European Journal of Medicinal Chemistry 45 (2010) 2132–2140

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

3D QSAR pharmacophore based virtual screening and molecular docking for identification of potential HSP90 inhibitors

Sugunadevi Sakkiah [a], Sundarapandian Thangapandian [a], Shalini John [a], Yong Jung Kwon [b], Keun Woo Lee [a,*]

[a] Division of Applied Life Science (BK21 Program), Environmental Biotechnology National Core Research Center (EB-NCRC), Plant Molecular Biology and Biotechnology Research Center (PMBBRC), Gyeongsang National University, Jinju 660-701, Republic of Korea
[b] Department of Chemical Engineering, Kangwon National University, Chunchon 200-701, Republic of Korea

ARTICLE INFO

Article history:
Received 8 October 2009
Received in revised form
22 December 2009
Accepted 9 January 2010
Available online 4 February 2010

Keywords:
HSP90
Pharmacophore
HYPOGEN
Molecular docking
LigandFit
Cancer

ABSTRACT

Chemical features based 3D pharmacophore models were developed for HSP90 based on the known inhibitors using Discovery Studio V2.1. An optimal pharmacophore model was brought forth and validated using a decoy set, external test set and Fischer's randomization method. The best five features pharmacophore model, Hypo1, includes two hydrogen bond acceptors, three hydrophobic features, which has the highest correlation coefficient (0.93), cost difference (73.88), low RMS (1.24), as well as it shows a high goodness of fit and enrichment factor. Hypo1 was used as a 3D query for virtual screening to retrieve potential inhibitors from Maybridge and Scaffold databases. The hit compounds were subsequently subjected to molecular docking studies and finally, 36 compounds were obtained based on consensus scoring function.

1. Introduction

Over the last few years molecular chaperones are emerging as a hottest and the most important topic in biology since its plays an essential role in apoptosis as well as in oncogenesis. Molecular chaperons are responsible for maintaining the appropriate folding and 3D conformation of proteins in the cell. They also play critical roles in controlling the balance between the synthesis and degradation of many proteins [1] and some chaperons are particularly important under conditions like where protein damage is induced. Heat shock protein 90 (HSP90), is an ATP-dependent molecular chaperone, which is an exciting new target for the development of innovative chemogenomics approach [2–6]. Among several vital chaperones and co-chaperones, it is one of the most widely analyzed [7–9] and highly conserved among various species [10–12]. In eukaryotic cells, HSP90 comprising 1–2% of cellular proteins under non-stress conditions [13] which can be found in cystosol, nucleoplasm, endoplasmic reticulum (ER), mitochondria and chloroplasts [14,15]. There are two major cytoplasmic isoforms of HSP90, they are HSP90a (inducible/major form) and HSP90b (constitutive/minor form) [16,17]. Its function is required for oncogenes maturation and it also plays a key role in the hallmark of malignancy like assisting survival, invasion, proliferation, metastasis and angiogenesis. Human HSP90 family includes 17 genes which fall into four classes: HSP90AA, HSP90AB, HSP90B and TRAP [13]. Its expression is associated with many types of tumors including breast cancer, pancreatic carcinomia, human leukemia and systemic lupus erythematosus as well as multidrug resistance. Raf-2, HERc2, Src-family kinases, steroid hormone receptors, polo-1-kinase, death domain kinase, protein kinase B, focal adhesion kinase, telomerase, hypoxia inducible factor and MET kinases are HSP90 dependent client proteins [18], dysregulation of these client protein pathways are associated with tumor pathology [19].

HSP90 forms homodimer, the contact sites are localized within the C-terminus. Three functional domains are involved in its mechanism: (i) highly conserved N-terminal domain (NTD) of ~25 kDa, which includes an ATP-binding cleft (ii) protein binding domain of ~40 kDa and (iii) C-terminal domain (CTD) of 12 kDa, responsible for the constitutive homodimerization domain, each of which play a crucial role in the protein [20]. Yeast and human N-terminal domains are very similar in structure and which belongs to GHKL super family [21]. N-terminal domain contains 9 α-helices and 8 anti-parallel β-pleated sheets, which combines to form

* Corresponding author. Tel.: +82 55 751 6276; fax: +82 55 752 7062.
E-mail address: kwlee@gnu.ac.kr (K.W. Lee).

0223-5234/$ – see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2010.01.016

several α/β sandwiches. The backbone of the N-terminal domain comprises of eight-stranded β-sheets, ATP-binding site which is formed by four α-helices and the overall structural configuration is referred to as a "Bergerat fold". N-terminal domain of HSP90 exhibits structural similarity to proteins such as bacterial DNA gyrase B, topoisomerase II, histidine kinase and MutL mismatch repair proteins [22]. Under normal condition HSP have vital functions like regulating the correct folding, degradation, localization and growth function [3–5]. There are accumulating evidences supports that HSPs may play an important role in cancer and particularly HSP90 is an important new target for cancer therapeutic intervention [4,5,23]. HSP90 inhibitors block the growth of cancers containing a range of genetic and molecular abnormalities, it can be speculated that they will have broad-spectrum anticancer activity against many tumors. Although, HSP90 is not a cancer gene, it's a promising molecular target in cancer biology because of the following reasons: (i) the key role of HSP90 activity involves in various oncogenic proteins pathway (ii) two different natural products (HSP90 inhibitors) which shows anticancer activity in animal models [18]. Therefore, it is believed to be an attractive and potential target in cancer biology.

Inhibition of HSP90 leads to deregulation of many crucial pathways like (i) self-sufficiency in growth signals (ii) tissues invasion/metastasis (iii) insensitivity to antigrowth signals (iv) sustained angiogenesis (v) evasion of apoptosis and (vi) limitless replicative potential which are responsible for the cancer cell survival [24]. Inhibition of HSP90 will be an important pharmacological platform for anticancer therapy [13,25]. Radicicol and geldanamycin are the two natural products known to be potent inhibitors of HSP90's chaperone activity. The goal of this study is to generate a 3D pharmacophore models based on the known HSP90 inhibitors, which can correctly reflect the structure–activity relationship (SAR) of the existing HSP90 inhibitors. Further employment of this pharmacophore model will be used as a 3D query for searching large databases to identify novel HSP90 inhibitors and also to utilize this pharmacophore model as a predictive tool for estimating biological activity of HSP90 inhibitors through virtual screening or molecular designing on the basis of structure–activity analysis.

