109.6	−73.6	198.7	345.6	2.035
24	2.072	−158.3	−27.4	171.3	511.1	1.840
25	0.954	−15.4	−116.9	960.4	727.2	0.722
26	0.602	−194.7	973.6	440.5	729.3	0.536
27	1.792	−129.9	−11.4	112.4	737.6	1.653
28	1.255	−151.8	−127.0	214.7	722.7	1.623
29	1.518	−109.1	−176.7	258.4	774.9	1.555
30	2.107	−85.7	−194.2	268.3	660.3	1.689
31	1.886	−117.2	36.6	67.6	667.2	1.755
32	1.919	−90.9	−40.1	77.1	667.8	1.804
33	2.167	−113.4	21.0	108.0	492.2	1.914

PCPDCF, percentage of cellular protein–DNA complex formation.

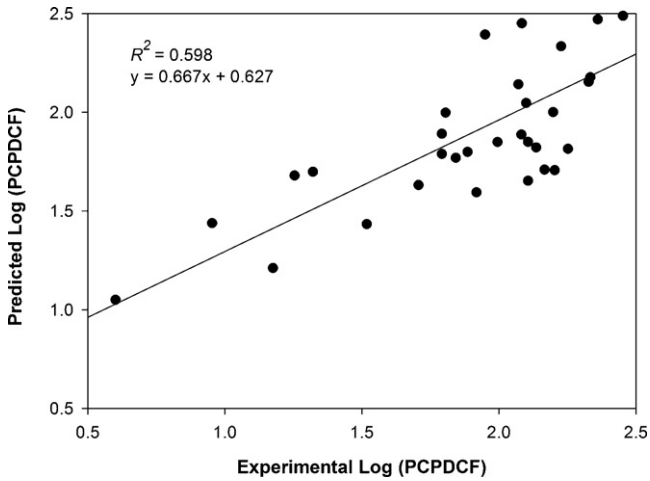


Fig. 8. Models for predicting binding affinity (log PCPDCF) of the epipodophyllotoxin derivatives based on Glide score for the test set.

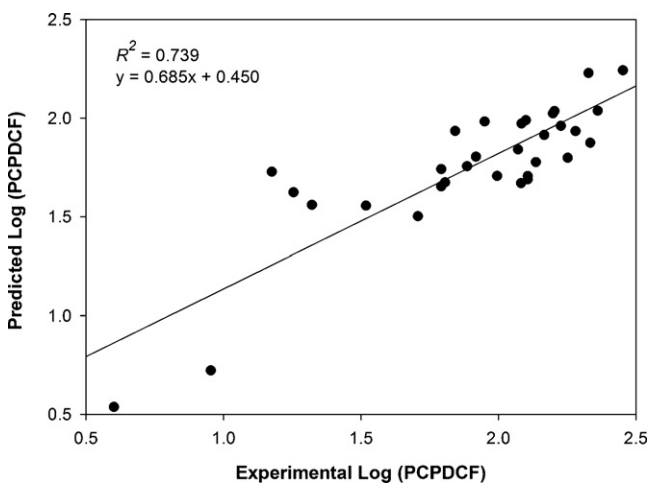


Fig. 9. Models for predicting binding affinity (log PCPDCF) of the epipodophyllotoxin derivatives based on eMBrAcE for the test set.

4. Conclusion

The binding structures of these ligands in human TP IIα were predicted by flexible docking simulations. They bind in a similar position inside the TP IIα drug binding site and try to fit the binding pocket well. The calculated FEB for these ligands reasonably predicted the activity of this set of ligands. The calculated activity has good correlation to experimental activity. The result shows that the linear combination of four energy terms: vdW, electrostatic, solvation (electrostatic part), and nonpolar energies optimized by regression has power to express the binding affinity of large set of ligands in receptor. The Docking and eMBrAcE demonstrates a good ability on the binding structure prediction and binding energy determination to produce reasonable energies. This work suggests that in the relative FEB calculation, which is major interest in drug design, the contribution of different energy terms can be scaled by a set of weight factors to reach a good correlation. In practice, it is know that same energy term plays different role in different type of systems. This is one of the reasons that a reasonable activity model can be obtained just based on some energy terms. The calculation of solvation effect upon a ligand binding in a protein is a challenge work. This work and many others have shown that solvation effect is an important driving force on ligand binding and a key factor in expression of activity of a set of ligands. In the work, GB and SASA methods were used to estimate the electrostatic and the nonpolar parts of solvation and produced satisfactory results in terms of good correlation with experimental activity.

