Combined Support-Vector-Machine-Based Virtual Screening and Docking Method for the Discovery of IMP-1 Metallo-β-Lactamase Inhibitors

Jiao Chen, Yifang Liu, Mi Fang, Hui Chen, Xingzhen Lao, Xiangdong Gao, Heng Zheng*, Wenbing Yao*

School of Life Science and Technology, China Pharmaceutical University, Nanjing, P.R. China

Received November 07, 2013; Revised November 13, 2013; Accepted November 15, 2013

Abstract Metallo-β-lactamases can hydrolyze a broad range of β-lactam antibiotics and no effective inhibitors could be used in the clinic. Therefore, the discovery of metallo-β-lactamase inhibitors has attracted much attention in recent years. In this study, a support vector machine (SVM) that separates compounds into positives and negatives, combined with docking method was employed for virtual screening of IMP-1 metallo-β-lactamase inhibitors. Eight of the twenty five selected compounds were purchased for in vitro assays. Among them, four compounds show inhibitory potency against IMP-1. Two of them are found to have novel scaffolds, implying a good potential for further optimization.

Keywords: IMP-1 metallo-β-lactamase inhibitors, support vector machine, docking, virtual screening, in vitro assays

Cite This Article: Jiao Chen, Yifang Liu, Mi Fang, Hui Chen, Xingzhen Lao, Xiangdong Gao, Heng Zheng, and Wenbing Yao, “Combined Support-Vector-Machine-Based Virtual Screening and Docking Method for the Discovery of IMP-1 Metallo-β-Lactamase Inhibitors.” American Journal of Biomedical Research 1, no. 4 (2013): 120-131. doi: 10.12691/ajbr-1-4-8.

1. Introduction

The production of β-lactamases is one of the most important resistant mechanisms in bacteria, which can inactivate β-lactams by hydrolyzing the amide bond of the four-membered β-lactam ring. These enzymes can be divided into serine-β-lactamases and metallo-β-lactamases (MβLs), which employ a serine and one or two zinc ions in the active site, respectively [1]. Except monobactams, the MBLs can degrade almost all classes of β-lactams including carbapenems, which are stable against the serine-β-lactamases. The currently clinically used β-lactamase inhibitors such as clavulanic acid, sulbactam, and tazobactam can poorly inhibit MBLs.

Imipenemase (IMP), one of the most dangerous metallo-β-lactamases, can hydrolyze a broad spectrum of lactams, including penicillins, cephalosporins, and carbapenems [2]. Accordingly, 47 variants of IMP MβLs have been identified (http://www.lahey.org/Studies/other.asp#table1, data collected before 29 Oct. 2013), and the number is rapidly growing. In the past few years, several types of compounds have been reported inhibitory activity against IMP-1, such as 2, 3-disubstituted succinic acids [3], thiols [4], thioesters [5], maleic acid derivatives [6], and carbapenem derivatives [7] etc. However, none of these inhibitors has fortunately passed clinical trials. Thus, the discovery of new types of MβL inhibitors has increasingly attracted public attention in recent years in order to recover the effectiveness of β-lactam antibiotics.

Several in silico methods including pharmacophore [8], docking [9], molecular dynamics [10] and our previously established 3D-QSAR models [11] have been used for facilitating the search and design of IMP MβL inhibitors. These methods have shown significant capability in the identification of potential IMP inhibitors. Molecular docking, which is the most widely used structure-based virtual screening method, considers sufficiently the match of three-dimensional geometrical shape between the ligand and the active site of receptor [12], but the problems such as the sparse chemical space, the highly consuming of time and computing resource, and even the limited diversity of training set may affect their applications. Thereby, it is highly desirable to obtain other computational model capable of screening large compound libraries rapidly and guiding the discovery of IMP inhibitors with new scaffolds.