2. Results and discussion

2.1. HYPOGEN model for HSP90 inhibitors

A training set of 16 compounds with diverse scaffold were collected from literature and tested with the same assay (FP assay) have been used in this study. This diverse training set includes the well known two natural (Radicicol and Geldanamycin) potent inhibitors of HSP90. Structures and biological activities of the training set compounds are shown in Fig. 1. *3D QSAR Pharmacophore Generation* module/Discovery Studio (DS) was used to construct pharmacophore model using hydrogen bond acceptor (***HBA***), hydrogen bond donor (***HBD***), hydrophobic (***H***) and ring aromatic (***RA***) chemical features. It produced ten top-scored hypotheses based on the activity values of the training set

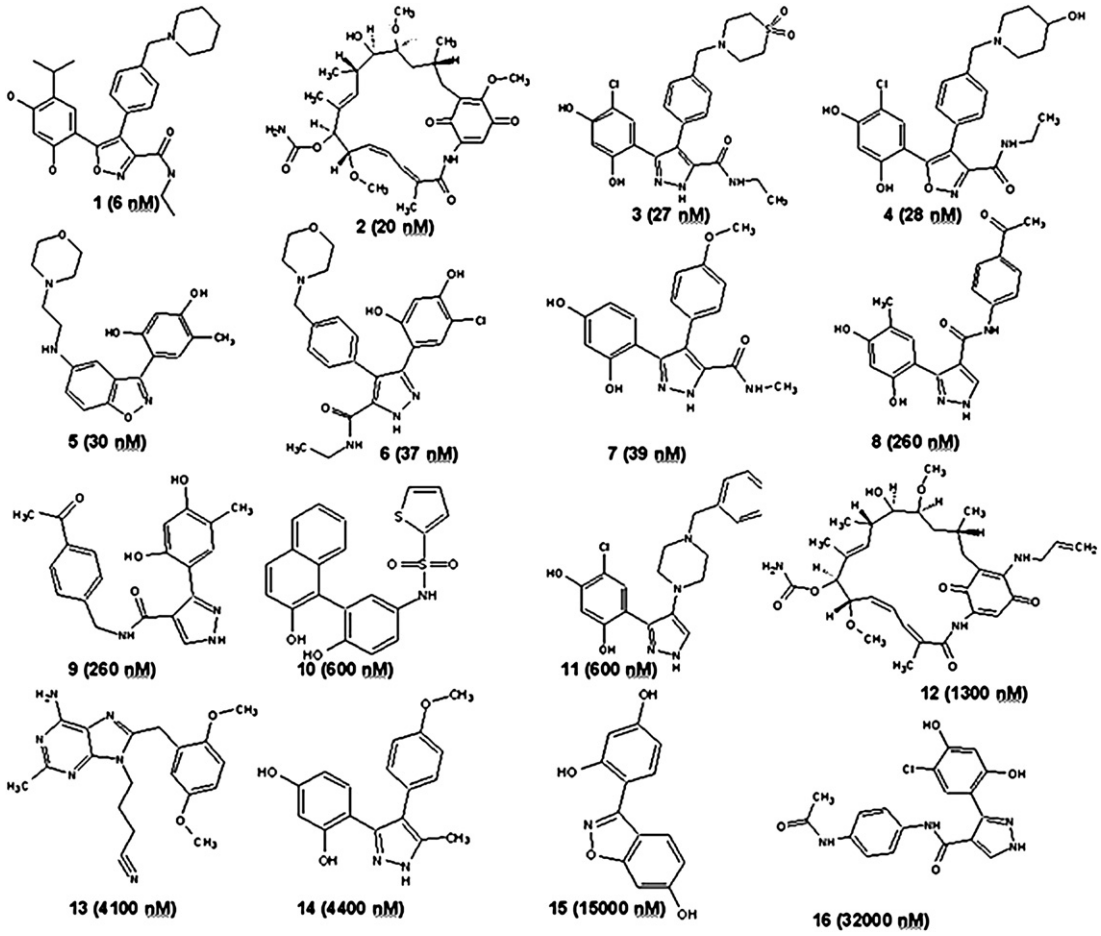

Fig. 1. Chemically diverse 16 compounds used as training set in 3D QSAR Discovery Studio/Pharmacophore generation. IC50 values are indicated in parentheses for each compound.

Table 1
Information of statistical significance and predictive power presented in cost values measured in bits for the top 10 hypotheses as a result of automated 3D QSAR pharmacophore generation.

Hypo no.	Total cost	Cost difference[a]	RMS[b]	Correlation	Features[b]	Max. fit
Hypo1	78.69	73.88	1.29	0.93	***HBA***, ***HBA***, ***3H***	11.78
Hypo2	95.31	57.27	1.95	0.84	***HBA***, ***HBA***, ***HBD***, ***H***	7.14
Hypo3	96.55	56.02	1.99	0.83	***HBA***, ***HBA***, ***HBA***, ***H***	7.42
Hypo4	98.44	54.13	2.01	0.83	***HBA***, ***HBD***, ***H***, ***H***, ***H***	7.40
Hypo5	98.96	53.61	2.05	0.82	***HBA***, ***HBA***, ***HBA***	7.15
Hypo6	99.24	53.33	2.08	0.82	***HBA***, ***HBD***, ***H***, ***H***	7.59
Hypo7	100.12	52.45	2.09	0.81	***HBA***, ***HBA***, ***HBA***, ***H***	6.72
Hypo8	100.63	51.94	2.09	0.81	***HBA***, ***HBA***, ***H***, ***H***	9.75
Hypo9	100.95	51.62	2.08	0.82	***HBA***, ***HBD***, ***H***, ***H***, ***H***	7.25
Hypo10	101.87	50.70	2.1	0.80	***HBA***, ***HBA***, ***HBA***, ***H***	9.04

[a] Cost difference between the null and the total cost. The null cost, the fixed cost and the configuration cost are 152.581, 64.5212 and 16.9475 respectively.
[b] Abbreviation used for features: RMS, root mean square deviation; ***HBA***, hydrogen bond acceptor; ***HBD***, hydrogen bond donor; ***H***, hydrophobic.

molecules. The best ten hypotheses contains only three features: ***HBD***, ***HBA*** and ***H***. Hypo1 consist of two ***HBA*** and three ***H*** which establishes the highest cost difference (73.88), lower errors (59.92), best correlation coefficient (0.93), maximum fit value (11.78) and lowest root mean square (RMS) of 1.2. The fixed and the null cost values are 64.52 and 152.58, respectively. Fixed total cost was dependent on summation of the cost components: weight cost, error cost and configuration cost. Two key values were used for cost analysis: one is the difference between the fixed and null costs and another one is the difference between null and total cost (cost difference). The fixed cost represents a cost of the theoretical ideal hypothesis, which could absolutely predict the activity of compounds in the training set with lowest deviation, while null cost represented the cost of hypothesis with no features that estimates every activity to be the average activity. The difference between these two costs should be $\geq$70 bits to show the over 90% statistical significance of the model. The cost difference should be greater than 60 bits to represent a true correlation data.

