Available online at www.sciencedirect.com

ScienceDirect

Journal of Molecular Graphics and Modelling 25 (2007) 813–823

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Pharmacophore identification and virtual screening for methionyl-tRNA synthetase inhibitors

Nagakumar Bharatham [1], Kavitha Bharatham [1], Keun Woo Lee *

Department of Biochemistry, Division of Applied Life Science, Environmental Biotechnology National Core Research Center, Gyeongsang National University, Jinju 660-701, Republic of Korea

Received 19 May 2006; received in revised form 3 August 2006; accepted 3 August 2006
Available online 14 August 2006

Abstract

Aminoacyl-tRNA synthetases (aaRSs) are essential enzymes involved in protein biosynthesis in all living organisms and are an unexploited antibacterial targets, as many strains of bacteria have become resistant to all established classes of antibiotics. Therefore, the main aim of this study is to discover new lead molecules which would be useful as anti-bacterial compounds. Pharmacophore models were developed by using *CATALYST HypoGen* with a training set of 29 diverse methionyl-tRNA synthetase (MetRS) inhibitors. The best quantitative pharmacophore hypothesis (Hypo1) obtained a correlation coefficient of 0.975, root mean square deviation (RMSD) of 0.55 and cost difference (null cost-total cost) of 70.32. This Hypo1 was validated by two methods, first by using 104 test set molecules which resulted a correlation of 0.926 between *HypoGen* estimated activities versus experimental activities and secondly by *Cat-Scramble* validation method. This validated pharmacophore model was further used for screening databases for discovery of new MetRS inhibitors. The new lead compounds were further analyzed for drug-like properties. Homology modeled structure of *Staphylococcus aureus* MetRS was built and molecular docking studies were performed with many inhibitors using the newly built protein structure. Finally, it was found that the new leads exhibited good estimated inhibitory activity, calculated binding properties similar to experimentally proven compounds and also favorable drug-like properties.

Keywords: MetRS; Pharmacophore model development; Cat-scramble; Database screening; *GOLD* molecular docking; Drug-like properties

1. Introduction

Translation is one of the most complex biological processes, involving diverse protein factors and enzymes as well as messenger and transfer RNAs. As this process is required for the basic operation of cells, many translational factors and enzymes are considered to be housekeeping proteins [1]. Aminoacyl-tRNA synthetases (aaRSs) catalyze the ligation of specific amino acids to their cognate tRNAs, which is the initial step in protein synthesis [2]. The aminoacylation reaction proceeds in two stages. First, aaRSs activate their substrate amino acids by forming aminoacyl adenylate. Second, the enzyme-bound reaction intermediates are transferred to the 3′ acceptor end of the tRNAs docking onto their active sites [3]. As tRNAs cannot distinguish amino acids, the correct recognition of specific amino acids and tRNAs by aaRS enzymes is a crucial determinant to maintain the fidelity of protein synthesis [4]. Therefore, aaRSs are the essential enzymes for biological cell growth. The twenty aaRSs are divided into two classes, I and II. The 10 class I aaRSs are considered to have a common catalytic domain structure based on the Rossmann fold, which is totally different from the class II catalytic domain structure [5–7]. The class I synthetases are further divided into three subclasses, a–c, according to sequence homology [8].

Staphylococcus aureus is known to cause many forms of infections like superficial skin lesions, deep-seated infections such as osteomyelitis and endocarditis, hospital acquired (nosocomial) infection of surgical wounds, food poisoning by releasing enterotoxins into food, toxic shock syndrome by release of superantigens into the blood stream, etc,. Inhibition of methionyl-tRNA synthetase (MetRS) can be a crucial step in the control of bacterial growth and can act as a novel bacterial target, thus not being susceptible to established mechanisms of

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

1093-3263/$ – see front matter
doi:10.1016/j.jmgm.2006.08.002

bacterial resistance. Previously a 3D-QSAR design was performed using *S. aureus* MetRS inhibitors while molecular docking studies were done with *E. coli* MetRS structure [9]. Unlike QSAR study, which is based on single scaffold containing molecules, pharmacophore model design is based on diverse existing inhibitors. Therefore, pharmacophore models were developed using the *HypoGen* module implemented in *CATALYST4.10* software [10] and molecular docking studies were done using homology modeled *S. aureus* MetRS structure. *CATALYST* is one of the leading automated drug design software [11] since a large number of successful applications were clearly demonstrated in medicinal chemistry [12–14].

The main aim of this study is to construct a pharmacophore model based on key chemical features of compounds with MetRS inhibitory activity covering five orders of magnitude. The obtained pharmacophore model can provide a rational hypothetical picture of the primary chemical features responsible for activity, and is expected to provide useful knowledge for developing new potentially active candidates targeting the MetRS which are useful as anti-bacterial agents. Hence, the best pharmacophore model was selected carefully and was validated. New compounds with similar features were obtained from Maybridge database and they were screened based on their estimated activity, GOLD docking scores using *S. aureus* MetRS homology model and calculated drug-like properties.

2. Materials and methods

2.1. Data collection and development of database

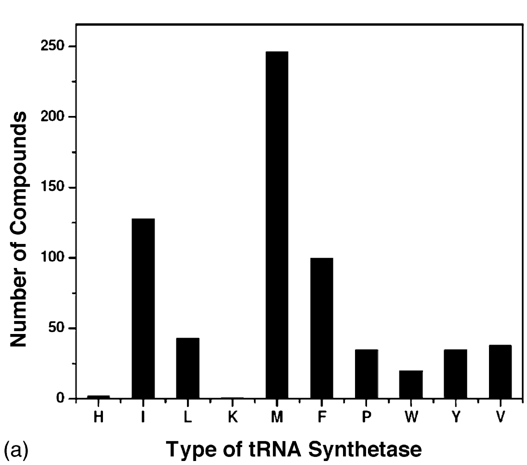
We have collected aaRSs inhibitors with their biological activity data from various medicinal chemistry as well as life science journals and developed a unique database using *MDL ISIS/Base* [15]. Our aaRS inhibitor database comprises of 595 compounds with experimental activities, out of which 246 compounds were against MetRS target (Fig. 1a). The highest number of compounds with MetRS inhibitory activity are in the molecular weight range 400–500 (Fig. 1b).

2.2. Training set selection and conformational generation

The training set comprising of 29 compounds (Nos. 1–29) was used to generate *HypoGen* hypotheses by considering structural diversity and wide coverage of activity range [16–20]. Conformationally restricted analogues of the central linker unit of *S. aureus* MetRS inhibitors [18] have been included in the training set. All the training set compounds from the literature have similar MetRS inhibitory assay. Before starting the pharmacophore generation process, conformational models for the molecules was developed by *poling algorithm*, which seeks to provide a broad coverage of conformational space using the best conformer generation method with a maximum conformational energy of 20 kcal/mol above the lowest energy conformation found [21,22]. The number of conformers generated for each compound was limited to a maximum number of 250.

2.3. Generation of pharmacophore models

Pharmacophore models were developed using *HypoGen* module implemented in *CATALYST* and the top 10 scoring hypotheses were exported. Analyses of the best ranking pharmacophore model revealed that four chemical features that are one H-bond donor (HBD), one hydrophobic aliphatic (HY-ALI) and two ring aromatic (RA) features could effectively map all the chemical features. The activity of each training set compound is estimated 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 greater the activity prediction of the compound. The fit function does not only check if the feature is mapped or not; 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 Error value shows the ratio of estimated activity to experimental activity. A positive error value indicates that the estimated IC_{50} is higher than the experimental IC_{50} while a negative error value indicates that the estimated IC_{50} is lower than the experimental IC_{50}.



