APPLICATION OF HIGH PERFORMANCE COMPUTING FOR DEVELOPMENT OF HIGHLY PREDICTIVE 3D-QSAR MODELS.

Jayendra B. Bhonsle, CPT. William McCalmont*, LTC. Michael Kozar and
COL Alan Magill
Walter Reed Army Institute of Research, Department of Medicinal Chemistry,
Division of Experimental Therapeutics, 503 Robert Grant Avenue Silver Spring, MD 20910. USA.

ABSTRACT

Infectious diseases such as malaria, leishmaniasis and a plethora of bacterial diseases have been and continue to be among the major problems for United States Military personnel deployed in disease endemic regions of the world. We currently employ computer-aided rational drug design and discovery methods to discover new and better drugs. Here, we compute the mathematical equation correlating the observed biological activity of the drug molecule to the various descriptors, such as physicochemical properties, electrostatic and steric fields and chemical functions of the drug molecules.

In brief, QSAR involves computation of the conformational model of the drug molecules, alignment of the conformers in a biologically meaningful way, computation of the descriptors, and lastly using statistical techniques such as linear regression analysis to compute the QSAR model. The traditional approach of global minimum energy conformation of the drug molecules fails to deliver good predictive QSAR models for flexible molecules. To address this issue we have developed a novel method viz. bioactive conformation mining, which consistently delivered good predictive QSAR models.

Development of Antimicrobial peptides (AMP) based antibacterials:

Antimicrobial peptides (AMP) are involved in the defense mechanism of animals against invading microorganisms. The mechanism of action for AMP is via disruption of cell membranes. We have developed a series of AMPs employing unnatural amino acids by strategically controlling the 3D-physicochemical properties to exhibit different in vitro activity against *Staphylococcus aureus* (SA) and *Mycobacterium ranae* (MR) bacteria. We present the PC based 3D-QSAR studies, which provide valuable insights in the design of novel AMPs and also the mechanism of action.

Development of novel DEET based insect repellents:

Mosquitoes transmit a variety of parasites and pathogens. Keeping the mosquitoes away using insect repellents is, therefore, a significant preventive approach against these deadly diseases. N,N-diethyl-3-methyl benzamide (DEET) is the most effective and widely used insect repellent. We computed a PC based 3D-QSAR model to assist in prediction of insect repellency protection time of novel DEET based insect repellents. The QSAR model also provides valuable insight into the mechanism of action of DEET analogs and derivatives.

1. INTRODUCTION

1.1 DEET based insect repellents 3D-QSAR

Mosquitoes and many other insects transmit a variety of parasitic and pathogenic diseases including malaria, yellow fever and viral encelphalitis.(Brewste 2001) Thus, using insect repellents for keeping the insects away is an important and significant strategy in the fight against these deadly diseases. Presently, the reported participating entities in the mechanism of action (Justice; Biessmann et al. 2003) of DEET based insect repellents are the odorant-binding protein (OBP), the neuronal G-protein coupled receptors (GPCRs) and the odorant degrading enzymes (ODEs). It is reported that the OBP binds to odorant which are typically hydrophobic and facilitates their movement through the hydrophilic hemolymph towards the olfactory neuronal GPCRs. Then, the OBP-odorant complex binds with the GPCR causing the repellency effect. The ODE is reported to degrade the odorants thereby preventing continued stimulation of the olfactory receptors.

1.2 AMP based antibacterials 3D-QSAR

Antimicrobial peptides (AMPs) have evolved in many classes of living organisms, as a host defence mechanism against invading micro-organisms.(Dennison; Wallace et al. 2005) AMPs may be divided into two super families as membrane-disruptors and non-membrane – disruptors based on their mechanism of action.(Brogden 2005) All membrane-disruptors are reported to follow specific steps in the process of binding to the target cells.(Blondelle; Lohner et al. 1999) The AMPs are first attracted to the surface of the membrane by the electrostatic interactions between the positively charged amino acids of the AMP and the negatively charged phospholipids of the cell membrane.(Dennison; Wallace et al. 2005) The next step involves the binding of the

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
DEC 2008	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Application Of High Performance Computing For Development Of Highly Predictive 3d-Qsar Models	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Walter Reed Army Institute of Research, Department of Medicinal Chemistry, Division of Experimental Therapeutics, 503 Robert Grant Avenue Silver Spring, MD 20910. USA.**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002187. Proceedings of the Army Science Conference (26th) Held in Orlando, Florida on 1-4 December 2008, The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**8**	

AMPs to the surface of the membrane. (Brogden 2005) Our guiding hypothesis based on the above assertion is as follows: the target cell membrane (bacterial or mammalian) interacts with the approaching AMP in a very specific way (via bioactive conformation) through the mutually complementary 3D-physicochemical surface properties and thus defining the resulting organism selectivity and potency. Here in we describe the computation of 3D-QSAR models for the *Staphylococcus aureus* ME/GM/TC resistant (ATCC 33592) (***SA***) and *Mycobacterium ranae* (ATCC 110) (***MR***) activity of AMPs.