In our result, except the fifth hypothesis, all other hypotheses were having the ***HBA*** and ***H*** group, which implies that ***HBA*** and ***H*** groups play an important role in HSP90 inhibition. The cost difference between null and fixed cost was found to be 73.88 and it was more than 70 bits. All hypotheses had a correlation coefficient of higher than 0.80, but Hypo1 showed the highest correlation coefficient values of 0.93, demonstrating good prediction ability of Hypo1. The fixed and total cost values of Hypo1 are 64.52 and 78.69, respectively shows very less difference (Table 1). Higher cost difference and correlation value with low RMS and error values have been observed for Hypo1 when compared with other hypotheses. Hence, Hypo1 was selected as a best hypothesis and employed for further analyses. Fig. 2 shows, the Hypo1 chemical features with its geometric parameters.

The most active and inactive compounds in the training set were aligned in Hypo1 was shown in Fig. 3. To verify, the prediction accuracy of Hypo1, training set was used and the activity of each compound in training set was estimated by regression analysis. Training set compounds were classified relatively into three sets based on their activity values: highly active $IC_{50} < 300$ nM = +++; moderately active $300 \leq IC_{50} < 3000$ nM = ++ and inactive $IC_{50} \geq 3000$ nM = +. Only one active compound was predicted as moderate and all the remaining highly active compounds and moderately active compounds in the training set were estimated correspondingly and all inactive compounds were estimated as inactive by Hypo1. Thus Hypo1 was able to estimate the activities of compounds in their own activity ranges. The experimental and estimated activities by Hypo1 for 16 training set compounds are shown in Table 2.

2.2. Validation of hypo1

Validating the hypothesis is one of the significant methods in pharmacophore generation. There are several methods to confirm

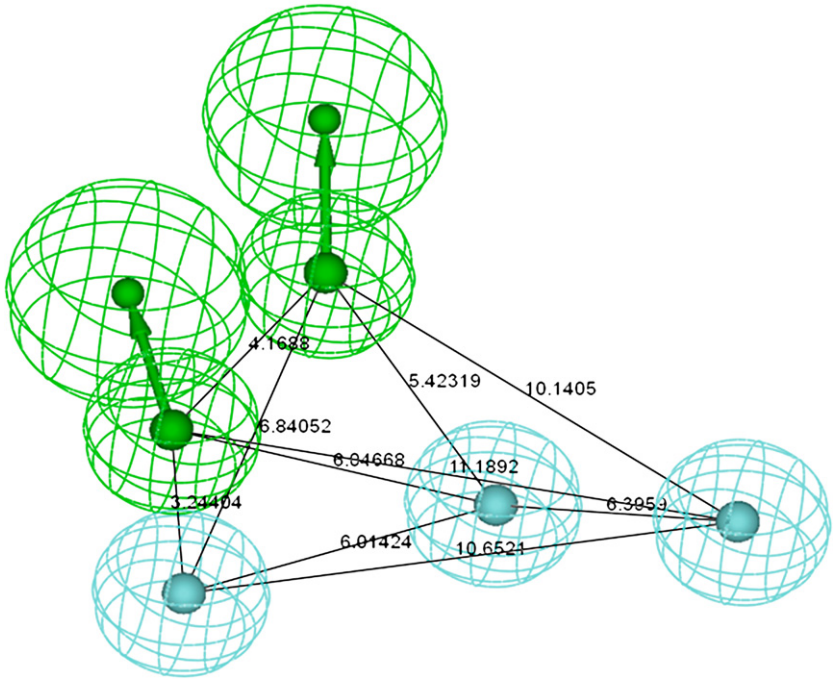

Fig. 2. CATALYST HYPOGEN pharmacophore model, where H and HBA are illustrated in cyan and green, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

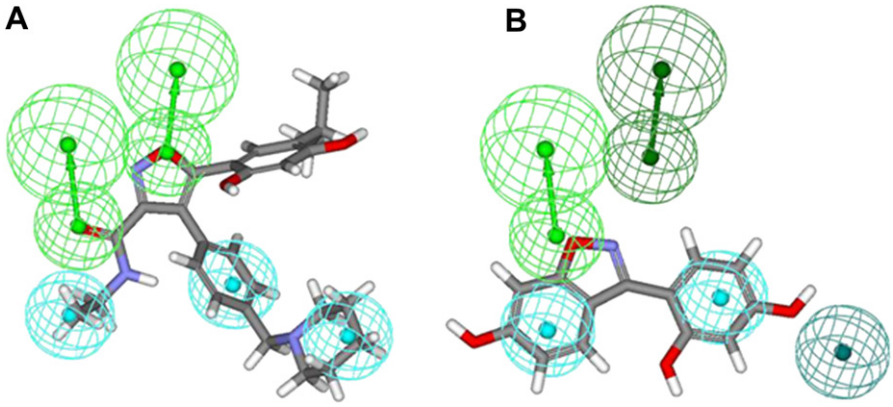

Fig. 3. Best pharmacophore model Hypo1 aligned to training set compound A) Active molecule isoxazole (IC50 6 nM) B) Inactive molecule benzisoxazole (IC50 15,200 nM). Pharmacophore features are colour coded (H, hydrobhobic, cyan and HBA, hydrogen bond acceptor, green). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the quality of pharmacophore like preparing test set, Fischer's randomization method, goodness of fit (**GF**) and enrichment factor (**EF**) etc. The generated hypotheses were mainly validated to check whether the best hypothesis selected the active compounds during the screening process such as the percentage of active compounds picked from dataset, correlation between the predicted and estimated values of test set along with its efficiency in reducing true negatives and false positives.

2.2.1. Fischer's randomization method

Fischer's randomization was used to evaluate the statistical significance of Hypo1. Validation was done by generating random spreadsheets for training set molecules, which randomly reassigned activity values to each compounds and subsequently generated the hypotheses using the same features and parameters originated for Hypo1. To achieve the confidence level of 95%, 19 random spreadsheets (random hypotheses) were generated. The significance of the hypotheses was calculated using the following formula $[1 - (1 + X)/Y] \times 100$, where X, total number of hypotheses having a total cost lower than Hypo X and Y, total number of Hypogen runs (initial + random runs). Here, $X = 0$ and $Y = (19 + 1)$, $S = [1 - ((1 + 0)/(19 + 1))] \times 100\% = 95\%$. Fig. 4 clearly shows that the Hypo1 hypothesis was not generated by chance, because its statistics are far more superior to all random hypotheses.

