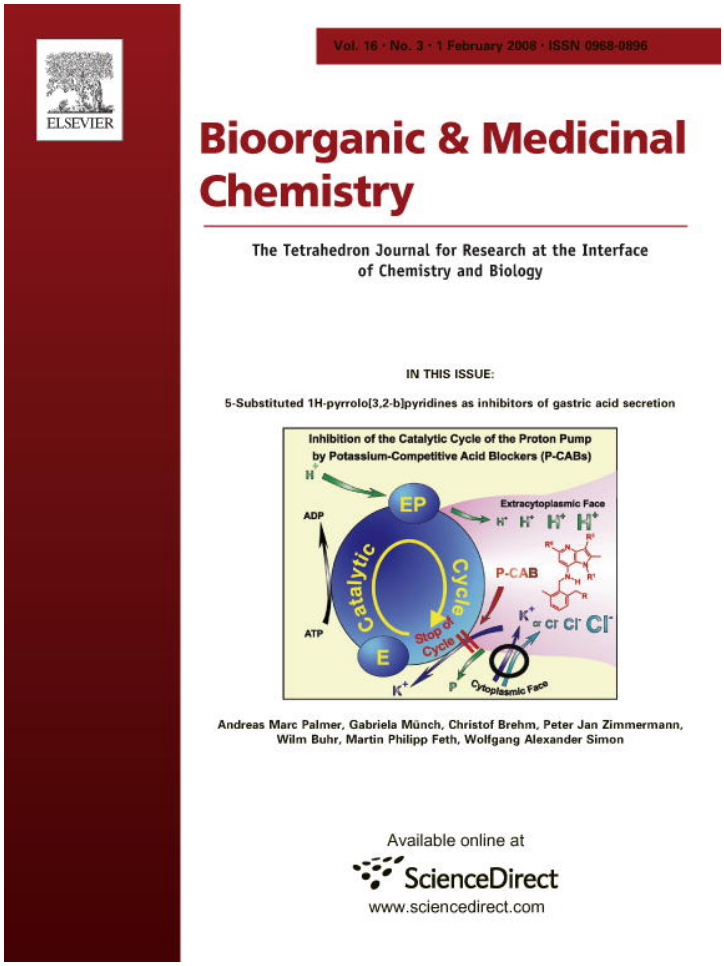


Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry
Bioorganic & Medicinal Chemistry 16 (2008) 1218–1235

Discovery of new MurF inhibitors via pharmacophore modeling and QSAR analysis followed by in-silico screening

Mutasem O. Taha,[a,*] Naji Atallah,[a] Amal G. Al-Bakri,[b] Catherine Paradis-Bleau,[d] Hiba Zalloum,[a] Khaled S. Younis[c] and Roger C. Levesque[d]

[a]*Department of Pharmaceutical Sciences, Faculty of Pharmacy, University of Jordan, Amman, Jordan*
[b]*Department of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmacy, University of Jordan, Amman, Jordan*
[c]*Department of Computer Engineering, Faculty of Engineering, University of Jordan, Amman, Jordan*
[d]*Microbiologie moléculaire et génie des proteins Faculíe de Médecine, Universíe Laval Que., Canada G1K 7P4*

Received 17 May 2007; revised 16 October 2007; accepted 23 October 2007
Available online 7 November 2007

Abstract—The pharmacophoric space of streptococcal MurF was explored using a set of 39 known inhibitors. Subsequently, genetic algorithm and multiple linear regression analysis were employed to select an optimal combination of pharmacophoric models and physicochemical descriptors that access self-consistent quantitative structure–activity relationship (QSAR) ($r^2 = 0.93, \mathrm{F} = 56.9, r^2_{\mathrm{LOO}} = 0.91, r^2_{\mathrm{PRESS}}$ against eight external test inhibitors = 0.75). Two orthogonal pharmacophores (of cross-correlation $r^2 = 0.26$) emerged in the QSAR equation suggesting the existence of at least two distinct binding modes accessible to ligands within MurF binding pocket. The validity of the QSAR equation and the associated pharmacophore models was experimentally established by the identification of three promising new MurF inhibitors retrieved from the NCI database. Docking studies conducted on active hits supported the binding modes suggested by the pharmacophore/QSAR analysis.

1. Introduction

The emergence of bacterial resistance is a major challenge to the antibacterial research community and clinicians worldwide. Among the widely prescribed antibiotics, resistance to β-lactams and macrolides is particularly high.[1,2] However, resistance to other antibacterial agents, for example, fluoroquinolones, is expected to increase with increased usage and over-prescription.[3,4] This situation has created an urgent need for the development of new antibacterial agents directed towards novel targets.[5]

One of the best known and most validated targets for antibacterial therapy is the machinery for peptidoglycan biosynthesis, an essential component of bacterial cell wall.[6–8] The biosynthesis of peptidoglycan is a complex process involving several steps.[6,9] Despite the existence of large number of antibacterial antibiotics that act by inhibiting the later steps in peptidoglycan biosynthesis, for example, β-lactams and glycopeptides, the earlier steps of making cytoplasmic peptidoglycan precursors are poorly exploited as antibacterial targets. None of the enzymes involved in these steps (also known as Mur enzymes) is inhibited by a known antibiotic or synthetic chemical of therapeutic usefulness, except for MurA, which is inhibited by fosfomycin (IC_{50} = 15 μM).[6,13]

The importance of Mur enzymes as valid targets is supported by the fact that their genes are essential for bacterial survival; their inactivation gives a lethal phenotype.[6,10–13] Furthermore, these proteins are highly conserved among various bacterial species, which should permit the development of broad spectrum bactericidal agents.[6] A particularly important enzyme of this family is the D-Ala-D-Ala adding enzyme MurF, which catalyzes the ATP-dependent formation of the UDP-*N*-acetylmuramic acid pentapeptide, an early critical component of the bacterial cell wall biosynthesis. Inhibition of this enzyme has the potential for the design of new wide spectrum antibacterial agents.[6] To date, this target received only limited medicinal chemistry attention.[6,14–18]

The main focus of current efforts toward the development of new MurF inhibitors concentrate on structure-based ligand design, particularly due to the recent

Keywords: MurF; Pharmacophore modeling; QSAR; In-silico screening; Experimental validation.
* Corresponding author. Tel.: +962 6 5355000x23305; fax: +962 6 5339649; e-mail: mutasem@ju.edu.jo

0968-0896/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2007.10.076

availability of two crystallographic structures, albeit at low resolutions (2.5 and 2.8 Å).[17] Still, all reported MurF inhibitors were found to be devoid of antibacterial activities, probably due to their poor penetrability into the bacterial cells.[18]

Although considered the most reliable structural information that can be used for drug design, crystallographic structures are limited by inadequate resolution[19] and crystallization-related artifacts of the ligand–protein complex.[20] Furthermore, crystallographic structures generally ignore structural heterogeneity related to protein anisotropic motion and discrete conformational substrates.[21] These factors may have contributed to the limited success in developing new MurF inhibitors through structure-based design efforts.

The recent interest in designing new antibacterial agents based on MurF inhibition, combined with the drawbacks of structure-based design prompted us to explore the possibility of developing ligand-based three-dimensional (3D) pharmacophore(s) integrated within self-consistent QSAR model. This approach avoids the pitfalls of structure-based techniques; furthermore, the pharmacophore model(s) can be used as 3D search query(ies) to mine 3D libraries for new MurF inhibitors.

We employed the HYPOGEN module from the CATALYST software package[22] to construct plausible binding hypotheses for MurF inhibitors. Subsequently, genetic function algorithm (GFA) and multiple linear regression (MLR) analyses were employed to search for an optimal QSAR that combines high-quality binding pharmacophores with other molecular descriptors and is capable of explaining bioactivity variation across a collection of diverse MurF inhibitors. The optimal pharmacophores were subsequently used as 3D search queries to screen the NCI compounds database for new MurF inhibitors.

CATALYST models drug–receptor interaction using information derived only from the drug structure.[22–24,31–33,35–37] HYPOGEN identifies a 3D array of a maximum of five chemical features common to active training molecules, which provides a relative alignment for each input molecule consistent with their binding to a proposed common receptor site. The chemical features considered can be hydrogen-bond donors and acceptors (HBDs and HBAs), aliphatic and aromatic hydrophobes, positive and negative charges, positive and negative ionizable groups and aromatic planes. The conformational flexibility of training ligands is modeled by creating multiple conformers, judiciously prepared to emphasize representative coverage over a specified energy range. CATALYST pharmacophores have been used as 3D queries for database searching and in 3D-QSAR studies.[25–33]

2. Results and discussion

2.1. Data mining and conformational coverage

The literature was extensively surveyed to collect as many structurally diverse MurF inhibitors as possible. Unfortunately, the number of publications that deal with this enzyme is rather limited restricting us with 39 inhibitors only (**1**–**39**, Table 1 and Fig. 1).[15,18] The 2D structures of the training inhibitors were imported into CATALYST and converted automatically into plausible 3D single conformer representations via the rule-based methods implemented within this package. The resulting single conformer 3D structures were later used as starting point for conformational analysis and in the determination of various molecular descriptors for QSAR modeling.

The conformational space of each inhibitor was extensively sampled utilizing the poling algorithm employed within CATALYST. Poling promotes conformational variation via employing molecular mechanical force field algorithm that penalizes similar conformers.[31–33] Conformational coverage was performed employing the 'Best' module to ensure extensive sampling of conformational space. Efficient conformational coverage guarantees minimum conformation-related noise during pharmacophore generation and validation stages. Pharmacophore generation and pharmacophore-based search procedures are known for their sensitivity to inadequate conformational sampling within the training compounds.[34]

Interestingly, despite the apparent 2D similarity among the training compounds, they differed significantly in their conformational spaces, which is considered promising in pharmacophore modeling.

2.2. Pharmacophore modeling

CATALYST–HYPOGEN enables automatic pharmacophore construction by using a collection of at least 16 molecules with bioactivities spanning over four orders of magnitude.[22–24,32,33,35,36] HYPOGEN implements an optimization algorithm that evaluates large number of potential models within the pharmacophoric 'space' of a particular target through fine perturbations to pharmacophore hypotheses that survived the subtractive and constructive phases (see Section 4.1.4 in Section 4).[32] The extent of the evaluated space is reflected by the configuration (config.) cost calculated for each modeling run. It is generally recommended that the config. cost of any HYPOGEN run does not exceed 17 (corresponding to 2^{17} hypotheses to be assessed by CATALYST), otherwise a thorough analysis of all models cannot be guaranteed.[33]

The size of the investigated pharmacophoric space is a function of training compounds, selected input chemical features, and other CATALYST control parameters such as feature tolerances and weights.[33] Restricting the extent of explored pharmacophoric space might improve the efficiency of optimization via allowing effective evaluation of limited number of pharmacophoric models. On the other hand, severe restrictions imposed on the pharmacophoric space might reduce the possibility of discovering optimal pharmacophoric hypotheses, particularly if they occur outside the 'boundaries' of the pharmacophoric space.

