Chem Biol Drug Des 2011; **78:** 289–300

Research Article

doi: 10.1111/j.1747-0285.2011.01127.x

Discovery of Potential Integrin VLA-4 Antagonists Using Pharmacophore Modeling, Virtual Screening and Molecular Docking Studies

Sundarapandian Thangapandian, Shalini John, Sugunadevi Sakkiah and Keun Woo Lee*

Department of Biochemistry and Division of Applied Life Science (BK21 Program), Environmental Biotechnology National Core Research Center, Gyeongsang National University, 900 Gazwa-dong, Jinju 660-701, Korea
**Corresponding author: Keun Woo Lee, kwlee@gnu.ac.kr*

Very late antigen-4 (VLA-4) is an integrin protein, and its antagonists are useful as anti-inflammatory drugs. The aim of this study is to discover novel virtual lead compounds to use them in designing potent VLA-4 antagonists. A best pharmacophore model was generated with correlation coefficient of 0.935, large cost difference of 114.078, comprising two hydrogen bond acceptors and three hydrophobic features. It was further validated and used in database screening for potential VLA-4 antagonists. A homology model of VLA-4 was built and employed in molecular docking of screened hit compounds. Finally, two compounds were identified as potential virtual leads to be deployed in the designing of novel potent VLA-4 antagonists.

Key words: chronic inflammation, integrins, molecular docking, pharmacophore modeling, very late antigen-4 antagonists

Abbreviations: ADMET, absorption, distribution, metabolism, excretion and toxicity; BLAST, basic local alignment tool; COPD, chronic obstructive pulmonary disease; DS, Discovery Studio; EGF, epidermal growth factor; FDA, Food and Drug Administration; GOLD, genetic optimization for ligand docking; *HBA*, hydrogen bond acceptor; *HBD*, hydrogen bond donor; *HYP*, hydrophobic; MIDAS, metal ion adjacent to the active site; *NI*, negative ionizable; PDB, Protein Data Bank; *RA*, ring aromatic; RMSD, root mean square deviation; VCAM-1, vascular cell adhesion molecule-1; VLA-4, very late antigen-4.

Received 27 September 2010, revised 19 February 2011 and accepted for publication 1 April 2011

Very late antigen-4 (VLA-4, $\alpha4\beta1$ integrin, or CD49d/CD29) is a member of integrins family and constitutively expressed on most mononuclear lymphocytes including eosinophils, monocytes, and basophils but not on circulating neutrophils (1,2). Excessive leukocyte infiltration that leads to tissue injury causes inflammatory diseases of the lung such as chronic obstructive pulmonary disease (COPD) and asthma (3). Integrins are $\alpha\beta$ heterodimeric cell surface receptor molecules found in many animal species ranging from sponges to mammals (4). These adhesion receptors are essential for mediating important bidirectional signals during morphogenesis, tissue modeling, and repair. Integrins are formed by non-covalent heteroassembling of α and β subunits. One of the various roles of integrins is integrating the extracellular matrix and cytoskeleton, hence the name. Each subunit is composed of a large extracellular region, a single transmembrane domain and a short intracellular domain (5). To date, 18 α subunits and 8 β subunits have been identified. On the basis of its structurally distinct β subunits, integrins are classified into three families: (i) very late activation (VLA) family, (ii) leucam family and (iii) cytoadhesion family consisting $\beta1$, $\beta2$, and $\beta3$ subunits, respectively. Each family is found to have diverse ligand specificity. Other β subunits 4–8 were not classified into any family yet. Pharmaceutical industries are extensively investigating these adhesion molecules of integrins families for the development of novel therapeutics. VLA-4 proteins are found normally at resting or inactivated state. Although the mechanism of activation is not yet clear, it is believed that VLA-4 proteins are activated upon the various stimuli including antigens, anti-T-cell antibodies, phorbol esters, the divalent cation manganese (Mn^{2+}), and selected $\beta1$-specific antibodies (6,7). VLA-4 majorly binds to vascular cell adhesion molecule-1 (VCAM-1), which is from the immunoglobulin superfamily found in endothelial cells and the alternatively spliced connecting segment-1 domain of fibronectin on the extracellular matrix. VLA-4 binds to these ligands only after its activation. The efficacy of VLA-4 monoclonal antibodies and small molecular VLA-4 antagonists has been demonstrated in a number of animal models for inflammatory diseases (8,9), experimental allergic encephalomyelitis (10), multiple sclerosis (11), rheumatoid arthritis (12), and inflammatory bowel disease (13). Natalizumab, a monoclonal antibody for the α-4 subunit of VLA-4, has been approved by Food and Drug Administration (FDA) to be used in multiple sclerosis and Crohn's disease, but its prolonged usage was recently observed to develop progressive multifocal leukoencephalopathy (14). Valategrast, a VLA-4 antagonist, is in clinical trials and has shown beneficial effects in the treatment for asthma (15). The active metabolite of valategrast found to have dual antagonism on activated states of $\alpha4\beta1$ and $\alpha4\beta7$, which are the only two known $\alpha4$ VLA-4 proteins with high affinity (16,17). Integrins are validated drug targets, as the US FDA approved integrin antagonists such as abciximab ($\alpha IIb\beta3$) (18) and efalizumab ($\alpha L\beta2$) (19) for acute coronary syndromes and psoriasis, respectively. However, none has been

approved for indications related to asthma or COPD (3). VLA-4 and VCAM-1 binding is considered to be the important step in cell–cell and cell–matrix adhesion. These cell adhesion interactions could be essential for the activation, migration, proliferation, and differentiation of lymphocytes during physiological processes. Inhibition of VLA-4 and VCAM-1 binding may have the effect on prolonged inflammatory response (20,21).

An attempt has been made to find compounds of novel scaffolds as virtual leads to design potent VLA-4 antagonists. In this study, we have developed a pharmacophore model and validated the model employing three different validation methods. A homology model of extracellular domains of both the subunits of VLA-4 was constructed and used in molecular docking studies with the hit compounds obtained from the virtual screening of chemical databases. Two novel compounds were brought to light as potential virtual leads.

Materials and Methods

Selection of dataset from collected compounds

We have searched for the different biological assays that are used to estimate the VLA-4 inhibitory activities of chemical compounds. Two assay procedures (VLA-4 protein assay and VLA-4 cell adhesion assay), among others, were used predominantly for this purpose (22). Various resources of medicinal chemistry and life science journals have been explored to collect the compounds with their VLA-4 inhibitory activities using either one of these two assay procedures. Finally, 110 compounds were collected with their activity values spanning over a wide range of 0.5–83 000 nM estimated using VLA-4 cell adhesion assay with same experimental conditions. These compounds include diverse compounds from different chemical scaffolds such as squaric acid (22), phenylalanine (23–25), and quinolinyl derivatives (26). The inhibitory activities of these 110 compounds were expressed as IC_{50} (i.e., concentration of a compound at which it inhibits 50% of VLA-4 activity) in nanomolar concentration. These diverse compounds with their experimentally known IC_{50} values were utilized in this study.

Dataset preparation and conformational models

Twenty diverse compounds along with their activity values were selected as training set (Figure 1), and the remaining 90 compounds were kept as test set to be utilized in pharmacophore validation. This was performed based on the diversity observed both in terms of structure and experimental activity values. The most and least active