References

[1] J.J. Champoux, DNA topoisomerases: structure, function, and mechanism, Annu. Rev. Biochem. 70 (2001) 369–413.
[2] J.C. Wang, Cellular roles of DNA topoisomerases: a molecular perspective, Nat. Rev. Mol. Cell. Biol. 3 (2002) 430–440.
[3] J.C. Wang, DNA topoisomerases, Annu. Rev. Biochem. 65 (1996) 635–692.
[4] D.A. Burden, N. Osheroff, Mechanism of action of eukaryotic topoisomerase II and drugs targeted to the enzyme, Biochim. Biophys. Acta 1400 (1998) 139–154.
[5] T. Goto, C. Holm, J.C. Wang, D. Botstein, DNA topoisomerase II is required at the time of mitosis in yeast, Cell 41 (1985) 553–563.
[6] A.H. Corbett, N. Osheroff, When good enzymes go bad: conversion of topoisomerase II to a cellular toxin by antineoplastic drugs, Chem. Res. Toxicol. 6 (1993) 585–597.
[7] S.J. Froelich-Ammon, N. Osheroff, Topoisomerase poisons: harnessing the dark side of enzyme mechanism, J. Biol. Chem. 270 (1995) 21429–21432.
[8] G. Capranico, M. Binaschi, DNA sequence selectivity of topoisomerases and topoisomerase poisons, Biochim. Biophys. Acta 1400 (1998) 185–194.
[9] G. Capranico, G. Zagotto, M. Palumbo, Development of DNA topoisomerase-related therapeutics: a short perspective of new challenges, Curr. Med. Chem. Anti-Cancer Agents 4 (2004) 335–345.
[10] K.R. Hande, Etoposide: four decades of development of a topoisomerase II inhibitor, Eur. J. Cancer 34 (1998) 1514–1521.
[11] K.R. Hande, Topoisomerase II inhibitors, Cancer Chemother. Biol. Response Modif. 21 (2003) 103–125.
[12] E.L. Baldwin, N. Osheroff, Etoposide, topoisomerase II and cancer, Curr. Med. Chem. Anti-Cancer Agents 5 (2005) 363–372.
[13] S.H. Kaufmann, Cell death induced by topoisomerase-targeted drugs: more questions than answers, Biochim. Biophys. Acta 1400 (1998) 195–211.
[14] O. Sordet, Q.A. Khan, K.W. Kohn, Y. Pommier, Apoptosis induced by topoisomerase inhibitors, Curr. Med. Chem. Anti-Cancer Agents 3 (2003) 271–290.
[15] J. Aisner, C.P. Belani, L.A. Doyle, Etoposide: current status and future perspectives in the management of malignant neoplasms, Cancer Chemother. Pharmacol. 34 (1994) S118.
[16] K.H. Lee, Y. Imakura, M. Haruna, S.A. Beers, L.S. Thurston, H.J. Dai, C.H. Chen, Antitumor Agents 170. New cytotoxic 4-alkylamino analogues of 4′-demethylepipodophyllotoxin as inhibitors of human DNA toposiomerase II, J. Nat. Prod. 52 (1989) 606–613.
[17] Z.Q. Wang, Y.H. Kuo, D. Schnur, J.P. Bowen, S.Y. Liu, F.S. Han, Y.C. Cheng, K.H. Lee, Antitumor agents 113. New 4-arylamino derivatives of 4′-O-demethyl-epipodophyllotoxin and related compounds as potent inhibitors of human DNA toposiomerase II, J. Med. Chem. 33 (1990) 2660–2666.
[18] K.H. Lee, S.A. Beers, M. Mori, Z.Q. Wang, Y.H. Kuo, I. Li, S.Y. Liu, J.Y. Cheng, F.S. Han, Y.C. Cheng, Antitumor agents. 111. New 4-hydroxylated and 4-halogenated anilino derivatives of 4′-demethylepipodophyllotoxin as potent inhibitors of Human DNA toposiomerase II, J. Med. Chem. 33 (1990) 1364–1368.
[19] X.M. Zhou, Z.Q. Wang, J.Y. Cheng, H.X. Chen, Y.C. Cheng, K.H. Lee, Antitumor agents. 120. New 4-substituted benzylamine and benzyl ether derivatives of 4′-O- demethyl-epipodophyllotoxin as potent inhibitors of human DNA toposiomerase II, J. Med. Chem. 33 (1991) 3346–3350.
[20] H. Hu, Z.Q. Wang, S.Y. Liu, Y.C. Cheng, K.H. Lee, Antitumor agents. 123. Synthesis and human DNA topoisomerase II inhibitory activity of 2′-chloro derivatives of etoposide and 4β-(arylamino)-4′-O-demethylpodophyllotoxins, J. Med. Chem. 35 (1992) 866–871.
[21] Z.Q. Wang, H. Hu, H.X. Chen, Y.C. Cheng, K.H. Lee, Antitumor Agents. 124. New 4β-substituted aniline derivatives of 6,7-O, O-demethylpodophyllotoxin as potent inhibitors of Human DNA toposiomerase II, J. Med. Chem. 35 (1992) 871–877.
[22] X.M. Zhou, Z.Q. Wang, H.X. Chen, Y.C. Cheng, K.H. Lee, Antitumor Agents. 125. New 4-benzoylamino derivatives of 4′-O-demethyl-4-desoxypodophyllotoxin and 4-benzoyl derivatives of 4′-O-demethylpodophyllotoxin as potent inhibitors of human DNA toposiomerase II, Pharm. Res. 10 (1993) 214–219.
[23] Z.Q. Wang, Y.C. Cheng, H.X. Chen, J.Y. Cheng, X. Guo, Y.C. Cheng, K.H. Lee, Antitumor Agents. 126. Novel 4β-substituted anilino derivatives of 3,4′-O, O-didemethylpodophyllotoxin as potent inhibitors of human DNA toposiomerase II, Pharm. Res. 10 (1993) 343–350.
[24] M. Miyahara, Y. Kashiwada, X. Guo, H.X. Chen, Y.C. Cheng, K.H. Lee, Nitrosourea derivatives of 3′,4′-dioxo-4-deoxypodophyllotoxin and urea derivatives of 4′-O-demethylpodophyllotoxin as potent inhibitors of Human DNA toposiomerase II, Heterocycles 39 (1994) 361–369.
[25] Y.L. Zhang, X. Guo, Y.C. Cheng, K.H. Lee, Antitumor Agents. 148.Synthesis and biological evaluation of novel 4-amino derivatives of etoposide with better pharmacological profiles, J. Med. Chem. 37 (1994) 446–452.
[26] Z. Ji, H.K. Wang, K.F. Bastow, X.K. Zhu, S.J. Cho, Y.C. Cheng, K.H. Lee, Antitumor agents. 177. Design, synthesis and biological evaluation of novel etoposide analogues bearing pyrrolecarboxamidino group as DNA topoisomerase II inhibitors, Bioorg. Med. Chem. Lett. 7 (1997) 607–612.
[27] X.K. Zhu, J. Guan, Y. Tachibana, K.F. Bastow, S.J. Cho, H.H. Cheng, Y.C. Cheng, M. Gurwith, K.H. Lee, Antitumor agents. 194. Synthesis and biological evaluation of