Therefore, it was decided to explore the pharmacophoric space of MurF inhibitors under reasonably imposed

Table 1. The structures of MurF inhibitors utilized in modeling[18]

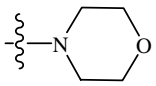
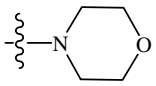
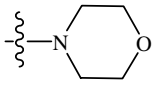
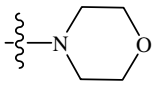
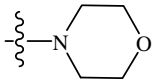
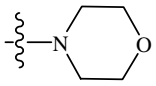
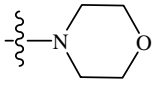
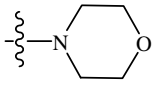
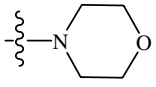
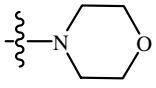
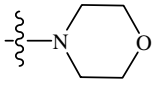
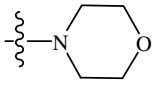
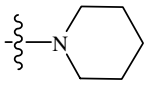
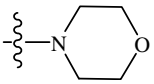
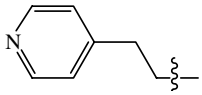
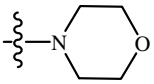
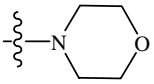
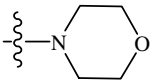
Compound	R1	R2	R3	R4	IC_{50} (μM)
1		Cl	H	CN	9.00
2[a]	–$NHCH_3$	Cl	H	CN	15.00
3	–NHPhenyl	Cl	H	CN	26.00
4	–$NH(CH_2)_3OH$	Cl	H	CN	8.00
5	—	Cl	H	CN	>100.00
6		Cl	F	CN	6.40
7	–$N(CH_2CH3)_2$	Cl	Cl	CN	5.20
8		Cl	Cl	CN	0.30
9		Cl	–$NHCH_3$	CN	62.00
10		Cl	–$NH(CH_2)_3OH$	CN	>100.00
11		Cl	–$NH(CH_2)_3N(CH_3)_2$	CN	>100.00
12	–$N(CH_2CH3)_2$	Cl	H	–$CONH_2$	66.00
13	–$N(CH_2CH3)_2$	Cl	H	–CH_2NH_2	>100.00
14[a]	–$N(CH_2CH3)_2$	Cl	H	–$CH_2NHCOCH_3$	>100.00
15	–$N(CH_2CH3)_2$	Cl	H	–$COOCH_2CH_3$	>100.00
16[a]		–$NH(CH_2)_3N(CH_3)_2$	H	CN	74.00
17		–$N(CH_3)_2$	H	CN	>100.00
18		–OCH_2CH_3	H	CN	>100.00
19		–OCH_2CHCH_3Phenyl	H	CN	>100.00
20		H	H	CN	>100.00
21		Br	H	CN	4.20
22	–$NH(CH_2)_4OH$	Cl	H	CN	22.00
23		Cl	H	CN	2.50
24	–NH_2	Cl	H	CN	78.00

Table 1 (*continued*)

Compound	R1	R2	R3	R4	IC_{50} (μM)
25[a]	$-N(CH_2CH_3)_2$	H	—	—	151.00
26	$-N(CH_2CH_3)_2$	Br	—	—	12.60
27			—	—	6.00
28[a]	C=O	$-CH_3$	—	—	>100.00
29	$-CH_2-$	H	—	—	>100.00
30	$-CH(CH_3)$	H	—	—	>100.00
31	$-N(CH_2CH_3)_2$	Cl	H	H	1.40
32		Cl	H	H	1.70
33[a]		Cl	Cl	–Phenyl	0.070
34	$-N(CH_2CH_3)_2$	Cl	H	–*p*-Phenol	0.054
35[a]	$-N(CH_2CH_3)_2$	Cl	H	$-COOCH_2CH_3$	6.00
36	$-N(CH_2CH_3)_2$	Cl	Cl	–*p*-Phenol	0.067
37		Cl	Cl	–*p*-Phenol	0.022
38[a]	$-(CH_2CH_3)_2$	—	—	—	3.40
39		—	—	—	4.20

The corresponding scaffolds are in Figure 1.

[a] These compounds were employed as the external testing subset in QSAR modeling.

'boundaries' through four HYPOGEN runs (Table 2). Accordingly, the software was restricted to explore pharmacophoric models incorporating from zero to three features of any particular selected feature type (i.e., HBA, HBD, hydrophobic and RingArom), i.e., instead of the default range of zero to five. Furthermore, only four- and five-featured pharmacophores were explored, i.e., models of lesser number of features were ignored. The later restriction has the dual advantage of narrowing the investigated pharmacophoric space and best-representing the feature-rich nature of MurF binding pocket. CATALYST–HYPOGEN can produce pharmacophore hypotheses of a maximum of five features.[32,33]

Clearly from Table 2, we allowed a maximum of 4–5 features per pharmacophore in runs 1 and 3; however, runs 2 and 4 were limited to five-featured pharmacophores only. Despite their bias toward five-featured hypotheses, the two approaches should explore different sections of the pharmacophoric space of MurF inhibitors as reflected by the diversity of the resulting models (see Supplementary figures) and their success criteria, in particular their config. costs (Table 3).

The input features were selected in agreement with published SAR studies and crystallographic data. For example, the fact that crystallographic information suggested the involvement of asparagine moieties (i.e., ASN326 and ASN328) in hydrogen-bonding with some ligands[15] prompted us to select hydrogen-bond acceptor and donor (HBA and HBD) functionalities as possible pharmacophoric features. In the same manner, the following features were fed into HYPOGEN: aromatic rings (RingArom) and hydrophobes.

To evaluate the optimal pharmacophoric inter-feature distances, we explored the pharmacophoric space of 10 and 300 pm, corresponding to 'narrow' and 'wide' inter-feature distances, respectively (Table 2).

Eventually, 10 optimal pharmacophoric hypotheses were generated for each run, yielding 40 models from four automatic runs. In each run, the resulting models were automatically ranked according to their corresponding 'total cost' values, which is defined as the sum of error cost, weight cost, and configuration cost (see Section 4.1.5 in Section 4).[22,23,32,33,35–37]

Table 3 shows the pharmacophoric features and success criteria of the generated pharmacophores. Clearly from the table, all models shared comparable features and satisfactory success criteria; in particular they illustrated

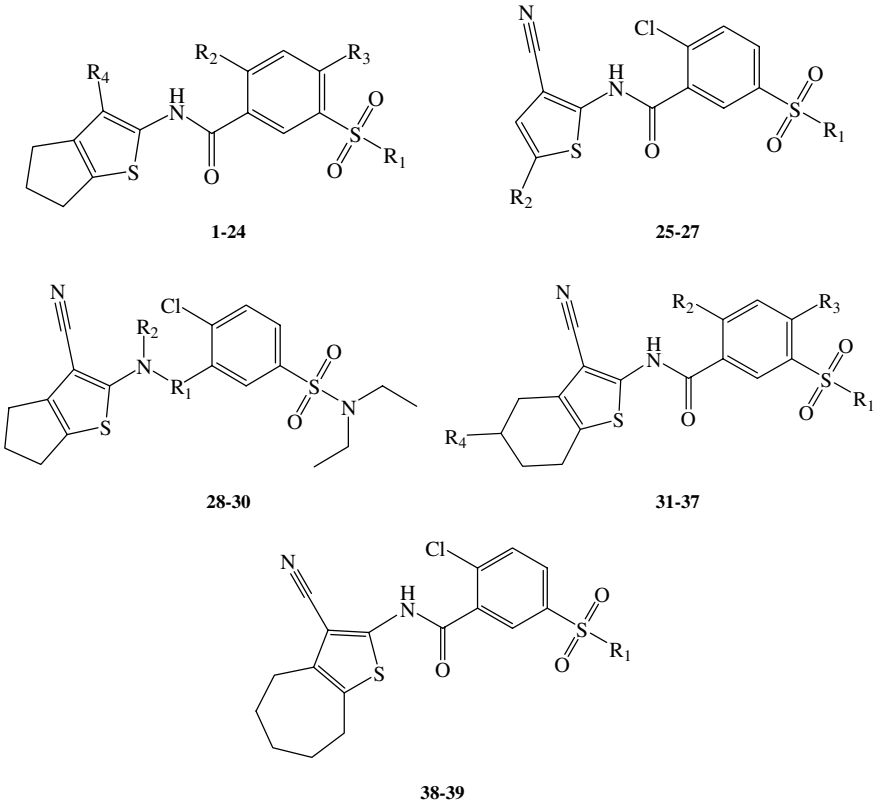


Figure 1. The chemical scaffolds of training compounds, the side chains are as in Table 1.

Table 2. Selected input features and CATALYST run parameters employed in exploring MurF inhibitors pharmacophoric space

Run	Selected input features: types and ranges[a]	Other run parameters[b,c]
1	Hbic (0–3), HBA (0–3), HBD (0–3), RingArom (0–3)	Max – Min: 5–4, spacing: 10
2	Hbic (0–3), HBA (0–3), HBD (0–3), RingArom (0–3)	Max – Min: 5–5, spacing: 10
3	Hbic (0–3), HBA (0–3), HBD (0–3), RingArom (0–3)	Max – Min: 5–4, spacing: 300
4	Hbic (0–3), HBA (0–3), HBD (0–3), RingArom (0–3)	Max – Min: 5–5, spacing: 300

[a] HBA, hydrogen-bond acceptor; HBD, hydrogen-bond donor; Hbic, hydrophobic feature; RingArom, ring aromatic feature. The allowed ranges of input features are in brackets.
[b] Min – Max, allowed minimum and maximum number of output features; Spacing, allowed minimal inter-feature distances in picometer.
[c] Unmentioned parameters were set to their default values.

acceptable Fischer confidence levels (85–95%).[37,38] However, despite our restrictions on the extent of evaluated pharmacophore space, the config. costs of the runs were generally high and exceeded the maximum limit of 17. This conduct is probably related to the pronounced flexibilities of the training compounds, which seem to increase the entropy of the hypotheses space. Nevertheless, the reasonable cost and confidence criteria of the resulting pharmacophore models should overshadow any drawbacks related to the less than optimal config. costs.

The resulting pharmacophoric hypotheses are shown in Supplementary figures. The figures also show how the models map compound **37** (the most potent inhibitor). Clearly from these figures, the resulting pharmacophores are quite diverse *vis-à-vis* feature types and spatial arrangements. However, two features seem to emerge frequently among the resulting pharmacophores: hydrophobic (occasionally replaced by ring-aromatic) and hydrogen-bond acceptor features mapping the terminal phenol ring and the sulfonyl oxygens of **37**, respectively. This is probably related to the fact that potent training inhibitors (e.g., **33**, **34**, **36**, and **37**) uniquely exhibit terminal phenol ring and *meta*-dichlorobenzene linker. The later seems to push the sulfone oxygens in a certain pharmacophorically crucial orientation allowing them to be picked by CATALYST–HYPOGEN as hydrogen-bond acceptor.

The emergence of numerous high-quality pharmacophore models suggest that MurF ligands assume multiple pharmacophoric binding modes within the binding pocket. Therefore, it is quite challenging to select any particular pharmacophore hypothesis as a sole representative of the ligand–protein binding process.

Table 3. The performance of different pharmacophoric hypotheses generated for MurF inhibitors employing different parameters (as in Table 2)