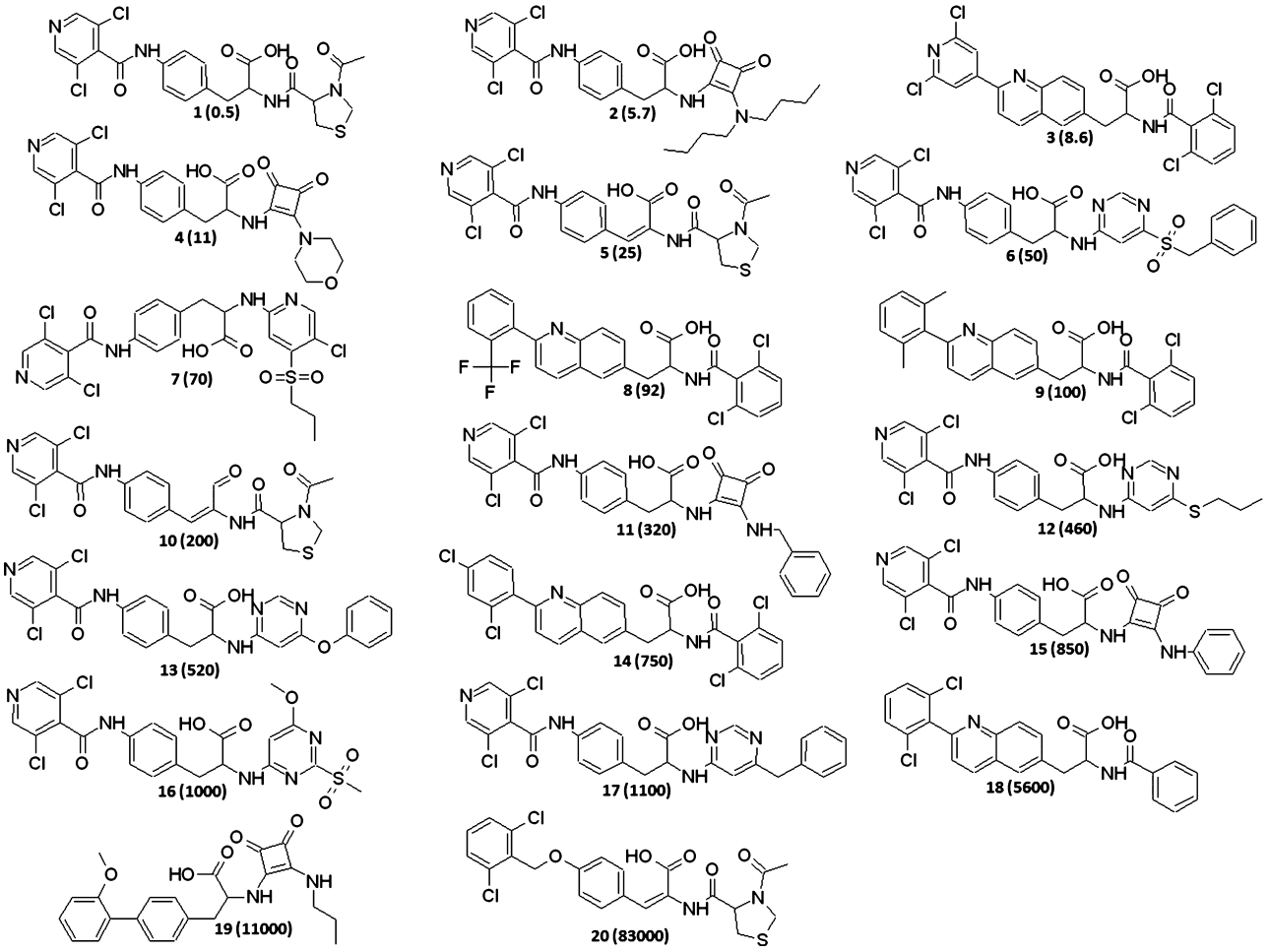

Figure 1: Training set compounds shown with their experimental activity values in nM concentration.

compounds were chosen to be in training set along with other compounds with diverse activity values. Preparation of training set has fulfilled the minimum requirements of *HypoGen*-automated pharmacophore generation to have at least 16 compounds with structural diversity and wide range of activity with a minimum of four orders of magnitude. Most active compound in the dataset has been included in the training set as it is considered inevitable for the automated generation of pharmacophore models using *HypoGen* (27,28) available in Accelrys Discovery Studio 2.1 (DS).[a] The E and Z isomers with considerable difference in their experimental activity values were also included in the training set to provide the isomeric information during pharmacophore generation. All training and test set compounds were sketched and saved in MDL-mol file format using ChemSketch, version 12 program.[b] Subsequently, they were exported to DS and saved in MDL-sd file format. The first step of the pharmacophore generation is to generate diverse conformations for every training set compound. Other drug design applications including protein docking, database searching, etc. also are critically dependent on using a good set of ligand conformations. It is well known that bioactive conformations often lie several kcal/mol higher in energy than the global minimum. Therefore, to get meaningful results, a set of diverse conformations within a reasonable energy window from the global minimum must be used. All the dataset compounds were converted into their standard 3D structures, and conformations were generated using *Diverse Conformation Generation* module of DS running with *Best conformations* option. This best conformation generation option utilizes CHARMm force field and Poling algorithm to ensure energy-minimized conformation for each compound and reduce conformational redundancies. Other parameters were kept default during conformation generation procedure. The conformations with energy higher than 20 kcal/mol from the global minimum were rejected, and the maximum number of conformations was set to 250. For compounds with unknown stereochemistry, all possible conformers were generated. Compounds with the generated conformations were used in pharmacophore generation. All the computations were carried out in DS unless it is mentioned.

Pharmacophore model generation

Two different methodologies (HipHop and HypoGen) are employed in automated generation of pharmacophore hypotheses (29). Whereas HipHop is used in generating pharmacophore models based on common chemical features present in most active compounds, HypoGen produces pharmacophore models based on structure–activity relationship in a set of compounds spanning activities of 4–5 orders of magnitude. It generates and ranks the pharmacophore models that correlate best the 3D arrangement of features in a given set of training compounds with the corresponding pharmacological activities (IC_{50} or K_i). This generation and ranking by HypoGen is executed in three steps: the constructive phase, the subtractive phase, and the optimization phase (30). In the constructive phase, hypotheses that are common to the most active set of compounds are identified. In the subtractive phase, all pharmacophore configurations that are also present in the least active set of molecules are removed. All compounds whose activity is by default 3.5 orders of magnitude less than that of the most active compound are considered to represent the least active molecules. During the optimization phase, the hypothesis score is improved. The optimization involves a variation of features and/or locations to optimize activity prediction via a simulated annealing approach. Various cost parameters were also calculated for each new hypothesis. When the optimization process no longer improves the score, HypoGen stops and reports the top-scoring 10 unique pharmacophore models. *HypoGen* implemented within DS has been employed to generate automated structure–activity relationship pharmacophore models using training set compounds. The *Feature Mapping* module available in DS computes all possible pharmacophore feature mappings for the selected ligands and features. An analysis performed beforehand using this module revealed that hydrogen bond acceptor (*HBA*), hydrogen bond donor (*HBD*), hydrophobic (*HYP*), ring aromatic (*RA*), and negative ionizable (*NI*) features map well with all the training set compounds. Thus, these features were selected and used to build pharmacophore models using *HypoGen* module of DS. Number of pharmacophore models have been generated to find the one with valuable statistics such as, high correlation coefficient between experimental and estimated activities of training set compounds, large cost difference between null cost and total cost, lowest RMSD to the conformations of training set compounds when projected on to the pharmacophore model and configuration cost, which is equal to the entropy of the hypothesis space. Hypothesis space is the conformational space accessible to the most active compounds in the training set. The best model is selected based on these statistical values. Having various parameters changed from their default values, we have found that 'Uncertainty factor, Minimum features and Minimum Interfeature Distance' had influenced the pharmacophore generation calculation. Finally, ten best pharmacophore models have been generated with the values 2, 4, and 4 given for 'Uncertainty, Minimum Features, and Minimum Interfeature Distance', respectively. Uncertainty factor was set to 2 instead of default value 3 (defined by *HypoGen* as the measured value being within two times higher or two times lower of the true value) as it correlates well the activity values of training set compounds (31). Indeed, HypoGen generates a chemical feature-based model on the basis of the most active compounds excluding features from inactive compounds within conformationally allowable regions of space. It also predicts the activity of training set compounds using regression parameters. The parameters are computed by the regression analysis using the relationship of geometric fit value versus the negative logarithm of activity. The greater the geometric fit, the higher the activity prediction of the compound. The fit function not only checks whether the feature is mapped or not, but it also contains a distance term, which measures the distance that separates the feature on the molecule from the centroid of the hypothesis feature. Both terms are used to calculate the geometric fit value. The *HypoGen* method calculates two important theoretical cost values, represented in bit units, for determining the success of any pharmacophore hypothesis. One is the 'fixed cost' also known as ideal cost, which represents the simplest model that fits all data perfectly, and another one is the 'null cost', also known as no correlation cost, which depicts the highest cost of a pharmacophore with no features and estimates activity to be the average of the activity data of the training set molecules. A meaningful pharmacophore hypothesis may result in a large difference between null and fixed cost values, in which a value of 40–60 bits may indicate 75–90% probability of correlating the data. Another parameter for estimating the model quality is the configuration cost or entropy cost,

which depends on the complexity of the pharmacophore hypothesis space and should have a value <17. The RMSD represents the quality of the correlation between the experimental and the estimated activity data. Total cost is another value calculated for every pharmacophore model. A significant pharmacophore model should have the total cost closer to fixed cost and away from null cost values thereby large difference between null and total cost values along with high correlation coefficient and lower RMSD values. All the cost values are represented in bit units. Ten statistically significant pharmacophore models were generated. The best model was selected and validated to be used in further studies.