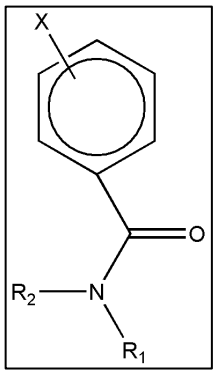
2. RESULTS & DISCUSSIONS

2.1 DEET based insect repellents 3D-QSAR

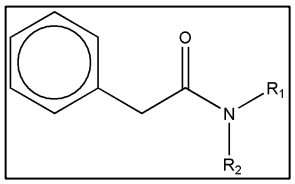
We chose a collection of forty benzamides, benzyl amides and cyclohexyl amide DEET (**4c**) analogs and derivatives that were reported earlier (Suryanarayana; Pandey et al. 1991) for this QSAR study. The chemical structures, vapor pressures @ 30deg C and their respective protection times are summarized in Table–1.

We used Cerius2 (C2) to build and minimize the molecular structures using the 3D-sketcher module. For the minimization process we employed Gasteiger-Marsali (Marsali; Gasteiger 1980) charges and Drieding force field (Mayo; Olafson et al. 1990). We computed the conformational models by performing exhaustive conformational search using the Grid Scan method (Accelrys 2005) followed by cluster analysis based on the root mean squares (RMS) differences of the torsion angles. We aligned the cluster nuclei using the amide group common core as the template.

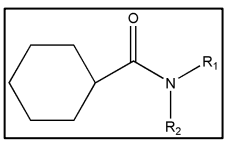
We found that the clusters with 20-25 nuclei showed good 3D sampling of the space around the amide, the putative pharmacophoric moiety, with little or no vacant volume and with much less crowding or over representation. The overlay of all conformers is depicted in Figure 1.



Cmpd 1 – 6



Cmpd 7



Cmpd 8

Compound Structures for C# in Table 1.

Table 1 Compounds Structure & Bioactivity Data

C#	X	R1	R2	PT Hrs	R/T	VP
1a	4-OCH_3	Et	H	0.08	T	0.0062
1b		CH_3	CH_3	1.00	R	0.0039
1c		Et	Et	1.00	R	0.0037
1d		iPr	iPr	1.17	R	0.0155
1e		C_5H_{10}		0.75	R	0.1486
2a	4-CH_3	Et	H	0.08	R	0.0063
2b		CH_3	CH_3	4.00	R	0.0110
2c		Et	Et	2.83	R	0.0244
2d		iPr	iPr	0.50	T	0.0159
2e		C_5H_{10}		1.00	R	0.0313
3a	H	Et	H	0.58	R	0.0015
3b		CH_3	CH_3	1.67	T	0.0015
3c		Et	Et	4.00	R	0.1015
3d		iPr	iPr	3.00	R	0.0116
3e		C_5H_{10}		3.00	T	0.0559
4a	3-CH_3	Et	H	0.67	R	0.0013
4b		CH_3	CH_3	3.00	R	0.0055
4c*		Et	Et	5.00	T	0.0260
4d		iPr	iPr	2.67	T	0.0151
4e		C_5H_{10}		1.42	R	0.0001
5a	2-Cl	Et	H	0.58	R	0.0006
5b		CH_3	CH_3	5.00	R	0.0076
5c		Et	Et	3.00	R	0.0602
5d		iPr	iPr	1.00	T	0.7728
5e		C_5H_{10}		1.00	R	0.0281
6a	2-OEt	Et	H	0.08	R	0.0003
6b		CH_3	CH_3	2.83	R	0.0264
6c		Et	Et	3.50	R	0.0012
6d		iPr	iPr	1.08	T	0.0144
6e		C_5H_{10}		1.33	R	0.0030
7a		Et	H	1.00	R	0.0058
7b		CH_3	CH_3	2.17	T	0.0020
7c		Et	Et	6.00	R	0.1043
7d		iPr	iPr	1.00	T	0.0014
7e		C_5H_{10}		2.58	R	0.1814
8a		Et	H	0.50	R	0.0168
8b		CH_3	CH_3	3.00	R	0.0136
8c		Et	Et	4.00	R	0.1638
8d		iPr	iPr	2.00	R	0.2843
8e		C_5H_{10}		2.00	R	0.0315

PT = Protection Time; **VP** = Vapor Pressure @ 60° C.
* = DEET.