Run[a]	Hypotheses[b]	Pharmacophoric features in generated hypotheses	Cost values					Cost of null hypothesis	Residual cost[e]	Fischer confidence[f]
			R[c]	Error	Weight	Config.[d]	Total			
1	1	2 × HBA, 2 × HBic	0.92	95.9	3.6	20.4	119.9	155.4	35.5	95%
	2	2 × HBA, 2 × HBic	0.92	95.4	4.4	20.4	120.2	155.4	35.2	95%
	3	HBA,3 × HBic, HBD	0.89	98.7	1.2	20.4	120.2	155.4	35.2	95%
	4[g]	2 × HBA, 3 × HBic	**0.89**	**98.7**	**1.3**	**20.4**	**120.4**	**155.4**	**35**	**95%**
	5	HBA, 3 × HBic, HBD	0.90	97.4	2.7	20.4	120.5	155.4	34.9	95%
	6	HBA, 2 × HBic, HBD	0.90	97.9	2.4	20.4	120.7	155.4	34.7	95%
	7	2 × HBA, HBic, HBD	0.90	97.9	2.5	20.4	120.8	155.4	34.6	95%
	8	2 × HBA, HBic, RingArom	0.90	97.4	3.1	20.4	120.9	155.4	34.5	95%
	9	2 × HBA, 3 × HBic	0.89	98.6	1.9	20.4	120.9	155.4	34.5	95%
	10	HBA, 3 × HBic, RingArom	0.90	97.5	3.1	20.4	120.9	155.4	34.5	95%
2	1	HBA, 3 × HBic, RingArom	0.94	93.2	3.1	18.7	115.0	155.4	40.4	95%
	2	HBA, 3 × HBic, HBD	0.89	98.7	1.2	18.7	118.6	155.4	36.8	95%
	3	HBA, 3 × HBic, RingArom	0.89	98.1	1.9	18.7	118.8	155.4	36.6	95%
	4	HBA, 3 × HBic, HBD	0.90	97.7	2.5	18.7	118.8	155.4	36.6	95%
	5	2 × HBA, 3 × HBic	0.89	99.0	1.2	18.7	118.9	155.4	36.5	95%
	6	2 × HBA, 3 × HBic	0.89	98.4	1.8	18.7	118.8	155.4	36.6	95%
	7	HBD, 3 × HBic, RingArom	0.88	99.1	1.1	18.7	118.9	155.4	36.5	95%
	8	HBD, 3 × HBic, RingArom	0.88	99.1	1.2	18.7	118.9	155.4	36.5	95%
	9	HBD, 3 × HBic, RingArom	0.90	97.3	3.1	18.7	119.0	155.4	36.4	95%
	10	HBA, 3 × HBic, RingArom	0.90	97.7	2.8	18.7	119.2	155.4	36.2	95%
3	1	2 × HBA, 3 × HBic	0.91	96.2	2.9	19.1	118.2	155.4	37.2	95%
	2	2 × HBA, HBic, RingArom	0.91	96.4	3.0	19.1	118.5	155.4	36.9	95%
	3	HBA, 3 × HBic, RingArom	0.90	97.5	2.7	19.1	119.3	155.4	36.1	85%
	4	2 × HBA, 3 × HBic	0.89	98.8	1.6	19.1	119.5	155.4	35.9	85%
	5	HBA, 2 × HBic, RingArom	0.90	97.6	3.1	19.1	119.8	155.4	35.6	85%
	6	HBA, HBic, 2 × RingArom	0.89	97.9	2.9	19.1	119.9	155.4	35.5	85%
	7	HBA, HBic, 2 × RingArom	0.90	97.7	3.3	19.1	120.0	155.4	35.4	85%
	8[g]	3 × HBA, HBic	**0.89**	**98.0**	**2.9**	**19.1**	**120.1**	**155.4**	**35.3**	**85%**
	9	HBA, HBic, 2 × RingArom	0.90	97.7	3.3	19.1	120.2	155.4	35.2	85%
	10	2 × HBA, HBic, RingArom	0.89	98.1	3.0	19.1	120.2	155.4	35.2	85%
4	1	HBA, 3 × HBic, RingArom	0.92	95.6	3.2	16.8	115.6	155.4	39.8	95%
	2	HBA, 3 × HBic, HBD	0.91	96.2	2.6	16.8	115.6	155.4	39.8	95%
	3	HBA, 3 × HBic, RingArom	0.92	96.0	3.7	16.8	116.5	155.4	38.9	95%
	4	HBA, 3 × HBic, RingArom	0.91	96.4	3.4	16.8	116.6	155.4	38.8	95%
	5	HBA, 3 × HBic, RingArom	0.91	96.7	3.2	16.8	116.7	155.4	38.7	95%
	6	2 × HBA, 3 × HBic	0.89	98.5	1.7	16.8	117.1	155.4	38.3	95%
	7	2 × HBA, 3 × HBic	0.91	96.8	3.6	16.8	117.2	155.4	38.2	95%
	8	HBA, 3 × HBic, RingArom	0.89	98.1	2.3	16.8	117.3	155.4	38.1	95%
	9	2 × HBA, 3 × HBic	0.90	97.3	3.1	16.8	117.3	155.4	38.1	95%
	10	2 × HBA, 3 × HBic	0.88	99.0	1.6	16.8	117.5	155.4	37.9	95%

[a] The automatic pharmacophore run employing the parameters and conditions in Table 2.
[b] MurF inhibition hypotheses from the corresponding runs. The models are ranked by CATALYST–HYPOGEN according to their cost criteria.
[c] The correlation coefficients between the bioactivity estimates and the bioactivities of the training compounds.
[d] Configuration costs.
[e] The difference between the total cost and the cost of the corresponding null hypotheses.
[f] Calculated employing the Cat. Scramble methods.[22,23,40]
[g] These pharmacophores appeared in the best QSAR equations (bolded).

2.3. QSAR modeling

Despite the undisputed significance of pharmacophoric hypotheses in understanding ligand-macromolecule affinity and as 3D search queries, their predictive value as 3D-QSAR models is generally limited by steric shielding and bioactivity-enhancing or -reducing auxiliary groups.[36] This point combined with the fact that our training list of MurF inhibitors furnished numerous reasonable binding hypotheses (Table 3) prompted us to employ classical QSAR analysis to search for the best combination of orthogonal pharmacophores and other structural descriptors (connectivity, topological, 2D, etc.) capable of explaining bioactivity variation across the training compounds. This task was performed via genetic function approximation-multiple linear regression-QSAR (GFA-MLR-QSAR) analysis (see Section 4.1.6 in Section 4).

Table 4 shows the statistical criteria of the resulting top-ranking QSAR models. Model B seems to have the best overall statistical qualities. Equation 1 shows this model, the 95% confidence limits are shown in brackets [±CL]. Figure 2 shows the corresponding scatter plots of

Table 4. The statistical results of the scanned QSAR models

Model	Terms[a]	$r^2_{(31)}$ [b]	*F*-value	$r^2_{(LOO)}$ [c]	$r^2_{(BS)}$ [d]	$r^2_{(PRESS)}$ [e]	PRESS[f]	Descriptors[g]
A	8	0.944	55.45	0.894	0.914	0.683	2.837	$^3\chi^v_P$, S_tN, S_sF, $^0\chi^v$, LUMO, Hypo3/8, Hypo1/4
B[h]	**7**	0.934	56.9	0.905	0.934	0.748	2.259	$^3\chi^v_P$, S_tN, LUMO, κ_1, Hypo3/8, Hypo1/4
C	6	0.918	55.96	0.893	0.918	0.697	2.711	$^3\chi^v_P$, LUMO, $\kappa_{1\alpha}$, Hypo3/8, Hypo1/4
D	5	0.887	50.86	0.840	0.887	0.639	3.233	$^3\chi^v_P$, $^3\chi_c$, LUMO, Hypo2/7
E	4	0.853	52.38	0.818	0.853	0.808	1.720	$^3\chi^v_P$, LUMO, Hypo2/7
F	3	0.795	34.90	0.752	0.795	—[i]	12.751	LUMO, Hypo4/8, Hypo2/5

[a] Number of explanatory terms including the intercept.
[b] Non-cross-validated correlation coefficient for 31 training compounds.
[c] Cross-validation correlation coefficients determined by the leave-one-out technique.
[d] Bootstrapping correlation coefficient.
[e] Predictive r^2 determined for the eight test compounds.
[f] The sum of squared deviations between predicted and actual activity values for every molecule in the test set of eight compounds.
[g] The descriptors in each successful QSAR model. LUMO is the energy of the lowest unoccupied molecular orbital, $\kappa 1$ is the first-order kappa shape index, Hypo3/8, Hypo1/4, Hypo2/7, Hypo4/8, and Hypo2/5 represent the fit values of the training compounds against the 8th, 4th, 7th, 8th, and 5th pharmacophores from the 3rd, 1st, 2nd, 4th, and 2nd modeling runs, respectively (Table 3), S_tN and S_sF are the electrotopological sum descriptors for tertiary nitrogen and fluorine atoms, respectively, $^3\chi^v_P$, $^0\chi^v$, $^3\chi_c$ are the third-order valence path, zero-order valence, and third-order cluster connectivity indices, respectively.[39]
[h] This QSAR equation was selected to predict the MurF inhibitory activities of the captured hits as it yielded the best statistical criteria.
[i] The r^2_{PRESS} value is negative.

experimental versus estimated bioactivities for the training and testing inhibitors.

$$\begin{aligned}\mathrm{Log}(1/\mathrm{IC}_{50}) &= -6.3[\pm 2.7] - 2.8[\pm 0.70]\mathrm{LUMO} - 0.3[\pm 0.10]\kappa_1 \\ &\quad + 0.14[\pm 0.10]\mathrm{Hypo3/8} + 0.19[\pm 0.10]\mathrm{Hypo1/4} \\ &\quad + 0.06[\pm 0.04]\mathrm{S\_tN} + 0.93[\pm 0.30]^3\chi^v_P \\ r^2_{31} &= 0.93, F\text{-statistic} = 56.9,\ r^2_{\mathrm{BS}} = 0.93, \\ r^2_{\mathrm{LOO}} &= 0.91,\ r^2_{\mathrm{PRESS}} = 0.748, \end{aligned} \quad (1)$$

where r^2_{31} is the correlation coefficient, r^2_{LOO} is the leave-one-out correlation coefficient, r^2_{BS} is the bootstrapping regression coefficient and r^2_{PRESS} is the predictive r^2 determined for the eight test compounds.[39–41] LUMO is the energy of the lowest unoccupied molecular orbital (calculated using single point MOPAC(AM1)-based semi-empirical quantum mechanical methods), κ_1 is the first-order kappa shape index, Hypo3/8 and Hypo1/4 represent the fit values of the training compounds against the 8th and 4th pharmacophores from the 3rd and 1st modeling runs, respectively (Fig. 3 and Tables 3 and 5), S_tN is the electrotopological sum descriptor for tertiary nitrogen atoms, $^3\chi^v_P$ is the third-order valence path connectivity index.[39]

Interestingly, the combination of Hypo1/4 and Hypo3/8 frequented in three out of the six highest-ranking QSAR equations, that is, models A, B and C in Table 4, which further supports their statistical significance. Emergence of two orthogonal pharmacophoric models (Hypo1/4 and Hypo3/8, cross-correlation $r^2 = 0.26$) in the highest-ranking models, including Equation 1, suggests they represent two complementary binding modes accessible to ligands within the binding pocket of MurF, that is, one of the pharmacophores can optimally explain the bioactivities of some training inhibitors, while other inhibitors are more appropriately explained by the second pharmacophore.[26]

Figure 3 shows the two pharmacophores and how they map to the most potent inhibitor **37** (IC_{50} = 0.022 μM), while Table 5 shows the *X*, *Y*, and *Z* coordinates of the two pharmacophores.

Emergence of topological and shape descriptors (i.e., $\kappa 1$, S_tN, and $^3\chi^v_P$) in Equation 1 illustrates certain role played by the ligands' topology in the binding process. However, despite the predictive significance of these descriptors, their information content is quite obscure. On the other hand, emergence of LUMO in Equation 1, in association with a negative regression coefficient, indicates a direct relationship between ligand/MurF affinity and ligands' electrophilicities. This trend can be clearly seen from the activity-enhancing effects of halogen and cyano substituents at the sulfonated aromatic ring (R2 and R3 in compounds **1**–**24** and **32**–**37**, Table 1 and Fig. 1), which is probably explainable by stacking against the aromatic ring of PHE31 in the binding pocket (see Section 2.6 below).

2.4. In-silico screening of NCI database and subsequent experimental evaluation

Despite the proposition that orthogonal pharmacophores complement each other in explaining bioactivity variations across the training compounds therefore represent different binding modes within the binding pocket, we believe that potent inhibitors can assume several high-affinity binding modes within the binding pocket, i.e., orthogonal pharmacophores should converge at high-affinity ligands. This is simply because all pharmacophores emerging in a QSAR model contribute to bioactivity. This theory is supported by the fact that upon screening **1**–**39** (Table 1 and Fig. 1) against the two pharmacophores, Hypo3/8 captured all listed compounds except **13**, **24**, **29**, and **30**, while Hypo1/4 captured only 6 compounds, namely, **14**, **27**, **34**, **35**, **36**, and **37**. This significant difference in their hit lists further supports the orthogonality of the two models, while their agreement on selecting potent inhibitors, namely, **27**, **34**, **35**, **36** and **37**, suggests their convergence at high-affinity ligands.