Pharmacophore model validation

The selected pharmacophore model was validated by three different validation methods. First, a test set comprising 90 diverse compounds including few isomers was used to validate the generated pharmacophore models. *Ligand pharmacophore mapping* module implemented in DS was used with *Best/Flexible search* option. Second, Fischer's randomization method was employed to prove the confidence level of best pharmacophore model that it has not been generated randomly. The confidence level of a particular pharmacophore model has been calculated using the following formula: $S = 100\ (1–(1 + X)/Y)$, where S = significance level or confidence level; X = the number of randomized hypotheses scored better than the selected pharmacophore model; Y = the number of random hypotheses generated. As a third method of validation, a small database of 38 active and 381 inactive compounds was generated and used to evaluate the capability of the generated pharmacophore model calculating enrichment factor (E) and goodness-of-hit (GH) values. This database was built using the datasets retrieved from QSAR-world,[c] binding database,[d] and patents (32,33). The reason for the validation is to cross check the pharmacophore model for its ability to predict the activity of external compounds accurately and thereby to pick the active compounds from the database.

Database searching

Selected pharmacophore model built within DS was used as query to search two databases of commercially available chemical compounds namely Maybridge[e] and Chembridge[f] containing 60 000 and 50 000 diverse compounds, respectively. Database screening serves the purpose of finding novel, potential virtual leads suitable for further optimization and provides the preference of providing easily available and/or synthesizable compounds as hits for further steps in drug development. A molecule of the database has to map all the pharmacophoric features of the pharmacophore model to be retrieved as a hit. In our study, all screenings were carried out using *Fast Flexible Search* method. Retrieved database compounds were subjected to various drug-like screening filters to select only the compounds with drug-like properties and reject the compounds with non-drug-like properties. Compounds those passed these screenings were considered for molecular docking.

Homology modeling

An extensive search using BLAST (BLASTP, Bethesda, MD, USA) program implemented in DS against the Protein Data Bank (PDB) database was made to identify homologous proteins that can be used as template in homology modeling study. Sequence alignment was carried out using the *Align multiple Sequences* module of DS to align template and target protein. Homology models for both the subunits of VLA-4 were constructed using the crystal structure of closely related $\alpha V\beta 3$ as a template (PDBID: 1L5G) (34). *Build Homology models* module available in DS was employed to construct homology model, and the final model was minimized using *Smart Minimizer* protocol of DS, which utilizes *Steepest Descent* and *Conjugate Gradient* algorithms, with the default parameters. The minimized structure was further validated using PROCHECK (35), WHATIF (36), and PROSA (37,38) analyses and utilized in molecular docking study.

Molecular docking

All the compounds retrieved from the databases along with training set compounds were docked as a further screening procedure based on the necessary protein–ligand interactions. Genetic Optimization for Ligand Docking (GOLD 4.1) from Cambridge Crystallographic Data Center, UK, has been employed in molecular docking (39). GOLD uses a genetic algorithm for docking flexible ligands into protein-binding sites to explore the full range of ligand conformational flexibility with partial flexibility of the protein. Coordinates of homology-modeled VLA-4 was exported to GOLD, and the active site was defined using the XYZ coordinates of oxygen atom attached with the phenyl ring of Y187, which is one of the important active site residues and located at the center, with the active site radius of 10 Å. The active site was defined using Y187 as there was no X-ray-based ligand binding information for VLA-4 protein is available. The annealing parameters of van der Waals and hydrogen bond interactions were considered within 4.0 and 2.5 Å, respectively. The total GOLD score, which is also known as 'Fitness', was calculated from the following equation:

$$\text{Fitness} = S_{\text{hb ext}} + 1.375 * S_{\text{vdw ext}} + S_{\text{hb int}} + S_{\text{vdw int}}$$

$S_{\text{hb ext}}$ and $S_{\text{vdw ext}}$ denote the contribution from hydrogen bonds and van der Waals interactions between protein and ligand, respectively. $S_{\text{hb int}}$ represents the contribution of intramolecular hydrogen bonds in the ligand, and $S_{\text{vdw int}}$ is the contribution from intramolecular strain in the ligand. The external vdw score is multiplied by a factor of 1.375 when the total fitness score is computed.[g] This is an empirical correction to encourage protein–ligand *HYP* contact. Molecular docking results from GOLD have been exported to DS and analyzed. Ligand and metal ion interactions have been observed in Molegro Virtual Docker program (40).

Results and Discussion

Cost analysis and training set prediction

A training set of 20 compounds tested against VLA-4 was used to develop a total of 10 pharmacophore models. Statistical parameters including cost functions calculated during the pharmacophore generation were used to evaluate the generated pharmacophore models. Overall, pharmacophore model 1 ('Hypo 1') was generated with the best statistical significance as its total cost (99.895 bits) is close to the fixed cost (75.874 bits), and a high cost difference (114.078

bits). These results are the indicators of greater predictability of the pharmacophore model, 'Hypo 1'. Top 10 ranked pharmacophore models as well as their statistical parameters are presented in Table 1.

In addition to the cost analysis, 'Hypo 1' was validated for its capacity to predict the activity of the training set compounds. Three-dimensional arrangements of features and their distance constraints of 'Hypo 1' have been shown in Figure 2. 'Hypo 1' was comprised of two *HBA* and three *HYP* features and has been used to estimate the activity of the compounds in the training set. All 10 pharmacophore models possessed at least one *HBA* and *HYP* features, while six of 10 were five feature pharmacophore models. Interestingly, although predicted with the *Feature Mapping* module, no HBD feature was present in any of the 10 hypotheses generated. 'Hypo 1' with the high cost difference has also scored the low RMSD of 1.412, high correlation coefficient of 0.935 and a configuration cost of 16.688. Configuration cost must be <17 bits for a good pharmacophore model (41). Figure 3 displays the mapping of most and least active compounds of the training set on the best pharmacophore model. The most active compound present in the training set maps all the features of 'Hypo 1', whereas the least active compound misses two *HBA* features. This explains well the difference in activities between the most and the least active compounds.

Training set compounds have roughly been classified into three categories on the basis of their activity values: active (IC_{50} < 20 nM, +++), moderately active (20 nM < IC_{50} ≤ 2000 nM, ++), and inactive (IC_{50} > 2000 nM, +). All the training set compounds except compound **7** were estimated correctly within their activity scales including the Z and E isomers (compound **5** and **10**), and the results have been tabulated in Table 2. Error values reported in Table 2 were the ratio of the estimated activity and experimental activity. Positive error value indicates that the estimated activity is greater than experimental activity, and negative error value indicates that the estimated activity is smaller than experimental activity. Correlation plot drawn between the experimental and estimated activities of the training set compounds is shown in Figure 4.

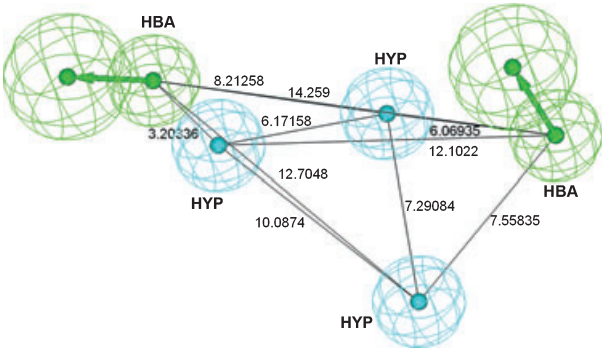

Figure 2: 'Hypo 1' is shown with distance constraints. 'Hypo 1' consists of two hydrogen bond acceptor (green) and three hydrophobic (cyan) features.

