

## A Combined Pharmacophore-Based Virtual Screening, Docking Study and Molecular Dynamics (MD) Simulation Approach to Identify Inhibitors with Novel Scaffolds for Myeloid cell leukemia (Mcl-1)

Guang-kai Bao, Lu Zhou,\* Tai-jin Wang, Lu-fen He, and Tao Liu

College of Chemical Engineering, Sichuan University, Chengdu 610065, Sichuan, China. \*E-mail: [zhoulu@scu.edu.cn](mailto:zhoulu@scu.edu.cn)

Received February 19, 2014, Accepted March 21, 2014

Chemical feature based quantitative pharmacophore models were generated using the HypoGen module implemented in DS2.5. The best hypothesis, Hypo1, which was characterized by the highest correlation coefficient (0.96), the highest cost difference (61.60) and the lowest RMSD (0.74), consisted of one hydrogen bond acceptor, one hydrogen bond donor, one hydrophobic and one ring aromatic. The reliability of Hypo1 was validated on the basis of cost analysis, test set, Fischer's randomization method and GH test method. The validated Hypo1 was used as a 3D search query to identify novel inhibitors. The screened molecules were further refined by employing ADMET, docking studies and visual inspection. Three compounds with novel scaffolds were selected as the most promising candidates for the designing of Mcl-1 antagonists. Finally, a 10 ns molecular dynamics simulation was carried out on the complex of receptor and the retrieved ligand to demonstrate that the binding mode was stable during the MD simulation.

**Key Words :** Mcl-1 inhibitors, Pharmacophore model, Molecular docking, Molecular dynamics simulation

### Introduction

Apoptosis, or termed as programmed cell death, is a critical cell process in the normal development and homeostasis of multi-cellular organisms. The Bcl-2 family of proteins is a central arbiter of apoptosis, which consists of anti-apoptotic proteins such as Bcl-2, Bcl-xL, Mcl-1 and pro-apoptotic proteins, namely, Bak, Bax, Bim, Bid and Bad.<sup>1-4</sup> The Bcl-2 family of proteins regulates apoptosis through the influence imposed on mitochondrial outer membrane (MOM) permeability and the release of cytochrome *c*.<sup>5-7</sup> Besides, it has been well demonstrated that the anti-apoptotic proteins and the pro-apoptotic proteins modulate their opposing functions by heterodimerization.<sup>8</sup> The anti-apoptotic proteins of the Bcl-2 family proteins are normally over-expressed in many human cancer cells, which account for the resistance of traditional anti-cancer drugs.<sup>1-4</sup>

Myeloid cell leukemia (Mcl-1) is one of the Bcl-2 family proteins, which is discovered to play an essential role in the regulation of apoptosis.<sup>9</sup> It is observed that Mcl-1 is over-expressed in a variety of human hematopoietic and lymphoid cancers including B cell lymphoma, multiple myeloma and chronic lymphocytic leukemia.<sup>10-15</sup> The C-terminus of Mcl-1 contains a hydrophobic transmembrane domain and the BH1, BH2, BH3 domains within the cytosolic region. On the other hand, the N-terminus of Mcl-1 is comprised of two prolines (P), glutamic acid (E), serine (S) and threonine (T) (PEST) sequences, which has been shown to be the target for degradation.<sup>16</sup> There are several phosphorylation sites in the PEST region of Mcl-1. For example, the half-life of the protein increases resulted from the phosphorylation of Thr163 by ERK.<sup>17</sup> Additionally, Mcl-1 doesn't have the domain of

BH4 whereas it is present in Bcl-2, Bcl-xL and Bcl-W.<sup>18</sup>

Mcl-1 is regulated at translational and transcriptional levels. Translationally, the mRNA of Mcl-1 can be spliced to remove exon2 leading to the generation of Mcl-1<sub>s</sub>, which is a shorten form of Mcl-1 and lacks the BH1, BH2 as well as the transmembrane domain. Another shorten form of Mcl-1, Mcl-1<sub>ES</sub>, has also been identified, which remains all the three BH domains and the C-terminal transmembrane domain.<sup>19</sup> These two forms of Mcl-1 induce apoptosis by inhibiting the full length Mcl-1 because they can't interact with the pro-apoptotic Bcl-2 family proteins.<sup>20</sup> At the transcriptional level, the regulation of Mcl-1 transcription is adjusted by a series of activated transcription factors, especially the STAT family and CAMP response element binding protein.<sup>21</sup> On the contrary, Mcl-1 can also be down-regulated upon the induction of apoptosis resulted from different treatments such as staurosporine.<sup>22</sup> While the down-regulation of Mcl-1 can induce apoptosis in a number of cancer cell lines,<sup>23</sup> hence Mcl-1 is emerging as an attractive target for the design of small molecule inhibitors in recent years.

The main aim of this research is to generate a 3D quantitative pharmacophore model based on a set of structurally diverse inhibitors. The reliability of the best hypothesis, Hypo 1, was seriously ascertained in terms of cost analysis, test set, Fischer's randomization method and GH test method. The validated hypothesis was used as a 3D search query to screen chemical databases, namely, Nature product and Asinex, for identifying inhibitors with novel scaffolds. The compounds, which mapped effectively on the chemical features of Hypo 1, were subjected to docking studies and visual inspection for the purpose of checking their shape fits in the binding site. Finally, to our best knowledge, the MD

simulation was carried out for the first time in order to monitor the stability of the protein with the presence of ligand.