Support Vector Machine (SVM), a supervised learning method used for classification, has been gaining popularity due to many attractive features and promising empirical performance in the fields of protein structure prediction [13,14], specific biochemical reaction sites identification [15,16,17], protein-ligand binding affinity prediction [18] and so on. Prof. Chen. Y. Z. et.al employed this method in virtual screening of kinase inhibitors [19], gamma-secretase inhibitors [20], HDAC inhibitors [21], new potential organocatalysts [22], Abl
inhibitors [23], and small molecule aggregators [24]. They demonstrated that SVM-based virtual screening has comparable and in some cases substantially higher yields than those of conventional virtual screening tools reported in the literature. Most importantly, it is in a timely, low-resource manner. SVM trains models and identifies active compounds based on the physicochemical properties rather than their structural similarity. So no knowledge of enzyme structure, no computation of structural flexibility, binding affinities, activity-related features and no inhibitory mechanisms are required. However, SVM-based virtual screening still perform imperfectly in some aspects, for example, in reducing the false positive rate, which might originates from the lack of consideration of the information about the macromolecule target. Therefore, it is expected the combined SVM-based virtual screening and docking method can lead to a high hit rate and a large enrichment factor, as well as high-speed [25].

In this work, we adopted the combined SVM-based virtual screening and docking method for new active inhibitor discovery. The accuracy of SVM is largely dependent on the descriptors used in the model. So we applied the automatic step-by-step descriptors selector (SDS-SVM) developed in house to select the molecular descriptors. The SVM model was built based on the known IMP-1 inhibitors and putative noninhibitors, which could rapidly classify the compounds in the library into potential IMP-1 inhibitors and noninhibitors. The hit compounds were subjected to subsequent docking screening. Finally, we selected 25 compounds and 8 were purchased for in vitro assays. The schematic workflow is shown in Figure 1.

2. Experimental Procedures

2.1. Data Sets

A total of 262 IMP-1 inhibitors were collected from different literature (Supplementary Data, Table S1) [3,4,5,6,7,26-38]. Scaffolds of the representative inhibitors are shown in Figure 2. As few IMP-1 noninhibitors have been reported, putative noninhibitors [39] were generated according to the method described in our previous study [40]. 10,605 putative noninhibitors were randomly selected from putative negative families for SVM modeling.

2.2. Descriptor Selection

Descriptor selection has an important impact on the improvement of classification performance in SVM modeling. In this work, we applied the automatic step-by-step descriptor selector (SDS-SVM) developed in house [40] based on Perl & R to select the molecular descriptors for SVM modeling.

The molecular descriptors used in our work were calculated in MOE2009 software. Redundant descriptors with > 90% identical values and those irrelevant to the
pharmacokinetic and toxicological properties were removed in the preprocessing. A stricter criterion was used to define the positive compound in order to select more representative features of the inhibitors sensitively. Thus, only 123 compounds with IC$_{50}$ $\leq$ 10 $\mu$M or Ki $\leq$ 2 $\mu$M among the total 262 IMP-1 inhibitors were considered as positives, whereas the other 139 compounds with IC$_{50}$ $> 10$ $\mu$M or Ki $> 2$ $\mu$M were treated as negatives during the descriptor selection step. After step-by-step eliminating, 24 descriptors were retained for the subsequent virtual screening. These descriptors characterize structural and physicochemical features of IMP-1 inhibitors.

2.3. SVM Modeling and Virtual Screening

In our SVM model, 225 IMP-1 inhibitors with IC$_{50}$ $\leq$ 200 $\mu$M or Ki $\leq$ 100 $\mu$M together with 1060 putative noninhibitors were divided into the training set (180 positives and 8484 putative noninhibitors) and test set (45 positives and 2121 putative noninhibitors) based on their structural diversity and biological activity. The model was built based on the training compounds and then tested using the independent test set. Subsequently, the merged compound library was subjected to virtual screening based on the selected 24 descriptors by this SVM model. 2861 compounds were hit.

2.4. Molecular Docking

The SVM-hit compounds were washed and the lowest energy conformer of each compound was generated in MOE2009 (Chemical Computing Group Inc., Montreal, Canada). Molecular docking studies were carried out by GOLD (Genetic Optimization of Ligand Docking) 3.0.1 [41]. GOLD adopts the genetic algorithm to dock flexible ligands into protein binding sites. The crystal structure of IMP-1 in complex with 2,3-bis-benzo[1,3]dioxol-5-ylmethyl- succinic acid (PDB entry: 1JJE) was used as the reference receptor for the docking studies. The crystal waters were deleted and hydrogen atoms were added to the protein. Residues that stay within 15 Å from the original ligand were defined as the binding site. A tetrahedral geometry was applied to Zn$_1$ (3H binding site), whereas a trigonal bipyramid was set to Zn$_2$ (DCH binding sites). The default values were used for the remaining parameters, and 10 Genetic Algorithm runs were performed for each ligand. GOLDScore was used to evaluate the docking results.