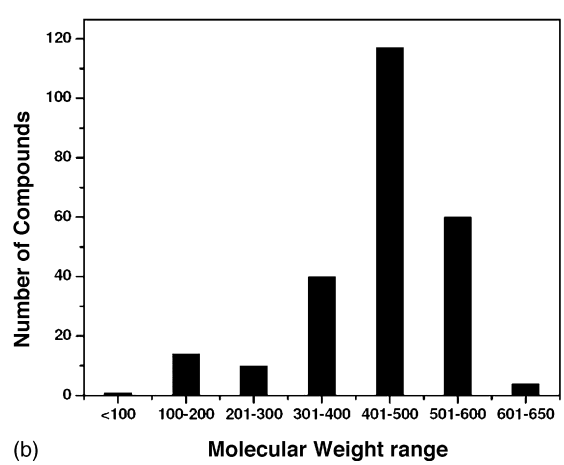


Fig. 1. Numbers of compounds for each type of tRNA synthetases present in our in-house database (a), molecular weight analyses for 246 MetRS inhibitors (b).

2.4. Pharmacophore model validation

The main purpose of validating a quantitative model is to determine whether our model is able to identify active structures and forecast their activity accurately. Therefore, two validation procedures were followed namely, test set prediction method and *Cat-Scramble* method. The 104 compounds were used as test set to validate the pharmacophore model. The *Cat-Scramble* validation procedure is based on Fischer's randomization test. The goal of this type of validation is to check whether there is a strong correlation between the chemical structures and the biological activity. This is done by randomizing the activity data associated with the training set compounds, generating pharmacophore hypotheses using the same features and parameters to develop the original pharmacophore hypothesis. The statistical significance is calculated as following formula:

$$\text{significance} = 100\left(1 - \frac{1+x}{y}\right)$$

where x is the total number of hypotheses having a total cost lower than HypoX (original hypothesis), and y is the total number of *HypoGen* runs (initial + random runs).

Thus 19 random spreadsheets (or 19 *HypoGen* runs) have to be generated for 95% confidence level. If the randomized data set results in the generation of a pharmacophore with similar or better cost values, RMSD, and correlation, then the original hypothesis is considered to have been generated by chance.

2.5. Database searching

CATALYST generated best pharmacophore model comprising of four chemical features was used as a query for searching Maybridge chemical database consisting of 60,000 structurally diversified small molecules. Virtual screening of such databases can serve two main purposes: first, validating the quality of the generated pharmacophore models by selective detection of compounds with known inhibitory activity, and second, finding novel, potential leads suitable for further development. *Best flexible search method* was used for database searching to retrieve new lead molecules.

2.6. Homology modeling

BLAST (blastp) was employed to search the relevant target or template proteins for building *S. aureus* MetRS protein structure. ClustalW multiple sequence alignment method was applied to compare the *S. aureus* MetRS sequence with other bacterial MetRS. The *MODELLER* module in *INSIGHTII* software was used to develop the homology model [23–25]. Sequence alignments were achieved by Align2d method and the final 3D model was validated by *PROCHECK* software [26].

2.7. Molecular docking

The program *GOLD 3* (Genetic Optimisation for Ligand Docking) from Cambridge Crystallographic Data Center, UK [27] uses 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. A pseudoatom was created at binding site region of modeled *S. aureus* MetRS whose coordinates were taken to define active site region with a active site radius of 8.0 Å. The annealing parameters of van der Waals and H-bond interactions were considered within 4.0 and 2.5 Å, respectively.

3. Results and discussion

3.1. Pharmacophore generation and validation

A set of 10 hypotheses was generated using a training set of 29 compounds (Fig. 2). The *HypoGen* module in *CATALYST* performs a 'fixed cost' calculation which represents the simple model that fits all data perfectly, and a 'null cost' calculation which presumes that there is no relationship in the dataset and that the experimental activities are normally distributed around their average value. A meaningful pharmacophore hypothesis may result when the difference between null and fixed cost value is large. In this study the null cost value of the top 10 hypotheses is 189.084, and the fixed cost value is 114.072 with a difference of 75.012 bits. The total cost of any pharmacophore hypothesis should be close to the fixed cost. All 10 hypotheses have a total cost close to the cost of the fixed hypothesis. The cost values, correlation coefficients (r), root mean square deviations (RMSDs), and pharmacophore features are listed in Table 1. Configuration cost value must be less than 17 for a good pharmacophore and accordingly 15.402 bits was obtained.

All the 10 hypotheses contain four features each and three of them, H-bond donor (HBD), hydrophobic aliphatic (HY-ALI) and ring aromatic (RA) features were common. Six of the 10 hypotheses had RA, three hypotheses had HY-ALI while only one had HBD as the fourth feature. Hypo1 consists of one HBD, HY-ALI and two RA features. The RMSD indicates the quality of prediction for the training set. In this case the RMSD value of the best hypothesis Hypo1 is 0.55, which represents a good prediction by Hypo1. The correlation coefficient for the Hypo1, 0.975 shows a good correlation by linear regression of the geometric fit index. All these results conclude that Hypo1 is the best ranking pharmacophore among the 10 obtained.

All compounds both in the training set and in the test set were classified into three activity scales: highly active (+++, $IC_{50} \leq$ 100 nM); moderately active (++, 1000 nM $> IC_{50} >$ 100 nM), and inactive compounds (+, $IC_{50} \geq$ 1000 nM). Activities were estimated for all compounds based on the best ranking pharmacophore (Hypo1). The experimental and estimated activities for 29 training set compounds are shown in Table 2. Out of 29 compounds one high active compound (+++) was estimated as moderately active (++), two moderately active (++) compounds were estimated as inactive (+) and one moderately active compound was estimated as highly active (+++). All inactive molecules were estimated as inactive only. Two of the highest active compounds (1 and 2) in the training set with an estimated activities of 3.5 nM (IC_{50} = 3.3 nM) and 5.5 nM