Validation of pharmacophore model

A test set containing 90 compounds with the diverse structures and activity values was used as an external set to verify the ability of the 'Hypo 1' on estimating the activity of compounds that are structurally different from that of training set. All ten pharmacophore models were evaluated to predict the activities for test set compounds. Among ten models, 'Hypo 1' has predicted the activity values of test set compounds accurately. Experimental and estimated activities of the test set compounds based on 'Hypo 1' are shown in Table 3 along with their error values. Correlation coefficient of 0.907 for 90 test set compounds was obtained. It indicates the high predictive ability of 'Hypo 1' on external chemical compounds and the correlation plot between experimental and estimated activities shown in Figure 4. Except four compounds in the test set all including E and Z isomers (compound **107** and **110**) were estimated with the error value <10, representing not more than one order of magnitude difference between experimental and estimated activity values. Compound **101**, which is the Z isomer of compound **20**, has also been predicted close to its experimental activity value. Overall, only 12 out of 90 test set compounds have been estimated in different activity scale (i.e., inactive compounds as moderately active) ensuring 86.66% success rate.

Table 1: Statistical results of top ten pharmacophore hypotheses generated

Hypo	Total cost	Cost difference[a]	RMSD[b]	Error cost	Correlation	Features[c]	Test set correlation
1	99.90	114.078	1.412	77.99	0.935	*HBA HBA HYP HYP HYP*	0.907
2	116.48	97.494	1.966	96.71	0.868	*HBA HBA HYP HYP NI*	0.896
3	118.05	95.923	2.014	98.63	0.861	*HBA HYP HYP HYP HYP*	0.884
4	118.51	95.467	2.027	99.16	0.859	*HBA HBA HYP HYP NI*	0.867
5	119.29	94.688	1.967	96.74	0.869	*HBA HYP HYP RA*	0.853
6	119.49	94.482	1.904	94.32	0.880	*HBA HBA HBA HYP*	0.885
7	121.89	92.081	1.941	95.73	0.875	*HBA HYP HYP HYP NI*	0.826
8	122.01	91.959	2.148	104.20	0.839	*HBA HBA HYP HYP NI*	0.810
9	122.94	91.029	2.015	98.65	0.863	*HBA HYP HYP RA*	0.778
10	123.16	90.811	2.130	103.41	0.843	*HBA HBA HYP RA*	0.767

[a]Cost difference, the difference between the null cost and the total cost, the null cost of the 10 top-scored hypotheses is 213.973, the fixed cost value is 75.874, and the configuration cost is 16.688. All costs are represented in bit units.
[b]RMSD, the deviation of the log (estimated activities) from the log (experimental activities) normalized by the log (uncertainties).
[c]*HBA*, hydrogen bond acceptor; *HYP*, hydrophobic; *NI*, negative ionizable; *RA*, ring aromatic.

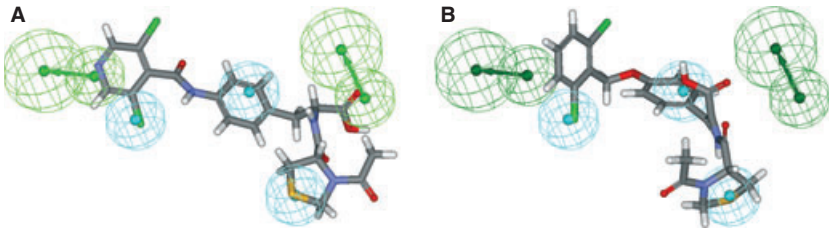

Figure 3: Most active compound **1** (A) and least active compound **20** (B) were overlaid upon the best pharmacophore model 'Hypo 1'. Missed hydrogen bond acceptor features are shown in dark green.

Table 2: Prediction of training set compounds using 'Hypo 1'

Name	Fit value	IC_{50}, nM		Error[a]	Activity scale[b]	
		Experimental	Estimated		Experimental	Estimated
1	12.84	0.5	1.4	2.8	+++	+++
2	11.82	5.7	14	2.5	+++	+++
3	11.80	8.6	15	1.8	+++	+++
4	11.83	11	14	1.3	+++	+++
5	11.90	25	12	−2.1	++	++
6	10.49	50	310	6.2	++	++
7	12.00	70	9.6	−7.3	++	+++
8	10.10	92	760	8.3	++	++
9	10.64	100	220	2.1	++	++
10	10.57	200	260	1.3	++	++
11	10.50	320	300	−1.1	++	++
12	10.40	460	380	−1.2	++	++
13	10.47	520	330	−1.6	++	++
14	10.41	750	370	−2	++	++
15	10.12	850	730	−1.2	++	++
16	10.41	1000	380	−2.7	++	++
17	10.30	1100	490	−2.3	++	++
18	9.64	5600	2200	−2.5	+	+
19	8.94	11 000	11 000	1	+	+
20	8.41	83 000	37 000	−2.2	+	+

[a]Value in the error column represents the ratio of the estimated activity to the experimental activity or its negative inverse if the ratio is <1.
[b]Activity scale: IC_{50} < 20 nM = +++ (Most active), 20 nM < IC_{50} ≤ 2000 nM = ++ (Moderate active), IC_{50} > 2000 nM = + (Inactive).

In addition, Fischer randomization test has been used to validate the statistical relevance of 'Hypo 1'. The purpose of this type of validation is to check whether there is a strong correlation between the chemical structures and the biological activity. According to the randomization procedure, the experimental activities of the compounds in the training set were scrambled randomly, and the resulting new training sets were used for a number of new *HypoGen* runs. The parameters used in running these randomization calculations were the same as employed in the initial *HypoGen* calculation. Pharmacophore generation was performed with 95% confidence level in this validation test. Nineteen random spread sheets were generated as result and have been compared with the statistical parameters of 'Hypo 1'. A significant pharmacophore should score the better statistical values such as low RMSD, high correlation, and large cost difference compared to any of the random pharmacophore models. Results of Fischer randomization test has been presented in Table 4. None of the randomly generated pharmacophore

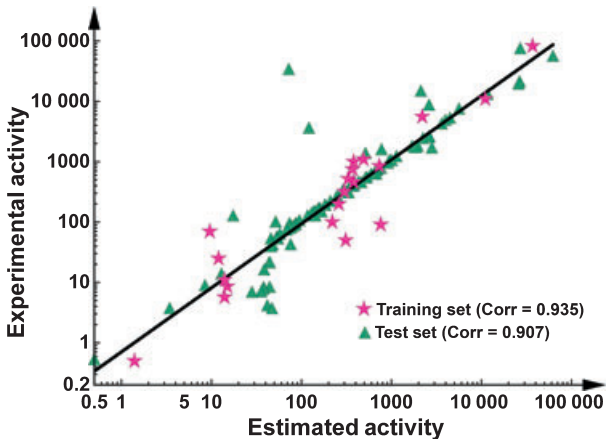

Figure 4: Correlation plot between experimental and estimated activities of training and test set compounds.

models has scored better than 'Hypo 1' in terms of total cost, RMSD, and correlation coefficient. Three of the random pharmacophore models have shown the correlation coefficient more than 0.9, but their total cost and RMSD values were higher than 'Hypo 1'. So they are not statistically robust when compared to 'Hypo 1' (42). Thus, the confidence level of the best pharmacophore model 'Hypo 1' was calculated to be 95%.

A database containing 419 of active and inactive compounds was used as decoy set to evaluate the ability of 'Hypo 1' to identify the active compounds from the inactive compounds. Various parameters such as percentage yield of actives, percentage ratio of actives, E, and GH values were calculated. *Ligand Pharmacophore Mapping* protocol as implemented in DS was employed to screen the compounds in the database. The calculated E value was 9.49 for the pharmacophore model, 'Hypo 1', indicating that the 'Hypo 1' is 9.49 times more probable to pick an active compound from the database than picking an inactive one. Parameters, the formulae used in calculation, and the results of this validation procedure are displayed in Table 5.