2.2.2. Test set

Secondly, test set was prepared using the same protocol as training set prepared and used to determine whether the hypothesis was able to predict the active compounds other than the training set molecules. The correlation coefficient (r) for the test set given by Hypo1 was 0.91. Test set compounds were classified relatively into three sets based on their activity values: highly active $IC_{50} < 300$ nM = +++; moderately active $300 \leq IC_{50} < 3000$ nM = ++ and inactive $IC_{50} \geq 3000$ nM = +. Except few compounds, all the remaining highly active compounds and moderately active compounds were estimated correspondingly and all inactive compounds were estimated as inactive by Hypo1. Thus Hypo1 was able to estimate the activities of compounds in their own activity ranges. The experimental and predicted activities of Hypo1 as applied to test set are shown in Table 3. This result was used for further legalization of Hypo1 and we also suggest that the Hypo1 not only fit for training set compounds but also for the external compounds.

Table 2
Actual and estimated activity of the training set molecules based on the pharmacophore model Hypo1.

Compound no.	Fit value[b]	Exp.IC50 nM	Predicted IC50 nM	Error[a]	Experimental scale[c]	Predicted scale[c]
1	10.24	6	2.5	−2.4	+++	+++
2	8.67	20	90	+4.5	+++	+++
3	9.15	27	30	+1.1	+++	+++
4	9.10	28	34	+1.2	+++	+++
5	8.49	30	140	+4.6	+++	+++
6	9.10	37	34	−1.1	+++	+++
7	9.05	39	38	−1.0	+++	+++
8	8.15	260	300	+1.2	+++	+++
9	7.56	260	1200	+4.6	+++	++
10	7.98	600	440	−1.4	++	++
11	8.31	600	210	−2.9	++	++
12	7.96	1300	460	−2.7	++	++
13	7.03	4100	3900	−1.0	+	+
14	6.85	4400	6000	+1.4	+	+
15	6.54	15,000	12,000	−1.2	+	+
16	6.86	32,000	5900	−5.3	+	+

[a] Difference between the predicted and experimental values. '+' indicates that the predicted IC50 is higher than the experimental IC50; '−' indicates that the predicted IC50 is lower than the experimental IC5; a value of 1 indicates that the predicted IC50 is equal to the experimental IC50.
[b] Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule. Fit Value = weight × [max(0.1 − SSE)] where $SSE = (D/T)^2$, D = displacement of the feature from the center of the location constraints and T = the radius of the location constraint sphere for the feature (tolerance).
[c] Activity scale: IC50 < 300 nM = +++ (highly active); 300 nM ≤ IC50 < 3000 nM = ++ (moderately active); IC50 ≥ 3000 nM = + (low active).

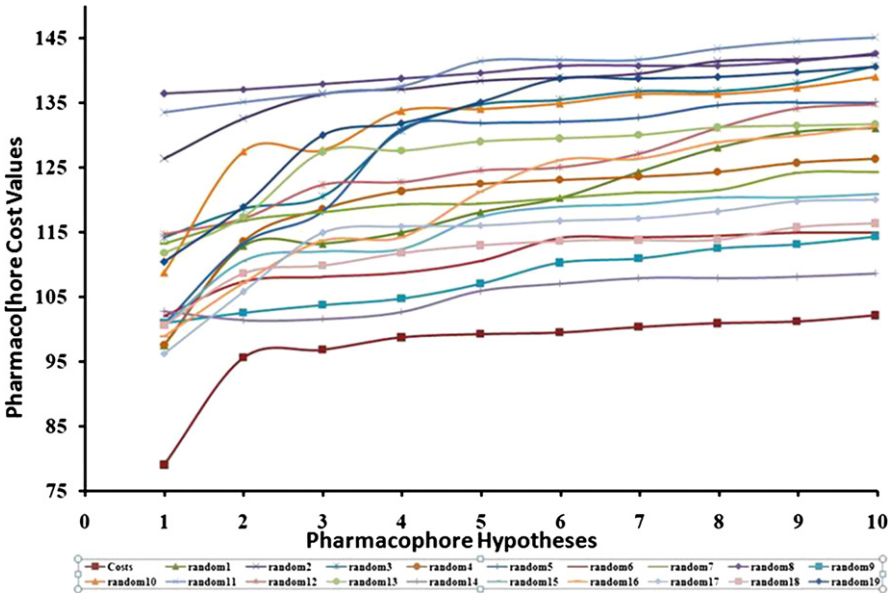

Fig. 4. The difference in costs between HYPOGEN runs and the scrambled runs. The 95% confidence level was selected.

2.2.3. Decoy set

Finally, Decoy set was generated to evaluate the efficiency of Hypo1 by computing **GF** and **EF**. Decoy set contains active and inactive compounds of HSP90 inhibitors. The screening was performed using the *Ligand Pharmacophore Mapping* module in DS, parameters such as total number of compounds in the hit list (**Ht**), number of active percent of yields (%**Y**), percent ratio of actives in the hit list (%**A**), **EF**, false negatives, false positives and **GF** were calculated. The false positives and true negatives are 3 and 8 respectively. **EF** and **GF** are 3.59 and 0.8655 respectively, are very good indications of the high efficiency of the screening (Table 4).

Table 3
Experimental and predicted IC50 data values of 30 test set molecules against Hypo1.

Compound no.	Fit value[b]	Exp.IC50 nM	Predicted IC50 nM	Error[a]	Experimental scale[c]	Predicted scale[c]
1	10.40	127	170	+1.3	+++	+++
2	10.97	239	160	−1.5	++	+++
3	11.40	343	130	−2.6	++	+++
4	10.13	431	140	−3.0	++	+++
5	11.67	231	130	−1.8	+++	+++
6	10.16	222	130	−1.7	+++	+++
7	09.13	258	520	+2.0	+++	++
8	10.59	146	65	−2.2	+++	+++
9	11.63	115	140	+1.2	+++	+++
10	10.31	37	120	+3.2	+++	+++
11	10.85	142	130	−1.1	+++	+++
12	10.31	57	130	+2.4	+++	+++
13	09.87	540	220	−2.4	++	+++
14	06.05	5350	5700	+1.1	+	+
15	06.00	5400	5300	−1.0	+	+
16	05.08	12,800	5900	−2.2	+	+
17	06.25	6000	3500	−1.7	+	+
18	08.63	190	150	−1.3	+++	+++
19	03.73	53,500	16,000	−3.4	+	+
20	08.68	800	340	−2.4	++	++
21	10.31	60	150	+2.5	+++	+++
22	10.34	39	100	+2.6	+++	+++
23	10.42	70	150	+2.2	+++	+++
24	10.02	120	170	+1.4	+++	+++
25	08.67	150	520	+3.5	+++	++
26	08.90	280	690	+2.5	+++	++
27	02.66	4400	1000	−4.2	+	++
28	08.44	62	150	+2.4	+++	+++
29	07.37	210	230	+1.1	++	++
30	10.90	39	78	+2.0	+++	+++

[a] Difference between the predicted and experimental values. '+' indicates that the predicted IC50 is higher than the experimental IC50; '−' indicates that the predicted IC50 is lower than the experimental IC5; a value of 1 indicates that the predicted IC50 is equal to the experimental IC50.