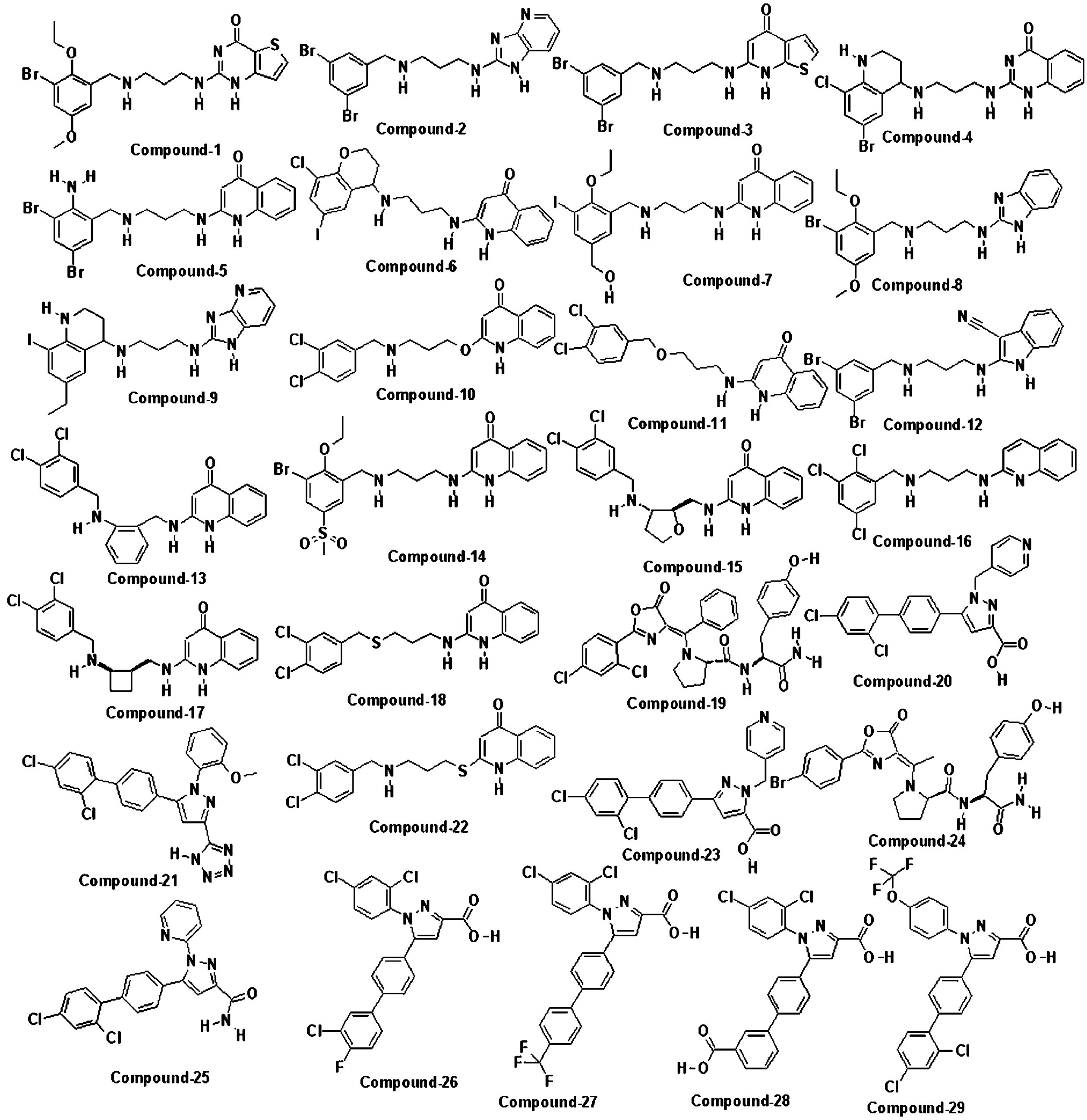


Fig. 2. Molecular structures of training set compounds: compounds 1–4, 8–10, 12, 16, 22 taken from Ref. [16], compounds 5–7, 11, 14, 18 from Ref. [17], compounds 13, 15, 17 from Ref. [18], compounds 20, 21, 23, 25–29 taken from Ref. [19], compounds 19 and 24 from Ref. [20]. Smiles format of these compounds are provided as supplementary information.

(IC_{50} = 5 nM) and fit values of 8.21 and 8.01, respectively, were mapped with Hypo1 (Fig. 3a and b).

In order to verify if the hypothesis can also estimate the activity of compounds that are structurally distinct from those included in the training set, we have applied a test set of 104 compounds, selected from literature representing diverse activity classes and different structural groups. All test molecules were imported into *CATALYST* from internal MetRS database and conformers were generated in a similar way as for the training set compounds. Hypo1 was used to estimate the activities for test set compounds using *score hypothesis* method. Analyses of the estimated activities of test set revealed interesting results, out of 104 compounds, 101 compounds had the error value of less than 10, representing a not more than one order difference between experimental and estimated activity. All 61 active compounds in the test set were estimated as highly active. The 60% of moderately active compounds were estimated appropriately. Among 29 inactive compounds, 4 compounds were estimated as moderately active. The predictive ability of Hypo1 was very impressive and the correlation graph between experimental and estimated activities (Fig. 4) achieved 0.926 correlation value. As very few compounds within the range of 30–130 nM were obtained from the literature, we used them in the training set. Thus equal

Table 1
Information of statistical significance and predictive power presented in cost values measured in bits for top 10 hypotheses[a]

Hypothesis	Total cost	Cost difference (null − total cost)	RMSD	Correlation	Features	Correlation value for 104 test set compounds
1	118.757	70.327	0.55	0.975	HBD, HY-ALI, RA, RA	0.926
2	119.569	69.515	0.587	0.972	HBD, HY-ALI, RA, RA	0.924
3	123.245	65.839	0.787	0.949	HBD, HY-ALI, RA, RA	0.890
4	124.103	64.981	0.827	0.944	HBD, HY-ALI, HY-ALI, RA	0.850
5	124.103	64.981	0.831	0.943	HBD, HY-ALI, RA, RA	0.907
6	124.301	64.783	0.836	0.943	HBD, HY-ALI, RA, RA	0.881
7	124.615	64.469	0.813	0.946	HBD, HY-ALI, HY-ALI, RA	0.839
8	124.984	64.1	0.847	0.941	HBD, HY-ALI, HY-ALI, RA	0.830
9	125.005	64.079	0.842	0.942	HBD, HBD, HY-ALI, RA	0.757
10	125.099	63.985	0.863	0.939	HBD, HY-ALI, RA, RA	0.889

[a] Null cost of top 10 score hypotheses is 189.084 bits; fixed cost is 114.072 bits; configuration cost is 15.402 bits. Abbreviation used for features: HBD, H-bond donor; HY-ALI, hydrophobic aliphatic; RA, ring aromatic.

distribution of test set compounds with activity, particularly in the mentioned range could not be obtained. Two highly active test set compounds, Test 8 and 46 ($IC_{50} = 7$ and 18 nM, respectively), were mapped with Hypo1 (Fig. 3c and d). Hypo1 features matched very well with the chemical groups of these two compounds with accurate estimated IC_{50} values of 6.2 and 18 nM, respectively.

Another approach to validate the quality of *HypoGen* hypothesis was to apply cross validation using the *Cat-Scramble* program available from *CATALYST*. In this validation test, we have selected 95% confidence level, and thus 19 spreadsheets (Table 3) were generated. The data of cross validation clearly indicates that all values generated after randomization, produced hypotheses with no predictive value

Table 2
Experimental biological activity data and estimated IC_{50} values of training set molecules based on pharmacophore model Hypo1

Compound	Experimental activity (nM)	Estimated activity	Error[a]	Fit value[b]	Activity scale[c]	Est. activity scale
1	3.3	3.5	1.1	8.21	+++	+++
2	5	5.5	1.1	8.01	+++	+++
3	6.2	15	2.4	7.58	+++	+++
4	8.2	7.6	−1.1	7.87	+++	+++
5	9.1	33	3.6	7.24	+++	+++
6	12	8.6	−1.4	7.82	+++	+++
7	14	12	−1.1	7.66	+++	+++
8	17	20	1.2	7.44	+++	+++
9	18	24	1.4	7.37	+++	+++
10	38	16	−2.4	7.56	+++	+++
11	50	30	−1.6	7.27	+++	+++
12	54	42	−1.3	7.14	+++	+++
13	56	62	1.1	6.96	+++	+++
14	72	89	1.2	6.81	+++	+++
15	78	89	1.1	6.81	+++	+++
16	100	310	3.1	6.26	+++	++
17	110	47	−2.4	7.09	++	+++
18	170	140	−1.2	6.61	++	++
19	430	1,400	3.3	5.61	++	+
20	500	600	1.2	5.98	++	++
21	530	360	−1.5	6.20	++	++
22	680	240	−2.8	6.37	++	++
23	910	1,100	1.2	5.73	++	+
24	1,700	2,200	1.3	5.42	+	+
25	2,900	3,000	1	5.28	+	+
26	17,000	32,000	1.8	4.26	+	+
27	34,000	20,000	−1.7	4.46	+	+
28	41,000	20,000	−2.1	4.46	+	+
29	47,000	20,000	−2.4	4.46	+	+

[a] + Indicates that the estimated IC_{50} is higher than the experimental IC_{50}; − indicates that the estimated IC_{50} is lower than the experimental IC_{50}.
[b] Fit value indicates how well the features in the pharmacophore overlap the chemical features in the molecule.
[c] MetRS activity scale: +++, $IC_{50} \leq 100$ nM (high active); ++, 1000 nM > IC_{50} > 100 nM (moderately active); +, ≥1000 nM (inactive).