Database searching

The validated pharmacophore model, 'Hypo 1', was used as a 3D query to search potential antagonists from chemical databases namely, Maybridge and Chembridge. *Ligand Pharmacophore Mapping* module available in DS with *Fast flexible search* option was

Table 3: Prediction of test set compounds using 'Hypo 1'

	IC_{50}, nM			Activity scale[d]			IC_{50}, nM			Activity scale	
Name	Expt.[a]	Estd.[b]	Error[c]	Expt.	Estd.	Name	Expt.	Estd.	Error	Expt.	Estd.
21	0.5	0.51	1.002	+++	+++	66	290	335.29	1.156	++	++
22	3.5	3.44	−1.019	+++	+++	67	300	319.76	1.066	++	++
23	3.5	47.14	13.467	+++	++	68	320	289.64	−1.105	++	++
24	3.9	41.87	10.736	+++	++	69	320	299.82	−1.067	++	++
25	6.5	28.22	4.341	+++	++	70	370	346.40	−1.068	++	++
26	6.6	35.73	5.413	+++	++	71	372	366.49	−1.014	++	++
27	7.3	37.77	5.174	+++	++	72	381	386.19	1.014	++	++
28	7.8	38.14	4.890	+++	++	73	425	393.40	1.080	++	++
29	7.9	44.36	5.615	+++	++	74	430	462.87	1.076	++	++
30	8.4	8.50	1.012	+++	+++	75	435	434.55	−1.001	++	++
31	13	12.93	−1.005	+++	+++	76	510	514.03	1.008	++	++
32	15.3	38.62	2.524	+++	++	77	520	531.94	1.023	++	++
33	20	43.84	2.192	++	++	78	590	676.19	1.146	++	++
34	20	44.62	2.231	++	++	79	590	593.88	1.007	++	++
35	37.2	46.16	1.241	++	++	80	670	674.70	1.007	++	++
36	40	47.88	1.197	++	++	81	690	675.41	−1.022	++	++
37	40	75.87	1.897	++	++	82	710	706.03	−1.006	++	++
38	50	45.82	−1.091	++	++	83	740	753.44	1.018	++	++
39	50	55.07	1.101	++	++	84	740	765.15	1.034	++	++
40	50	52.49	1.050	++	++	85	900	915.91	1.018	++	++
41	53.7	53.38	−1.006	++	++	86	990	1010.63	1.021	++	++
42	60	57.86	−1.037	++	++	87	1000	961.51	−1.040	++	++
43	70	68.65	−1.019	++	++	88	1160	1135.39	−1.022	++	++
44	75	75.57	1.008	++	++	89	1300	514.48	−2.527	++	++
45	80	80.14	1.002	++	++	90	1500	775.17	1.935	++	++
46	85	85.02	1.000	++	++	91	1600	2806.33	1.754	++	+
47	90	73.91	−1.218	++	++	92	1620	1845.07	1.139	++	++
48	90	86.36	−1.042	++	++	93	1700	1958.25	1.152	++	++
49	95	51.72	−1.837	++	++	94	1700	1691.53	−1.005	++	++
50	100	94.79	−1.055	++	++	95	2300	2240.40	−1.027	+	+
51	120	116.82	−1.027	++	++	96	2400	2619.79	1.092	+	+
52	120	142.76	1.190	++	++	97	2500	2623.21	−1.049	+	+
53	120	17.49	−6.859	++	+++	98	3400	121.45	−27.994	+	++
54	130	128.72	−1.010	++	++	99	4000	3617.35	−1.106	+	+
55	140	141.88	1.013	++	++	100	4600	3979.63	−1.156	+	+
56	140	136.21	−1.028	++	++	101	5000	4466.85	−1.119	+	+
57	140	167.08	1.193	++	++	102	7100	5631.85	−1.261	+	+
58	153.5	153.29	−1.001	++	++	103	8200	2614.63	−3.136	+	+
59	155	153.25	−1.011	++	++	104	12400	11700.50	−1.060	+	+
60	160	160.29	1.001	++	++	105	14000	2121.34	−6.599	+	+
61	180	183.89	1.022	++	++	106	18300	25835.90	1.412	+	+
62	180	181.809	1.010	++	++	107	20000	26585.60	1.329	+	+
63	208.5	209.05	1.003	++	++	108	32000	73.02	−438.230	+	++
64	240	243.46	1.014	++	++	109	53000	62178.00	1.173	+	+
65	250	250.62	1.002	++	++	110	71000	26968.30	−2.633	+	+

[a]Expt., Experimental activity.
[b]Estd., Estimated activity.
[c]Value in the error column represents the ratio of the estimated activity to the experimental activity or its negative inverse if the ratio is <1.
[d]Activity scale: IC_{50} < 20 nM = +++ (Most active), 20 nM < IC_{50} ≤ 2000 nM = ++ (Moderate active), IC_{50} > 2000 nM = + (Inactive).

used to screen the databases with a *Maximum Omitted Features* value of '0' (43). The query identified 3274 and 2303 hit compounds from Maybridge and Chembridge databases, respectively. Of these, total of 392 compounds were found to fit the query with fit value more than 10. These top-scored compounds have been subjected to various screenings using Lipinski's rule of five and drug-like absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties (44,45) employing human intestinal absorption, aqueous solubility, blood brain barrier penetration, hepatotoxicity, and CYP450 2D6 inhibition models available in DS. The 186 of 392 compounds were found to violate any one of the Lipinski's rules. Thus, remaining 206 compounds were considered for further screening based on their ADMET properties (Table S1). Finally, after rejecting 194 compounds that were predicted with unfavorable ADMET

Table 4: Validation results for 'Hypo 1' using Fischer randomization test

Validation no.	Total cost	Fixed cost	RMSD	Correlation	Configuration cost
Results for unscrambled					
Hypo 1	99.895	75.874	1.412	0.935	16.688
Results for scrambled					
Trial 1	107.645	75.385	1.700	0.903	16.199
Trial 2	136.599	75.077	2.415	0.793	15.891
Trial 3	128.015	76.674	2.249	0.822	17.488
Trial 4	131.611	75.915	2.176	0.840	16.729
Trial 5	138.173	76.056	2.481	0.778	16.870
Trial 6	127.238	76.290	2.173	0.837	17.104
Trial 7	147.542	75.924	2.586	0.760	16.738
Trial 8	132.573	75.801	2.185	0.840	16.616
Trial 9	131.881	76.607	2.282	0.817	17.421
Trial 10	134.355	74.425	2.296	0.818	15.239
Trial 11	129.351	75.414	2.101	0.853	16.228
Trial 12	131.907	76.078	2.290	0.816	16.892
Trial 13	129.380	76.153	2.282	0.816	16.967
Trial 14	141.553	75.068	2.457	0.787	15.882
Trial 15	133.547	75.566	2.124	0.853	16.380
Trial 16	108.724	75.884	1.631	0.913	16.698
Trial 17	105.794	77.560	1.579	0.918	18.374
Trial 18	119.198	76.660	2.060	0.853	17.474
Trial 19	121.798	77.331	1.913	0.879	18.145

All costs are represented in bit units.

Table 5: Enrichment factor and goodness of hit score validation for 'Hypo 1'

Parameters	Values
Total molecules in database (D)	419
Total number of actives in database (A)	38
Total hits (Ht)	43
Active hits (Ha)	37
% Yield of actives [(Ha/Ht)*100]	86.05
% Ratio of actives [(Ha/A)*100]	97.36
Enrichment factor (E) [(Ha*D)/(Ht*A)]	9.49
False negatives [A − Ha]	1
False positives [Ht − Ha]	6
Goodness of hit score (GH)[a]	0.88

[a][(Ha/4HtA) (3A + Ht))*(1 − ((Ht − Ha)/(D − A))]; GH Score of >0.7 indicates a very good model.

properties, 12 compounds with favorable ADMET properties, such as good intestinal absorption, optimal and good aqueous solubility, low blood brain barrier penetration, non-inhibition of CYP450 2D6 enzyme, and with no hepatotoxicity were selected and docked into the active site of the homology modeled VLA-4.