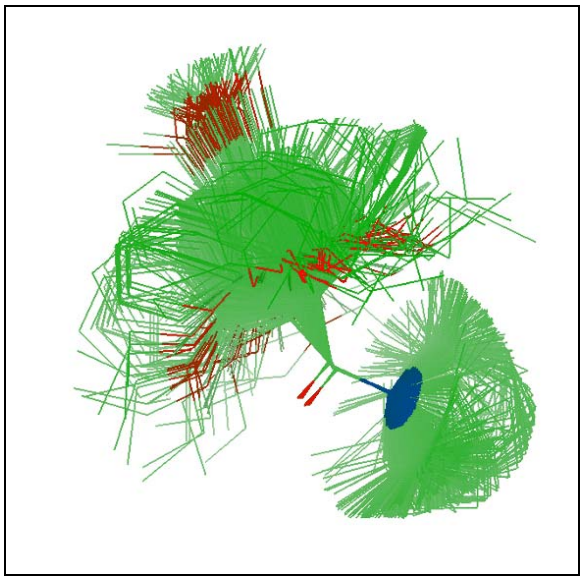


Fig 1 Alignment overlay of all 940 conformers.

The data set was divided in two sets as training and test set of thirty and ten compounds respectively. We computed a total of 127 descriptors comprising of ADME, electrotopological state (Kier; Hall 1992), thermodynamic (Ghose; Crippen 1986), Ghosh and Crippen atom types, Kiers shape indices (Kier 1985), Jurs (Stanton; Jurs 1990) partially charged surface areas, shadow indices (Rohrbaugh; Jurs 1987) and quantum chemical descriptors. The descriptor selection was performed (Yao; Lopes et al. 2003) by first discarding all descriptors with poor correlation with bioactivity ($|r| < 0.1$) followed by discarding the highly collinear descriptors with cross correlation coefficients greater than 0.9. To juxtapose the traditional 3D-QSAR methodology of the global minimum conformation with our novel methodology we computed 3D-QSAR models using the global minimas of the training set compounds with 127 descriptors and also with 30 selected descriptors using genetic function algorithm (GFA), partial least square (PLS) and genetic partial least square (G/PLS) methods. This effort furnished models with non-validated R2 (nvR2) ranging from 0.792 to 0.935 and internal cross validation tests, leave-one-out (q^2_{LOO}), leave-10%-out and leave-20%-out greater than 0.7. However, all models performed poorly when subjected to the rigorous external validation with the test set compounds, they yielding a predictive r2 of 0.349 or less. The contemporary approach of using the global minimum conformation does not furnish good QSAR models probably because the bioactive conformations are quite different than the global minimum conformations. Thus, the novel methodology we have devised to discover the bioactive conformation is by mining through a set of conformations within the energy range of 20 Kcals/mol of the global minimum such that the conformations have a good representation in the 3D space around some putative pharmacophoric moiety for all of the compounds in the training set. The set of all 20-25 conformers of all of the 30 training set compounds totaled to 706 conformations.

The first generation 3D-QSAR model based on the selected 30 descriptors using PLS method for the 706 conformations gave a model with nvR^2 of 0.883, q^2_{LOO} of 0.877 and prediction error sum of squares (PRESS) of 200.06. The predicted residual values of several conformers showed identical values and on closer examination of the descriptor values they were also almost identical. On removal of such 'duplicate' conformers we got a set of 501 conformers, which on PLS analysis furnished the second generation 3D-QSAR model with nvR^2 of 0.879, q^2_{LOO} of 0.869 and PRESS of 135.01. The conformers selected for all of the subsequent generation models were the ones with least residual (Predicted – Actual PT) values. The next generations QSAR models were built by selecting aforementioned number of conformers from their respective previous generation QSAR models. Thus, 10 conformers for the IIIrd generation (300 conformers), 5 conformers for the IVth generation (150 conformers) and 2 conformers for the Vth generation (60 conformers) QSAR models were selected to give nvR^2 of 0.921, 0965 & 0.988, q^2_{LOO} of 0.911, 0.956 & 0.977 and PRESS values of 60.43, 15.12 & 3.10 respectively. For the VIth generation QSAR model the data was divided into two sets with most active PT cut off value of 3.0 hrs and not active PT values of less than 3.0. Thus, for the 9 compounds viz. **C#**(PT): **2b**(4.0), **3c**(4.0), **3d**(3.0), **5b**(5.0), **5c**(3.0), **6c**(3.5), **7c**(6.0), **8b**(3.0) and **8c**(4.0) two conformers were retained and for the remaining 21 training set compounds, the least residual value conformer were selected for the VIth generation 3D-QSAR model. The VIth generation 3D-QSAR model showed nvR^2 of 0.991, q^2_{LOO} of 0.974 and PRESS of 2.565. The final VIIth generation QSAR model can be computed by choosing either one conformer for the nine most active compounds in 2^9 or 512 different ways. The computation of 512 3D-QSAR models using a TCL-based Cerius2 script yielded six VIIth generation models with q^2_{LOO} of 0.67 or larger. The best VIIth generation 3D-QSAR model showed nvR^2 of 0.989, q^2_{LOO} of 0.701 and PRESS value of 20.37. Figure 2 shows the observed and predicted activity plot for the best VII generation QSAR model. The final 3D-QSAR model showed and excellent predictive r2 of 0.845.