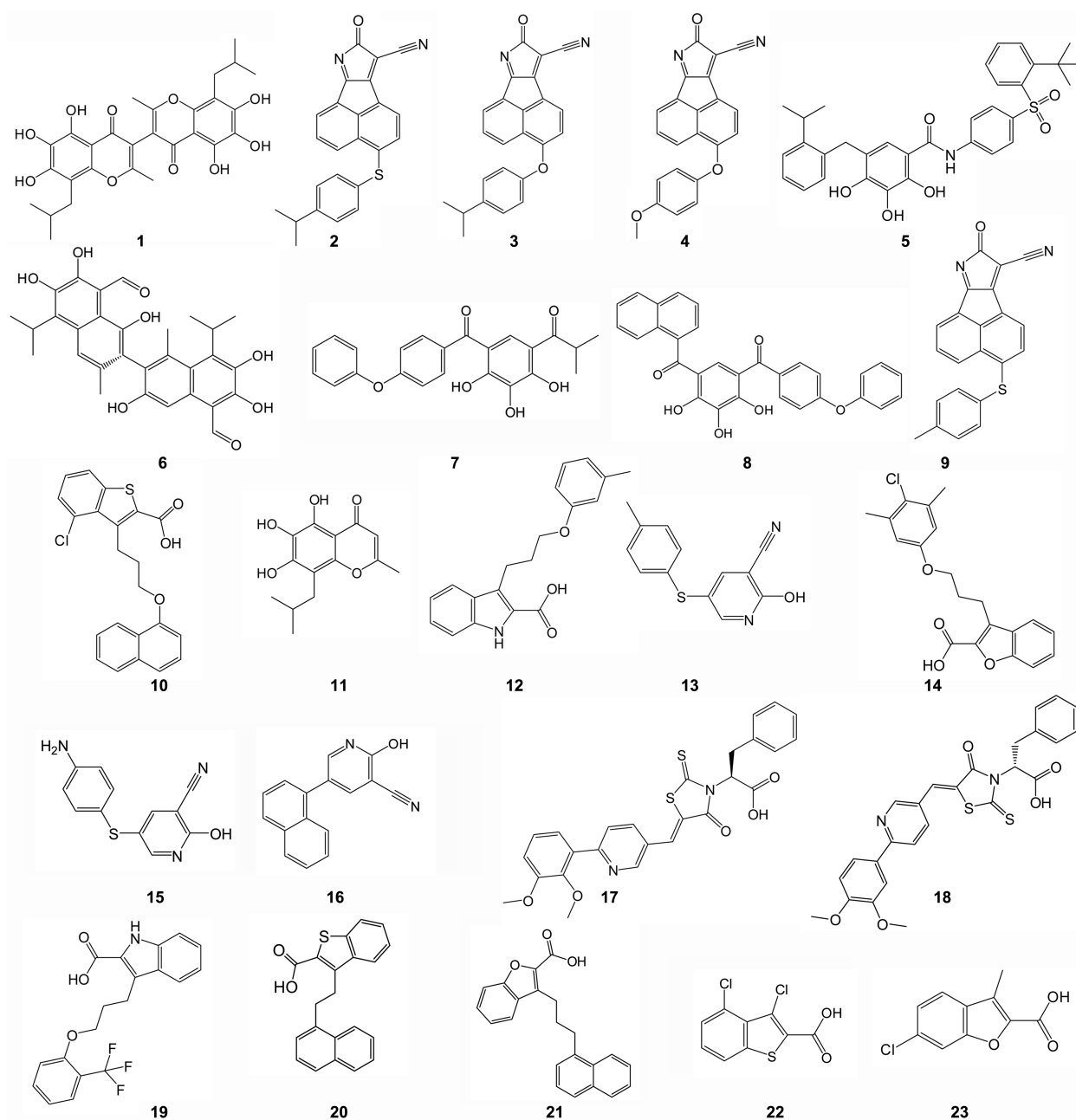
### Experimental

**Data Preparation.** A set of 64 Mcl-1 inhibitors were collected from the recently published literatures by considering structural diversity and wide coverage activity range. The bioactivities were represented as  $K_i$  values determined under the same conditions and procedures.<sup>24-29</sup> Initially, the three-dimensional chemical structures of the collected inhibitors were sketched using ChemBio3D12.0 and subsequently, were subjected to energy minimization applying the CHARMM

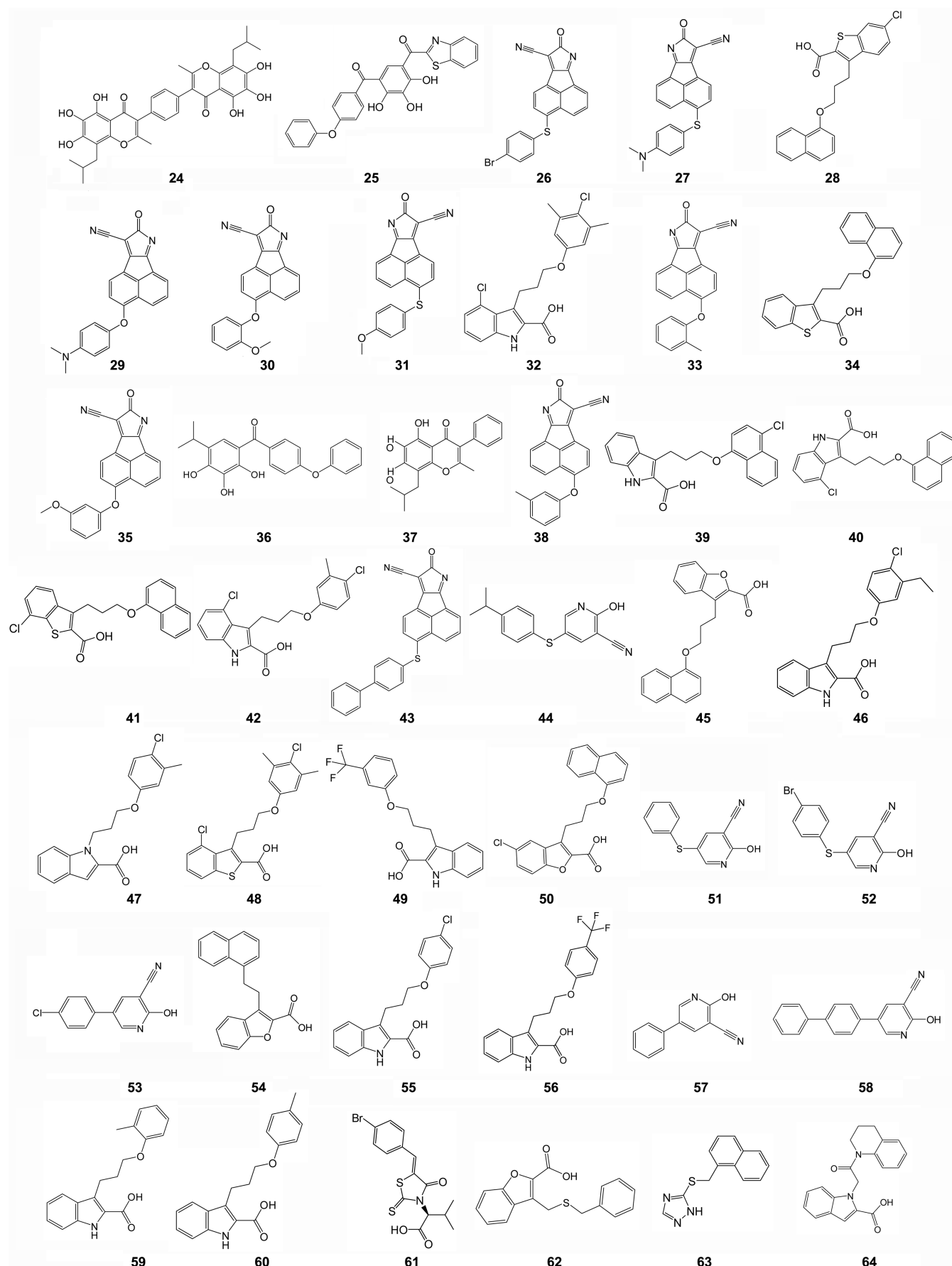
force field implemented within DS2.5.

Selection of suitable training set, which is responsible for the quality of the generated hypothesis, is the most important step in the pharmacophore modeling. A training set of 23 structurally diverse molecules (Fig. 1) with values of  $K_i$  spanning four orders of magnitude were selected for the generation of pharmacophore models, while the remaining 41 molecules were designed as a test set of external data (Fig. 2) to validate the resultant pharmacophore models. For the purpose of estimation, the activity values were classified into three categories, namely, highly active ( $K_i < 1 \mu\text{M}$ , represented as +++), moderately active ( $1 \mu\text{M} \leq K_i \leq 10 \mu\text{M}$ , represented as ++) and inactive ( $K_i > 10 \mu\text{M}$ , represented as +).

To take into consideration of the molecular flexibility, the



**Figure 1.** Chemical structures of the 23 training molecules applied in the generation of pharmacophore models.

**Figure 2.** Chemical structures of the 41 molecules in the test set.

flexibility of each molecule of the data sets was constructed by making multiple conformers within a specific energy range. The protocol "Diverse Conformation Generation" in DS2.5 was utilized for the generation of a representative family of conformation models using Poling algorithm. It is a method for modeling conformational variations of molecule that forces similar conformers away from each other.<sup>30,31</sup> The maximum number of conformers was specified at 255 using an energy range of 20 Kcal/mol to ensure maximum coverage of the conformational space and the remaining parameters were kept at their default values.