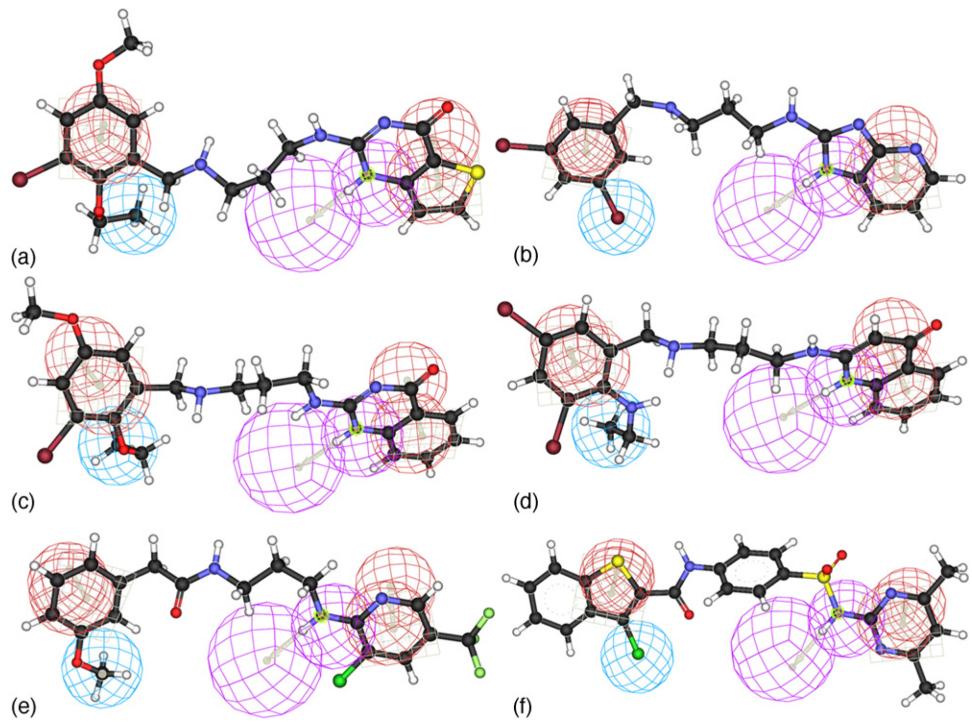


Fig. 3. Mapping of top scoring Hypo1 on high active training set compound 1 (a), training set compound 2 (b), Test 8 (c), Test 46 (d), Maybridge compound AW01179 (e) and Maybridge compound BTB00521 (f). In hypothesis orange color represents ring aromatic (RA), magenta represents H-bond donor (HBD) and light blue represents hydrophobic aliphatic (HY-ALI).

similar or near to that of Hypo1. Out of the 19 runs, only four had a correlation between 0.70 and 0.80, but the RMSD were high and total costs were close to the null cost, which is not desirable for a good hypothesis. Thus these two validation procedures provided strong confidence on the initial pharmacophore hypothesis, Hypo1.

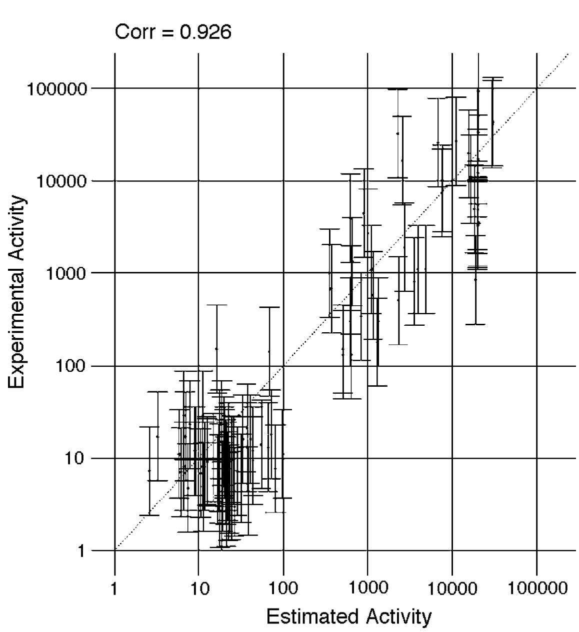


Fig. 4. Correlation graph between experimental and Hypo1 estimated activities.

3.2. Search for new lead compounds and their screening

The validated pharmacophore model, Hypo1, was used as a search query to retrieve molecules with novel and desired chemical features from multi-conformational Maybridge

Table 3
Results from cross-validation using *Cat-Scramble* in *CATALYST*[a]

Trial no.	Total cost	Fixed cost	RMSD	Correlation (r)
Results for unscrambled				
	118.757	114.072	0.55	0.975
Results for scrambled				
1	169.967	111.084	1.962	0.635
2	160.643	114.606	1.735	0.727
3	155.348	113.433	1.684	0.743
4	157.174	112.151	1.611	0.788
5	172.594	110.524	2.061	0.573
6	180.066	110.18	2.117	0.587
7	168.208	112.57	1.905	0.660
8	174.533	111.872	2.061	0.575
9	157.866	111.586	1.756	0.717
10	170.297	111.756	1.995	0.609
11	178.043	111.517	2.141	0.522
12	164.116	111.204	1.909	0.65
13	181.69	110.472	2.213	0.473
14	165.514	111.027	1.919	0.647
15	172.402	113.654	2.008	0.601
16	174.081	112.563	2.058	0.573
17	161.695	111.731	1.849	0.677
18	184.774	112.412	2.229	0.462
19	166.052	112.816	1.894	0.659

[a] Null cost is 189.084.