2.5. In vitro IMP-1 Inhibitory Assay

The IMP-1 gene lacking the first 18 residues was synthesized chemically and then cloned between BamH and XhoI restriction sites into the pET-28a plasmid harboring a kanamycin resistance gene. The forward primer was 5’- CCGggatccgCAGAGTCTTTGCCAGATTTAAAAA-3’ and the reverse primer was 5’-CCCGcctcgagTTAGTTGCTTGGTTTAGATGGTTTTT-3’. The constructed plasmids encoding IMP-1 was transferred into E.coli BL21 (DE3). The enzymes were expressed and purified as previously described [42].

Inhibition of IMP-1 was monitored by following the absorbance variation that resulted from the hydrolysis of the β-lactam ring. All the kinetic assays were performed at 30°C in the buffer of 50mM Hepes, 100µM ZnCl$_2$ and 0.1M NaCl at pH 7.5. The hydrolysis of ampicillin was measured as the substrate at 235nm, using a UV-1800 spectrophotometer (MAPADA). The enzyme and inhibitor were preincubated at 30°C for 20min before the substrate was added. The concentration of the tested compound which exhibited 50% inhibition of IMP-1 was set as IC$_{50}$.

3. Results and Discussion

3.1. Descriptor Selection

All the collected 262 IMP-1 inhibitors retrieved from the literature were used for descriptor selection by our SDS-SVM program. The inhibitors were divided into 123 positives and 139 negatives (Supplementary Data, Table S1). After SDS-SVM elimination, 24 descriptors survived for SVM modeling and virtual screening. The distribution of them can be seen in Table 1. These descriptors cover eight classes including physical properties, subdivided surface areas, atom counts and bond counts, adjacency and distance matrix descriptors, pharmacophore feature descriptors, partial charge descriptors, potential energy descriptors and surface area, volume and shape descriptors. Among the 24 selected descriptors, 9 belong to the class of surface area, volume and shape descriptors, which depend on the structure connectivity and conformation (dimensions are measured in Å), indicating that this class contributes the most to the SVM model we will build.

### Table 1. Molecular Descriptors Selected by the SDS Algorithm for SVM-based VS

| Descriptor class       | Number of descriptors in each class | Descriptors |
|------------------------|------------------------------------|-------------|
| Physical Properties    | 2                                  | FCharge, rsynth |
| Subdivided Surface Areas | 3                                         | SlogP\_VSA8, SlogP\_VSA9, SMR\_VSA7 |
| Atom Counts and Bond Counts | 2                                         | a\_ICM, opt\_leadlike |
| Adjacency and Distance Matrix Descriptors | 2                                         | BCUT\_PEOE\_0, petitjean |
| Pharmacophore Feature Descriptors | 2                                         | vsa\_acc, vsa\_other |
| Partial Charge Descriptors | 3                                         | PEOE\_VSA+2, PEOE\_VSA-4, PEOE\_VSA-6 |
| Potential Energy Descriptors | 1                                         | E\_stab |
| Surface Area, Volume and Shape Descriptors | 9                                         | rygr, vsurf\_DW13, vsurf\_EWmin3, vsurf\_W4, vsurf\_HB1, vsurf\_HB4, vsurf\_EDmin2, vsurf\_ID4, vsurf\_DD12 |

Given a set of training compounds that are divided into two categories (IMP-1 inhibitors and noninhibitors), an
SVM training algorithm builds a model that predicts whether a new compound falls into one category or the other [43]. In our SVM model, the training set comprises 8,664 compounds, including 180 known IMP-1 inhibitors (positives) and 8,484 putative noninhibitors (negatives). The 24 selected descriptors were used to build the SVM model (Table 2). The accuracies for predicting IMP-1 inhibitors (SE) and noninhibitors (SP) are 89.44% and 99.98%, respectively. The overall prediction accuracy Q and Matthews correlation coefficient (MCC) are 99.76% and 0.94, respectively.

Fivefold cross-validation was performed to evaluate the generated SVM model. As shown in Table 2, the average accuracies for predicting IMP-1 inhibitors (SE) and noninhibitors (SP) are 81.67% and 99.92%, respectively. The average overall prediction accuracy Q and MCC are 99.54% and 0.88, respectively. These indicate that the generated SVM model is reliable for the prediction of training set agents.