4-β-Mono-,-Di-,and –trisubstituted aniline-4′-demethyl-podophyllotoxin and related compounds with improved pharmacological profiles, J. Med. Chem. 42 (1999) 2441–2446.
[28] D. Leroy, A.V. Kajava, C. Frei, S.M. Gasser, Analysis of etoposide binding to subdomains of human DNA topoisomerase II in the absence of DNA, Biochemistry 40 (2001) 1624–1634.
[29] D. Fass, C.E. Bogden, J.M. Berger, Quaternary changes in topoisomerase II may direct orthogonal movement of two DNA strands, Nat. Struct. Biol. 6 (1999) 322–326.
[30] Prime Version 1.5, Macromodel Version 9.1, Schrodinger, LLC, New York, NY, 2005.
[31] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[32] G.N. Ramachandran, C. Ramakrishnan, V. Sasisekharan, Stereochemistry of polypeptide chain configurations, J. Mol. Biol. 7 (1963) 95–99.
[33] D. Eisenberg, R. Luthy, J.U. Bowie, VERIFY3D: assessment of protein models with three-dimensional profiles, Methods Enzymol. 277 (1997) 396–404.
[34] M.J. Robinson, A.H. Corbett, N. Osheroff, Effects of topoisomerase II targeted drugs on enzyme-mediated DNA cleavage and ATP hydrolysis: evidence for distinct drug interaction domains on topoisomerase II, Biochemistry 32 (1993) 3638–3643.
[35] G. Capranico, G. Zagotto, M. Palumbo, Development of DNA topoisomerase-related therapeutics: a short perspective of new challenges, Curr. Top. Med. Chem. 4 (2004) 335–345.
[36] Schrodinger LLC., http://www.schrodinger.com, Schrodinger, Inc., Portland, OR, 2000 (accessed: 24. 04.2007).
[37] R.A. Friesner, J.L. Banks, R.B. Murphy, T.A. Halgren, J.J. Klicic, D.T. Mainz, M.P. Repasky, E.H. Knoll, M. Shelley, J.K. Perry, D.E. Shaw, P. Francis, P.S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring 1. Method and assessment of docking accuracy, J. Med. Chem. 47 (2004) 1739–1749.
[38] M.D. Eldridge, C.W. Murray, T.R. Auton, G.V. Paolini, R.P. Mee, Empirical scoring functions: the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes, J. Comput-Aided Mol. Des. 11 (1997) 425–445.
[39] O. Guvench, J. Weiser, P.S. Shenkin, I. Kolossváry, W.C. Still, Application of the frozen atom approximation to the GB/SA continuum model for solvation free energy, J. Comput. Chem. 23 (2002) 214–221.
[40] X. Wu, J.L.S. Milne, M.J. Borgnia, A.V. Rostapshov, S. Subramaniam, B.R. Brooks, A core-weighted fitting method for docking atomic structures into low-resolution maps: application to cryo-electron microscopy, J. Struct. Biol. 141 (2003) 63–76.
[41] N.P. Todorov, R.L. Mancera, P.H. Monthoux, A new quantum stochastic tunneling optimisation method for protein-ligand docking, Chem. Phys. Lett. 369 (2003) 257–263.
[42] C.H. Reynolds, Estimating liphophilicity using GB/SA continuum solvation model: a direct method for computing partition coefficients, J. Chem. Inf. Comput. Sci. 35 (1995) 738–742.