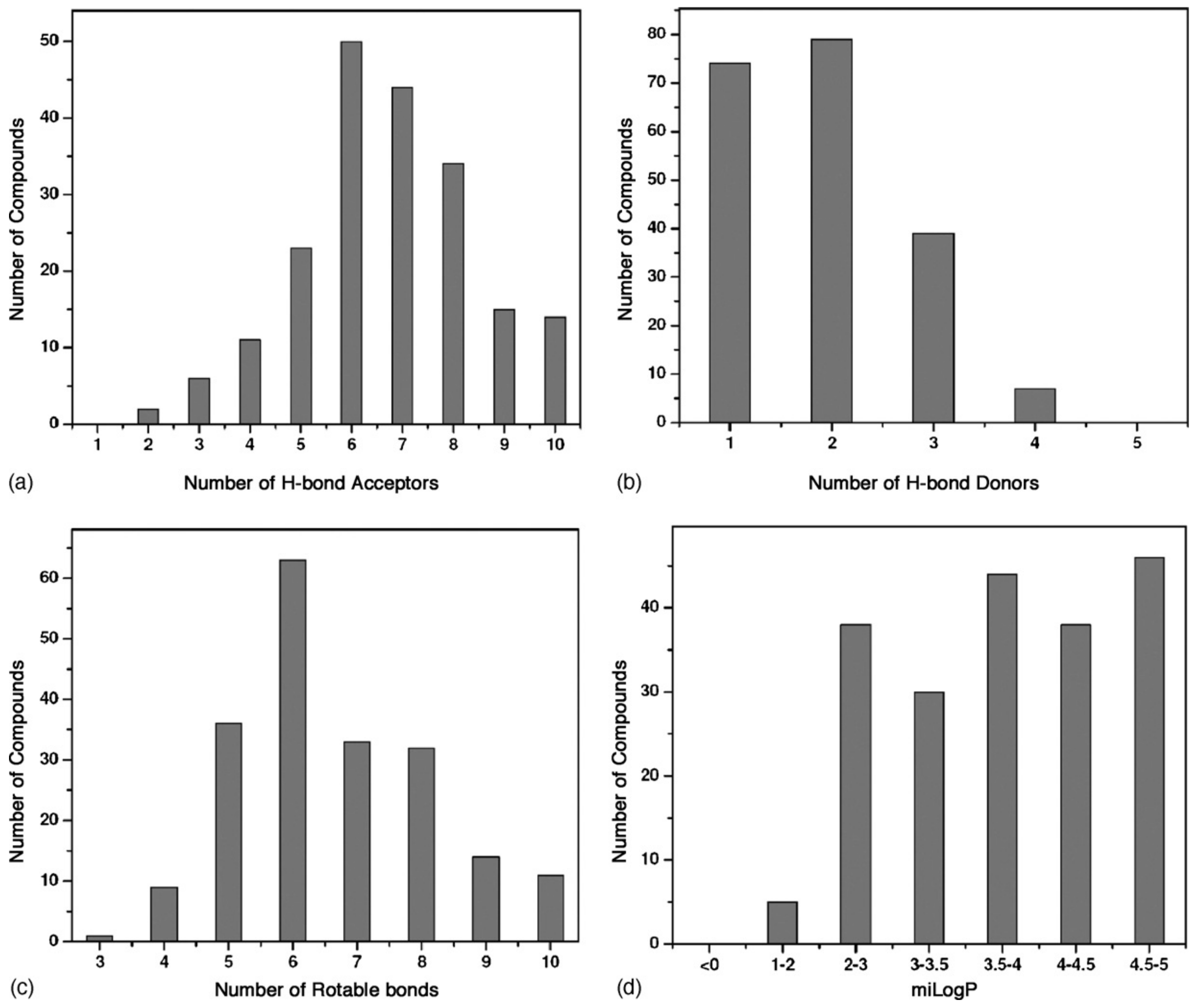


Fig. 5. Properties of compounds (199) which have satisfied Lipinski's rule of five as well as rotatable bonds (<10) have been shown as column graphs. Number of H-bond acceptors (a), H-bond donors (b), rotatable bonds (c) and miLogP (d).

chemical database consisting of 60,000 compounds. Inhibitory activities were estimated for the 5439 compounds that were retrieved from Maybridge database. Compounds that had *HypoGen* estimated activity ≤50 nM were considered as best active compounds. The 400 compounds satisfied the specified cutoff value and hence proceeded for further evaluation. The mapping of Hypo1 on top two hits found in Maybridge database search named as AW01179 and BTB00521 are represented in

```
1A8H   :  MEKVFYVTTPIYYVNAEPHLGHAYTTVVADFLARWHRLDGYRTFFLTGTDEHGETVYRAAQAAGEDP
Subj   : MAKETFYITTPIYYPSGNLHIGHAYSTVAGDVIARYKRMQGYDVRYLTGTDEHGQKIQEKAQKAGKTE
              ** ******     * **** **  *  **  *  **    ********      ** **
1A8H   : KAFVDRVSGRFKRAWDLLGIAYDDFIRTTEERHKKVVQLVLKKVYEAGDIYYGEYEGLYCVSCERFYT
Subj   : IEYLDEMIAGIKQLWAKLEISNDDFIRTTEERHKHVVEQVFERLLKQGDIYLGEYEGWYSVPDETYYT
             *      *  *  * *  ************ **  *       **** ***** * *  *  **
1A8H   : EKELVE----------GLCPIHGRPVERRKEGNYFFRMEKYRPWLQEYIQENPDLIRPEGYRNEVLAM
Subj   : ESQLVDPQYENGKIIGGKSPDSGHEVELVKEESYFFNISKYTDRLLEFYDQNPDFIQPPSRKNEMINN
         *  **           *  *  *  **  **  ***   **   * *    *** * *    **
1A8H   : LAEP-IGDLSISRPKSRVPWGIPLPWDENHVTYVWFDALLNYVSALDYP--EGEAYRTFWPHAWHLIG
Subj   : FIKPGLADLAVSRTS--FNWGVHVPSNPKHVVYVWIDALVNYISALGYLSDDESLFNKYWPADIHLMA
            *   **  **      **   *    ** *** *** ** *** *           **   **
1A8H   : KDILKPHAVFWPTMLKAAGIPMYRHLNVGGFLLGPDGRKMSKTLGNVVDPFALLEKYGRDALRYYLLR
Subj   : KEIVRFHSIIWPILLMALDLPLPKKVFAHGWILMKDG-KMSKSKGNVVDPNILIDRYGLDATRYYLMR
         * *   *   **  * *   *        *  *  ** ****  ******  *   ** ** **** *
1A8H   : EIPYGQDTPVSEEALRTRYEADLADDLGNLVQRTRAMLFRFAEGRIPE------PVAGEELAEGTGLA
Subj   : ELPFGSDGVFTPEAFVERTNFDLANDLGNLVNRTISMINKYFDGELPAYQGPLHELDEEMEAMALETV
         * * * *     **   *   *** ****** **  *      *  *           *  *
1A8H   : GRLRPLVRELKFHVALEEAMAYVKALNRYINEKKPWELFK--KEPEEARAVLYRVVEGLRIASILLTP
Subj   : KSYTESMESLQFSVALSTVWKFISRTNKYIDETTPWVLAKDDSQKDMLGNVMAHLVENIRYAAVLLRP
                  * * ***          * ** *  ** * *          *    **  * *  ** *
1A8H   : AMPDKMAELRRALGLKEE--VRLEEAERWG-LAEPRPIPEEAPVLFPKK------------------
Subj   : FLTHAPKEIFEQLNINNPQFMEFSSLEQYGVLTEPIMVTGQPKPIFPRLDSEAEIAYIKESMQPPAT
                *    *             *  * **          **
```

Fig. 6. Sequence alignment result between the template protein (MetRS of *T. thermophilus* represented with PDBID, 1A8H*)* and the target protein (MetRS of *S. aureus* represented as subject). The asterisks (*) indicate the conserved residues between the two proteins.

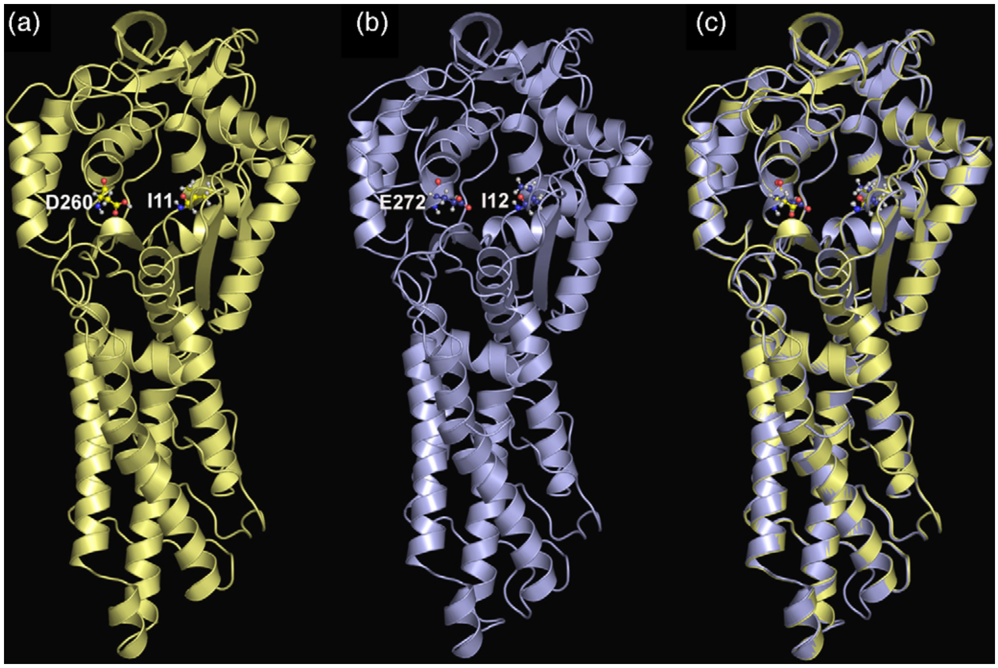


Fig. 7. Ribbon diagrams of the template structure of *T. thermophilus* MetRS (a), the homology modeled structure *S. aureus* MetRS (b), and the aligned structures of both the proteins (c). The figures a–c are showing catalytic residues which form H-bond interactions with ligands.