Subsequently, the established SVM classification model was further validated by an independent test set, with the purpose of assessing the predictive power of SVM model to the external compounds that are not contained in the training set. The independent test set contains 45 inhibitors and 2,121 noninhibitors. Of the 45 inhibitors, 40 were correctly predicted, indicating a prediction accuracy of 88.89% (SE, Table 2) for the inhibitors. For the 2,121 noninhibitors, all of them were properly predicted, indicating a prediction accuracy of 100% (SP, Table 2). The overall prediction accuracy Q and MCC are 99.77% and 0.94, respectively, which is comparable with the training set. All of these results demonstrate that our SVM model is reliable in differentiating the IMP-1 inhibitors and noninhibitors in both the training set and the external test set.

### Table 2. Validation results for the SVM model by a fivefold cross-validation method and an independent test set method

| Method                        | TP  | FN  | SE(%) | TN  | FP  | SP(%) | Q(%) | MCC |
|-------------------------------|-----|-----|-------|-----|-----|-------|------|-----|
| SVM model                     | 161 | 19  | 89.44 | 8483| 2   | 99.98 | 99.76| 0.94|
| 1                             | 28  | 8   | 77.78 | 1696| 0   | 100.00| 99.54| 0.88|
| 2                             | 31  | 5   | 86.11 | 1694| 2   | 99.88 | 99.60| 0.90|
| 3                              | 29  | 7   | 80.56 | 1692| 4   | 99.76 | 99.36| 0.84|
| 4                             | 26  | 10  | 72.22 | 1696| 0   | 100.00| 99.42| 0.85|
| 5                              | 33  | 3   | 91.67 | 1695| 1   | 99.94 | 99.77| 0.94|
| **Average**                   |     |     |       |     |     |       |      | 0.88|
| **Independent validation test**| 40  | 5   | 88.89 | 2121| 0   | 100.00| 99.77| 0.94|

Virtual screening of the merged library with 1,257 million compounds by the established SVM model generated 6,452 hits, which were later subjected to a more accurate docking program.

### 3.3. Molecular Docking

Docking is the most popular structure-based virtual screening method. Among various docking programs, GOLD has been proved to be a better program in the docking of IMP-1 inhibitors, giving a good correlation between the experimentally determined affinities and the GOLD scores [9]. Therefore, we chose GOLD to re-screen the SVM-hit 2,861 compounds. Twenty five compounds were selected based on experience and the GOLD scores for further investigation. Eight of them were purchased for in vitro assays, since the others were not commercially available. The docking scores and calculated MM/GBVI binding free energy values are shown in Table 3.

### Table 3. Docking scores, MM/GBVI and IC_{50} values of the assayed compounds

| Compound | GoldScore | MM/GBVI (kcal/mol) | IC_{50} (μM) |
|----------|-----------|--------------------|-------------|
| 1        | 71.27     | -124.36            | 124         |
| 2        | 73.85     | -25.62             | 227         |
| 3        | 75.57     | -124.94            | 254         |
| 4        | 85.86     | -119.29            | 288         |
| 5        | 80.07     | -113.79            | 7800        |
| 6        | 73.65     | -25.86             | ND*         |
| 7        | 73.08     | -127.70            | ND          |
| 8        | 99.60     | -31.89             | ND          |

*Not detected because of poor solubility.*

### 3.4. In vitro Assay

IMP-1 was expressed and purified, and then confirmed by SDS-PAGE (Figure 3). Inhibitory activities of the purchased compounds were determined using ampicillin as report substrate by the methods described above. Their structures together with the measured IC_{50} values are shown in Figure 4.
solution; lane 5, crude proteins after ultrasonication and centrifugation in precipitation; lane 6, IMP-1 after Ni-NTA affinity chromatography

The docking scores and calculated binding free energies are not always consistent with the experimental bioactivities. The values of MM/GBVI are comparable only when the evaluated compounds have similar binding manner, for example, the MM/GBVI values of compound 1, 3, 4, 5 and 7 with two carboxyl groups in their structures, are distinct from the values of compound 2, 6 and 8 without carboxyl groups. Table 3 also demonstrates that the top ranked compound 5 with GoldScore 80.07 has almost no inhibitory capacity against IMP-1, while for compound 1, whose GoldScore is only 71.27, possesses much higher inhibitory capacity with IC₅₀ value of 124 μM. The results suggest that it is not always successful to select the inhibitor solely based on the docking score or binding energy. The better way is to combine different methods to improve the accuracy and sensitivity.