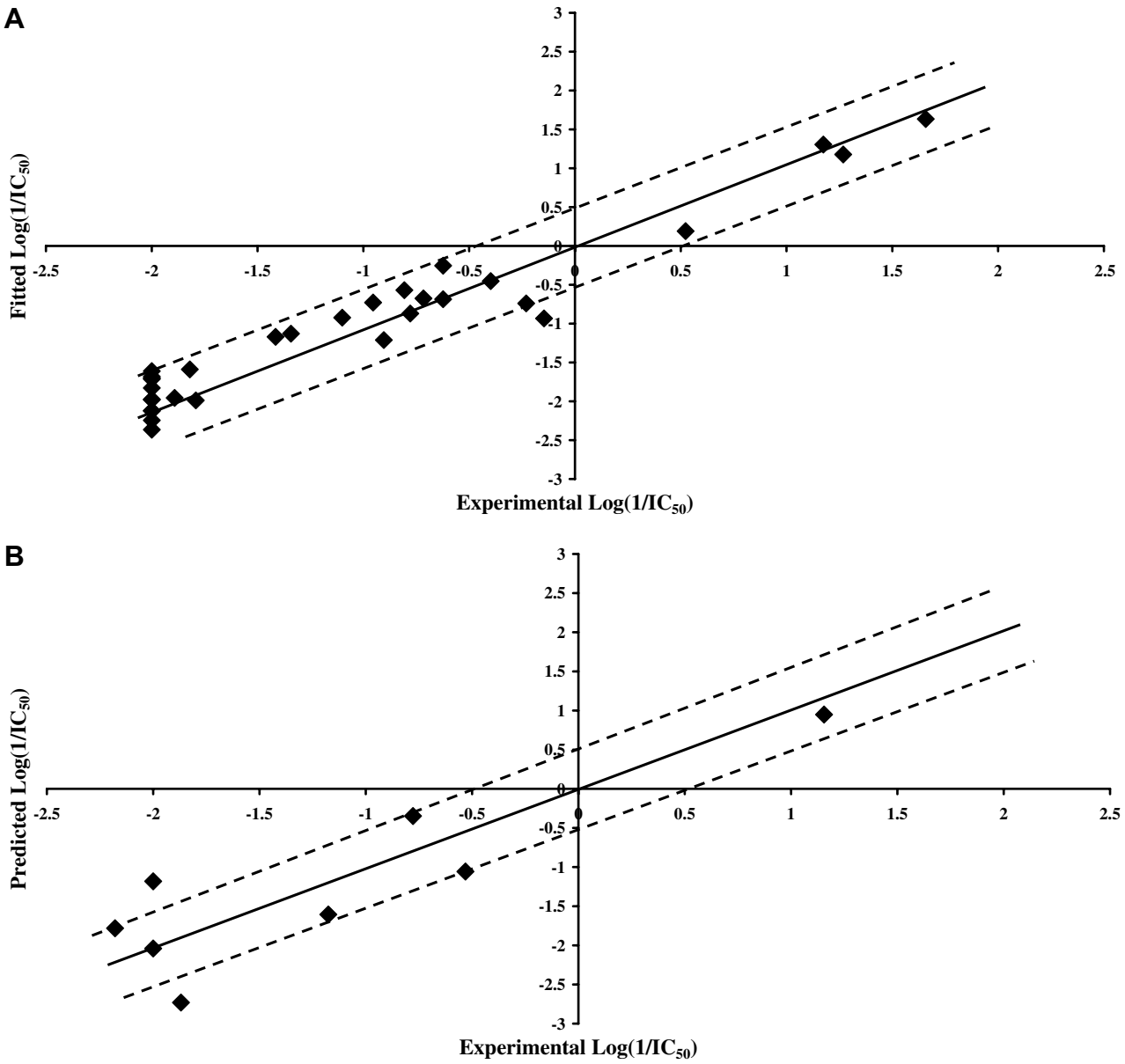


Figure 2. Experimental versus fitted (A, 31 compounds, $r^2_{LOO} = 0.905$) and predicted (B, 8 compounds, $r^2_{PRESS} = 0.748$) bioactivities calculated from the best QSAR model Equation 1. The solid lines are the regression lines for the fitted and predicted bioactivities of training and test compounds, respectively, whereas the dotted lines indicate the 0.5 log point error margins.

Therefore, we were prompted to select common hits captured by Hypo1/4 and Hypo3/8 upon screening the NCI database (a multiconformer database supplied with CATALYST and includes 238,819 compounds stored at the National Cancer Institute[22]) for subsequent processing via the optimal QSAR model Equation 1. Hits are defined as those compounds that have their chemical moieties spatially overlap (map) with corresponding features in the pharmacophoric query.

Hypo3/8 and Hypo1/4 captured 36481 and 5811 hits, respectively. However, the two pharmacophores shared 4700 common hits. These were subsequently filtered based on Lipinski's rule of five,[42] except for the molecular weight limit which we raised to 600D, to yield 2633 molecules. The remaining hits were subsequently filtered based on the number of rotatable bonds, i.e., hits of > 12 rotatable bonds were eliminated, leaving eventually 904 hits. We believe loosening Lipinski's and Veber's[50] pre-filters (i.e., allowing hits of molecular weights > 500 D and rotable bonds > 10, respectively) allows reasonable structural diversity among the hits particularly under the combined restrictions of two orthogonal searching pharmacophores and selection via a QSAR model. Lax in drug-likeness pre-filters in this case is not unreasonable as it is highly unexpected that such studies (i.e., in-silico screening followed by in-vitro evaluation) can identify drugs or even drug candidates. Furthermore, strict adherence to oral bioavailability indicators (i.e., Lipinski's and Veber's rules) is not expected to significantly improve the likelihood of finding antibacterial hits due to the significant differences in the nature of cellular barriers between prokaryotic and eukaryotic cells. Nevertheless, enforcing some kind of drug-likeness pre-filters

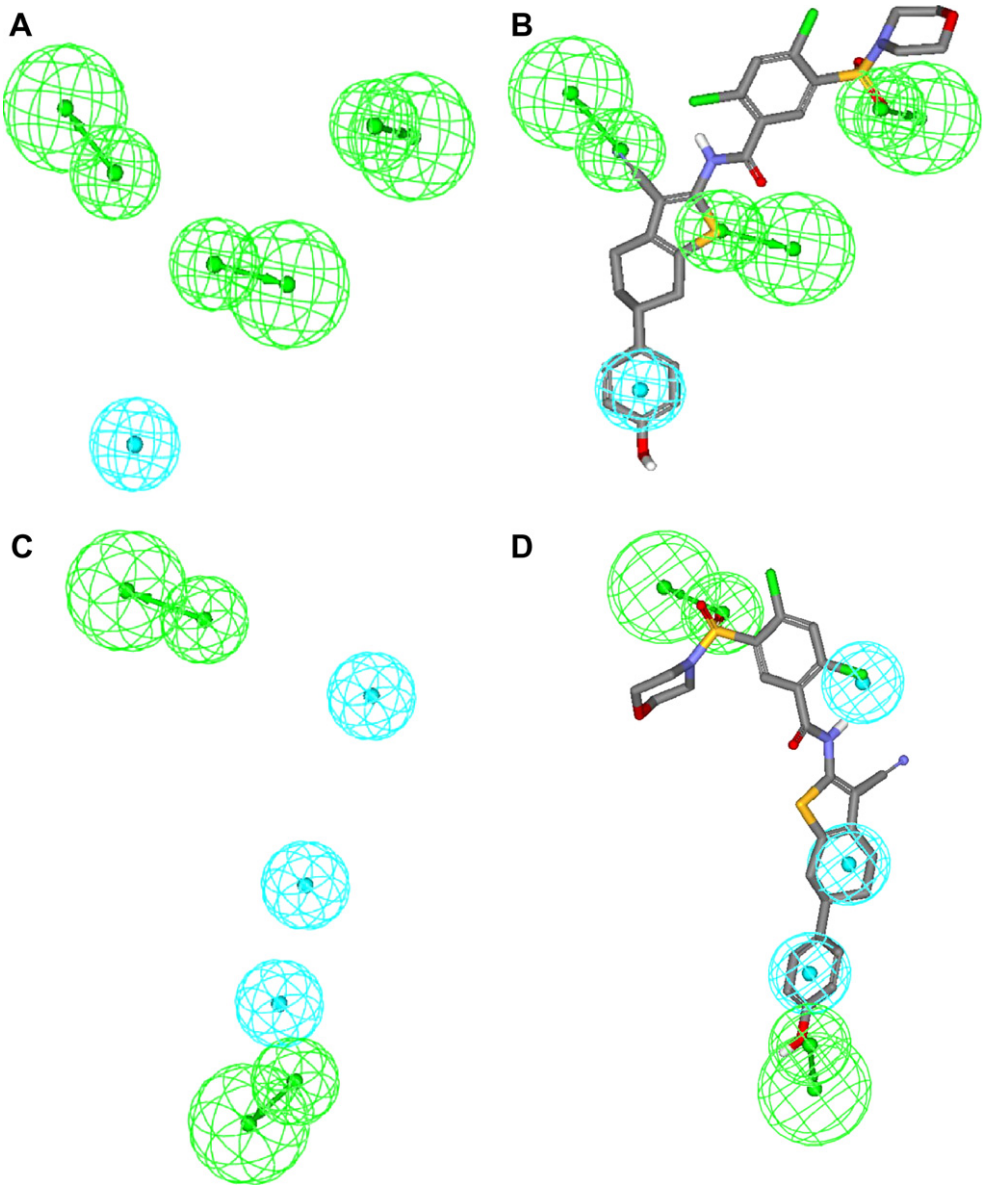


Figure 3. The binding pharmacophore hypotheses emerging in the optimal QSAR models (hydrogen-bond acceptor as green vectored spheres, hydrophobic features as blue spheres): (A) Hypo3/8, (B) Hypo3/8 fitted against the most potent inhibitor **37** (Table 1 and Fig. 1, IC_{50} = 0.022 μM, predicted IC_{50} = 0.025 μM based on QSAR Equation 1), (C) Hypo1/4, and (D) Hypo1/4 mapped against **37**.

Table 5. Pharmacophoric features and corresponding weights, tolerances and 3D coordinates of Hypo1/4[a] and Hypo3/8[b]

Model	Definitions		Chemical features								
			HBA		HBA		HBA		Hbic	Hbic	Hbic
Hypo1/4[a]	Weights		1.84188		1.84188		—		1.84188	1.84188	1.84188
	Tolerances		1.6	2.2	1.6	2.2	—	—	1.6	1.6	1.6
	Coordinates	*X*	−8.88	−10.81	5.67	7.42	—	—	1.02	−4.62	−6.92
		Y	−5.74	−6.64	1.70	0.51	—	—	4.36	−0.32	−4.08
		Z	−2.22	−0.11	2.78	4.95	—	—	−1.52	−0.08	−0.90
Hypo3/8[b]	Weights		2.58464		2.58464		2.58464		2.58464	—	—
	Tolerances		1.6	2.2	1.6	2.2	1.6	2.2	1.6	—	—
	Coordinates	*X*	−1.92	−1.82	3.55	2.84	−0.04	1.52	−8.44	—	—
		Y	−0.69	1.04	4.11	6.51	−3.12	−4.36	−2.50	—	—
		Z	0.22	2.68	1.18	−0.55	−3.93	−6.20	0.06	—	—

[a] Hypo1/4 is hypothesis number 4 generated in run number 1.
[b] Hypo3/8 is hypothesis number 8 generated in run number 3.