Fig. 3e and f. The estimated activities for these two compounds are 2.8 and 4.5 nM with fit values of 8.31 and 8.10, respectively.

Molecular properties were calculated for all 400 hits retrieved from the multi-conformational Maybridge chemical database. Lipinski's rule-of-five is a simple model to forecast the absorption and intestinal permeability of a compound [28,29]. *Molinspiration* online database (www.molinspiration.com) gives information of miLogP, molecular weight,

```
S. aureus         --MAKETFYITTPIYYPSGNLHIGHAY-STVAGDVIARYKRMQGYDVRYLTGTDEHGQKIQ
T. thermophilus   ---MEKVFYVTTPIYYVNAEPHLGHAY-TTVVADFLARWHRLDGYRTFFLTGTDEHGETVY
A. aeolicus       -MTLMKKFYVTTPIYYVNDVPHLGHAY-TTIAADTIARYYRLRDYDVFFLTGTDEHGLKIQ
E. coli           MTQVAKKILVTCALPYANGSIHLGHML-EHIQADVWVRYQRMRGHEVNFICADDAHGTPIM
P. abyssi         ----MVRYMVTSALPYANGPIHAGHLAGAYLPADIFVRYLRLKGEDVVFICGTDEHGTPIS
                           :* .: * .   * **     : .*  .*: *: .  . :: . * **  :
S. aureus         EKAQKAGKTEIEYLDEMIAGIKQLWAKLEISNDDFIRTTEERHKHVVEQVFERLLKQGDIY
T. thermophilus   RAAQAAGEDPKAFVDRVSGRFKRAWDLLGIAYDDFIRTTEERHKKVVQLVLKKVYEAGDIY
A. aeolicus       KKAEELGISPKELVDRNAERFKKLWEFLKIEYTKFIRTTDPYHVKFVQKVFEECYKRGDIY
E. coli           LKAQQLGITPEQMIGEMSQEHQTDFAGFNISYDNYHSTHSEENRQLSELIYSRLKENGFIK
P. abyssi         FRALKEGRSPREIVDEFHEQIKITFQRAKISFDFFGRTELPIHYKLSQEFFLKAYENGHLV
                    *   *      :..     :  :    *    :  *    : :. : .  .  : * :
S. aureus         LGEYEGWY-------------------SVPDETYYTESQLVDPQYENGKIIGGKSPDSGHE
T. thermophilus   YGEYEGLY-------------------CVSCERFYTEKELVEGLC----------PIHGRP
A. aeolicus       LGEYEGWY-------------------CVGCEEFKSEAELAEDHT---------CPIHQKK
E. coli           NRTISQLYDPEKGMFLPDRFVKGTCPKCKSPDQYGDNCEVCGATYSPTELIEPKSVVSGAT
P. abyssi         KKVTKQAYCEHDKMFLPDRFVIGTCPYCGAEDQKGDQCEVCGRPLTPEILINPRCAICGRP
                      .  *                   .    :    : ::
S. aureus         VELVKEESYFFNISKYTDRLLEFYDQNPDFIQPPSRKNEMINNFIKPGLADLAVSR--TSF
T. thermophilus   VERRKEGNYFFRMEKYRPWLQEYIQENPDLIRPEGYRNEVLAMLAEP-IGDLSISRPKSRV
A. aeolicus       CEYIKEPSYFFRLSKYQDKLLELYEKNPEFIQPDYRRNEIIS-FVKQGLKDLSVTRPRSRV
E. coli           PVMRDSEHFFFDLPSFSEMLQAWTRSGA---LQEQVANKMQEWFESGLQQWDISRDAPY--
P. abyssi         ISFRDSAHYYIKMQDFAERLKRWIEKQP---WKPNVKNMVLSWIEEGLEERAITRDLNWGI
                      ..  ::: : .:   *     . .         * :   : .
S. aureus         NWGVHVPSNPKHVVYVWIDALVNYISALGYLSD---DESLFNKYWP------ADIHLMAKE
T. thermophilus   PWGIPLPWDENHVTYVWFDALLNYVSALDYP-----EGEAYRTFWP------HAWHLIGKD
A. aeolicus       KWGIPVPFDPEHTIYVWFDALFNYISAL---------EDKVEIYWP------ADLHLVGKD
E. coli           -FGFEIPNAPGKYFYVWLDAPIGYMGSFKNLCDKRGDSVSFDEYWK-KDSTAELYHFIGKD
P. abyssi         PVPLDEEDMKGKVLYVWFEAPIGYISITIEHFKRIGKPNEWKKYWLNIDGQTRVIHFIGKD
                      .      :  ***::* ..*:.                 :*          *::.*:
S. aureus         IVRFHSIIWPILLMALDLPLP----------KKVFAHGWILMKDG
T. thermophilus   ILKPHAVFWPTMLKAAGIPMY----------RHLNVGGFLLGPDG
A. aeolicus       ILRFHTVYWPAFLMSLGYELP----------KKVFAHGWWTVEGK
E. coli           IVYFHSLFWPAMLEGSNFRK----------PSNLFVHGYVTVNGA
P. abyssi         NIPFHAIFWPAFLMAYGKYKDEEVEAEWNLPYDIPANEYLTLEGK
                   :  *:: ** :* . .               .: .  :    .
```

Fig. 8. Sequence alignment of *S. aureus* MetRS with other bacterial MetRS. Catalytic site amino acids are highlighted with color representation. Yellow color represents conserved and same amino acids present in all types of bacterial MetRS. Green represents conserved and similar type of amino acids (1–307 residues of *S. aureus* MetRS sequence with others were shown).

number of H-bond donors, acceptors and rotatable bonds based on the chemical structure [30,31]. According to the rule-of-five model, compounds were considered likely to be well-absorbed when they possess LogP < 5, molecular weight < 500, number of H-bond donors < 5, and number H-bond acceptors < 10. Number of rotatable bonds were also restricted to 10 as previous studies [32] suggest that candidate design directed at reduced flexibility (rotatable bonds < 10) and satisfying Lipinski's rule would increase the success in achieving compounds with high oral bioavailability. The 199 compounds have satisfied Lipinski's rule among 400 compounds. Analyses of these 199 new hits which satisfied H-bond acceptors, donors, miLogP as well as rotatable bonds criteria are provided in Fig. 5.