### 3.5. Analysis of Active Compounds

An in vitro IMP-1 inhibitory assay was carried out for the purchased 8 compounds. The inhibitory potency was tested with ampicillin as the substrate. Four compounds showed IC₅₀ < 300 μM and three compounds were not tested because of their poor solubility (Figure 4). Among them, compound 1 and 3 have new scaffolds whereas compound 2 possesses common pharmacophore of sulfonamide with the reported inhibitor of DansylCnSH, and compound 4 is a new derivative of succinic acid. These four active compounds were redocked into IMP-1 metallo-β-lactamase (PDB code: 1JJE). Their individual binding interactions are shown in Figure 5. The residues were numbered according to Galleni et al. [44]. Docking results reveal that compound 1, 3 and 4 share similar binding mode with two carboxyl groups coordinating to the two zinc ions of IMP-1. Zn2 coordinates to both of the two carboxyl groups while Zn1 coordinates to only one of them. The sidechains of Asn233 and Lys224 act as hydrogen bond donor to the carboxyl groups in compound 1, 3 and 4. Arene-arene interactions were generated between imidazole group of His in IMP-1 and aromatic rings in the active compounds. For compound 2, the furan ring also formed arene-arene interaction with the indole group of Trp64. Zn1 of IMP-1 coordinates to the oxygen of sulfonamide and Zn2 coordinates to both of the oxygen and nitrogen of sulfonamide in compound 2.

For the sake of increasing inhibitory activity of these compounds, chemical modifications are underway. Protein-ligand interaction fingerprint (PLIF) analysis of the crystal structures of IMP-1 in complex with inhibitors [40] suggests that the hydrogen bond between the inhibitor and Asn233 is an important interaction. However, for compound 1, no hydrogen bond is formed with Asn233. Substitutions like –OH on the benzene ring at R2 might increase the potential inhibitory activity of compound 1. Take compound 3 as another example, our 3D-QSAR model [40] suggested that modifications with
some hydrophobic substituent groups might enhance its inhibitory activity against IMP-1.

4. Conclusion

MβLs have gained a lot of attention owing to their broad hydrolytic spectrum on β-lactams and the current lack of effective inhibitors. In addition, continuous evolution of drug-resistant bacteria leads to the emergence of new MβL variants, which makes the situation more discouraging [45]. In our study, an SVM model of IMP-1 inhibitors and putative noninhibitors was established based on the 24 SDS-SVM selected descriptors and validated by fivefold cross-validation and an independent test set. Then the established SVM classification model combined with docking method were used to virtual screen a merged compound library for new IMP-1 inhibitors. Finally, 25 compounds were selected and 8 were purchased for in vitro IMP-1 inhibitory assays. 4 compounds were found to have IC₅₀ < 300 μM and 2 of them have novel scaffolds, which worth further optimization. The 17 commercial unavailable compounds might also have potential inhibitory activity against IMP-1, and the synthetic work is in progress.

This work demonstrates that combined SVM-based virtual screening and docking method considerably increase hit rate and enrichment factor. Most importantly, it is in a timely, low-resource manner compared with the individual methods.

Figure 5. Binding interactions of the active compounds with IMP-1 (PDB code: 1JJE)
Acknowledgments

This work has been supported by the Fundamental Research Funds for the Central Universities of China (JKZZ2009007), the National Natural Science Foundation of China (Grant No. 31300643), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

References

[1] Aigner, R.P. “The structure of beta-lactamases.” Philos Trans R Soc Lond B, 289(1036). 321-331. May. 1980.

[2] Iyobe, S., Kasadoshoro, K., Ozaki, J., Matsumura, N., Minami, S., Haruta, S., Sawai, T., O’Hara, K. “Amino acid substitutions in a variant of IMP-1 metallo-beta-lactamase.” Antimicrob Agents Chemother, 44(8). 2023-2027. Aug. 2000.