Homology modeling

Very late antigen-4 receptors are cell surface receptors come under the family of integrins, which are known for cell–cell and cell–matrix interactions. Integrins are heterodimers formed by assembling of α and β subunits. VLA-4 constitutes of $\alpha 4$ and $\beta 1$ subunits. Some research groups have tried building the model for VLA-4 using its closest homologous protein that shares maximum identity as template (5,46). Crystal structure of $\alpha V\beta 3$ (PDB ID: 1L5G), after its deposition in PDB in 2002, was predominantly used as a template to model other integrins including VLA-4. Whole extracellular domains including the head piece of both α and β subunits (except EGF-1 and EGF-2 domains of β subunit) of VLA-4 using the crystal structure of $\alpha V\beta 3$ as template have been modeled in this study. The structural information of EGF-1 and EGF-2 domains was not available in the crystal structure of $\alpha V\beta 3$ because of the insufficient/poor electron density (4). So the part corresponding to these domains in VLA-4 could not be modeled and truncated from the primary sequence of β subunit ($\beta 1$) of VLA-4 after the sequence alignment. The models that have been reported previously in the literature comprised only the headpiece of α subunit of VLA-4 (5). Figures S1 and S2 show the sequence alignment between the template ($\alpha V\beta 3$) and the target sequence ($\alpha 4\beta 1$). *Align Multiple Sequences* module of DS showed 24.3% identity and 46.7% similarity between α subunits and 44.3% identity and 64.3% similarity between β subunits of both the template and target proteins. *Build homology models* module of DS has been employed to build the homology models. DS has some useful options such as listing out (i) all the cysteine amino acids and making them available to users to define the necessary disulfide bonds to be formed during the modeling process and (ii) all the available ligands and metal ions in the template to be copied during the modeling process. All the cysteine residues that have been listed in the SwissProt page of the human VLA-4 (both α and β subunits) known to be engaged in disulfide bonds were chosen to form disulfide bonds in the homology model. All the divalent metal (Mn^{2+}) ions that are cocrystallized with template protein also have been copied in to the model, during the homology modeling process. Three independent tests were performed to validate the constructed homology model before using in molecular docking experiments. As the first test of validation, the PROCHECK analysis predicted 98.4% of amino acid residues (total number of amino acids is equal to 1494) in favorable regions where 75%, 18.7% and 4.6% of amino acids were predicted in core, allowed, and generously allowed regions. The template protein had only 69.7% of its amino acids in core regions. The 93.6% and 83.6% of main-chain bond lengths and main-chain bond angles were predicted within limits, and 99.6% of planar groups were also within limits. The second validation of the homology model was performed using WHATIF program. This result was used to evaluate the packing quality of the model compared to that of template. The overall quality of the structures is evaluated based on two *Z*-scores, namely structure-*Z*-scores, which are based on how well the model conforms to common refinement restraint values, and RMS-*Z*-scores, which are restraint-independent quality indicators. In terms of the structure-*Z*-scores, the model and template were very similar with each other. For instance, the packing quality of the model and template were −1.978 and −1.853, respectively. RMS-*Z*-scores of the model were better than that of the template (Table S2). As a third validation test, PROSA program was used to investigate the overall quality of the VLA-4 model. The *Z*-score computed by PROSA indicates the overall quality and its value for the respective model is related to the *Z*-scores of all the experimentally determined protein structures in the current database indicating whether the *Z*-score of a particular structure is within the range of reasonable scores (Figure S3). The α subunit of VLA-4 has scored a

Figure 5: Colored ribbon representation of (A) αV (green) and β3 (pink) subunits of the crystal structure of the template (B) α4 (orange) and β1 (blue) subunits of homology modeled very late antigen-4; eight divalent metal (Mn^{2+}) ions shown in spheres (C) overlay between the template and the homology model. Ligand cocrystallized with $\alpha V\beta$3 in the crystal structure (1L5G) shown in spheres at the active site.

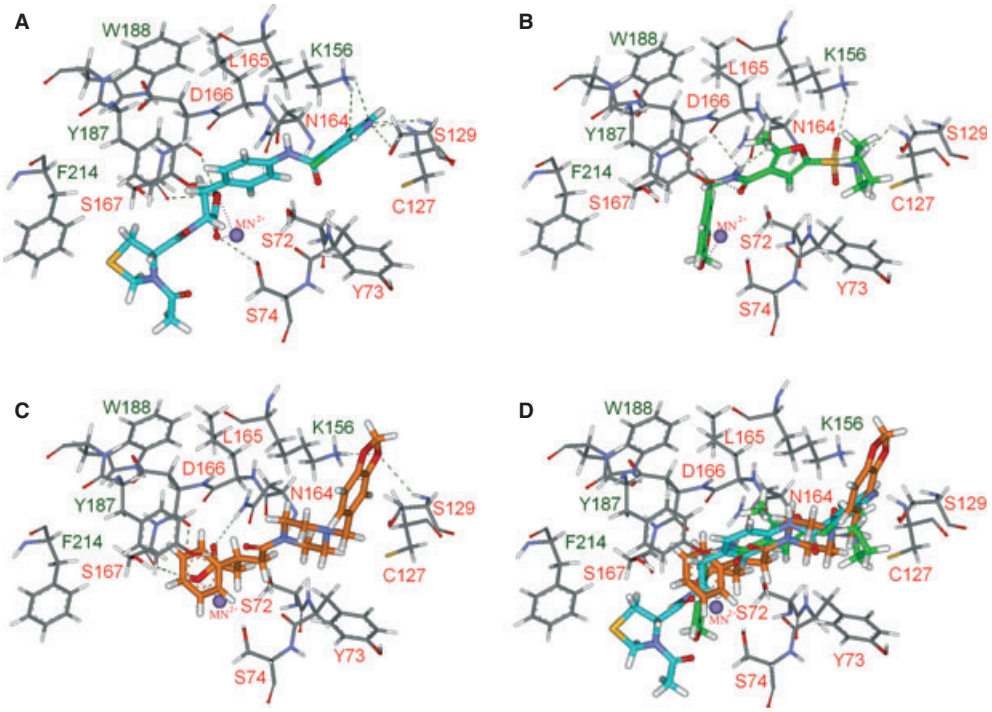

Figure 6: Docked conformations in the active site (A) compound **1**, most active compound in the training set; colored cyan (B) and (C) final hit compounds HTS 02885 and HTS 05216; colored green and orange, respectively, (D) overlay of all the compounds shown with no hydrogen bonds. Residues from α and β subunits labeled in green and red, respectively. Metal ion interactions are shown in dark pink dotted lines, while other hydrogen bond interactions are shown in green dashed lines.

Z-score value of −6.79, which is a better score for a 941 amino acid long protein. In the other end, 553–amino acid–long β subunit of VLA-4 model has scored a *Z*-score value of −3.34. These results from the validation analyses have concluded that the model represents an acceptable quality when compared to the X-ray structures available in PDB. Thus, the final structure was considered reliable and was used in further studies. Final homology model and its overlay with the template structure are shown in Figure 5. Binding site of VLA-4 ligands is located at the headpiece of the protein particularly at the junction of α and β subunits. The βA domain of the β subunit and β-propeller domain of the α subunit jointly make the head piece of the VLA-4 protein (4). All of the active site residues have been predicted to be in the most favored region by PROCHECK analysis. The final validated model has been selected and used in molecular docking studies.

Molecular docking

Molecular docking experiments were carried out using GOLD docking program. Training set of 20 compounds and 12 new hit compounds retrieved from Maybridge and Chembridge databases after drug-likeness screenings were chosen for molecular docking. The ligand-binding site was defined as a collection of amino acids enclosed within a sphere of 10 Å radius around the hydroxyl group of active site tyrosine (Y187) residue. GOLD score that ranks the molecules based on their binding was calculated for all the molecules. Early termination option was chosen that if five conformations of a particular compound are within the RMSD of 1.5 Å, the GOLD docking run would either be stopped or moved to the next compound. It has been reported in previous docking studies of VLA-4 that α Y187 (Y226), α F214 (F253), β S72 (S152), β S74 (S154), β C127 (C207), β N164 (N244), and β S167 (S247) are the important active site amino acids and involved in

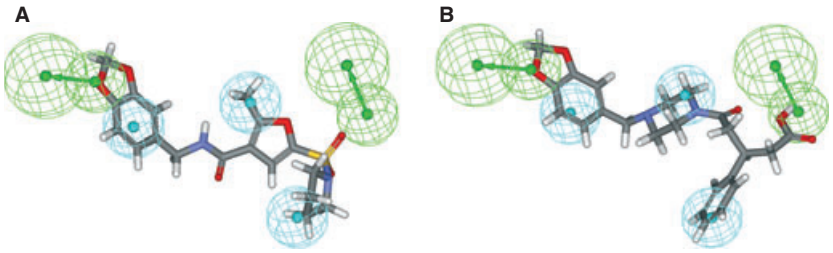

Figure 7: Overlay of final database hit compounds HTS 02885 (A) and HTS 05216 (B) on the best pharmacophore model, 'Hypo 1'.