**Pharmacophore Generation.** A pharmacophore model is comprised of 3D arrangement of a collection of features, which account for the biological activities of the inhibitors. Prior to performing the quantitative pharmacophore modeling, an analysis was carried out for molecules presented in the training set by using the "Feature Mapping" module as available in DS2.5, the purpose of which was to identify the suitable chemical features necessary for potent Mcl-1 inhibitors. The result revealed that hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), hydrophobic (H) and ring aromatic (RA) were effectively mapped with molecules of training set and thus these features were used as input for the generation of hypothesis. The minimum and maximum number of the selected features was set to 0 and 5, respectively. Uncertainty is one of the key parameter in pharmacophore generation, which represents the ratio range of uncertainty in the activity value of a particular agonist. The uncertainty value was preferred to 2 over its default value of 3 as the activity range in the training set molecules barely met with the minimum requirement of four orders of magnitude.<sup>32-33</sup>

The hypotheses were built using the "3D-QSAR Pharmacophore Generation" protocol implemented in DS2.5 using HypoGen algorithm.<sup>34</sup> HypoGen algorithm generates hypothesis by giving more priority to chemical features that are common to active compounds but excluding common features from inactive compounds. During the generation of hypothesis, the correlation between structure and activity of each molecule in the training set was rigorously examined. The activities of training set compounds were estimated and the ratio of estimated and experimental activities was termed as error value. The process of hypothesis generation accomplished in three steps, namely, a constructive step, a subtractive step and an optimization step. The constructive step was an initial stage in which the active compounds were identified. Subsequently, the inactive compounds were removed in the subtractive step and the consequential hypothesis was optimized finally to improve model quality. The hypothesis generation was ceased when there was no better hypothesis could be accomplished.<sup>35</sup> Ten hypotheses with significant statistical parameters were generated for each HypoGen run, from which the one with the highest correlation coefficient, lowest total cost and RMSD was chosen as the best one. The reliability of the best hypothesis was validated on the basis of cost functions and other statistical parameters calculated during the generation of hypotheses.

**Pharmacophore Validation.** The validation of hypothesis was carried out in order to determine whether it was capable of identifying active compounds from a database as well as predicting their activities accurately or not. The statistical significance of the hypothesis can be evaluated in terms of fixed cost, null cost, total cost and other statistical parameters. The fixed cost is the simplest model that fits all the data perfectly, while null cost represents the highest cost of a hypothesis which estimates activity to be the average of activities of training set compounds. The total cost for each hypothesis should be close to the fixed cost as well as far away from the null cost to provide a valuable model. The hypothesis is not generated by chance if the difference between fixed hypothesis and null hypothesis cost values is high. It has been observed that a cost difference of above 60 bits would entail the probability of 90% for correlating the experimental and predicted activity values. Besides, there are two other parameters that play an important role in evaluating the quality of hypothesis, namely, configuration cost and error cost. The configuration cost, which depends on the complexity of the hypothesis space and should not be greater than 17, is represented as  $\log_2 P$ , where  $P$  denotes the number of hypothesis created in the constructive phase and survived in the subtractive phase. The error cost is solely determined by RMSD, which explains the deviations between the predicted and experimental activities of training set compounds. In addition to the cost analysis, a test set of external compound, Fischer's randomization method and GH test method were also used to assess the best hypothesis.

**Test Set Prediction Method.** To validate whether the best hypothesis was capable of estimating the activities of external compounds accurately, 41 structurally diverse compounds with a wide range of activities, which were not included in the training set, were chose as test set. All the test set compounds were sketched and the conformations of them were generated using the same way as described for training set molecules. The performance of the hypothesis was ascertained by applying it to regress against the test set compounds. "Ligand Pharmacophore Mapping" protocol implemented in DS was used with Rigid/BEST search option in this procedure to map test set compounds on the selected hypothesis.

**Fischer's Randomization Method.** Fischer's randomization methodology was employed as a validation procedure with the hope of figuring out whether there was a sturdy correlation relationship between the chemical features and biological activities.<sup>35</sup> The randomized data sets were generated by shuffling randomly and then reassigning the activity values from one molecule to another in the training set. Since the confidence level was set to 95%, 19 different random spreadsheets were developed to build hypotheses using exactly the same features and parameters as described in the original hypothesis generation. The hypotheses generated with the randomized data sets should not have similar or better cost values, otherwise the original hypothesis would be considered to be generated by chance.

**GH Test Method.** GH test method was utilized to demonstrate whether the hypothesis can identify the active compounds during the virtual screening process.<sup>36</sup> Therefore, the decoy set consisting of 1632 molecules, out of which 26 experimentally known inhibitors were included, was used to figure out the ability of hypothesis to discriminate the active from the inactive compounds. Various statistical parameters, such as goodness of fit score (GH) and enrichment factor (E value), were calculated during the screening procedure, and they determined the capability of the generated hypothesis. The GH score ranged from 0 to 1 and it must be greater than 0.5 for a reasonable model.

**Virtual Screening and Drug-Likeness Prediction.** Virtual screening of chemical databases is a complementary approach to identify the novel and potential leads which are suitable for the further optimization. In the present study, the databases of Nature product and Asinex were initially screened with the Lipinski's rule of five. Compounds were considered to be drug-like only if MW < 500, logP < 5, hydrogen bond donor < 5 and hydrogen bond acceptor < 10.<sup>37</sup> Subsequently, the best hypothesis, Hypo 1, was used as a 3D search query to retrieve novel inhibitors with Rigid/FAST search option as provided in DS2.5. Only the compounds which mapped effectively upon the features of hypothesis as well as had a fit value of above 5.32 (highest fit value for the compounds in the data set) were identified as hits. The retrieved hits were further screened by ADMET properties which refer to the adsorption, distribution, metabolism, excretion and toxicity of a molecule within an organism. Finally, the compounds, which passed the above various filters, were subjected to molecular docking studies and visual inspection.

**Molecular Docking.** The retrieved hits were further refined by the molecular docking studies using GOLD docking suit from Cambridge Crystallographic Data Center, UK.<sup>38</sup> GOLD uses a genetic algorithm to dock the small molecules into the active site of protein allowing for a full range of flexibility of the small molecules as well as partial flexibility of protein. The GOLD fitness score is calculated from the contributions of hydrogen bonds and van der Waals between

the protein and ligand together with intra-molecular hydrogen bonds and strains of ligand.<sup>39</sup> The crystal structure of Mcl-1 (PDB code: 4HW3) was downloaded directly from the Protein Data Bank (www.rcsb.org). The binding site of protein was defined as a collection of amino acids within 10 Å around the co-crystallized ligand in the X-ray structure. The molecular docking procedures were performed with the same parameters as follows: population size 100, selection pressure 1.1, operation 100000, islands 5, niche size 2, migration 10, crossover 95, mutation 95. The resultant molecules were further checked by visualization to figure out the ones with reasonable binding mode and critical interactions with the key amino acids.

**Molecular Dynamics Simulation.** The docking complex which obtained the highest score was used for the molecular dynamics (MD) simulation.

MD simulation was carried out using the GROMACS4.5.5 package with the standard GROMOS96 43a1 force field.<sup>40</sup> The molecular topology and coordinate files of the ligand in protein were generated by the PRODRG program.<sup>41</sup> Periodic boundary condition was applied in all three directions of space to minimize edge effects in the system and the simple point charge (SPC) water model was used to represent the water molecules. Additionally, the minimum distance between the box wall and protein was set to 10 Å for avoiding the interaction of protein and its own periodical images. The system was neutralized by adding four Cl<sup>-</sup> counter ions. Prior to the simulation, the system was subjected to energy minimization in order to relax any steric conflicts using the steepest descent method for 5000 steps. A NVT ensemble was employed with a coupling constant of 0.1 ps as well as the time duration of 100 ps to reach a constant temperature of 300 K. This was followed by an equilibration under NPT condition, in which a constant pressure of 1 bar was obtained with a coupling constant of 2.0 ps and time duration of 1 ns. Electronic interactions were calculated using the partial mesh Ewald (PME) method and LINCS algorithm was applied for constraining all covalent bonds.<sup>42,43</sup> Cutoff values for the calculation of Coulomb and van der Waals interactions were set to 1.0 and 1.4 nm, respectively. Finally, a 10

**Table 1.** Statistical parameters of the top 10 pharmacophore models generated by the training set molecules using HypoGen algorithm

Hypo No.	Total cost	Cost difference <sup>a</sup>	RMSD <sup>b</sup>	Correlation	Features <sup>b</sup>
Hypo1	90.65	61.60	0.74	0.96	HBA, HBD, H, RA
Hypo2	91.65	60.60	0.79	0.96	HBA, HBD, H, RA
Hypo3	92.49	59.76	0.84	0.95	HBA, H, RA
Hypo4	93.68	58.57	0.90	0.94	HBA, H, RA
Hypo5	94.06	58.19	0.91	0.94	HBA, H, RA
Hypo6	94.31	57.94	0.91	0.94	HBA, H, RA
Hypo7	96.72	55.53	1.03	0.93	HBA, H, RA
Hypo8	97.10	55.15	1.02	0.93	HBA, HBD, H, RA
Hypo9	97.11	55.14	1.02	0.93	HBA, HBD, H, RA
Hypo10	98.40	53.85	1.07	0.92	HBA, HBD, H, RA

<sup>a</sup>Cost difference between the null cost and the total cost. The null cost, the fixed cost and the configuration cost are 152.25, 84.36 and 16.46, respectively. All costs are in units of bit. <sup>b</sup>Abbreviation used for features: RMSD, root mean square deviation; HBA, hydrogen bond acceptor; HBD, hydrogen bond donor; H, hydrophobic; RA, ring aromatic.

ns simulation was performed with a time step of 2 fs. The MD simulation can provide a set of configurations that can be used to investigate the stability and behavior of the protein beyond the crystal structure.