3.3. 3D structure of S. aureus MetRS was built by homology modeling method

There is no binding mechanism or any docking studies done for *S. aureus* MetRS until now although we have effective inhibitors for MetRS. This is due to the lack of an appropriate crystal structure. Therefore, we have developed 3D-model structure using homology modeling method. Fig. 6 shows the sequence alignment between the target protein, *S. aureus* MetRS, and the template protein (Fig. 7a), *T. thermophilus* MetRS (PDB ID: 1A8H). *BLAST* (blastp) showed about 38% sequence identity between the two sequences. The final homology modeling calculation using *MODELLER* generated a very reliable 3D structure of *S. aureus* MetRS (Fig. 7b and c) since *PROCHECK*, protein structure validation program, predicted that the 90.8% of residues of the 3D structure lied in most favored regions unlike the template protein which has 87.6% of residues in most favored regions. The RMSD between the template and the target structures is 0.37 Å. Multiple sequence alignment using ClustalW revealed the conserved aminoacids present in the modeled and other similar class proteins (Fig. 8). The methionine binding pocket of *E. coli* is formed by key aminoacids such as L13, Y15, H23, D52, W253, A256, Y260 G294, K295, D296, H301, and W305. Most of the residues are conserved in all five bacterial MetRS. Similar residues like I12, A270 and E272 of *S. aureus* were present in the place of L13, G294, D296 residues of *E. coli*, respectively, while the rest of the binding pocket residues were the same.

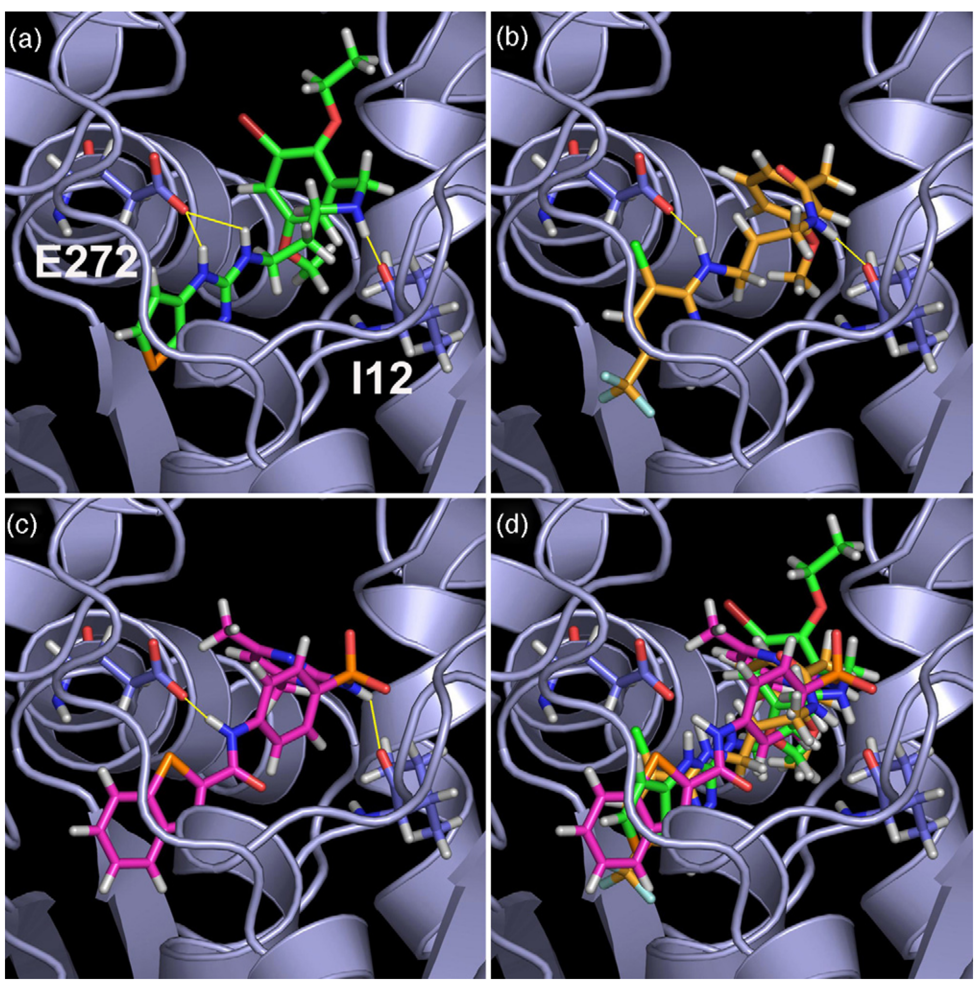
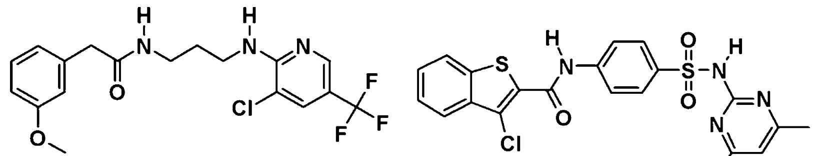


Fig. 9. The molecular docking results. The docked compound 1 of training set (a), Maybridge compound AW01179 (b), Maybridge compound BTB00521 (c), all three ligands at the binding pocket (d) are shown with the two catalytic residues (I12 and E272) of the modeled *S. aureus* MetRS structure.



AW01179

BTB00521

Fig. 10. Molecular structures of compound AW01179 and BTB00521.

3.4. Molecular docking

Training set of 29 compounds as well as all 199 new hits retrieved from the multi-conformational Maybridge chemical database which have satisfied drug-like properties were docked in the modeled *S. aureus* MetRS active site region using *GOLD* docking software. *GOLD* dock score which distinguishes molecules based on their interacting ability is calculated for all molecules. Compound 1, the most active molecule in training set showed 58.19 dock score and formed H-bonding with the two active site residues of I12 and E272. Latest docking studies on *E. coli* MetRS revealed that L13 and E296 are key residues in active site [33]. Some of the hits retrieved in database search also showed good dock scores and formed similar type of interactions with these two active site amino acids. Fig. 9a represents the H-bond interaction between compound 1 and active site amino acids. The 77 out of 199 compounds which obtained a *GOLD* dock score $\geq$55 were considered as final hits for further evaluation. Compound AW01179 retrieved from the Maybridge database have good *HypoGen* estimated activity (2.8 nM) as well as good *GOLD* fitness score (55.15) and it also formed similar interactions with I12 and E272 (Fig. 9b). Compound BTB00521 also showed good estimated activity (4.5 nM) and better *GOLD* fitness score, 62.97 than the best active molecule from the training set (Fig. 9c). These two molecules (Fig. 10) showed extraordinary results with respect to all properties like estimated activity, binding affinity, calculated drug-like properties and thus can be treated as good leads in the design of potent inhibitors of *S. aureus* MetRS.

4. Conclusions

Our pharmacophore hypothesis was able to accurately estimate the activities of known inhibitors with a correlation factor of 0.926. The mapping information based on the pharmacophore model we developed is now being taken advantage in the identification of novel lead compounds with improved inhibitory activity through 3D database searches. The homology model structure of *S. aureus* MetRS was very much helpful in exploring the binding site interactions and evaluating compounds with good dock score. The 77 new hits showed good estimated activities, higher *GOLD* dock scores as well as drug-like properties. These compounds calculated binding properties are very similar with experimentally proved compounds. Some of these compounds may have better in vitro activity against *S. aureus* MetRS. Therefore, our pharmacophore model is able to search new hits in any chemical databases and give good molecules which may have good anti-bacterial activity.

Acknowledgements

NB and KB thank Dr. J.A.R.P. Sarma (Head, Informatics division) and Mr. G.V. Sanjay Reddy, CEO GVK Biosciences Pvt. Ltd. for giving a great chance to work. NB and KB were recipients of fellowships from the BK21 Programs of Korean Ministry of Education & Human Resources Development. This work was supported by grants from the MOST/KOSEF for the Environmental Biotechnology National Core Research Center (grant #: R15-2003-012-02001-0) and for the Basic Research Program (grant #: R01-2005-000-10373-0).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2006.08.002.