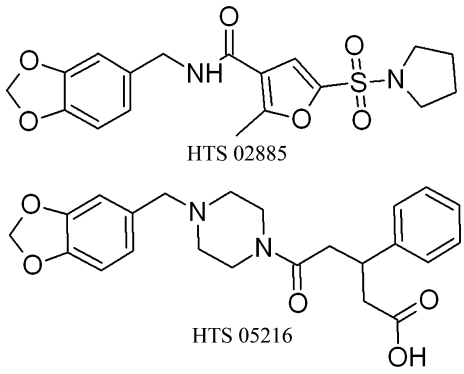

Figure 8: Two-dimensional representation of final hit compounds.

antagonist binding. Amino acids numbered in parentheses were according to Macchiarulo *et al.* (5). Compound **1** from the training set scored a GOLD fitness score of 54.439 and showed interactions with residues α K156, α S167, α Y187, β S74, β C127, β S129, β N164, and also with metal ion (Mn^{2+}) adjacent to the active site (MIDAS) (4,5,47). Figure 6A represents the binding conformation and the hydrogen bond interactions between compound **1** and active site amino acids. The GOLD fitness scores of all training set compounds are shown in Table S3. Only four of 12 hit compounds have scored better GOLD fitness score than the most active training set compound **1**. These compounds were considered for further evaluation. Compound HTS02885 that was found as a hit from Maybridge database scored a GOLD fitness score of 59.392 and also found to interact with most of the active site amino acids and MIDAS (Figure 6B). The another hit compound named HTS05216 also scored better GOLD score (57.337) and showed good interactions with active site amino acids and MIDAS (Figure 6C). Figures 7 and 8 show the pharmacophore mapping and 2D representation of two final compounds, respectively. These two of four top-scored compounds showed good fit on the pharmacophore model, higher GOLD scores compared to the most active compound in the training set, hydrogen bond interactions with the active site amino acids and drug-like properties. Finally, Scifinder Scholar[h] and Pubchem structure[i] searches proved that these hit compounds have not been reported earlier for VLA-4 antagonism. Thus, these hit compounds can be treated as good virtual leads in designing of novel potent VLA-4 antagonists.

Conclusion

The development of potent VLA-4 antagonists is a challenging area for the treatment of many diseases including inflammation. As an attempt to understand the molecular interaction between the available antagonists and human VLA-4, a pharmacophore model, 'Hypo 1', was developed. 'Hypo 1' that consists of five chemical features (two *HBA* and three *HYP* features) was validated with a test set containing 90 compounds, Fischer validation method and a database of active and inactive compounds. The selected 'Hypo 1' model predicted the test set with a high correlation value of 0.907 and also the activity values of E and Z isomers accurately. The validated pharmacophore model was used as a 3D query to search two databases (Maybridge and Chembridge) containing totally 110 000 compounds. Database search retrieved 5577 compounds that fit the model 'Hypo 1' well. Fit value, drug-likeness, and ADMET properties screening helped us to reject the non-drug-like compounds. Finally, 12 drug-like compounds were identified through all the screenings and were subjected to molecular docking study. A 3D homology model of the extracellular domain of human VLA-4 was modeled using the crystal structure of $\alpha V\beta 3$ as template and was validated to use it further in molecular docking study. Out of four compounds that showed GOLD fitness score higher than the most active training set compounds, two were found to have good interactions with the active site residues of human VLA-4 model. Combining the results of pharmacophore, drug-likeness, ADMET, molecular docking studies, and the novelty search, we have found two compounds and listed them as possible virtual leads to design novel human VLA-4 antagonists.

Acknowledgments

Students were recipients of fellowships from the BK21 Programs. This work was also supported by grants from the MEST/NRF to the Environmental Biotechnology National Core Research Center (20090091489) and Pioneer Research Center Program (2009-0081539).

References

1. Hynes R.O. (1987) Integrins: a family of cell surface receptors. Cell;48:549–554.
2. Yang G.X., Hagmann W.K. (2003) VLA-4 antagonists: potent inhibitors of lymphocyte migration. Med Res Rev;23:369–392.
3. Woodside D.G., Vanderslice P. (2008) Cell adhesion antagonists: therapeutic potential in asthma and chronic obstructive pulmonary disease. mechanisms and targets. BioDrugs;22:85–100.
4. Xiong J.P., Stehle T., Diefenbach B., Zhang R., Dunker R., Scott D.L., Joachimiak A., Goodman S.L., Arnaout M.A. (2001) Crystal

structure of the extracellular segment of integrin $\alpha V\beta 3$. Science;294:339–345.
5. Macchiarulo A., Costantino G., Meniconi M., Pleban K., Ecker G., Bellocchi D., Pellicciari R. (2004) Insights into phenylalanine derivatives recognition of VLA-4 integrin: from a pharmacophoric study to 3D-QSAR and molecular docking analyses. J Chem Inf Comput Sci;44:1829–1839.
6. Lin L.S., Lanza T.Jr., Jewell J.P., Liu P., Jones C., Kieczykowski G.R. *et al.* (2009) Discovery of *N*-{*N*-[(3-Cyanophenyl)sulfonyl]-4(R)-cyclobutylamino-(L)-prolyl}-4-[(3′,5′-dichloroisonicotinoyl)amino]-(L)-phenylalanine (MK-0668), an extremely potent and orally active antagonist of very late antigen-4. J Med Chem;52:3449–3452.
7. Masumoto A., Hemler M.E. (1993) Multiple activation states of VLA-4. Mechanistic differences between adhesion to CS1/fibronectin and to vascular cell adhesion molecule-1. J Biol Chem;268:228–234.
8. Abraham W.M., Gill A., Ahmed A., Sielczak M.W., Lauredo I.T., Botinnikova Y., Lin K., Pepinsky B., Leone D.R., Lobb R.R., Adams S.P. (2000) A small-molecule, tight-binding inhibitor of the integrin $\alpha 4\beta 1$ blocks antigen-induced airway responses and inflammation in experimental asthma in sheep. Am J Respir Crit Care Med;162:603–611.
9. Lin K. *et al.* (1999) Selective, tight-binding inhibitors of integrin $\alpha 4\beta 1$ that inhibit allergic airway responses. J Med Chem;42: 920–934.
10. Yednock T.A., Cannon C., Fritz L.C., Madrid F., Steinman L., Karin N. (1992) Prevention of experimental autoimmune encephalomyelitis by antibodies against $\alpha 4\beta 1$ integrin. Nature;356: 63–66.
11. Piraino P.S., Yednock Y.A., Freedman S.B., Messersmith E.K., Pleiss M.A., Vandevert C., Thorsett E.D., Karlik S.J. (2002) Prolonged reversal of chronic experimental allergic encephalomyelitis using a small molecule inhibitor of $\alpha 4$ integrin. J Neuroimmunol;131:147–159.
12. Laffon A., Garcia-Vicuna R., Humbria A., Postigo A.A., Corbi A.L., De Landazuri M.O., Sanchez-Madrid F. (1991) Upregulated expression and function of VLA-4 fibronectin receptors on human activated T cells in rheumatoid arthritis. J Clin Invest;88:546–552.
13. Powrie F., Leach M.W. (1995) Genetic and spontaneous models of inflammatory bowel disease in rodents: evidence for abnormalities in mucosal immune regulation. Ther Immunol;2:115–123.
14. Wenning W., Haghikia A., Laubenberger J., Clifford D.B., Behrens P.F., Chan A., Gold R. (2009) Treatment of progressive multifocal leukoencephalopathy associated with natalizumab. N Engl J Med;361:1075–1080.
15. Hijazi Y., Welker H., Dorr A.E., Tang J., Blain R., Renzetti L.M., Abbas R. (2004) Pharmacokinetics, safety, and tolerability of R411, a dual $\{\alpha\}4\beta 1$-$\{\alpha\}4\beta 7$ integrin antagonist after oral administration at single and multiple once-daily ascending doses in healthy volunteers. J Clin Pharmacol;44:1368–1378.
16. Tilley J.W. (2008) Very late antigen-4 integrin antagonists. Ther Expert Opin Pat;18:841–859.
17. Ozdemir C., Akdis D.S. (2007) Discontinued drugs in 2006: pulmonary-allergy, dermatological, gastrointestinal and arthritis drugs. Expert Opin Invest Drugs;16:1327–1344.
18. Romagnoli E., Francesco B., Trani C., Biondi-Zoccai G.G.L., Giannico F., Crea F. (2007) Rationale for intracoronary administration of abciximab. J Thromb Thrombolysis;23:57–63.
19. Yassky E.G., Vugmeyster Y., Lowes M.A., Chamian F., Kikuchi T., Kagen M., Gilleaudeau P., Lee E., Hunte B., Howell K., Dummer W., Bodary C., Krueger J.G. (2008) Blockade of CD11a by efalizumab in psoriasis patients induces a unique state of T-cell hyporesponsiveness. J Invest Dermatol;128:1182–1191.
20. Lin K., Castro A. (1998) Very late antigen 4 (VLA-4) antagonists as anti-inflammatory agents. Curr Opin Chem Biol;2:453–457.
21. Makagiansar H.Y., Anderson M.E., Yakovleva T.V., Murray J.S., Siahaan T.J. (2002) Inhibition of LFA-1/ICAM-1and VLA-4/VCAM-1 as a therapeutic approach to inflammation and autoimmune diseases. Med Res Rev;22:146–167.
22. Porter J.R., Archibald S.C., Childs S.C., Critchley D., Head J.C., Linsley J.M., Parton T.A.H., Robinson M.K., Shock A., Taylor R.J., Warrellow G.J., Alexander R.P., Langham B. (2002) Squaric acid derivatives as VLA-4 integrin antagonists. Bioorg Med Chem Lett;12:1051–1054.
23. Porter J.R., Archibald S.C., Brown J.A., Childs K., Critchley D., Head J.C., Hutchinson B., Parton T.A.H., Robinson M.K., Shock A., Warrellow G.J., Zomaya A. (2002) Discovery and evaluation of *N*-(triazin-1,3,5-yl) phenylalanine derivatives as VLA-4 integrin antagonists. Bioorg Med Chem Lett;12:1591–1594.
24. Porter J.R., Archibald S.C., Brown J.A., Childs K., Critchley D., Head J.C., Hutchinson B., Parton T.A.H., Robinson M.K., Shock A., Warrellow G.J., Zomaya A. (2002) Discovery and evaluation of *N*-(triazin-1,3,5-yl) phenylalanine derivatives as VLA-4 integrin antagonists. Bioorg Med Chem Lett;12:1595–1598.
25. Porter J.R., Archibald S.C., Brown J.A., Childs K., Critchley D., Head J.C., Parton T.A.H., Robinson M.K., Shock A., Taylor R.J., Warrellow G.J. (2003) Dehydrophenylalanine derivatives as VLA-4 integrin antagonists. Bioorg Med Chem Lett;13:805–808.
26. Lassoie M. *et al.* (2007) 2,6-Quinolinyl derivatives as potent VLA-4 antagonists. Bioorg Med Chem Lett;17:142–146.
27. Krovat E.M., Langer T. (2003) Non-peptide angiotensin II receptor antagonists: chemical feature based pharmacophore identification. J Med Chem;46:716–726.
28. Chen Y., Jiang Y., Zhou J., Yu Q., You Q. (2008) Identification of ligand features essential for HDACs inhibitors by pharmacophore modeling. J Mol Graph Model;26:1160–1168.
29. Thangapandian S., John S., Sakkiah S., Lee K.W. (2010) Ligand and structure based pharmacophore modeling to facilitate novel histone deacetylase 8 inhibitor design. Eur J Med Chem;45:4409–4417.
30. Kurogi Y., Guner O.F. (2001) Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. Curr Med Chem;8:1035–1055.
31. Santo R.D., Fermeglia M., Ferrone M., Paneni M.S., Costi R., Artico M., Roux A., Gabriele M., Tardif K.D., Siddiqui A., Pricl S. (2005) Simple but highly effective three-dimensional chemical-feature-based pharmacophore model for diketo acid derivatives as hepatitis C virus RNA-dependent RNA polymerase inhibitors. J Med Chem;48:6304–6314.
32. Miguel J.M.J., Jordi B.T., Maria O.O.H., Eloisa N.R. (2004) Urea derivatives as integrin alpha 4 antagonists. Patent US 2004/0142982 A1.