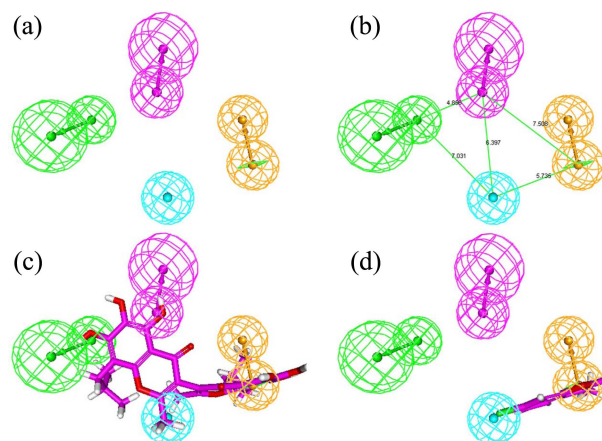
## Results and Discussion

**Pharmacophore Generation.** A set of 10 chemical-feature-based hypotheses were generated using the training set of 23 structurally diverse molecules. The top 10 ranked hypotheses along with their statistical parameters were presented in (Table 1). Intriguingly, all the generated hypotheses possessed a common set of pharmacophobic features, namely, hydrogen bond acceptor, aromatic ring and hydrophobic features, which may imply that hydrogen bond acceptor, aromatic ring and hydrophobic features were essential for inhibiting the biological functions of Mcl-1.

As shown in (Table 1), the null cost, the fixed cost and the configuration cost for the top 10 ranked hypotheses were 152.25, 84.36 and 16.46, respectively. The first pharmacophore model, Hypo1, was characterized by the highest cost difference (61.60), lowest RMSD (0.74) along with the best correlation coefficient (0.96) and thus was selected for further analysis. It consisted of spatial arrangement of four chemical features: hydrogen bond acceptor, hydrogen bond donor, hydrophobic and ring aromatic. The 3D space and distance constraints of these features were depicted in (Fig.

3(a)) and (Fig. 3(b)).

As for the best hypothesis, the cost difference between the null cost and the total cost was 61.60, which demonstrated that it had the probability of 90% to correlate the data



**Figure 3.** Pharmacophore models of Mcl-1 inhibitors generated by HypoGen. (a) The chemical features of the best hypothesis, Hypo1. (b) 3D spatial relationship and geometric parameters of Hypo1. (c) Hypo1 aligned with the most active compound **1** of the training set ( $K_i$ : 0.025  $\mu\text{M}$ ). (d) Hypo1 aligned with the least active compound **23** of the training set ( $K_i$ : 23  $\mu\text{M}$ ). Pharmacophore features are color-coded with blue for hydrophobic (H), orange for ring aromatic (AR), green for hydrogen bond acceptor (HBA) and magenta for hydrogen bond donor (HBD).

**Table 2.** The experimental and predicted activities of the training set compounds on the basis of Hypo1

Compound No.	Fit value <sup>a</sup>	Exp. $K_i$ (nm)	Est. $K_i$ (nm)	Error <sup>b</sup>	Exp. scale <sup>c</sup>	Est. scale <sup>c</sup>
1	6.50	0.025	0.017	-1.47	+++	+++
2	5.46	0.082	0.18	+2.20	+++	+++
3	5.57	0.099	0.14	+1.41	+++	+++
4	5.11	0.25	0.41	+1.64	+++	+++
5	5.30	0.26	0.27	+1.04	+++	+++
6	5.24	0.28	0.31	+1.10	+++	+++
7	5.27	0.34	0.28	-1.21	+++	+++
8	5.21	0.50	0.33	-1.52	+++	+++
9	4.52	1.51	1.62	+1.07	++	++
10	4.35	1.60	2.34	+1.46	++	++
11	4.81	1.69	0.83	-2.04	++	+++
12	4.08	1.90	4.40	+2.32	++	++
13	4.18	2.52	3.53	+1.60	++	++
14	3.86	3.00	7.41	+2.47	++	++
15	4.04	5.31	4.89	-1.09	++	++
16	3.85	6.32	7.50	+1.19	++	++
17	4.21	6.90	3.31	-2.08	++	++
18	3.85	8.00	7.48	-1.07	++	++
19	4.17	8.80	3.57	-2.46	++	++
20	3.85	13.00	7.51	-1.73	+	++
21	3.85	14.60	7.45	-1.96	+	++
22	3.11	22.00	41.13	+1.87	+	+
23	3.50	23.00	16.66	-1.38	+	+

<sup>a</sup>Fit value indicates how well the features in the pharmacophore overlap the corresponding chemical features in the molecule. <sup>b</sup>Difference between the predicted and experimental values. '+' indicates that the predicted  $K_i$  is higher than the experimental  $K_i$ ; '-' indicates that the predicted  $K_i$  is lower than the experimental  $K_i$ ; a value of 1 indicates that the predicted  $K_i$  is equal to the experimental  $K_i$ . <sup>c</sup>Activity scale:  $K_i < 1 \mu\text{M}$  = +++ (highly active);  $1 \mu\text{M} \leq K_i \leq 10 \mu\text{M}$  = ++ (moderately active);  $K_i > 10 \mu\text{M}$  = + (inactive).