References

[1] S.W. Lee, B.H. Cho, S.G. Park, S. Kim, Aminoacyl-tRNA synthetase complexes: beyond translation, J. Cell Sci. 117 (2004) 3725–3734.

[2] S. Bernier, P.M. Akochy, J. Lapointe, R. Chenevert, Synthesis and aminoacyl-tRNA synthetase inhibitory activity of aspartyl adenylate analogs, Bioorg. Med. Chem. 13 (2005) 69–75.

[3] A.K. Forrest, R.L. Jarvest, L.M. Mensah, P.J. O'Hanlon, A.J. Pope, R.J. Sheppard, Aminoalkyl adenylate and aminoacyl sulfamate intermediate analogues differing greatly in affinity for their cognate *Staphylococcus aureus* aminoacyl tRNA synthetases, Bioorg. Med. Chem. Lett. 10 (2000) 1871–1874.

[4] N. Shen, L. Guo, B. Yang, Y. Jin, J. Ding, Structure of human tryptophanyl-tRNA synthetase in complex with tRNATrp reveals the molecular basis of tRNA recognition and specificity, Nucl. Acids Res. 34 (2006) 3246–3258.

[5] M.W. Davis, D.D. Buechter, P. Schimmel, Functional dissection of a predicted class-defining motif in a class II tRNA synthetase of unknown structure, Biochemistry 33 (1994) 9904–9911.

[6] M. Delarue, D. Moras, The aminoacyl-tRNA synthetase family: modules at work, Bioessays 15 (1993) 675–687.

[7] Y.M. Hou, C. Francklyn, P. Schimmel, Molecular dissection of a transfer RNA and the basis for its identity, Trends Biochem. Sci. 14 (1989) 233–237.

[8] I. Sugiura, O. Nureki, Y. Ugaji-Yoshikawa, S. Kuwabara, A. Shimada, M. Tateno, B. Lorber, R. Giege, D. Moras, S. Yokoyama, M. Konno, The 2.0 A crystal structure of Thermus thermophilus methionyl-tRNA synthetase reveals two RNA-binding modules, Structure 8 (2000) 197–208.

[9] S.Y. Kim, J. Lee, 3-D-QSAR study and molecular docking of methionyl-tRNA synthetase inhibitors, Bioorg. Med. Chem. Lett. 11 (2003) 5325–5331.

[10] CATALYST 4.10 User Guide, Accelrys Inc., San Diego, CA, USA, 2005.

[11] Y. Kurogi, O.F. Guner, Pharmacophore modeling and three-dimensional database searching for drug design using catalyst, Curr. Med. Chem. 8 (2001) 1035–1055.

[12] A.K. Debnath, Pharmacophore mapping of a series of 2,4-diamino-5-deazapteridine inhibitors of *Mycobacterium avium* complex dihydrofolate reductase, J. Med. Chem. 45 (2002) 41–53.

[13] P. Kahnberg, M.H. Howard, T. Liljefors, M. Nielsen, E.O. Nielsen, O. Sterner, I. Pettersson, 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 (2004) 253–261.

[14] J. Faragalla, J. Bremner, D. Brown, R. Griffith, A. Heaton, Comparative pharmacophore development for inhibitors of human and rat 5-alpha-reductase, J. Mol. Graph. Model. 22 (2003) 83–92.

[15] J. Shen, HAD: an automated database tool for analyzing screening hits in drug discovery, J. Chem. Inf. Comput. Sci. 43 (2003) 1668–1672.

[16] R.L. Jarvest, S.A. Armstrong, J.M. Berge, P. Brown, J.S. Elder, M.J. Brown, R.C. Copley, A.K. Forrest, D.W. Hamprecht, P.J. O'Hanlon, D.J. Mitchell, S. Rittenhouse, D.R. Witty, Definition of the heterocyclic pharmacophore of bacterial methionyl tRNA synthetase inhibitors: potent antibacterially active non-quinolone analogues, Bioorg. Med. Chem. Lett. 14 (2004) 3937–3941.

[17] R.L. Jarvest, J.M. Berge, M.J. Brown, P. Brown, J.S. Elder, A.K. Forrest, C.S. Houge-Frydrych, P.J. O'Hanlon, D.J. McNair, S. Rittenhouse, R.J. Sheppard, Optimisation of aryl substitution leading to potent methionyl tRNA synthetase inhibitors with excellent gram-positive antibacterial activity, Bioorg. Med. Chem. Lett. 13 (2003) 665–668.

[18] R.L. Jarvest, J.M. Berge, P. Brown, C.S. Houge-Frydrych, P.J. O'Hanlon, D.J. McNair, A.J. Pope, S. Rittenhouse, Conformational restriction of methionyl tRNA synthetase inhibitors leading to analogues with potent inhibition and excellent gram-positive antibacterial activity, Bioorg. Med. Chem. Lett. 13 (2003) 1265–1268.

[19] J. Finn, K. Mattia, M. Morytko, S. Ram, Y. Yang, X. Wu, E. Mak, P. Gallant, D. Keith, Discovery of a potent and selective series of pyrazole bacterial methionyl-tRNA synthetase inhibitors, Bioorg. Med. Chem. Lett. 13 (2003) 2231–2234.

[20] M. Tandon, D.L. Coffen, P. Gallant, D. Keith, M.A. Ashwell, Potent and selective inhibitors of bacterial methionyl tRNA synthetase derived from an oxazolone-dipeptide scaffold, Bioorg. Med. Chem. Lett. 14 (2004) 1909–1911.

[21] A. Smellie, S.D. Kahn, S.L. Teig, Analysis of Conformational Coverage. 1. Validation and Estimation of Coverage, J. Chem. Inf. Comput. Sci. 35 (1995) 285–294.

[22] A. Smellie, S.D. Kahn, S.L. Teig, Analysis of conformational space. 2. Applications of conformational models, J. Chem. Inf. Comput. Sci. 35 (1995) 295–304.

[23] A. Sali, T.L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, J. Mol. Biol. 234 (1993) 779–815.

[24] A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, Protein Sci. 9 (2000) 1753–1773.

[25] InsightII, Version 2005.3L, Accelrys Inc., San Diego (www.accelrys.com) 2005.

[26] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.

[27] G. Jones, P. Willett, R.C. Glen, A.R. Leach, R. Taylor, Development and validation of a genetic algorithm for flexible docking, J. Mol. Biol. 267 (1997) 727–748.

[28] C.A. Lipinski, Drug-like properties and the causes of poor solubility and poor permeability, J. Pharmacol. Toxicol. Meth. 44 (2000) 235–249.

[29] C.A. Lipinski, F. Lombardo, B.W. Dominy, P.J. Feeney, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, Adv. Drug Deliv. Rev. 46 (2001) 3–26.

[30] P. Ertl, B. Rohde, P. Selzer, Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties, J. Med. Chem. 43 (2000) 3714–3717.

[31] J.J. Irwin, B.K. Shoichet, ZINC—a free database of commercially available compounds for virtual screening, J. Chem. Inf. Model. 45 (2005) 177–182.

[32] D.F. Veber, S.R. Johnson, H.Y. Cheng, B.R. Smith, K.W. Ward, K.D. Kopple, Molecular properties that influence the oral bioavailability of drug candidates, J. Med. Chem. 45 (2002) 2615–2623.

[33] S.Y. Kim, Y.S. Lee, T. Kang, S. Kim, J. Lee, Pharmacophore-based virtual screening: The discovery of novel methionyl-tRNA synthetase inhibitors, Bioorg. Med. Chem. Lett. 16 (2006) 4898–4907.