[b] Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule. Fit = weight × [max (0.1 − SSE)] where SSE = $(D/T)^2$, D = displacement of the feature from the center of the location constraints and T = the radius of the location constraint sphere for the feature (tolerance).

[c] Activity scale: IC50 < 300 nM = +++ (highly active); 300 nM ≤ IC50 < 3000 nM = ++ (moderately active); IC50 ≥ 3000 nM = + (low active).

Table 4
Statistical parameter from screening test set molecules.

No	Parameter	Values
1	Total number of molecules in database (**D**)	1500
2	Total number of actives in database (**A**)	38.00
3	Total number of hit molecules from the database (**Ht**)	33.00
4	Total number of active molecules in hit list (**Ha**)	30.00
5	% Yield of actives [(**Ha**/**Ht**) × 100]	90.90
6	% Ratio of actives [(**Ha**/**A**) × 100]	78.94
7	Enrichment Factor (**EF**)	3.50
8	False negatives [**A** − **Ha**]	8
9	False Positives [**Ht** − **Ha**]	3
10	Goodness of fit score[a] (**GF**)	0.86

[a] [(**Ha**/4**HtA**)(3**A** + **Ht**) × (1 − ((**Ht** − **Ha**)/(**D** − **A**)))].

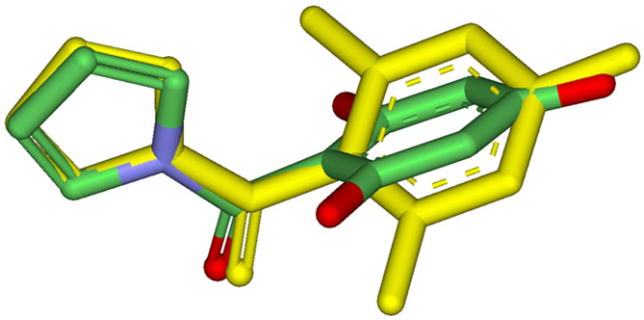

Fig. 6. Structure overlap between co-crystal and docked orientation.

2.3. Databases screening

Virtual screening of small-molecule libraries forms one aspect of a sophisticated approach to drug discovery [20]. Hypo1 was used to screen the various databases like Maybridge and Scaffold, which consists of 60,000 and 100,677 compounds respectively. In drug discovery process virtual screening of databases is an effective alternative to high throughput screening (HTS). Totally, 31,390 compounds (10,390 Maybridge and 20,770 Scaffold) satisfied all the critical features in Hypo1 and 2650 compounds (75 Maybridge and 2575 Scaffold) were considered for further analyses based on the maximum fit value of 11. Drug-likeness properties are an important indicator for selecting the compounds for in vitro studies, which includes molecular or physicochemical properties that contribute to favourable Lipinski's rule of five. So, we further sorted these 2650 compounds using the Lipinski's rule of five and finally 1150 (74 Maybridge and 1076 Scaffold) compounds were further considered for docking studies. Fig. 5 shows the mapping of hit compounds in Hypo1. To evaluate the fit of the pharmacophore to the binding site of the crystal structure of HSP90 (PDB ID: 3EKO), the pharmacophore model Hypo1 was compared with the bound conformation of the compound *PYU* (2-(1H-pyrrol-1-ylcarbonyl) benzene-1, 3, 5-triol) complexed with HSP90. The *PYU* was sketched and the energy-minimized structure was screened using the *Ligand Pharmacophore model* from DS using Hypo1. The root mean square distance (RMSD) between the heavy atom positions of the bound and the pharmacophore screened conformation was 0.75 Å both the conformations are quite well overlaid as shown in Fig. 6. The bound conformation of the compound *PYU* is able to successfully fit all chemical features in Hypo1. To further assess the Hypo1, we compared the pharmacophore model with the active site of HSP90 crystal structure. A summary of the mappings of the pharmacophore Hypo1 to the active site of HSP90 co-crystal structure (PDB ID: 3EKO) is shown in Fig. 7 and it clearly shows a good agreement with the target based pharmacophore.

2.4. Molecular docking studies of HSP90

Molecular docking is a computational technique that samples conformations of small compounds in protein binding sites; scoring functions are used to assess which of these conformations were

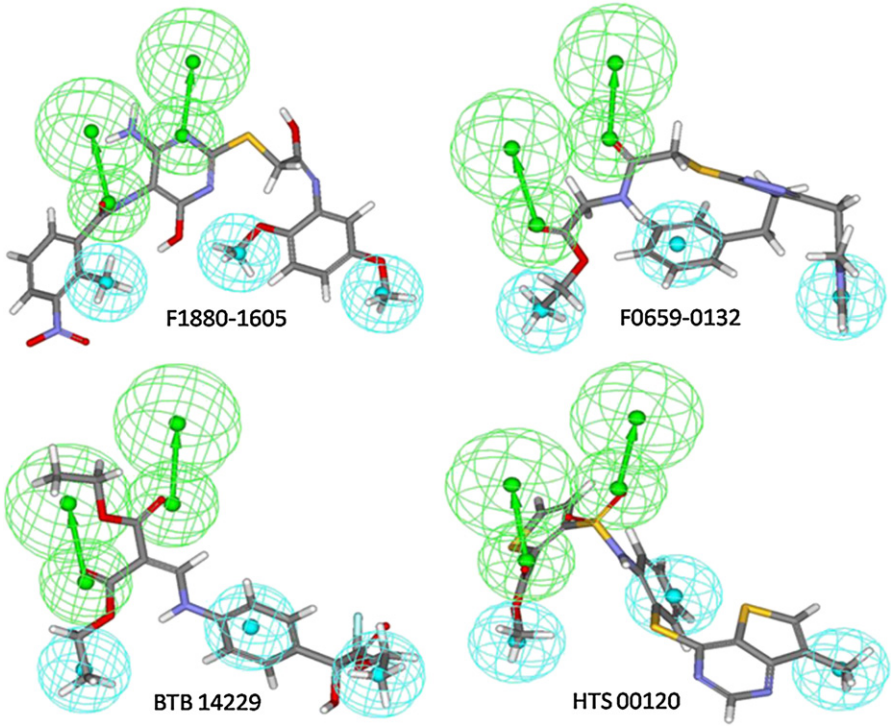

Fig. 5. Mapping of hit molecules from Hypo1. F1880-1605 and F0659-0132 are from Scaffold database and BTB 14229 and HTS 00120 are from Maybridge database.