**Table 3.** The experimental and estimated activities of the test set compounds against Hypo1

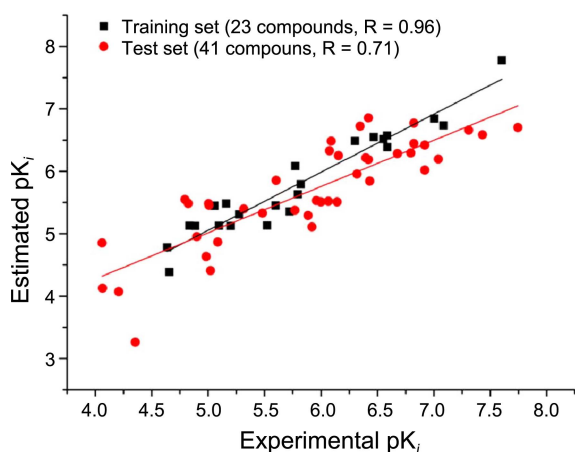
Compound No.	Fit value <sup>a</sup>	Exp. K <sub>i</sub> (μM)	Est. K <sub>i</sub> (μM)	Error <sup>b</sup>	Exp. scale <sup>c</sup>	Est. scale <sup>c</sup>
24	5.41	0.018	0.20	+11.11	+++	+++
25	5.31	0.037	0.26	+7.03	+++	+++
26	5.38	0.049	0.22	+4.49	+++	+++
27	4.92	0.091	0.64	+7.03	+++	+++
28	4.74	0.12	0.96	+8.00	+++	+++
29	5.15	0.12	0.38	+3.17	+++	+++
30	5.49	0.15	0.17	+1.13	+++	+++
31	5.17	0.15	0.36	+2.40	+++	+++
32	5.02	0.16	0.51	+3.19	+++	+++
33	5.01	0.21	0.52	+2.48	+++	+++
34	4.57	0.37	1.44	+3.89	+++	++
35	5.58	0.38	0.14	-2.71	+++	+++
36	4.91	0.38	0.65	+1.71	+++	+++
37	4.94	0.40	0.61	+1.53	+++	+++
38	5.45	0.45	0.19	-2.37	+++	+++
39	4.68	0.48	1.11	+2.31	+++	++
40	4.97	0.70	0.56	-1.25	+++	+++
41	4.23	0.72	3.10	+4.31	+++	++
42	5.21	0.81	0.33	-2.45	+++	+++
43	5.05	0.84	0.47	-1.79	+++	+++
44	4.25	0.86	3.02	+3.51	+++	++
45	4.23	1.00	3.11	+3.11	++	++
46	4.25	1.10	2.96	+2.69	++	++
47	3.83	1.20	7.77	+6.48	++	++
48	4.02	1.30	5.12	+3.94	++	++
49	4.10	1.70	4.22	+2.48	++	++
50	4.58	2.50	1.40	-1.79	++	++
51	4.05	3.29	4.69	+1.43	++	++
52	4.12	4.83	4.02	-1.20	++	++
53	3.59	8.19	13.6	+1.67	++	+
54	3.13	9.50	39.1	+4.12	++	+
55	4.18	9.80	3.49	-2.81	++	++
56	4.20	9.90	3.31	-2.99	++	++
57	3.36	10.41	23.23	+2.23	+	+
58	3.68	12.60	11.15	-1.13	+	+
59	4.21	15.00	3.27	-4.59	+	++
60	4.27	16.00	2.84	-5.63	+	++
61	1.99	44.00	546.32	+12.42	+	+
62	2.80	62.00	84.87	+1.37	+	+
63	2.85	86.00	74.8	-1.15	+	+
64	3.58	87.00	14.08	-6.18	+	+

<sup>a</sup>Fit value indicates how well the features in the pharmacophore overlap the corresponding chemical features in the molecule. <sup>b</sup>Difference between the predicted and experimental values. '+' indicates that the predicted K<sub>i</sub> is higher than the experimental K<sub>i</sub>; '-' indicates that the predicted K<sub>i</sub> is lower than the experimental K<sub>i</sub>; a value of 1 indicates that the predicted K<sub>i</sub> is equal to the experimental K<sub>i</sub>. <sup>c</sup>Activity scale: K<sub>i</sub> < 1 μM = +++ (highly active); 1 μM ≤ K<sub>i</sub> ≤ 10 μM = ++ (moderately active); K<sub>i</sub> > 10 μM = + (inactive).

accurately. Additionally, the fixed cost value of 84.36 was very close to the total cost value of 90.65. The configuration value was 16.46 for Hypo1 signifying the model was not obtained by chance.

To evaluate the predictive power of Hypo1, it was used to estimate the activities of training set molecules. The estimated activity values and the corresponding error values were listed in (Table 2).

The results showed that out of the 11 moderately active compounds **10** were predicted as moderately active compounds with exception of compound **11**, which was over-estimated as highly active compound. All the highly active compounds were predicted correctly. In terms of the four inactive compounds, 2 of which were predicted on the same order of magnitude, while the remaining were over-estimated as moderately active compounds. The error values reported in



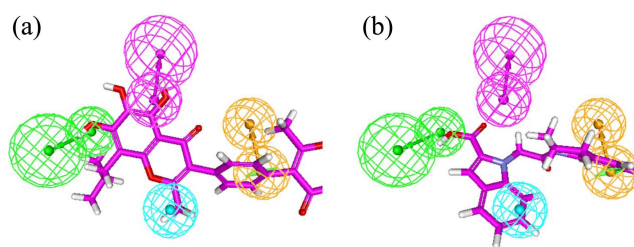
**Figure 4.** The regression of experimental and estimated activities for the test set of 41 molecules against Hypo1 along with 23 training set molecules.

(Table 2) are the ratio between the estimated and experimental activity values. The error values for all the training set compounds were below 3, demonstrating that a reasonable consistency between the estimated and experimental values. Furthermore, Hypo1 aligned with the most active compound **1** and the least active compound **23** in the training set, as showed in (Fig. 3(c)) and (Fig. 3(d)). The functional groups in compound **1** were perfectly mapped on the corresponding chemical features of Hypo1, while the least active compound **23** mapped on aromatic ring and hydrophobic features but failed to fit the HBD and HBA features.

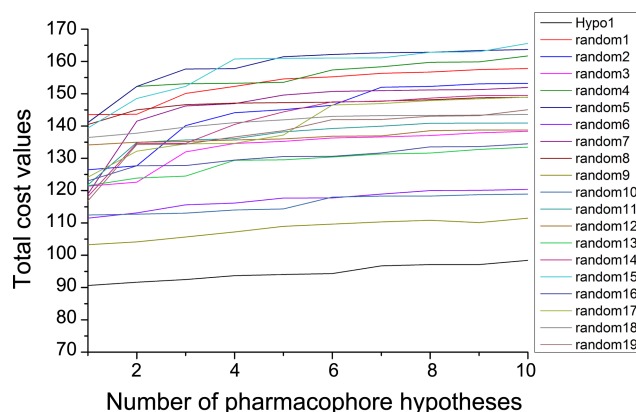
#### Pharmacophore Validation.

**Test Set Prediction Method:** With the purpose of assessing whether Hypo1 can also be capable of predicting the activities of external compounds accurately or not, an independent test set of 41 structurally diverse compounds, which were not included in training set, were used to evaluate the predictability of Hypo1 by applying the same approach as described in the training set. The estimated and experimental activities were summarized in (Table 3).

Analysis of the activity values presented in Table 3 revealed that out of the 21 highly active compounds, 4 compounds were under-estimated as moderately active compounds. For the 12 moderately active compounds, 10 compounds were classified as moderately active compounds and 2 compounds were under-estimated as inactive compounds. In the case of 8 inactive compounds, 6 of which were properly estimated as inactive compounds whereas the remaining 2 compounds were over-estimated as moderately active compounds. Subsequently, as shown in (Fig. 4), Hypo1 was regressed against the test set molecules and achieved a correlation coefficient of 0.71, which denoted the reasonable predictability of Hypo1. The most active compound **24** in the test set mapped effectively on the chemical features of Hypo1, as represented in (Fig. 5(a)). Besides, (Fig. 5(b)) depicted the alignment of compound **64** (the least active compound in the test set) on the Hypo1, which fitted the features of HBA, hydrophobic and ring aromatic but missed the HBD. On these bases, it



**Figure 5.** Pharmacophore mapping of the most active and least active compounds from the test set on Hypo1. (a) Pharmacophore mapping of the most active compound **24** ( $K_i$ : 0.018  $\mu$ M) on Hypo1. (b) Pharmacophore mapping of the least active compound **64** ( $K_i$ : 87  $\mu$ M) on Hypo1.



**Figure 6.** The difference in total cost of hypotheses between the initial hypothesis (Hypo1) spreadsheet and 19 random spreadsheets after Fischer's randomization run.

can be concluded that Hypo1 can predict the activities of external compounds accurately.

**Fischer's Randomization Method:** To further verify the statistical relevance of Hypo1, the cross-validation based Fischer's randomization method implemented in DS was employed. This technique randomly shuffles the activity values of training set molecules and then new spreadsheets are generated to validate the correlation between chemical structures and biological activities. In the present research, the confidence level was set to 95% and thus 19 spreadsheets were generated which were used as input for the generation of hypotheses using exactly the same parameters as used in modeling the original hypothesis. The statistical results of the generated hypotheses and Hypo 1 were represented in (Fig. 6).

It was observed that the original hypothesis, Hypo 1, was far more superior as compared with the 19 randomly generated hypotheses, which indicated that there is a 95% probability that Hypo1 represented a true correlation in the activity values of training set molecules.