[3] Toney, J.H., Hammond, G.G., Fitzgerald, P.M., Sharma, N., Balkovec, J.M., Rouen, G.P., Olson, S.H., Hammond, M.L., Greeene, M.L., Gao, Y.D. “Sucinonic acids as potent inhibitors of plasmid-borne IMP-1 metallo-beta-lactamase.” J Biol Chem, 276(34). 31913-31918. Jun. 2001.

[4] Siemann, S., Clarke, A.J., Viswanatha, T., Dmitrienko, G.I. “Thiols as classical and slow-binding inhibitors of IMP-1 and other binuclear metallo-beta-lactamases.” Biochemistry, 42(6). 1673-1683. Feb. 2003.

[5] Hammond, G.G., Huber, J.L., Greeene, M.L., La Fuente, J., Rees, M., Rivera-Sagredo, A. “Identification of a series of tricyclic natural products as potent broad-spectrum metallo-beta-lactamase inhibitors.” J Infect Dis, 189(4). 2627-2635. Apr. 2004.

[6] Oelsen, L., Patterson, I., Hemmingsen, L., Jorgensen, F.S. “New leads of metallo-beta-lactamase inhibitors from structure-based pharmacophore design.” Bioorg Med Chem, 14(8). 2627-2635. Apr. 2006.

[7] Du, H., Wang, J., Zhang, X., Hu, Z. “A novel quantitative structure-activity relationship method to predict the affinities of MT3 melatonin binding site.” Eur J Med Chem, 43(12). 2861-2869. Dec. 2008.

[8] Ma, X.H., Wang, R., Tan, C.Y., Jiang, Y.Y., Lu, T., Rao, H.B., Li, X.Y., Go, M.I., Low, B.C., Chen, Y.Z. “Virtual screening of selective multitarget kinase inhibitors by combinatorial support vector machines.” Mol Pharm, 7(5). 1545-1560. Oct. 2010.

[9] Yang, X.G., Lv, W., Chen, Y.Z., Xue, Y. “In silico prediction and screening of gamma-secretase inhibitors by molecular descriptors and machine learning methods.” J Comput Chem, 31(6). 1249-1258. Apr. 2010.

[10] Liu, X.H., Song, H.Y., Zhang, J.X., Han, B.C., Wei, X.N., Ma, X.H., Cui, W.K., Chen, Y.Z. “Identifying Novel Type ZBGs and Nonhydroxamate HDAC Inhibitors Through a SVM Based Virtual Screening Approach” Mol Inf, 29(5). 407-420. May. 2010.

[11] Liu, X.H., Song, H.Y., Ma, X.H., Lear, M.J., Chen, Y.Z. “Virtual Screening Prediction of New Potential Organo catalysts for Direct Aldol Reactions.” J Mol Catal A: Chem, 319(1-2). 114-118. Mar. 2010.

[12] Liu, X.H., Ma, X.H., Tan, C.Y., Jiang, Y.Y., Go, M.I., Low, B.C., Chen, Y.Z. “Virtual screening of Abi inhibitors from large compound libraries by support vector machines.” J Chem Inf Model, 49(9). 2101-2110. Sep. 2009.

[13] Rao, H., Li, Z., Li, X., Ma, X., Ung, C., Li, H., Liu, X., Chen, Y. “Identification of small molecule aggregators from large compound libraries by support vector machines.” J Comput Chem, 31(4). 752-763. Mar. 2010.

[14] Ren, J.X., Li, L.L., Zheng, R.L., Xie, H.Z., Cao, Z.X., Feng, S., Pan, Y.L., Chen, X., Wei, Y.Q., Yang, S.Y. “Discovery of novel Pim-1 kinase inhibitors by a hierarchical multistage virtual screening approach based on SVM model, pharmacophore, and molecular docking.” J Chem Inf Model, 51(6). 1364-1375. Jun. 2011.

[15] Toney, J.H., Cleary, K.A., Hammond, G.G., Yuan, X., May, W.J., Hutchins, S.M., Ashton, W.T., Vanderwall, D.E. “Structure-activity relationships of biphenyl tetraroles as metallo-beta-lactamase inhibitors.” Bioorg Med Chem, 9(18). 2741-2746. Sep. 1999.