33. Miguel J.M.J., Laura V.G., Graham W. (2007) *N*-(2-phenylethyl)sulfonamide derivatives as integrin alpha4 antagonists. Patent US 2007/0179183 A1.
34. Xiong J.P., Stehle T., Zhang R., Joachimiak A., Frech M., Goodman S.L., Arnaout M.A. (2002) Crystal structure of the extracellular segment of integrin $\alpha V\beta 3$ in complex with an Arg-Gly-Asp Ligand. Science;296:151–155.
35. Laskowski R.A., MacArthur M.W., Moss D., Thornton J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Cryst;26:283–291.
36. Vriend G. (1990) WHAT IF: a molecular modelling and drug design program. J Mol Graph;8:52–56.
37. Wiederstein M., Sippl M.J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Res;35:W407–W410.
38. Sippl M.J. (1993) Recognition of errors in three-dimensional structures of proteins. Proteins;17:355–362.
39. Gareth J., Peter W., Robert C.G., Andrew R.L., Robin T. (1997) Development and validation of a genetic algorithm for flexible docking. J Mol Biol;267:727–748.
40. Thomsen R., Christensen M.H. (2006) MolDock: a new technique for high-accuracy molecular docking. J Med Chem;49:3315–3321.
41. Sutter J., Guner O.F., Hoffmann R., Li H., Waldman M. (2000) Pharmacophore Perception, Development, and Use in Drug Design. CA, USA: International University Line; p. 501–511.
42. Lu I., Tsai K., Chiang Y., Jiaang W., Wu S., Mahindroo N., Chien C., Lee S., Chen X., Chao Y., Wu S. (2008) A three-dimensional pharmacophore model for dipeptidyl peptidase IV inhibitors. Eur J Med Chem;43:1603–1611.
43. Kahnberg P., Howard M.H., Liljefors T., Nielsen M., Nielsen E., Sterner O., Pettersson I. (2004) The use of a pharmacophore model for identification of novel ligands for the benzodiazepine binding site of the GABAA receptor. J Mol Graph Model;23:253–261.
44. Duchowicz P.R., Talevi A., Bellera C., Bruno-Blanch L.E., Castro E.A. (2007) Application of descriptors based on Lipinski's rules in the QSPR study of aqueous solubilities. Bioorg Med Chem;15:3711–3719.
45. Venkatapathy R., Moudgal C.J., Bruce R.M. (2004) Assessment of the oral rat chronic lowest observed adverse effect level model in TOPKAT, a QSAR software package for toxicity prediction. J Chem Inf Comput Sci;44:1623–1629.
46. You T.J., Maxwell D.S., Kogan T.P., Chen Q., Li J., Kassir J., Holland G.W., Dixon R.A.F. (2002) A 3D structure model of integrin $\alpha 4\beta 1$ complex: I. Construction of a homology model of $\beta 1$ and ligand binding analysis. Biophys J;82:447–457.
47. Xiong J., Stehle T., Diefenbach B., Zhang R., Joachimiak A., Frech M., Goodman S.L., Arnaout M.A. (2002) Crystal structure of the extracellular segment of integrin $\alpha 4\beta 3$ in complex with an Arg-Gly-Asp Ligand. Science;296:339–345.

Notes

[a]Discovery Studio, version 2.1, San Diego, CA, USA: Accelrys Inc.
[b]ACD/ChemSketch Freeware, version 12.00, Toronto, ON, Canada: Advanced Chemistry Development, Inc., available at: http://www.acdlabs.com.
[c]QSAR World, available at: http://www.qsarworld.com.
[d]Binding Database, available at: http://www.bindingdb.org.
[e]Maybridge, Cornwall, UK: Maybrdige Chemical Co., available at: http://www.maybridge.com.
[f]Chembridge, San Diego, CA, USA: ChemBridge Corp, available at: http://www.chembridge.com.
[g]GOLD user guide, available at: http://www.ccdc.cam.ac.uk/support/documentation/gold/3_2/doc/portable_html/gold_portable-3-077.html.
[h]Scifinder Scholar, available at: http://www.cas.org/products/scifindr/index.html.
[i]PubChem structure search, available at: http://pubchem.ncbi.nlm.nih.gov/.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Sequence alignment between α subunits of template (αV) and target (α4) sequences.

Figure S2. Sequence alignment between the β subunits of template ($\beta 3$) and target ($\beta 1$) sequences.

Figure S3. PROSA plot displaying the *Z*-scores for the α and β subunits of very late antigen-4 model in black and red color spots, respectively.

Table S1. Calculated absorption, distribution, metabolism, excretion and toxicity properties of 206 identified compounds after Lipinski prediction.

Table S2. WHATIF results for the constructed model and template used in homology modeling.

Table S3. Comparison of experimental activity and genetic optimization for ligand docking score values of training set compounds.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.