**GH Test Method:** A small database containing 1606 molecules, spiked with different classes of 26 known inhibitors of Mcl-1, was used to evaluate whether Hypo1 was capable of discriminating the active from inactive compounds or not. The GH score ranges from 0 to 1, which indicates a null model and an ideal model, respectively. For the model to be



**Table 4.** The statistical parameters resulted from the GH test method

No.	Parameter	Values
1	Total number of molecules in the database ( <b>D</b> )	1632
2	Total number of actives in database ( <b>A</b> )	26
3	Total number of hit molecules from database ( <b>Ht</b> )	32
4	Total number of active molecules in hit list ( <b>Ha</b> )	22
5	%Yield of actives [ $(\text{Ha}/\text{Ht}) \times 100$ ]	68.75
6	%Ratio of actives [ $(\text{Ha}/\text{A}) \times 100$ ]	84.62
7	False negatives [ $\text{A} - \text{Ha}$ ]	4
8	False positives [ $\text{Ht} - \text{Ha}$ ]	10
9	Enrichment Factor <sup>a</sup> ( <b>EF</b> )	43.15
10	Goodness of fit score <sup>b</sup> ( <b>GH</b> )	0.72

GH score of 0.7-0.8 indicates a very good model. <sup>a</sup> $[(\text{Ha}/\text{Ht})/(\text{A}/\text{D})]$ .  
<sup>b</sup> $[(\text{Ha}/4\text{HtA})(3\text{A} + \text{Ht}) \times (1 - ((\text{Ht} - \text{Ha})/(\text{D} - \text{A})))]$ .

considered as reliable, the value of GH score must be greater than 0.5. The parameters, such as hit list, number of active percent of yields (%Y), percent ratio of active compounds in the hit list (%A), enrichment factor (EF), false positive, false negative and GH score, were calculated and reported in (Table 4).

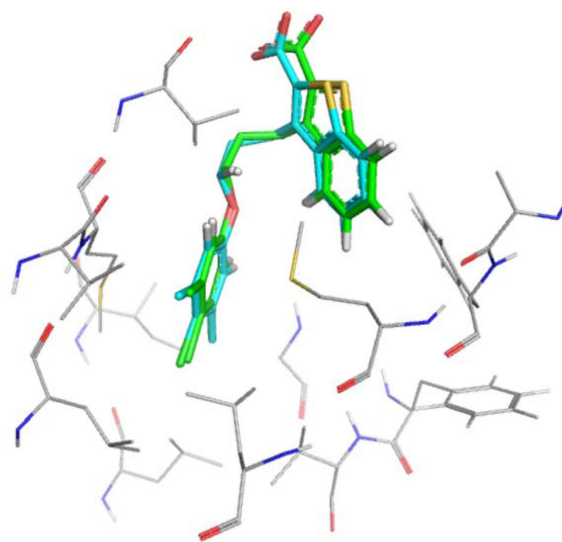
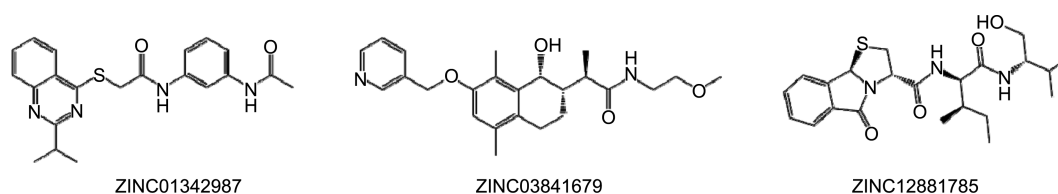
As depicted in (Table 4), the false positives and false negatives were 10 and 4, respectively and Hypo1 had successfully retrieved 68.75% of the active compounds, which indicated the competence of Hypo1 in figuring out active compounds properly. The EF and GH were found to be 43.15 and 0.72 indicating that Hypo 1 had a high efficiency in identifying active compounds during the virtual screening process.

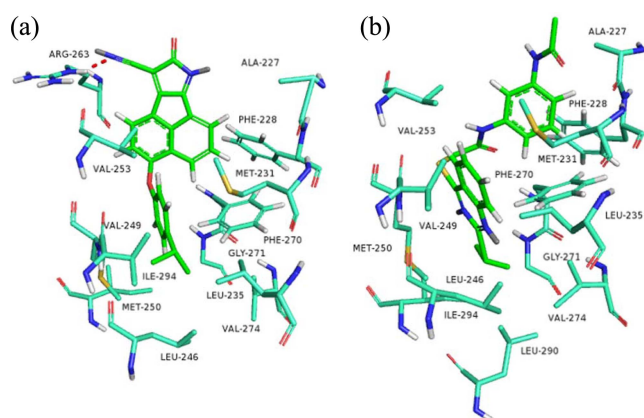
**Virtual Screening.** Since the pharmacophore model gives a methodology on how to get the necessary molecular dimension required for the inhibitor toward the target protein, the validated Hypo1 was used as a 3D search query to search the databases of Nature product and Asinex, which contains 141,050 and 145,633 molecules, respectively. Lipinski's rule of five was initially applied as a filter which retrieved a set of 107,898 compounds. This was followed by the virtual screening procedure using Hypo 1 as the search query, which yielded a total of 414 compounds based on the fact that they were perfectly mapped on all the chemical features of Hypo 1 as well as had a fit value of greater than 5.32 (the highest fit value for the active compounds in the data set). In addition, the hits should have good oral bioavailability, no toxicity and be capable of penetrating the blood brain barrier. Therefore, these hits were further screened for better ADMET properties and 107 hits were obtained. Finally, these hits which passed all the above filters were subjected

to molecular docking studies and visual inspection to explore the critical interactions between the essential amino acids within the binding site and the hits.

**Molecular Docking Studies.** Since that pharmacophore based virtual screening and molecular docking can be combined to identify novel and potential leads which were suitable for medicinal chemist to make them into drugs, thus all the obtained hits were further evaluated by docking studies to reduce the false positives. Initially, the reliability of the docking protocol was checked using the crystal structure of Mcl-1 in complex with the co-crystallized ligand Q19 (PDB code: 4HW3). The co-crystal ligand Q19 was re-docked into the binding site which was defined as a collection of amino acids within a sphere of 10 Å around the co-crystal ligand. The root mean square deviation (RMSD) between the co-crystal ligand and the best re-docked conformation was found to be 0.46. As showed in (Fig. 7), the best docked pose had all the interactions as the co-crystal ligand had, indicating the reliability of the docking protocol in terms of reproducing the experimentally determined binding mode.

Subsequently, the 107 hits along with the 23 training set compounds were docked into the active site of Mcl-1. According the crystal analysis reported by Anders Friberg, the pocket was linked with a collection of non-polar side chains, such as M231, L235, L246, V249, M250, L267,

**Figure 7.** The superimposition of the re-docked pose and co-crystal pose of the inhibitor within the active site of Mcl-1. The co-crystal conformation and re-docked conformation were showed in green and greencyan colors, respectively.**Figure 8.** The 2D structures of the three hits retrieved from virtual screening procedure.

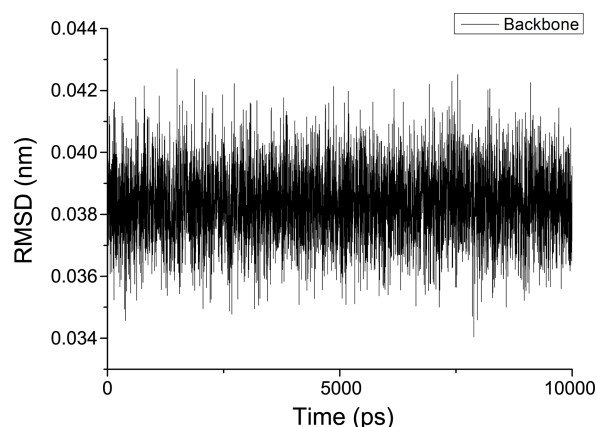


**Figure 9.** (a) Binding orientations of compound **3**, which is one of the most active compound in the training set. (b) Binding interactions of the best virtual hit (ZINC01342987) in the active cavity of Mcl-1. Hydrogen bond is shown in red dashed line.