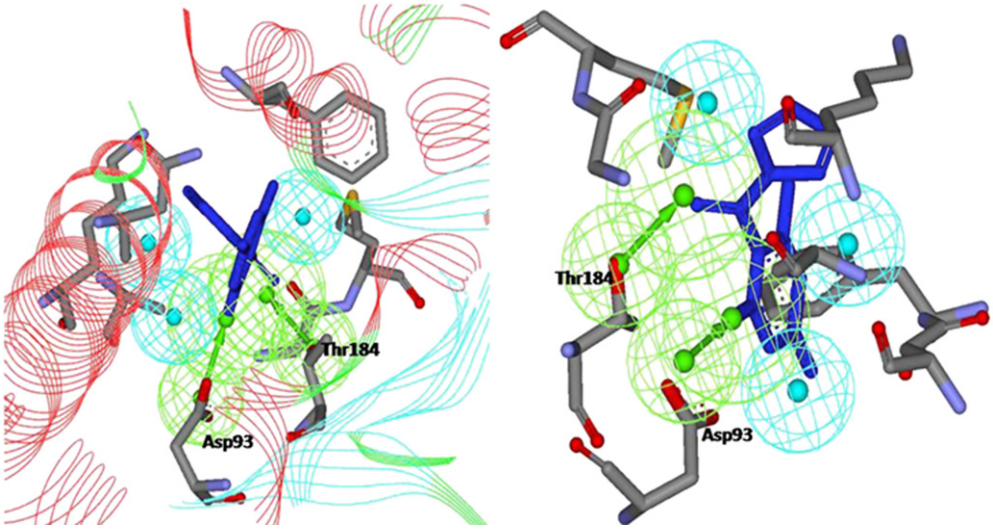

Fig. 7. Comparison and superimpose of Hypo1 in active site of HSP90 (PDB ID: 3EKO).

best complements to the protein binding site. There are two main aspects to assess the quality of docking methods: (i) docking accuracy, which recognizes the true binding mode of the ligands to the target protein and (ii) screening enrichment which measures the relative improvement in the identification of true binding ligands using a docking method versus random screening [26]. Training set of 16 compounds as well as 1150 (74 Maybridge and 1076 Scaffold) hit compounds retrieved from the databases which satisfied drug like properties were docked in the active site of HSP90 using *LigandFit*. Hypo1 was characterized with two ***HBA*** and three ***H*** was superimposed on the active site of HSP90 to find whether the two ***HBA*** group in Hypo1 was mapped on Asp93 and Thr184 residues (Fig. 7).

Totally, 73 compounds show hydrogen bond and hydrophobic interactions with Asp93, Thr184 and Asn51, Ala55, Lys58, Gly97, Met98 and Phe138, respectively. Fig. 8 represents the binding orientation of the hit compounds and also shows how well the compounds fit into Hypo1. The compounds were further sorted

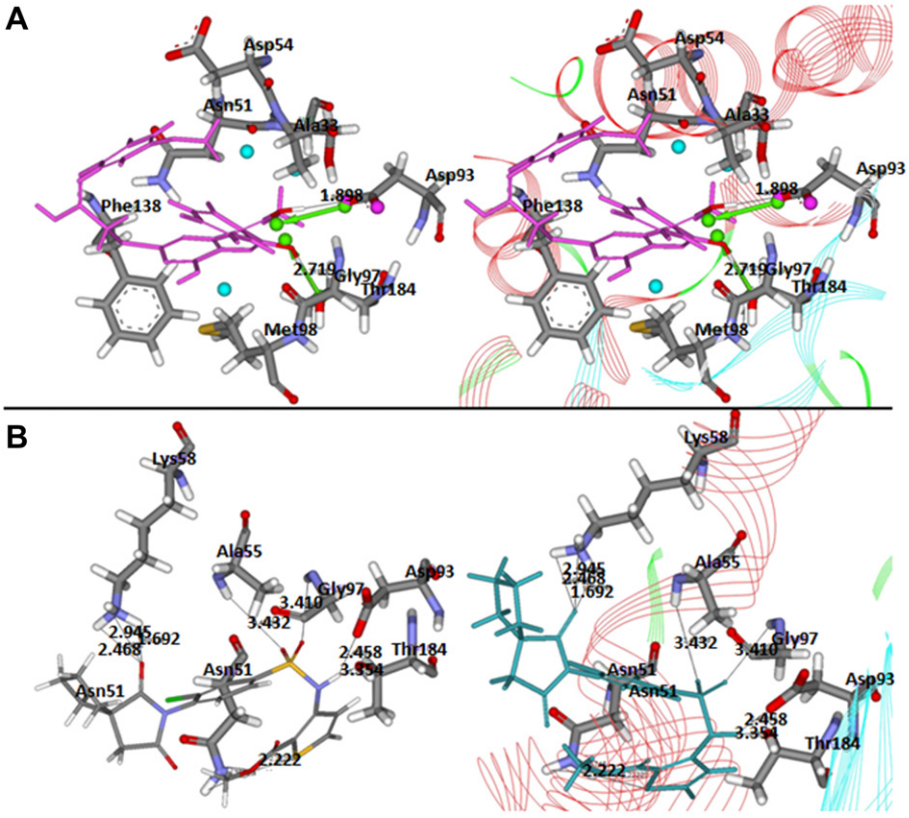

Fig. 8. A) Comparison of the hit molecule and the pharmacophore in the active site of HSP90. B) Compound bound to HSP90 (PDB ID: 3EKO). Hydrogen bonds in black are shown. (Model of 4-amino) bound to HSP90 (PDB ID: 3EKO) HB in black are shown.

based on the consensus scoring function. The training set compounds showed the dock score greater than 80 and the maximum fit value of 10. All of the hit compounds possessed the maximum fit value of 11 and the dock score of more than 85. Some of the active compounds in the training set show hydrogen bonding and hydrophobic interaction with the two active site residues of Asp93, Thr184 and with few hydrophobic residues (Asn51, Ala55, Lys58, Met98, and Phe138) were considered to be crucial for inhibitory activity. The hit compounds showed very good interactions with the critical residues of Asp93 and Thr184. Based on the consensus scoring function finally 36 compounds were sorted for further in-vivo studies.

3. Conclusion

The purpose of this study was not only to construct the pharmacophore model to predict the estimated activity of the compounds, but also to employ the hypothesis on virtual screening to search novel scaffolds. In this work, we have built 3D pharmacophore models from 16 HSP90 inhibitors and the best quantitative pharmacophore model, Hypo1, consisted of five features like two ***HBA*** and three ***H***. The features were perfectly complementary to the HSP90 active site and directed to relative protein residues, showing that Hypo1 represents the characteristic of HSP90 active site. For predicting activity, the correlation coefficient of Hypo1 with training and test sets were 0.93 and 0.91 respectively, suggesting a good predictive power of the hypothesis for the majority of HSP90 inhibitors. Hypo1 was used as a 3D query for screening large databases like Maybridge and Scaffold. Finally, 36 compounds were selected as potent HSP90 inhibitors based on the consensus scoring function. From the overall analyses, we conclude that the Hypo1 pharmacophore truly reflects the features of HSP90 inhibitors and this pharmacophore could be used as fast and accurate tool to assist discovery of novel HSP90 inhibitors.