[16] Payne, D.J., Hueso-Rodriguez, J.A., Boyd, H., Concha, N.O., Janson, C.A., Gilpin, M., Bateson, J.H., Cheever, C., Niconovich, N.L., Pearson, S., Rittenhouse, S., Tew, D., Diez, E., Perez, P., De La Fuente, J., Rees, M., Rivera-Sagredo, A. “Identification of a series of tricyclic natural products as potent broad-spectrum inhibitors of metallo-beta-lactamases.” Antimicrob Agents Chemother, 46(6). 1880-1886. Jun. 2002.

[17] Siemann, S., Evanoff, D.P., Marrone, L., Clarke, A.J., Viswanatha, T., Dmitrienko, G.I., “N-arylsulfonyl hydrazones as inhibitors of IMP-1 metallo-beta-lactamase.” Antimicrob Agents Chemother, 46(8). 2450-2457. Aug. 2002.

[18] Jin, W., Arakawa, Y., Yasuzawa, H., Taki, T., Hashiguchi, R., Mitsutani, K., Shoga, A., Yamaguchi, Y., Kurosaki, H., Shibata, N., Ohta, M., Goto, M. “Comparative study of the inhibition of metallo-beta-lactamase (IMP-1 and VIM-2) by thiol compounds that contain a hydrophobic group.” Biochim Biophys Acta, 1757(1). 137-143. Jun. 2007.

[19] Kurosaki, H., Yamaguchi, Y., Higashi, T., Soga, K., Matsueda, S., Yumoto, H., Misumi, S., Yamagata, Y., Arakawa, Y., Goto, M. “Irreversible inhibition of metallo-beta-lactamase (IMP-1) by 3-(3-mercaptopropionylsulfanyl)propionic acid pentfluorophenyl ester.” Angew Chem Int Ed Engl, 44(25). 3861-3864. Jun. 2005.

[20] Kurosaki, H., Yamaguchi, Y., Yasuzawa, H., Jin, W., Yamagata, Y., Arakawa, Y. “Probing, inhibition, and crystallographic characterization of metallo-beta-lactamase (IMP-1) with fluorescent agents containing dansyl and thiol groups.” ChemMedChem, 1(9). 969-972. Sep. 2006.

[21] Horsfall, L.E., Garau, G., Lienard, B.M., Dideberg, O., Schofield, C.J., Freire, J.M., Galleni, M. “Competitive inhibitors of the CphA metallo-beta-lactamase from Aeromonas hydrophila.” Antimicrob Agents Chemother, 51(6). 2136-2142. Jun. 2007.
Supplementary Data

Table S1 Structures of the 262 IMP-1 inhibitors (in SMILE format) used in the study with their experimental bioactivities (IC50 or Ki in μM)

| Index | Molecule | Activity (IC50 or Ki in μM) |
|-------|----------|-----------------------------|
| 1  | s1eCc(SC=2(C(3N(=O)C(OC)OC)=2(O)=OC)c2ccc1c2ccc2) | 0.70 |
| 2  | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 10 |
| 3  | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 10 |
| 4  | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 2.40 |
| 5  | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 10 |
| 6  | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.10 |
| 7  | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.1 |
| 8  | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.22 |
| 9  | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.22 |
| 10 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.15 |
| 11 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 1.40 |
| 12 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 1.80 |
| 13 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.10 |
| 14 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.40 |
| 15 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 5.00 |
| 16 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.70 |
| 17 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.1 |
| 18 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.40 |
| 19 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 5.00 |
| 20 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 6.20 |
| 21 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 9.00 |
| 22 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 9.00 |
| 23 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 10 |
| 24 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.1 |
| 25 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.1 |
| 26 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.80 |
| 27 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.80 |
| 28 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.1 |
| 29 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.1 |
| 30 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 0.1 |
| 31 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 32 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 33 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 34 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 35 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 36 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 37 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 38 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 39 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 40 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 200 |
| 41 | s1eCcSc1C3C(=O)C(OC)OC)c1C(=O)Oc2ccc1c2ccc2) | 65 |
Table S2. Molecular Descriptors used in this work