F270, V273, L290 and I294.<sup>26</sup> Additionally, the pocket was divided into two sub-pockets, namely, the upper part and the lower part. Compound **3**, which was one of the most active compounds in the training set, had a GOLD fitness score of 63.30 and showed very good interaction with essential amino acids within the binding site. Hence, a GOLD fitness score of 63.0 was chose as a cutoff value to figure out the screened hits and 25 compounds turned out to comply with this criteria. These 25 molecules were further subjected to visual inspection on the basis of the binding mode and structural diversity. As a result, three hits (ZINC01342987, ZINC03841679 and ZINC12881785), which were different in their scaffolds and the 2D structures were depicted in (Fig. 8), were selected as the most promising hits.

As showed in (Fig. 9(a)), the nitrile group attached at the five-membered heterocyclic ring formed a hydrogen bond with R263. The benzene ring together with the isopropyl occupied the lower part of the deep pocket which was comprised of L235, L246, V249, M250, G271, V274 and I294, while the naphthalene was located in the upper part of the pocket and also established hydrophobic interaction with residues of A227, F228, M231, V253 and F270. (Fig. 9(b)) represented the binding mode of ZINC01342987, which scored the highest value of 72.80. It clearly indicated that it had the similar binding orientation at the active site.

The benzene ring along with the six-membered heterocyclic ring were positioned in the lower part of the pocket and were surrounded by lipophilic side chains, namely, L235, L246, V249 and M250, while the isopropyl equally formed hydrophobic interaction with residues of G271, V274, L290 and I294 within the lower part of the pocket. Additionally, the benzene ring between the peptide bonds occupied the upper part of the pocket, which contained A227, F228, M231, V253 and F270 likely. Obviously, an additional L290 was involved in the hydrophobic interaction as compared with the case of Compound **3**, which may attribute to the increased binding affinity. These three retrieved compounds showing higher GOLD fitness score as



**Figure 10.** Graphical representations of the RMSD of the protein backbone atoms versus the MD simulation time in the simulated complex.

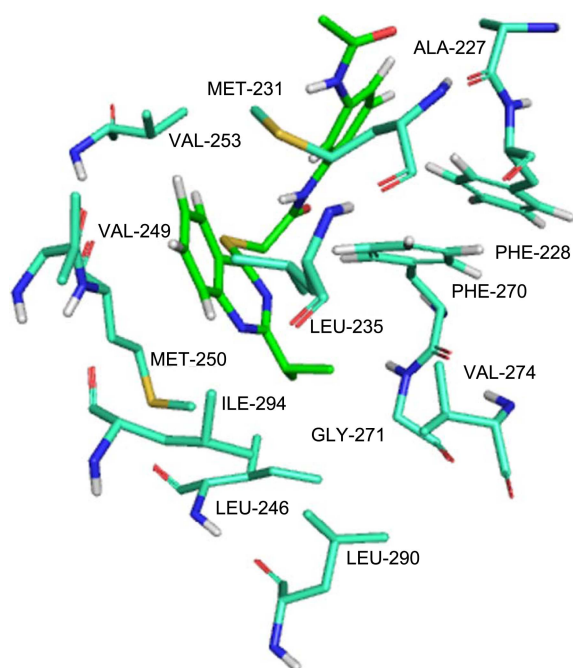
well as favorable interactions would be considered as potential leads in Mcl-1 inhibitors design.

**Molecular Dynamics Simulation.** Although molecular docking offers reasonable information on the binding mode of the investigated ligand, the MD simulation can figure out the smallest variances. Presently, a 10 ns simulation of the docked complex structure of Mcl-1 with ZINC01342987 was carried out in the aqueous solution from which we can obtain information on the conformational changes of the studied system. The convergence of the system equilibration was evaluated by the structural qualities during MD simulation. (Fig. 10) showed the root-mean-square deviation (RMSD) of the trajectory for the backbone atoms of the protein with respect to the initial structure.

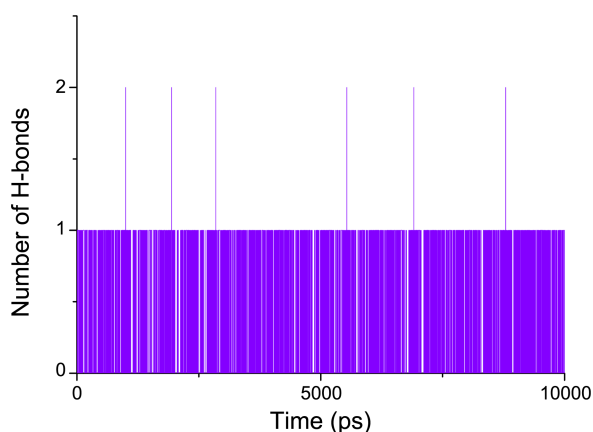
As seen from the figure, it can be clearly observed that the RMSD of the protein backbone atoms reached about 0.383 Å since the first picoseconds of the simulation, which led to the conclusion that the simulation produced stable structures, thus providing suitable basis for the subsequent analysis.

The binding orientation of the ligand in the active vicinity of the target protein, which was derived from the MD simulation, was depicted in (Fig. 11).

It could be noticed from this figure that the residues involved in the interaction between the protein and ligand did not change, suggesting the rationality and validity of the docking model. However, the binding conformation suffered from some movements as compared with the initial docking structure. It was obvious that the ligand in the cavity had a tendency to approach the surface of the protein leading to the result that the residue V253 was involved in the hydrophobic interaction with the benzene ring, which was attached on the six-membered heterocyclic ring. It could be explained by the small size of the ligand as well as the large hydrophobic space around the ligand. This may provide critical information for the optimization of this hit by increasing the hydrophobic volume reasonably. Additionally, the benzene moiety between the two amides of the ligand was anchored into the upper site of the pocket via lipophilic interaction involving side



**Figure 11.** The three-dimensional representation of interaction between the conformation of ZINC01342987 and the key residues within the active site of Mcl-1 after the MD simulation.



**Figure 12.** The observed intermolecular hydrogen bonds between ZINC01342987 and the active site residues monitored during the 10 ns MD simulation.

chains of A227, F228, M231 and F270. This interaction was also observed in the docking study leading to define such an interaction as the essential contact between the protein and ligand.

Finally, the number of hydrogen bonds formed during the MD simulation was monitored, as shown in (Fig. 12).

It clearly indicated that one hydrogen bond was formed mostly through the evolution of MD simulation, which may act as an essential role in anchoring the ligand in the active binding site. From the above analysis, we can conclude that the MD simulation can reveal some information about the binding conformation that the docking studies can't tell and the information may be critical for the further optimization of the ligand.

## Conclusion

In the present study, we firstly developed quantitative pharmacophore models based on a set of 23 structurally diverse compounds in order to understand the key chemical features for Mcl-1 inhibition. The best model, Hypo1, which was characterized by the lowest total cost (90.65), the lowest RMSD (0.74) and the best correlation coefficient (0.96), consisted of one hydrogen bond acceptor, one hydrogen bond donor, one hydrophobic and one ring aromatic. The statistical significance of Hypo1 was demonstrated by external test set of 41 compounds, which showed the correlation coefficient of 0.71, and GH test method, which achieved a GH score value of 0.72. Besides, Hypo1 was cross validated by Fischer's randomization method with confidence level of 95%. Before the virtual screening was carried out, the compounds of Nature product and Asinex, which contained 141,050 and 145,633 molecules, respectively, were initially filtered by applying the Lipinski's rule of five to improve the rate of drug-likeness. The drug-like hits retrieved from the screening procedure, in which the Hypo1 was used as a 3D search query, were further refined by applying ADMET, docking studies and visual inspection. Combining all these examinations, three compounds with new scaffolds were represented as possible lead candidates to reduce the over-expression of Mcl-1. Finally, the complex of the target protein with ZINC01342987, which had the highest GOLD score of 72.80, was subjected to the MD simulation with the time duration of 10 ns and the result revealed some valuable information for the further optimization of ZINC01342987 into a drug targeting Mcl-1.