4. Materials and methods

4.1. Pharmacophore

Pharmacophore modelling is one of the most potent and rapid method to discover a novel scaffold and it can be generated either based on ligands or on the active site of proteins. Two different methods were applied for the ligand based pharmacophore model: HipHop and HYPOGEN: HipHop also known as common features hypothesis, which is generated based on the common features present in the training set molecules. HYPOGEN uses the activity values of the small compounds in the training set to generate the hypothesis. In this study, HYPOGEN algorithm [27] was applied to build the 3D QSAR pharmacophore models using DS V2.1 software (Accelrys Inc., San Diego, CA, USA).

4.2. Test and training set preparation

For generating hypotheses, training set molecules ought to satisfy certain set of laws like it must be broadly colonized (minimum 16 compounds) by structurally diverse representatives and wrap an activity range of atleast four order magnitude. All the biologically relevant data must be obtained by homogeneous processes along with the most active compounds have to be necessarily be integrated in training set. For this study, 16 diverse compounds with activity value (IC_{50}) between 6 nM and 32,000 nM were selected as training set from various literatures, which span over four orders of magnitude, was used to engender the hypotheses which span over four orders of magnitude. To validate the hypothesis the test set was prepared using the same protocol as training set. Test set contains 30 structurally different compounds from the training set with wide range of activity values.

MDL-ISIS Draw 2.5 was used to sketch the two-dimensional (2D) chemical structures of all compounds which were converted into 3D structures using DS. Maximum numbers of 250 conformations were generated for each compound using the *Best Conformation* model generation method based on CHARMm force field and Poling algorithm to assure the energy-minimized conformation. The conformations with energy higher than 20 kcal/mol from the global minimum were rejected. Compounds with their conformational models were then submitted to DS for generating hypothesis.

4.3. Pharmacophore generation using HYPOGEN

An automated 3D QSAR pharmacophore was created by comparing the activity values of compounds in the training set, which could be used for quantitative assessment of activities while screening large databases [28]. Selecting the chemical feature is one of the most important steps in generating pharmacophore. *Feature Mapping* module from DS was used to select the chemical features for hypothesis generation. While generating hypotheses ***HBA***, ***HBD***, ***H*** and ***RA*** features were selected based on the training set compounds with the minimum of 0 to maximum of 5 features were selected. The 'Uncertainty' values for the 16 compounds in the training set were set to 2, and the default values for other parameters were kept

Table 5
Analyses of critical amino acids for HSP90 inhibition from 40 co-crystal structures deposited in protein databank.

PDB ID	Asn51	Ala55	Lys58	**Asp93**	Met98	Phe138	Gly97	**Thr184**
2BYI	✓	✓	✓	✓	✓	✓	✓	✓
1UY6	✓	✓		✓	✓	✓		✓
1UY7	✓	✓		✓	✓			✓
1YET	✓	✓	✓	✓		✓		✓
2BYH	✓	✓		✓	✓	✓	✓	✓
2CCS	✓	✓		✓	✓	✓	✓	✓
1OSF	✓	✓	✓	✓		✓		✓
2QF6		✓		✓	✓	✓		✓
3EKR	✓	✓		✓	✓		✓	✓
2QG0	✓	✓	✓	✓	✓		✓	✓
2QG2	✓			✓		✓	✓	✓
3D0B		✓		✓	✓			✓
2QFO	✓	✓		✓	✓		✓	✓
3BM9			✓	✓	✓	✓	✓	✓
3BMY	✓	✓		✓		✓	✓	✓
3EKO	✓	✓	✓	✓	✓			✓
2VCJ	✓	✓		✓	✓		✓	✓
2CCU		✓		✓	✓		✓	✓
1UYM		✓		✓		✓		✓
2CCT	✓		✓	✓	✓	✓	✓	✓
1UYI	✓	✓		✓	✓	✓	✓	✓
1UYK	✓	✓		✓	✓		✓	✓
1UYH	✓	✓		✓	✓	✓	✓	✓
1UYD		✓		✓	✓	✓		✓
1YC3	✓	✓	✓	✓	✓	✓	✓	✓
2FWY	✓	✓		✓	✓	✓		✓
2FWZ	✓	✓		✓	✓	✓		✓
2BSM	✓	✓	✓	✓	✓	✓	✓	✓
2H55	✓	✓		✓	✓	✓	✓	✓
2VCI	✓	✓	✓	✓	✓	✓	✓	✓
1UYG	✓	✓		✓	✓	✓	✓	✓
1UYF	✓	✓		✓	✓	✓	✓	✓
1UYE		✓		✓	✓	✓		✓
1UY9		✓		✓	✓	✓		✓
1UYC	✓	✓		✓	✓	✓		✓
1UY8		✓		✓	✓	✓		✓
2JJC		✓		✓	✓			✓
2BZ5	✓	✓	✓	✓	✓	✓	✓	✓
2BT0	✓	✓		✓		✓	✓	✓
2CDD	✓	✓	✓	✓	✓	✓	✓	✓
2UWD		✓		✓		✓	✓	✓

constant. Subsequently, pharmacophore models were computed using *3D QSAR Pharmacophore* module and the top 10 scoring hypotheses were collected. The qualities of the hypotheses were dependent on the fixed cost, null cost, and total cost values [29].

4.4. Assessment of pharmacophore quality and database screening

Three different methods were used to evaluate the quality of the pharmacophore. Initially, cross validation was performed by randomizing the data using the Fischer's randomization test. The result confirms that the generated hypotheses from the training set are reasonable. Secondly, the prediction of test set showed a correlation value of 0.91 between the experimental and predicted activities. All queries were performed using the *Ligand Pharmacophore Mapping* protocol. In order to further validate the Hypo1, **GF** and **EF** were calculated [30] using the formula $((\mathbf{Ha}/4\mathbf{HtA})(3\mathbf{A} + \mathbf{Ht})) \times (1-((\mathbf{Ht} - \mathbf{Ha})/(\mathbf{D} - \mathbf{A})))$ and (Ha/Ht)/(A/D), respectively. Finally, Hypo1 was selected as the best hypothesis and used as a 3D query to retrieve a novel scaffold for HSP90 inhibitors from various databases like Scaffold and Maybridge.

4.5. Structure based molecular docking

Molecular docking and pharmacophore model are the two potent methods in drug discovery process. Virtual screening followed by docking has become one of the reputable methods for drug discovery and enhancing the efficiency in lead optimization. The main advantage of pharmacophore based docking was to focus on specific key interaction for protein–ligand binding. Ameliorate the selection of active compounds it is optimal to use both methods like molecular docking and pharmacophore [31–37].