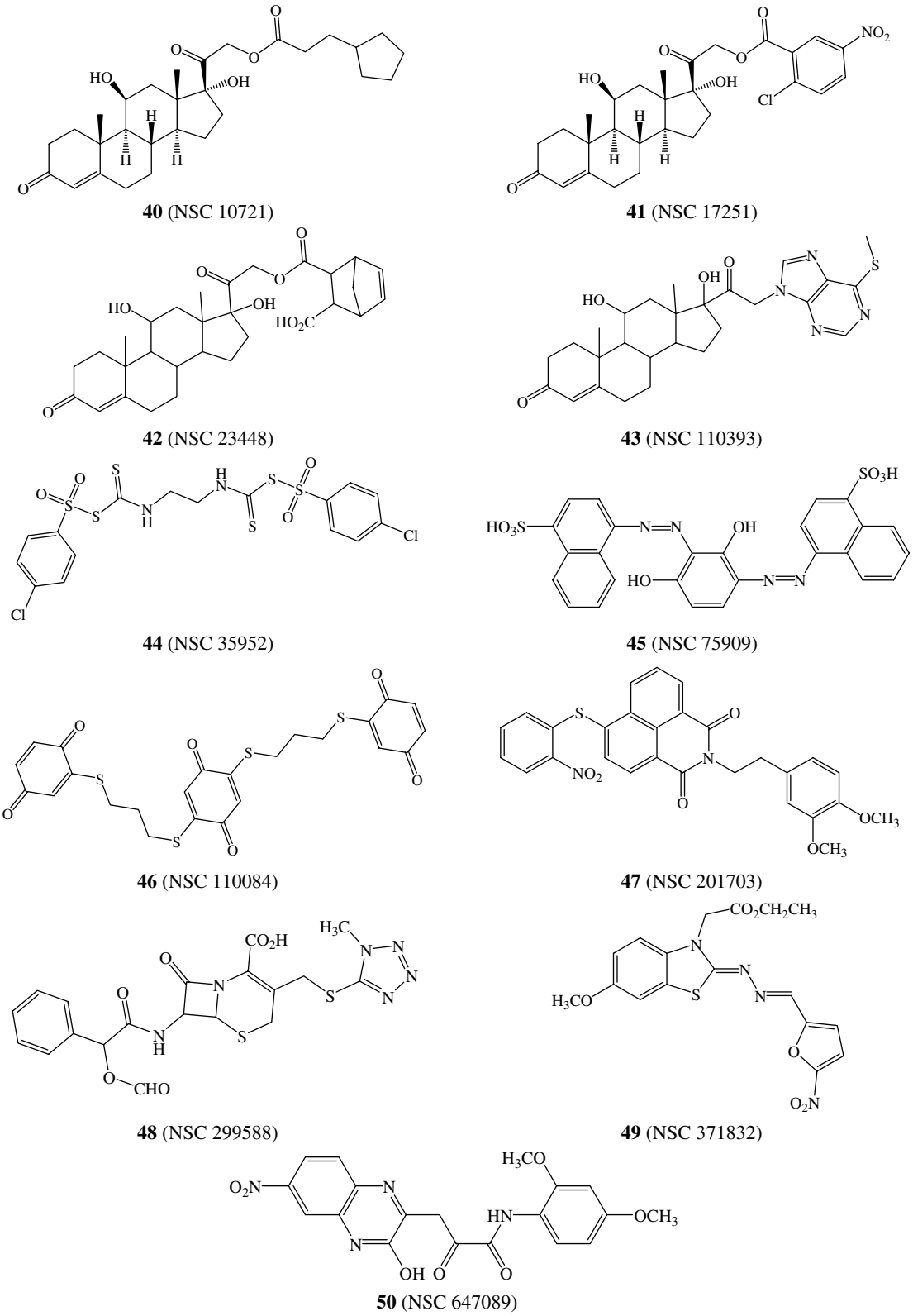


Figure 4. The chemical structures of the tested highest-ranking hits (as suggested by the best QSAR model and the associated pharmacophores).

should help in finding hits more amenable for subsequent optimization into leads.

Anyhow, the selected hits were fitted against the two pharmacophores (see the fit Eq. 5 in Section 4) and their fit values were substituted in QSAR Equation 1 to determine their predicted bioactivities. It must be remembered that it is not necessary that search hits will always optimally fit the corresponding pharmacophore query: a poor fit value for a hit molecule means that

although the chemical functionalities of the compound overlap with the corresponding pharmacophoric features, the centers of its functional groups are displaced from the centers of the corresponding pharmacophoric features, such that the term Σ (disp/tol)^2 in Equation 5 (Section 4.1.5 in Section 4) approaches 1.0 and the overall fit value approaches zero.[32,33,35]

Interestingly, the predicted Log(1/IC_{50}) values exceeded the upper and lower bioactivity limits of the training compounds. Therefore, in order to minimize the impact of any possible extrapolatory prediction errors[49] on decisions regarding which hits merit subsequent in vitro testing, we employed Log(1/IC_{50}) predictions merely to rank the corresponding hits. The highest-ranking 40 hits were requested from the NCI for experimental validation. Unfortunately, only 11 compounds were provided, three of which were found to possess promising in vitro inhibitory actions against MurF, namely: hits **46**, **47**, and **50**. Figure 4 shows the hit molecules including the three active compounds, while Table 6 shows the NCI codes of the tested hits and their estimated and experimental anti-MurF bioactivities. Figure 6 shows how Hypo1/4 and Hypo3/8 fit the most potent hit **50**.

2.5. Inhibition of MurF by hit compounds

We tested the hits against MurF enzyme from *Pseudomonas aeruginosa* as we lacked streptococcal MurF. We hypothesize that the same inhibitory capacity would be observed using MurF from different bacteria in terms of inhibitor:MurF molecule ratios because MurF enzymes are conserved among bacterial species and have similar substrates.[6] However, MurF from *P. aeruginosa* is generally less efficient than its counterparts from other bacterial species. The optimal specific activity of *P. aeruginosa* MurF was found to be 4.4 μmole/min/mg (this is in good agreement with the published value of 3.4 μmol/min/mg[51]) compared to around 16 μmol/min/mg for MurF from *Escherichia coli*.[52,53] Accordingly, we were prompted to employ a relatively high concentration of pseudomonal enzyme, i.e., 40 μM, for testing the selected hits.

At a concentration of 5 mM, the 11 hit compounds gave percentages of residual activity of MurF that varied from 0% to 100% (Table 6). The IC_{50} values of compounds **46**, **47**, and **50** were estimated to be 300, 325 and 125 μM, respectively. Despite their apparently high IC_{50} values, which are probably attributable to the high concentration of MurF enzyme, compounds **46** and **47** gave 50% inhibition with a ratio of eight molecules for each MurF enzyme; whereas **50** would give the same inhibitory percentage with a ratio of three molecules for each MurF enzyme, indicating promising inhibitory profiles. The inhibition curves displayed a nearly linear dose–response trend having variable slopes as shown in Figure 5. The reactions were also tested with an excess of bovine serum albumin, suggesting that the three compounds were specific to MurF and inhibited MurF ATPase activity.

Clearly from Table 6, the bioactivities of some hits were erroneously overestimated by Equation 1 (e.g., **41**, **44**, and **46**, Table 6), which seems to be related to the large contributions of LUMO and $^3\chi_P^v$ in the QSAR model. Hits that combine excessive electrophilicities (large negative LUMO) and elongated shapes (large positive $^3\chi_P^v$) tend to have their log(IC_{50}) values overestimated by Equation 1. Another source of prediction error could be the existence of certain steric clashes ignored by Hypo1/4 and Hypo3/8 and the associated QSAR model. Unfortunately, it was not possible to define the steric constrains of the binding pocket using the currently available list of training inhibitors, i.e., to add steric exclusion volumes to Hypo1/4 and Hypo3/8 using CATALYST, due to the lack of inactive training compounds that have their poor potencies due to steric bumping within the binding pocket.[30]

The active hits **46**, **47**, and **50** were tested against four bacterial species: *E. coli*, *Saccharomyces aureus*, *Saccharomyces pneumonia*, and *P. aeruginosa*. Unfortunately, all three compounds were devoid of any antibacterial activities; probably due to their failure in penetrating the bacterial membranes and/or cell walls.

Table 6. The hit molecules captured by Hypo3/8 and Hypo1/4 and their corresponding QSAR estimates from Equation 6 and their in vitro bioactivities

Tested hits[a]	NCI codes	Fit values against[b]		QSAR-based estimates log(1/IC_{50})	In vitro Anti-MurF Activity	
		Hypo3/8	Hypo1/4		Residual activity with 5.0 mM of hit (%)	IC_{50} (μM)
40	10721	10.0	2.6	−2.8	86	>5000
41	17251	8.8	8.2	2.6	67	>5000
42	23448	9.9	6.5	−1.9	42	4300
43	110393	10.0	1.2	−0.4	48	4800
44	35952	10.0	2.5	7.2	71	>5000
45	75909	9.9	6.5	−0.4	100	>5000
46	110084	10.1	8.4	1.8	0	300
47	201703	10.1	5.7	−0.3	0	325
48	299588	10.1	4.6	−1.9	84	>5000
49	371832	10.0	0.2	−3.1	51	>5100
50	647089	10.2	4.7	−2.2	0	125

[a] Compound numbers as in Figure 4.
[b] Best-fit values against each binding hypothesis calculated by Equation 1.

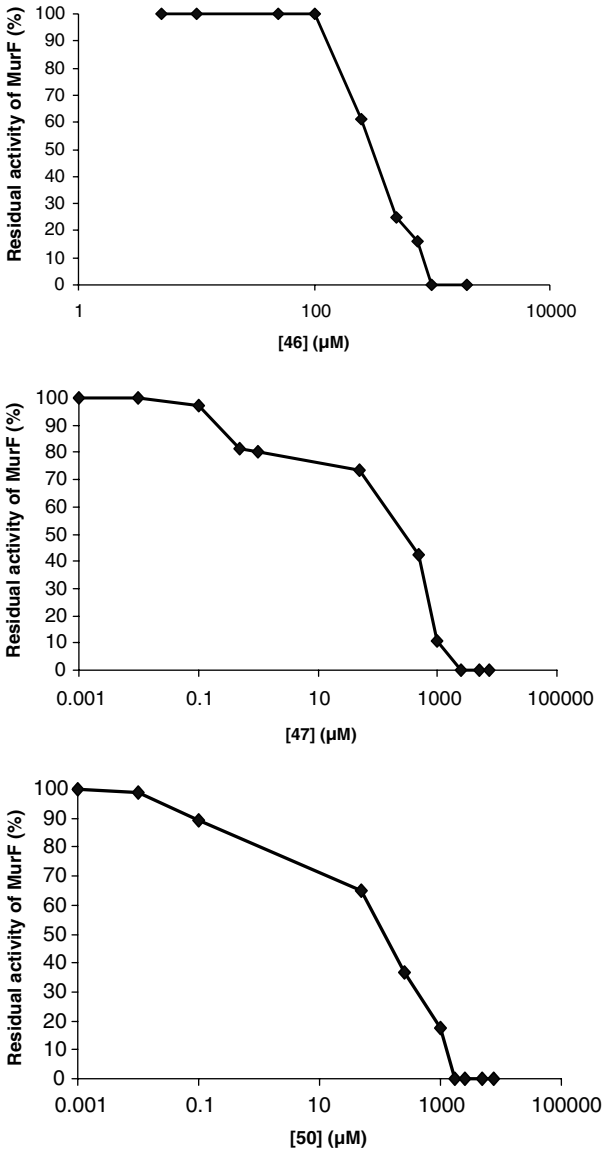


Figure 5. The inhibition curves of hit compounds **46**, **47** and **50** against MurF enzyme.

2.6. Comparison of Hypo3/8 and Hypo1/4 with the binding site of MurF

Despite the problems of crystallographic structures (see 1),[19–21] pharmacophore features obtained by pharmacophore/QSAR modeling can be compared with the structure of MurF binding site to identify probable residues important for inhibition. The features in Hypo1/4 and Hypo3/8, as well as the alignment of **50** (the most potent among the tested hits) as proposed by both pharmacophores, were compared with the corresponding structure docked into the binding pocket of MurF (PDB code: 2AM1, resolution 2.5 Å), as in Figure 6. The docking experiment was performed employing LIGANDFIT docking engine and through default docking parameters and consensus scoring function.[43–45] A marked similarity was observed between the features proposed by the pharmacophore models and the ligand binding features in the docked structures.

In the highest-ranking docked pose of **50** (Fig. 6C) the nitro oxygen atoms are hydrogen-bonded to the amidic nitrogens of ASN326 and ASN137. This interaction corresponds to a HBA feature in Hypo1/4 mapping the same nitro group (Fig. 6A). In the same docked pose, stacking the phenyl of PHE31 against the quinoxaline of **50** nicely corresponds to mapping the quinoxaline system with a hydrophobic feature in Hypo1/4 (Fig. 6A and C). Moreover, mapping the terminal phenyl ring of **50** and its *o*-methoxy substituent with two hydrophobic features in Hypo1/4 (Fig. 6A) correlates with hydrophobic interactions binding the same groups and the phenyl of PHE54 in the optimally docked pose (Fig. 6C). Finally, mapping the *p*-methoxy of the same ring (i.e., in **50**) by a HBA feature in Hypo1/4 (Fig. 6A) suggests the existence of corresponding hydrogen-bonding in the binding pocket. However, this proposition seems not to be supported by any analogous interaction in the docked pose in Figure 6C. Still, we believe peptidic NH of GLY361 can flip to become at close proximity with the *p*-methoxy of **50**causing their mutual hydrogen-bonding (Fig. 6C). Similarly, mapping **50** against Hypo3/8 seems to agree with one of the high-ranking docked poses of this compound: The HBA feature mapping the nitro group in **50** (Fig. 6B) corresponds to hydrogen-bonding with ASN326 in the docked pose (Fig. 6D). The close proximity of the quinoxaline hydroxyl of **50** to the peptidic carbonyl of LYS46 in docked structure (Fig. 6D) suggests their mutual hydrogen-bonding. This proposition is supported by a hydrogen-bonding feature mapping the same hydroxyl in Hypo3/8 (Fig. 6B). Similarly the amidic carbonyl of **50** is mapped by a HBA feature in Hypo3/8 (Fig. 6B) corresponding to hydrogen-bonding interaction with the amidic side chain of ASN328 in the corresponding docked pose (Fig. 6D). Finally, the *p*-methoxy group of **50** is docked near to PRO329 (Fig. 6D) suggesting the existence of significant mutual van der Waals' interactions, which correlate with a hydrophobic feature mapping the *p*-methoxy group in Hypo3/8 (Fig. 6B).