**Acknowledgments.** Publication cost of this paper was supported by the Korean Chemical Society.

## References

- Adams, J. M.; Cory, S. *Science* **1998**, *281*, 1322-1326.
- Reed, J. C. *Adv. Pharmacol.* **1997**, *41*, 501-532.
- Chao, D. T.; Korsmeyer, S. J. *Annul. Rev. Immunol.* **1998**, *16*, 395-419.
- Minn, A. J.; Swain, R. E.; Ma, A.; Thompson, C. B. *Adv. Immunol.* **1998**, *70*, 245-279.
- Kelly, P. N.; Strasser, A. *Cell. Death. Differ.* **2011**, *18*, 1414.
- Korsmeyer, S. J. *Cancer. Res.* **1999**, *59*, 1693s-1700s.
- Youle, R. J.; Strasser, A. *Nat. Rev. Mol. Cell. Biol.* **2008**, *9*, 47.
- Johnstone, R. W.; Ruefli, A. A.; Lowe, S. W. *Cell.* **2002**, *108*, 153-164.
- Warr, M. R.; Shore, G. C. *Curr. Mol. Med.* **2008**, *8*, 138-147.
- Toumi, S. W.; Robillard, N.; Gomez, P.; Moreau, P.; Gouill, S. L.; Loiseau, H. A.; Harousseau, J. L.; Amiot, M.; Bataille, R. *Leukemia*. **2005**, *19*, 1248-1252.
- Kitada, S.; Andersen, J.; Akar, S.; Zapata, J. M.; Takayama, S.; Krajewski, S.; Wang, H. G.; Zhang, X.; Bullrich, F.; Croce, C. M.; Rai, K.; Hines, J.; Reed, J. C. *Blood*. **1998**, *91*, 3379-3389.
- Vega, J. H. C.; Rassidakis, G. Z.; Admirand, G. H.; Oyarzo, M.; Ramalingam, P.; Paraguya, A.; McDonnell, T. J.; Amin, H. M.; Medeiros, L. J. *J. Hum. Pathol.* **2004**, *35*, 1095-1100.
- Derenne, S.; Monia, B.; Dean, N. M.; Taylor, J. K.; Rapp, M. J.; Harousseau, J. L.; Bataille, R.; Amiot, M. *Blood*. **2002**, *100*, 194-199.

14. Khoury, J. D.; Medeiros, L. J.; Rassidakis, G. Z.; McDonnell, T. J.; Abruzzo, L. V.; Lai, R. J. *Pathol.* **2003**, *199*, 90-97.
  15. Aichberger, K. J.; Mayerhofer, M.; Krauth, M. T.; Skvara, H.; Florian, S.; Sonneck, K.; Akgul, C.; Derdak, S.; Pickl, W. F.; Wacheck, V.; Selzer, E.; Monia, B. P.; Moriggl, R.; Valent, P.; Sillaber, C. *Blood*. **2005**, *105*, 3303-3311.
  16. Rechsteiner, M.; Rogers, S. W. *Trends. Biochem. Sci.* **1996**, *21*, 267-271.
  17. Domina, A. M.; Vrana, J. A.; Gregory, M. A.; Hann, S. R.; Craig, R. W. *Oncogene*. **2004**, *23*, 5301.
  18. Kozopas, K. M.; Yang, T.; Buchan, H. L.; Zhou, P.; Craig, R. W. *Proc. Natl. Acad. Sci. USA*. **1993**, *90*, 3516-3520.
  19. Kim, J. H.; Sim, S. H.; Ha, H. J.; Ko, J. J.; Lee, K.; Bae, J. *FEBS. Lett.* **2009**, *583*, 2758-2764.
  20. Bae, J.; Leo, C. P.; Hsu, S. Y.; Hsueh, A. J. W. *J. Biol. Chem.* **2000**, *275*, 25255-25261.
  21. Wang, J. M.; Chao, J. R.; Chen, W.; Kuo, M. L. *Mol. Cell. Biol.* **1999**, *19*, 6195-6206.
  22. Liu, X. H.; Yu, E. Z.; Li, Y. Y.; Kagan, E. J. *Cell. Biochem.* **2006**, *97*, 755-765.
  23. Moulding, D. A.; Giles, R. V.; Spillers, D. G.; White, M. R. H.; Tidd, D. M.; Edwards, S. W. *Blood*. **2000**, *96*, 1756-1763.
  24. Zhang, Z. C.; Yang, H.; Wu, G.; Li, Z. Q.; Song, T.; Li, X. Q. *Eur. J. Med. Chem.* **2011**, *46*, 3909-3916.
  25. Zhang, Z. C.; Liu, C. W.; Li, X. Q.; Song, T.; Wu, Z. Y.; Liang, X. M.; Zhao, Y.; Shen, X. Y.; Chen, H. B. *Eur. J. Med. Chem.* **2013**, *60*, 410-420.
  26. Friberg, A.; Vigil, D.; Zhao, B.; Daniels, R. N.; Jason, J. P.; Burke, P.; Barrantes, P. M. G.; Camper, D.; Chauder, B. A.; Lee, T.; Olejniczak, E. T.; Fesik, S. W. *J. Med. Chem.* **2013**, *56*, 15-30.
  27. Tang, G. Z.; Ding, K.; Coleska, Z. N.; Yang, C. Y.; Qiu, S.; Shangary, S.; Wang, S. M. *J. Med. Chem.* **2007**, *50*, 3163-3166.
  28. Zhang, Z. C.; Wu, G.; Xie, F.; Song, T.; Chang, X. *J. Med. Chem.* **2011**, *54*, 1101-1105.
  29. Bernardo, P. H.; Sivaraman, T.; Wan, K. F.; Xu, J.; Krishnamoorthy, J.; Song, C. M.; Tian, L.; Chin, J. S. F.; Chai, C. L. *J. Med. Chem.* **2010**, *53*, 2314-2318.
  30. Smellie, A.; Teig, S. L.; Towbin, P. *J. Comput. Chem.* **1995**, *16*, 171-187.
  31. Smellie, A.; Kahn, S. D.; Teig, S. L. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 285-304.
  32. Kansal, N.; Silakari, O.; Ravikumar, M. *Eur. J. Med. Chem.* **2010**, *45*, 393-404.
  33. Kurogi, Y.; Guner, O. F. *Curr. Med. Chem.* **2001**, *8*, 1035-1055.
  34. Sprague, P. W. *Perspect. Drug. Discovery* **1995**, *3*, 1-20.
  35. Sakkiath, S.; Thangapandian, S.; John, S.; Kwon, Y. J.; Lee, K. W. *Eur. J. Med. Chem.* **2010**, *45*, 2132-2140.
  36. Clement, O. O.; Freeman, C. M.; Hartmann, R. W.; Handratta, V. D.; Vasaitis, T. S.; Brodie, A. M. *J. Med. Chem.* **2003**, *46*, 2345-2351.
  37. Lipinski, C. A.; Lombarde, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug. Deliv. Rev.* **2001**, *46*, 3-26.
  38. Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. *J. Mol. Biol.* **1997**, *267*, 727-748.
  39. Cheng, F.; Wang, Q.; Chen, M.; Quicho, F. A.; Ma, J. *Proteins* **2008**, *70*, 1228-1234.
  40. Lindahl, E.; Hess, B.; Spoel, V. D. *J. Mol. Mod.* **2001**, *7*, 306-317.
  41. Aalten, D. M. F.; Bywater, R.; Findlay, J. B. C.; Hendlish, M.; Hooft, R. W. W.; Vriend, G. *J. Comput. Aided. Mol. Des.* **1996**, *10*, 255-262.
  42. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.
  43. Hess, B.; Bekker, H.; Berendsen, H. J. C. *J. Comput. Chem.* **1997**, *18*, 1463-1472.
-