Molecular docking was executed for accurate docking of ligands into protein active sites using *LigandFit* module in DS. There are three stages in *LigandFit* protocol: (i) Docking: attempt is made to dock a ligand into a user defined binding site (ii) In-Situ Ligand Minimization and (ii) Scoring: various scoring functions were calculated for each pose of the ligands. The protein complexes were selected from protein databank (PDB, www.rcsb.org). Till date, there are many HSP90 complexes were reported, among them PDB ID: 3EKO was selected based on the resolution of the complex and the size of the co-crystal. For docking study, initially protein was prepared by removing all water compounds and CHARMm force field was applied using *Receptor–Ligand Interactions* tool in DS.

After the protein preparation, the active site of the protein has to be identified. The active site of the protein was represented as binding site; it's a set of points on a grid that lie in a cavity. Two methods were applied to define the binding site for a protein: (i) Firstly, binding sites were identified based on the shape of the receptor using "eraser" algorithm and (ii) secondly, volume occupied by the known ligand pose already in an active site. In this study, binding site was defined using second method and the critical amino acids were identified by analysing the protein–ligand interactions from 40 HSP90 co-crystal structures which were deposited in PDB. All the ligands in the complex structures showed the hydrogen bond interactions with Asp93, Thr184 and hydrophobic interactions with Asn51, Ala55, Lys58, Gly97, Met98 and Phe138. This clearly indicates that these two hydrogen bonded amino acids play a crucial role in HSP90 inhibitions and the result was shown in Table 5.

During the docking process top 10 conformations was generated for each ligand based on dock score value after the energy minimization using the smart minimizer method, which begins with *steepest descent* method and followed by the *conjugate gradient* method. Ligand binding affinity was computed based on the dock dcore and other scoring functions like LigScore1, LigScore2, Jain, Potential of Mean Force (PMF), Piecewise Liner Potential (PLP) ligand.

Acknowledgement

All students were supported by a scholarship from the BK21 Program, the Ministry of Education, Science and Technology, Korea and this work was supported by grants from the MOST/KOSEF for the Environmental Biotechnology National Core Research Center (grant #: R15-2003-012-02001-0).

References

[1] W. Paul, Cancer Detect. Prev. 26 (2002) 405–410.
[2] T. Scheibel, J. Buchner, Biochem. Pharmacol 56 (1998) 675–682.
[3] D.F. Smith, L. Witesell, E. Katsanis, Pharmacol. Rev. 50 (1998) 493–513.
[4] L. Neckers, T.W. Schulte, E. Mimnaugh, Invest. New Drugs 17 (1999) 361–373.
[5] A. Maloney, P. Workman, Expert. Opin. Biol. Ther. 2 (2002) 3–24.
[6] J.S. Isaacs, W. Xu, L. Neckers, Cancer Cell 3 (2003) 213–217.
[7] D. Picard, Cell. Mol. Life Sci. 59 (2002) 1640–1648.
[8] Prodromou, Curr. Cancer Drug Targets 3 (2003) 301–323.
[9] L.H. Pearl, C. Prodromou, Ann. Rev. Biochem. 75 (2006) 271–294.
[10] Z. Yin, C.Ellen Henry, A.Thomas Gasiewicz, Biochemistry 48 (2009) 180–189.
[11] R. Bagatell, L. Whitesell, Mol. Cancer. Ther 3 (2004) 1021–1030.
[12] K. Terasawa, M. Minami, Y. Minami, J. Biochem. 137 (2005) 443–447.
[13] W.H. Piel, L. Gui, E. Bruford, A. Monteiro, Genomics 86 (2005) 627–637.
[14] S.J. Felts, J. Biol. Chem. 275 (2000) 3305–3312.
[15] P. Krishna, G. Gloor, Cell Stress Chaperones. 6 (2001) 238–246.
[16] Y.P. Cserme, T. Schnaider, C. Soti, Z. Prohaszka, G. Nardai, Pharmacol. Ther. 79 (1998) 129–168.
[17] A.S. Sreedhar, E. Kalmar, P. Csermely, S. Yu-Fei, FEBS Lett. 562 (2004) 11–15.
[18] H. Zhang, F. Burrows, J. Mol. Med. 82 (2004) 488–499.
[19] Hanahan, R.A. Weinberg, Cell 100 (2000) 57–70.
[20] M.G. Marcu, A. Chadli, I. Bouhouche, N. Catelli, L.M. Neckers, J. Biol. Chem. 275 (2000) 37181–37186.
[21] M. Dutta, M. Inouye, Trends Biochem. Sci. 25 (2000) 24–28.
[22] P. Workman, Cancer Lett. 206 (2004) 149–157.
[23] R.I. Morimoto, Genes Dev. 12 (1998) 3788–3796.
[24] Chiosis, M. Vilenchik, J. Kim, D. Solit, Drug Discov. Today 9 (2004) 881–888.
[25] A.S. Sreedhar, E. Kalmar, P. Csermely, Y.F. Shen, FEBS Lett. 563 (2004) 11–15.
[26] D. Muthas, A.S. Yogesh, M. Lundborg, A. Karlen, J. Mol. Graph. Model. 26 (2008) 1237–1251.
[27] Li, J. Sutter, R. Hoffmann, Iul Biotechnology Series, vol. 2, International University Line, La Jolla, CA, 2000, pp. 171–189.
[28] Catalyst, Version 4.10. Accelrys Inc., San Diego, CA, USA, 2005.
[29] A.K. Debnath, J. Med. Chem. 45 (2002) 41–53.
[30] F.G. Osman, R.H. Douglas, in: F.Guner Osman (Ed.), International University Line, La Jolla, CA, 2000, pp. 193–210 (Chapter 11).
[31] E. Perola, Proteins: Struct. Funct. Bioinform. 64 (2006) 422–435.
[32] V. Gopalakrishnan, J. Aparna, M. Jeevan, G.R. Ravi, J. Desiraju, Chem. Inf. Mod. 45 (2005) 1101–1108.
[33] M. Claussen, V. Gastreich, J. Apelt, S.A. Greene, C. Hindle, Lemmen, Curr. Drug Discov. Technol. 1 (2004) 49–60.
[34] A.C. Good, D.L. Cheney, D.F. Sitkoff, J.S. Tokarski, T.R. Stouch, D.A. Bassolino, S.R. Krystek, Y. Li, J.S. Mason, T.D.J. Perkins, J. Mol. Graph. Model. 22 (2003) 31–40.
[35] X. Fradera, R.M.A. Knegtel, J. Mestres, Proteins: Struct. Funct. Bioinform. 40(200) 623–636.
[36] S.A. Hindle, M. Rarey, C. Buning, T. Lengauer, J. Comput. Aid. Mol. Des. 16 (2002) 129–149.
[37] G. Klebe, Drug Discov. Today 11 (2006) 580–594.