To further emphasize the validity of our pharmacophore/QSAR modeling approach, we compared the crystallographic structure of a MurF/ligand complex (PDB code: 2AM1, resolution 2.5 Å) with Hypo1/4 and Hypo3/8. Interestingly, the complex correlated nicely with Hypo1/4. Figure 7 shows the chemical structure of the complexed ligand and a comparison between the complex and the way how the bound ligand maps Hypo1/4. Mapping the morpholine oxygen of the ligand with a hydrogen-bond acceptor correlates with hydrogen-bonding to the amidic side chain of ASN137. Similarly, mapping the aromatic chloro, the thiophene sulfur, and the terminal cyclohexyl of the ligand with three hydrophobic features corresponds to the way how these functionalities hydrophobically interact with the side chains of LEU334, ASN328, and PHE54, respectively. However, the complexed structure seems to lack any functional group capable of mapping the terminal hydrogen-bond acceptor in the Hypo1/4.

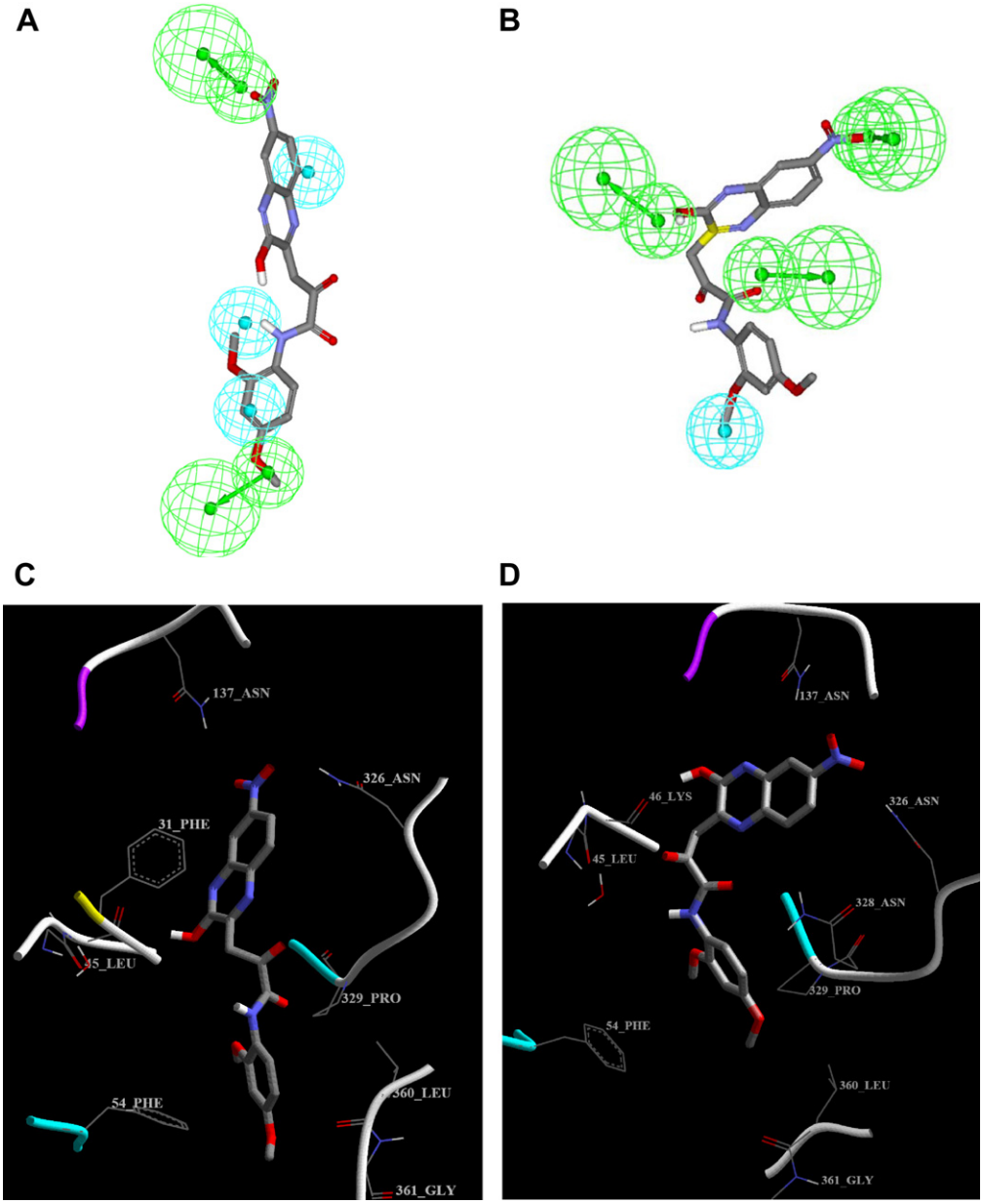


Figure 6. The mapping of Hypo3/8 and Hypo1/4 against hit **50**: (A) Hypo1/4 mapped against **50**, (B) Hypo3/8 mapped against **50**, (C) a successful docking pose of **50** resembling Hypo1/4, (D) a successful docking pose of **50** resembling Hypo3/8.

3. Conclusions

This work includes elaborate pharmacophore exploration of MurF inhibitors utilizing CATALYST–HYPOGEN. QSAR analysis was employed to select the best combination of molecular descriptors and pharmacophore models capable of explaining bioactivity variation across an informative list of training compounds. The best binding hypotheses were used as 3D search queries to screen the NCI database for new MurF inhibitors. The resulting hits were filtered according to Lipinski's rule of five and evaluated using the best QSAR equation. Eleven of the high-ranking hits were acquired, three of which were found to possess promising inhibitory IC_{50} values against MurF. Interestingly, the pharmacophoric features of the optimal models correlated well with the binding features proposed by docking evaluation.

The combination of pharmacophore modeling and QSAR analysis proved successful in enriching the hit list despite the less-than optimal diversity of the training compounds.

4. Experimental

4.1. Molecular modeling

4.1.1. Software and hardware. The following software packages were utilized in the present research

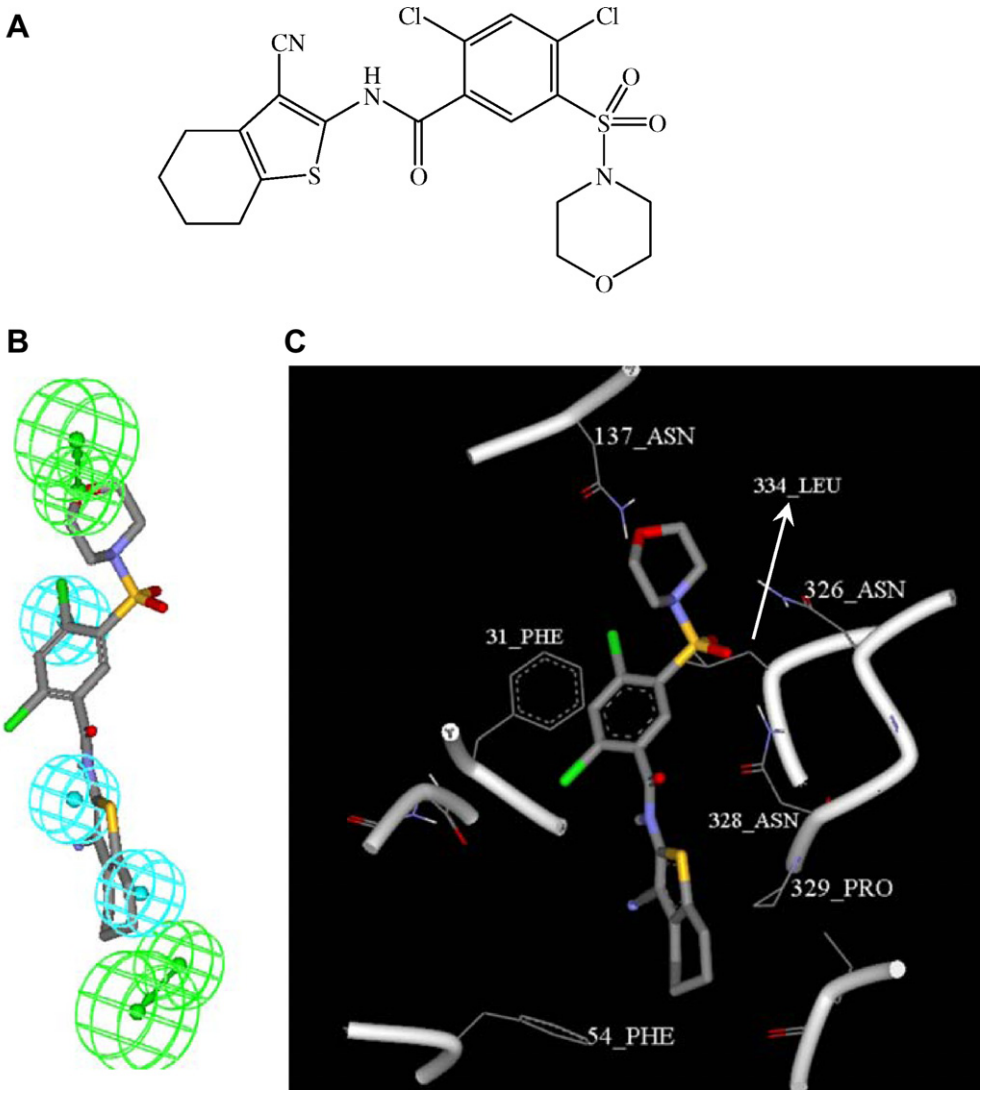


Figure 7. Compound 1LG of the published X-ray structure (PDB code: 2AM1, resolution 2.5 Å)[18]: (A) the chemical structure of 1LG, (B) mapping of Hypo1/4 against the crystallographic structure of 1LG (rigid mapping), (C) the binding pocket of MurF with the complexed ligand (1LG).

- CATALYST (Version 4.11), Accelrys Inc. (www.accelrys.com), USA.
- CERIUS2 (Version 4.10), Accelrys Inc. (www.accelrys.com), USA.
- CS ChemDraw Ultra 7.01, Cambridge Soft Corp. (http://www.cambridgesoft.com), USA.

Pharmacophore modeling, QSAR analysis and docking studies were performed using CATALYST (HYPOGEN module) and CERIUS2 software suites from Accelrys Inc. (San Diego, California, www.accelrys.com) installed on a Silicon Graphics Octane2 desktop workstation equipped with a 600 MHz MIPS R14000 processor (1.0 GB RAM) running the Irix 6.5 operating system.

4.1.2. Data set. The structures of 39 streptococcal (*Streptococcus pneumonia*) MurF enzyme inhibitors (Table 1 and Fig. 1) were collected from published literature.[15,18] The in-vitro bioactivities of the collected inhibitors were expressed as the concentration of the test compound that inhibited the activity of MurF enzyme by 50% (IC_{50}). Figure 1 and Table 1 show the structures and IC_{50} values of the considered inhibitors. The logarithm of measured IC_{50} (μM) values was used in pharmacophore modeling and QSAR analysis, thus correlating the data linear to the free energy change In cases where IC_{50} is expressed as being higher than 100 μM (e.g., **10**, **11**, **13**, **14**, and **15**), we assumed it equals 100 μM, this assumption is necessary to allow statistical correlation and QSAR analysis. The logarithmic transformation of IC_{50} values should minimize any potential errors resulting from this assumption.