| Descriptor class | Number of descriptors in each class | Descriptors |
|------------------|-------------------------------------|-------------|
| Physical Properties | 15 | apol, bpol, density, FCharge, mr, SMR, Weight, logP(o/w), logS, rsynth, SlogP, TPSA, vdw_vol, vdw_area, reactive |
| Subdivided Surface Areas                      | 18 | SlogP_VSA0, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA7, SlogP_VSA8, SlogP_VSA9, SMR_VSA0, SMR_VSA1, SMR_VSA2, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7 |
|---------------------------------------------|----|--------------------------------------------------------------------------------|
| Atom Counts and Bond Counts                 | 36 | a_aro, a_count, a_heavy, a_ICM, a_IC, a_nH, a_nC, a_nN, a_nO, a_nF, a_nS, a_nI, a_nCl, b_1rotN, b_1rotR, b_ar, b_count, b_double, b_heavy, b_rotR, b_rotN, b_single, chiral, chiral_u, lip_acc, lip_don, lip_drglike, lip_violation, opr_brigid, opr_leadlike, opr_nring, opr_rotR, opr_violation, rings, VAdjMa, VAdjEq |
| Kier&Hall Connectivity and Kappa Shape Indices | 16 | chi0, chi0_C, chi1, chi1_C, chi0v, chi0v_C, chi1_v, chi1v_C, Kier1, Kier2, Kier3, KierA1, KierA2, KierA3, KierFlex, zagreb |
| Adjacency and Distance Matrix Descriptors   | 33 | balabanJt, BCUT_PEOE_0, BCUT_PEOE_1, BCUT_PEOE_2, BCUT_PEOE_3, BCUT_SLOGP_0, BCUT_SLOGP_1, BCUT_SLOGP_2, BCUT_SLOGP_3, BCUT_SMIR_0, BCUT_SMIR_1, BCUT_SMIR_2, BCUT_SMIR_3, diameter, petitjean, GCUT_PEOE_0, GCUT_PEOE_1, GCUT_PEOE_2, GCUT_PEOE_3, GCUT_SLOGP_0, GCUT_SLOGP_1, GCUT_SLOGP_2, GCUT_SLOGP_3, GCUT_SMIR_0, GCUT_SMIR_1, GCUT_SMIR_2, GCUT_SMIR_3, petitjeansC, radius, VDistEq, VDistMa, wienerPath, wienerPol |
| Pharmacophore Feature Descriptors          | 10 | a_acc, a_base, a_don, a_hyd, vsa_acc, vsa_don, vsa_hyd, vsa_other, vsa_pol, vsa_base |
| Partial Charge Descriptors                  | 32 | PEOE_PC+, PEOE_PC-, PEOE_RPC+, PEOE_RPC-, Q_VSA_POS, PEOE_VSA_POS, PEOE_VSA_NEG, PEOE_VSA_PPOS, PEOE_VSA_PNEG, Q_VSA_HYD, PEOE_VSA_HYD, PEOE_VSA_POL, PEOE_VSA_FPOS, PEOE_VSA_FNEG, PEOE_VSA_FPPOS, PEOE_VSA_FPNEG, PEOE_VSA_FHYD, PEOE_VSA_FPOL, PEOE_VSA_F-6, PEOE_VSA_F-5, PEOE_VSA_F-4, PEOE_VSA_F-3, PEOE_VSA_F-2, PEOE_VSA_F-1, PEOE_VSA_F-0, PEOE_VSA_F-0, PEOE_VSA_F-1, PEOE_VSA_F-2, PEOE_VSA_F-3, PEOE_VSA_F-4, PEOE_VSA_F-5, PEOE_VSA_F-6 |
| Potential Energy Descriptors                | 10 | E, E_ang, E_ele, E_nb, E_sol, E_stb, E_str, E_strain, E_tor, E_vdw |
| Surface Area, Volume and Shape Descriptors  | 86 | ASA, dens, pmI, pmI_X, pmI_Y, pmI_1, pmI_2, pmI_3, prp1, prp2, rgyr, std_dim1, std_dim2, vol, VSA, vsurf_V, vsurf_S, vsurf_R, vsurf_G, vsurf_W*(8 descriptors), vsurf_IW*(8 descriptors), vsurf_CW*(8 descriptors), vsurf_EWmin*(3 descriptors), vsurf_DW*(3 descriptors), vsurf_EDmin*(3 descriptors), vsurf_DD*(3 descriptors), vsurf_HL*(2 descriptors), vsurf_A, vsurf_CP, vsurf_WP*(3 descriptors), vsurf_HB*(8 descriptors) |
| Conformation Dependent Charge Descriptors   | 1  | ASA_H |