The two-dimensional (2D) chemical structures of the inhibitors were sketched using ChemDraw Ultra, installed on a PC, and saved in MDL-mol file format. Subsequently, they were imported into CATALYST, converted into corresponding standard 3D structures and energy minimized to the closest local minimum

using the molecular mechanics CHARMm force field implemented in CATALYST. The resulting 3D structures were utilized as starting conformers for conformational analysis.

4.1.3. Conformational analysis. Molecular flexibility was taken into account by considering each compound as a collection of conformers representing different areas of the conformational space accessible to the molecule within a given energy range. Accordingly, the conformational space of each inhibitor (**1**–**39**, Fig. 1 and Table 1) was explored adopting the 'best conformer generation' option within CATALYST which is based on the generalized CHARMm force field implemented in the program. Default parameters were employed in the conformation generation procedure, i.e., conformational ensembles were generated with an energy threshold of 20 kcal/mol from the local minimized structure and a maximum limit of 250 conformers per molecule. This search procedure will probably identify the best 3D arrangement of chemical functionalities explaining the activity variations among the training set.[22–24,31–37]

4.1.4. Generation of pharmacophoric hypotheses. All 39 molecules with their associated conformational models were regrouped into a spreadsheet. The biological data of the inhibitors were reported with an 'Uncertainty' value of 3, which means that the actual bioactivity of a particular inhibitor is assumed to be situated somewhere in an interval ranging from one-third to three-times the reported bioactivity value of that inhibitor.[32,33,35–37] Subsequently, a subset from the collected inhibitors was employed as training set for pharmacophore modeling, namely, **1**, **2**, **4**, **5**, **7**, **9**, **10**, **11**, **12**, **15**, **16**, **17**, **19**, **22**, **23**, **24**, **26**, **30**, **33**, **34**, **35**, **36**, **37**, **38** and **39**. Typically, CATALYST requires informative training sets that include at least 16 compounds of evenly spread bioactivities over at least four orders of magnitude.[32,33,35–37]

The training compounds were utilized to conduct **4** modeling runs to explore the pharmacophoric space of MurF inhibitors. Different hypotheses were generated by altering the interfeature spacing and the number of allowed features in the resulting pharmacophores (Table 2).

Pharmacophore modeling employing CATALYST proceeds through three successive phases: the constructive phase, subtractive phase and optimization phase.[32,33,35–37] During the constructive phase, CATALYST generates common conformational alignments among the most active training compounds. Only molecular alignments based on a maximum of five chemical features are considered. The program identifies a particular compound as being within the most active category if it satisfies Equation 2.[32,33,35–37]

$$(\mathrm{MAct} \times \mathrm{UncMAct}) - (\mathrm{Act}/\mathrm{UncAct}) > 0.0, \qquad (2)$$

where 'MAct' is the activity of the most active compound in the training set, 'Unc' is the uncertainty of the compounds, and 'Act' is the activity of the training compounds under question. Accordingly, compounds **33**, **34**, **36**, and **37** (Fig. 1 and Table 1) are considered within the most active category.

In the subsequent subtractive phase, CATALYST eliminates some hypotheses that fit inactive training compounds. A particular training compound is defined as being inactive if it satisfies Equation 3[32,33,35–37]:

$$\mathrm{Log}(\mathrm{Act}) - \log(\mathrm{MAct}) > \mathrm{BS}, \qquad (3)$$

where 'BS' is the bioactivity spread (equals 3.5 by default). Therefore, compounds **5**, **10**, **11**, **15**, **16**, **17**, **19**, **24**, and **30** (of bioactivities > 70 μM, Fig. 1 and Table 1) are categorized within the inactive group.

Among the inactives, some inhibitors were reported to have $IC_{50} > 100$ μM (e.g., **10**, **11**, and **15**). These were fed into CATALYST with IC_{50} values = 100 μM. This assumption should have minimal impact on the pharmacophore modeling process as inactive compounds are used to filter out poor pharmacophores regardless of their actual bioactivities.

In the optimization phase, CATALYST applies fine perturbations in the form of vectored feature rotation, adding new feature and/or removing a feature, to selected hypotheses that survived the subtractive phase to find new models of enhanced bioactivity-to-mapping correlations. Eventually, CATALYST selects the highest-ranking models (10 by default) and presents them as the optimal pharmacophore hypotheses resulting from the particular automatic modeling run.[23]

4.1.5. Assessment of the generated hypotheses. When generating hypotheses, CATALYST attempts to minimize a cost function consisting of three terms: weight cost, error cost, and configuration cost.[23,32,33,35–37] Weight cost is a value that increases as the feature weight in a model deviates from an ideal value of 2. The deviation between the estimated activities of the training set and their experimentally determined values adds to the error cost.[23,32,33,35–37] The activity of any compound can be estimated from a particular hypothesis through Equation 4.[23]

$$\mathrm{Log}(\text{Estimated activity}) = I + \mathrm{Fit}, \qquad (4)$$

where, I = the intercept of the regression line obtained by plotting the log of the biological activity of the training set compounds against the Fit values of the training compounds. The Fit value for any compound is obtained automatically by employing Equation 5.[32,33,35]

$$\mathrm{Fit} = \Sigma\text{mapped hypothesis features} \times W[1 - \Sigma(\mathrm{disp}/\mathrm{tol})^2], \qquad (5)$$

where Σ mapped hypothesis features represents the number of pharmacophore features that successfully superimpose (i.e., map or overlap with) corresponding chemical moieties within the fitted compound, W is the weight of the corresponding hypothesis feature spheres. This value is fixed to 1.0 in CATALYST-generated models. disp is the distance between the center of a particular pharmacophoric sphere (feature centroid) and the center of the corresponding superimposed chemical moiety of the fitted compound; tol is the radius of the pharmacophoric feature sphere (known as Tolerance, equals to 1.6 Å by default). $\Sigma(\mathrm{disp}/\mathrm{tol})^2$ is the summa-

tion of $(disp/tol)^2$ values for all pharmacophoric features that successfully superimpose corresponding chemical functionalities in the fitted compound.[32,33,35]

The third term, that is, the configuration cost, penalizes the complexity of the hypothesis, i.e. the configuration cost. This is a fixed cost, which is equal to the entropy of the hypothesis space. The more the numbers of features (a maximum of five) in a generated hypothesis, the higher is the entropy with subsequent increase in this cost. The overall cost (total cost) of a hypothesis is calculated by summing over the three cost factors. However, error cost is the main contributor to total cost.

CATALYST also calculates the cost of the null hypothesis, which presumes that there is no relationship in the data and that experimental activities are normally distributed about their mean. Accordingly, the greater the difference from the null hypothesis cost (residual cost, Table 3), the more likely that the hypothesis does not reflect a chance correlation.[23,24,32,33,35–37]

In a successful automatic modeling run, CATALYST® ranks the generated models according to their total costs.[32,33,35]

An additional approach to assess the quality of CATALYST–HYPOGEN pharmacophores is to cross-validate them using the Cat-Scramble program implemented in CATALYST.[37,38] This validation procedure is based on Fischer's randomization test.[38] In this validation test, we selected a 95% confidence level, which instructs CATALYST to generate 19 random spreadsheets by the Cat-Scramble command. Subsequently, CATALYST–HYPOGEN is challenged to use these random spreadsheets to generate hypotheses using exactly the same features and parameters used in generating the initial unscrambled hypotheses.[46] Success in generating pharmacophores of comparable cost criteria to those produced by the original unscrambled data reduces the confidence in the training compounds and the unscrambled original pharmacophore models.[37,38]

Table 3 shows the pharmacophoric features and success criteria of the generated binding hypotheses, including the cost values (error, configuration, weight and total costs) of the resulting pharmacophores, as well as the costs of the corresponding null hypotheses. The table also shows the corresponding Fischer (Cat.scramble) confidence for each pharmacophore.

4.1.6. QSAR modeling. A subset of 31 compounds was employed as the training set for QSAR modeling. However, since it is essential to access the predictive power of the resulting QSAR models on an external set of inhibitors, the remaining eight molecules (ca 20% of the data set) were employed as an external test subset for validating the QSAR models. The test molecules were selected as follows: the 39 inhibitors were ranked according to their IC_{50} values, then every fifth compound was selected for the test set starting from the high-potency end. This selection considers the fact that the test molecules must represent a range of biological activities similar to that of the training set. The selected test inhibitors are: **2**, **14**, **16**, **25**, **28**, **33**, **35**, and **38** (numbers are as in Table 1 and Fig. 1). The chemical structures of the inhibitors were imported into CERIUS2 as standard 3D single conformers (of the lowest energy within the conformational ensemble generated by CATALYST) representations in SD format. Upon importing the structures into CERIUS2, we switched off the 'Clean and Minimize' options in CERIUS2 to preserve the conformers generated by CATALYST. Subsequently, different descriptor groups were calculated for each compound employing the C2.DESCRIPTOR module of CERIUS2. The calculated descriptors included various simple and valence connectivity indices, electro-topological state indices, single point quantum–mechanical descriptors (via the AM1 model) and other molecular descriptors (e.g., logarithm of partition coefficient, polarizability, dipole moment, molecular volume, molecular weight, molecular surface area, etc., the detailed molecular descriptors are listed under Supplementary material).[39] Furthermore, the training compounds were fitted against the 40 different pharmacophore hypotheses, generated by the CATALYST–HYPOGEN automatic runs (shown in Tables 2 and 3), and their fit values (produced by the best-fit command within CATALYST via Equation 5) were added as additional molecular descriptors.

Genetic function approximation (GFA) was employed to search for the best possible QSAR regression equation capable of correlating the variations in biological activities of the training compounds with variations in the generated descriptors, i.e., multiple linear regression modeling (MLR). GFA techniques rely on the Darwinian evolutionary operations of 'crossover and mutation' to select optimal combinations of descriptors (i.e., chromosomes) capable of explaining bioactivity variation among training compounds from a large pool of possible descriptor combinations, i.e., chromosomes population. Each chromosome is associated with a fitness value that reflects how good it is compared to other solutions. The fitness function employed herein is based on Friedman's 'lack-of-fit' (LOF).[39]

Our preliminary diagnostic trials suggested the following optimal GFA parameters: Explore linear equations at mating and mutation probabilities of 50%; population size = 500; number of genetic iterations = 30,000, and LOF smoothness parameter = 1.0. However, to determine the optimal number of explanatory terms (QSAR descriptors), it was decided to scan and evaluate all possible QSAR models resulting from 3 to 8 explanatory terms.

All QSAR models were validated employing leave-one-out cross-validation (r^2_{LOO}), bootstrapping (r^2_{BS})[39,40], and predictive r^2 (r^2_{PRESS}) calculated from the test subsets. The predictive r^2_{PRESS} is defined as:[39–41]

$$r^2_{PRESS} = \text{SD-PRESS}/\text{SD}, \quad (6)$$

where SD is the sum of the squared deviations between the biological activities of the test set and the mean

activity of the training set molecules, PRESS is the squared deviations between predicted and actual activity values for every molecule in the test set. The descriptor-scanning procedure identified Equation 1 (QSAR model B in Table 4) as the best equation. However, two pharmacophore hypotheses emerged in this model and in the other two highest-ranking QSAR equations (models A and C in Table 4), namely, Hypo1/4 and Hypo3/8 (Fig. 3 and Table 5). Table 4 shows the statistical criteria of the scanned QSAR models, while Figure 2 shows the plots of experimental versus fitted (training set) and predicted (testing set) bioactivities calculated from the best QSAR model (Eq. 1 or model B in Table 4):

4.1.7. In-silico screening of the NCI Database for new MurF inhibitors. Hypo1/4 and Hypo3/8 were employed as 3D search queries against the NCI structural library using the 'Best Flexible Database Search' option within CATALYST. Common hits were filtered based on Lipinski's rule of five and number of rotable bonds (see Section 2.4). The remaining hits (904) were fitted against Hypo1/4 and Hypo3/8 using the Best-Fit approach within CATALYST. Subsequently, the fit values together with the relevant molecular descriptors were substituted in Equation 1 (QSAR model B in Table 4) to determine their predicted anti-MurF IC_{50} values.

4.2. MurF enzyme inhibition assays

4.2.1. Preparation of MurF and its UDP-MurNAc-Ala-Glu-meso-A2 pm substrate. The MurF protein from *P. aeruginosa* was over-expressed, purified, sequenced, and dialyzed, and UDP-MurNAc-Ala-Glu-meso-A2pm was synthesized, purified, and analyzed as previously described.[47]

4.2.2. Quantification of MurF ATPase activity in a spectrophotometric assay. The ATPase activity of MurF was quantified by a colorimetric assay that measures the release of inorganic phosphate as previously described.[48] UDP-MurNAc-Ala-Glu-meso-A2pm was used at 75 μM instead of 100 μM, 25% DMSO was added to the reaction mixture, and MurF was incubated 30 min at room temperature in buffer C (100 mM Tris–HCl, pH 8.6, 40 mM KCl, 500 mM NaCl, 1 mg/ml bovine serum albumin, and 10 mM $MgCl_2$) before adding the substrates to start the enzymatic reaction.

4.2.3. MurF inhibition by hit compounds. The tested compounds were provided as dry powders in 20 mg quantities. They were initially dissolved in 1.0 ml DMSO and subsequently diluted to the required concentrations. The inhibition of MurF ATPase activity by the hit compounds was measured using the spectrophotometric assay described above. MurF was pre-incubated with 5 mM of each compound for 30 min before adding the three substrates. The final concentration of DMSO was always adjusted to 25%. The percentage of residual activity of MurF was determined for each compound by comparing the ATPase activity of MurF with and without the compound. Negative controls were tested to assess whether MurF is inhibited by 25% DMSO. The enzyme was not affected by DMSO. The concentration required to give 50% inhibition (IC_{50}) was determined for the three compounds having the best inhibitory activities. MurF was pre-incubated with different concentrations of compounds **46**, **47** and **50** (Fig. 4) and the percentages of residual activity of MurF data were used to evaluate the IC_{50} values.

Acknowledgments

This project was partially sponsored by the Faculty of Graduate Studies (M.Sc. Thesis of Naji Atallah Al-Gouleh). The authors thank the Deanship of Scientific Research and Hamdi-Mango Center for Scientific Research at the University of Jordan for their generous funds. R. C. Levesque is a scholar of exceptional merit from Le Fonds de la recherche en santé du Québec (FRSQ) and Catherine-Paradis-Bleau was awarded a studentship from the INSERC and from FRSQ. Work in R.C. Levesque's laboratory was funded by a team grant from Le Fonds Québécois de la recherche sur la nature et les technologies.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2007.10.076.

References and notes

1. Doern, G. V.; Heilmann, K. P.; Huynh, H. K.; Rhomberg, P. R.; Coffman, S. L.; Brueggemann, A. B. *Antimicrob. Agents Chemother.* **2001**, *45*, 1721.
2. Doern, G. V. *Clin. Infect. Dis.* **2001**, *33*, S187.
3. Goldsmith, C. E.; More, J. E.; Murphy, P. G.; Ambler, J. E. J. *Antimicrob. Agents Chemother.* **1998**, *41*, 420.
4. Ho, P. L.; Que, T. L.; Tsang, D. N. C.; Ng, T. K.; Chow, K. H.; Seto, W. H. *Antimicrob. Agents Chemother.* **1999**, *43*, 1310.
5. Black, M. T.; Hodgson, J. *Adv. Drug Delivery Rev.* **2005**, *57*(10), 1528.
6. El Zoeiby, A.; Sanschagrin, F.; Levesque, R. C. *Mol. Microbiol.* **2003**, *47*(1), 1.
7. Wong, K. K.; Pompiano, D. L. In Rosen, B. P., Mobashery, S., Eds.; Resolving the Antibiotic Paradox: Progress in Understanding Drug Resistance and Development of New Antibiotics; Plenum: NY, 1998; pp 197–217.
8. Goldman, R. C.; Branson, A. *Curr. Pharm. Design* **1999**, *5*, 473.
9. Lin, Y. I.; Li, Z.; Francisco, G. D.; McDonald, L. A.; Davis, R. A.; Singh, G.; Yang, Y.; Mansour, T. S. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2591.
10. Isono, K.; Uramoto, M.; Kusakabe, H.; Kimura, K.; Izaki, K.; Nelson, C. C.; McCloskey, J. A. *J. Antibiotics* **1985**, *38*, 1617.
11. Kimura, K.; Ikeda, Y.; Kagami, S.; Yoshihara, M. *J. Antibiotics* **1998**, *51*, 1099.
12. Green, D. W. *Expert Opin. Ther. Targets* **2002**, *6*, 1.
13. Kahan, F. M.; Kahan, J. S.; Cassidy, P. J.; Kropp, H. *Ann. N. Y. Acad. Sci.* **1974**, *235*(0), 364.

14. Comess, K. M.; Schurdak, M. E.; Voorbach, M. J.; Coen, M.; Trumbull, J. D.; Yang, H.; Gao, L.; Tang, H.; Cheng, X.; Lerner, C. G.; McCall, J. O.; Burns, D. J.; Beutel, B. A. *J. Biomol. Screen.* **2006**, *11*, 743.
15. Stamper, G. F.; Longenecker, K. L.; Fry, E. H.; Jakob, C. G.; Florjancic, A. S.; Gu, Y. G.; Anderson, D. D.; Cooper, C. S.; Zhang, T.; Clark, R. F.; Cia, Y.; Black-Schaefer, C. L.; McCall, J. O.; Lerner, C. G.; Hajduk, P. J.; Beutel, B. A.; Stoll, V. S. *Chem. Biol. Drug Des.* **2006**, *67*, 58.
16. Baum, E. Z.; Crespo-Carbone, S. M.; Abbanat, D.; Foleno, B.; Maden, A.; Goldschmidt, R.; Bush, K. *Antimicrob. Agents Chemother.* **2006**, *50*, 230.
17. Longenecker, K. L.; Stamper, G. F.; Hajduk, P. J.; Fry, E. H.; Jakob, C. G.; Harlan, J. E.; Edalji, R.; Bartley, D. M.; Walter, K. A.; Solomon, L. R.; Holzman, T. F.; Gu, Y. G.; Lerner, C. G.; Beutel, B. A.; Stoll, V. S. *Protein Sci.* **2005**, *14*, 3039.
18. Gu, Y. G.; Florjancic, A. S.; Clark, R. F.; Zhang, T.; Cooper, C. S.; Anderson, D. D.; Lerner, C. G.; McCall, J. O.; Cai, Y.; Black-Schaefer, C. L.; Stamper, G. F.; Hajduk, P. J.; Beutel, B. A. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 267.
19. Beeley, N. R. A.; Sage, C. *GPCRs: Targets* **2003**, *2*, 19.
20. Waszkowycz, B. In *Advances in Drug Discovery Techniques*; Harvey, A. L., Ed.; John Wiley & Sons: Chihester, UK, 1998; pp 150–153.
21. DePristo, M. A.; de Bakker, P. I. W.; Blundell, T. L. *Structure* **2004**, *12*, 831.
22. Catalyst User Guide, Accelrys Software Inc., San Diego, 2005.
23. Sprague, P. W.; Hoffmann, R. In Waterbeemd, H., Van de Testa, B., Folkers, G., Eds.; Computer Assisted Lead Finding and Optimization—Current Tools for Medicinal Chemistry; VHCA: Basel, 1997; pp 230–240.
24. Barnum, D.; Greene, J.; Smellie, A.; Sprague, P. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 563.
25. Singh, J.; Chuaqui, C. E.; Boriack-Sjodin, P. A.; Lee, W.-C.; Pontz, T.; Corbley, M. J.; Cheung, H.-K.; Arduini, R. M.; Mead, J. N.; Newman, M. N.; Papadatos, J. L.; Bowes, S.; Josiah, S.; Ling, L. E. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4355.
26. Taha, M. O.; Qandil, A. M.; Zaki, D. D.; AlDamen, M. A. *Eur. J. Med. Chem.* **2005**, *40*, 701.
27. Keller, P. A.; Bowman, M.; Dang, K. H.; Garner, J.; Leach, S. P.; Smith, R.; McCluskey, A. *J. Med. Chem.* **1999**, *42*, 2351.
28. Karki, R. G.; Kulkarni, V. M. *Eur. J. Med. Chem.* **2001**, *36*, 147.
29. Taha, M. O.; Al-Bakri, A. G.; Zalloum, W. A. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 5902.
30. Taha, M. O.; Bustanji, Y.; Al-Bakri, A. G.; Yousef, M.; Zalloum, W. A.; Al-Masri, I. M.; Atallah, N. *J. Mol. Graphics Model.* **2007**, *25*, 870.
31. Smellie, A.; Teig, S.; Towbin, P. *J. Comput. Chem.* **1995**, *16*, 171.
32. Li, H.; Sutter, J.; Hoffmann, R. In Güner, O. F., Ed.; Pharmacophor Perception, Development, and Use in Drug Design; International University Line: California, 2000; pp 173–189.
33. Sutter, J.; Güner, O.; Hoffmann, R.; Li, H.; Waldman, M. In *Pharmacophore Perception, Development, and Use in Drug Design*; International University Line: California, 2000; pp 501–511.
34. Sheridan, R. P.; Kearsley, S. K. *Drug Discov. Today* **2002**, *7*, 903.
35. Kurogi, Y.; Güner, O. F. *Curr. Med. Chem.* **2001**, *8*, 1035.
36. Bersuker, I. B.; Bahçeci, S.; Boggs, J. E. In Güner, O. F., Ed.; Pharmacophore Perception, Development, and Use in Drug Design; International University Line: California, 2000; pp 457–473.
37. Poptodorov, K.; Luu, T.; Hoffmann, R. In *Methods and principles in Medicinal Chemistry, Pharmacophores and Pharmacophores Searches*; Langer, T., Hoffmann, R. D., Eds.; Wiley-VCH: Weinheim, 2006; Vol. 2, pp 17–47.
38. Fischer, R. *The Principle of Experimentation Illustrated by a Psycho-Physical ExpeHafner Publishing Co*, 8th ed.; Hafner Publishing: New York, 1966, Chapter II.
39. CERIUS2 QSAR Users' Manual, Accelrys Inc., San Diego, CA, 2005, pp 43–88, 221–235, 237–250.
40. Tropsha, A.; Gramatica, P.; Gombar, V. K. *Quant. Struct.-Act. Relat. Comb. Sci.* **2003**, *22*, 69.
41. Ramsey, L. F.; Schafer, W. D. *The Statistical Sleuth*, 1st ed.; Wadesworth Publishing Company: USA, 1997.
42. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Delivery Rev.* **2001**, *46*, 3.
43. Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. *J. Mol. Graphics Model.* **2003**, *21*, 289.
44. CERIUS2 4.10 LigandFit User Manual, Accelrys Inc.: San Diego, CA, 2005, pp 3–48.
45. Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. *Chem. Biol.* **1995**, *2*, 317.
46. Krovat, E. M.; Langer, T. *J. Med. Chem.* **2003**, *46*, 716.
47. El Zoeiby, A.; Sanschagrin, F.; Havugimana, P.; Garnier, A.; Levesque, R. C. *FEMS Microbiol. Lett.* **2001**, *201*, 229.
48. Paradis-Bleau, C.; Beaumont, M.; Boudreault, L.; Lloyd, A.; Sanschagrin, F.; Bugg, T. D.; Levesque, R. C. *Peptides* **2006**, *27*, 1693.
49. Cronin, M. T. D.; Schultz, T. W. *J. Mol. Struct. (Theochem.)* **2003**, *622*, 39.
50. Veber, D. F.; Johnson, S. R.; Cheng, H. Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. *J. Med. Chem.* **2002**, *45*, 2615.
51. Azzolina, B. A.; Yuan, X.; Anderson, M. S.; El-Sherbeini, M. *Protein Expr. Purif.* **2001**, *21*, 393.
52. Anderson, M. S.; Eveland, S. S.; Onishi, H. R.; Pompliano, D. L. *Biochemistry* **1996**, *35*, 16264.
53. Duncan, K.; van Heijenoort, J.; Walsh, C. T. *Biochemistry* **1990**, *29*, 2379.