The gradual refinement of successively generated 3D-QSAR models computed by selecting the least residual value conformers gives the conformations that best correlate with the observed bioactivity. Thus, we argue that these are indeed the bioactive conformations of the respective compounds. The shapes of these selected 'bioactive conformers' allude to the roles of the various moieties around the putative amide pharmacophore in the mechanism of action as also in the structure activity

relationship. There are three important conclusions about the role of DEET analogs and derivatives in the insect repellency mechanism of action, viz.

1) The 3D-spatial location of the group (phenyl, benzyl & cyclohexyl) attached to the carbonyl C does not have significant effect on the bioactivity, which probably dock with the OBP to form the complex.
2) There is a preferential positioning of the methyl, ethyl, isopropyl etc moieties on the amidic N within a narrow range of 60° to 70°, which probably interacts with the neuronal GPCR in the rate limiting step.
3) The compounds with poor hydrophobic group (e.g. para or ortho methoxy / phenyl / benzyl) cannot dock effectively with the OBP and thus irrespective of the groups on the amidic N exhibit poor repellency activity, which probably also alludes to the competing nature of the OBP and ODE.

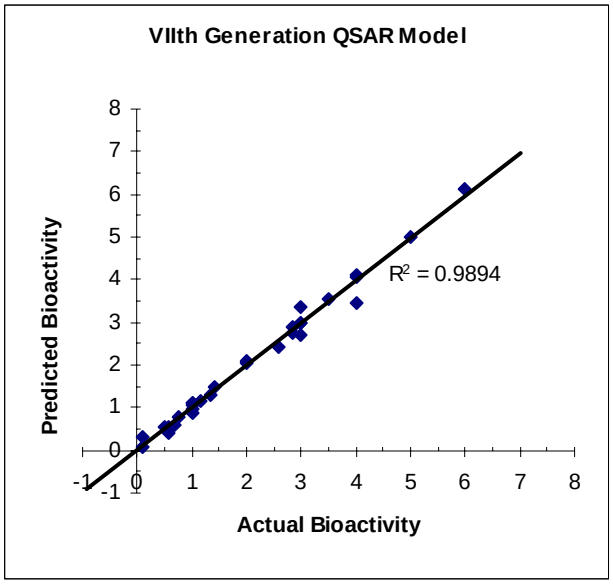


Fig 2 Observed and predicted Bioactivity plot for the best VIIth generation 3D-QSAR model.

The following equation describes the best 3D-QSAR model:

Pred Bioactivity = 0.538 * ADME_Absorption_T2_2D - 0.682 * ADME_BBB_2D - 0.042 * Energy - 0.689 * ADME_BBB_Level_2D - 0.531 * S_dssC - 1.209 * ADME_Solubility_Level - 0.192 * S_aasC - 0.367 * S_ssNH + 0.054 * S_ssO + 0.531* Jurs-FNSA-2 + 0.001 * LUMO_MOPAC + 0.433 * DIPOLE_MOPAC + 0.004 * HF_MOPAC - 0.0001 * Jurs-DPSA-2 - 0.014 * Jurs-DPSA-3 + 1.288 * Jurs-FPSA-1 + 66.492 * Jurs-FPSA-3 + 0.536 * Jurs-RPCS + 12.508 * Jurs-RASA - 0.008 * Shadow-XY - 0.531 * Shadow-nu - 0.285 * Shadow-Xlength - 0.057 * Shadow-Zlength + 0.312 * Density - 0.001 * PMI-mag - 0.074 * Atype_C_5 + 0.196 * Atype_H_47 + 0.097 * Fh2o + 1.515 * JX + 1.299 * Kappa-3-AM - 12.4913

The value and sign of the 3D-QSAR equation coefficients provide a qualitative insight into the correlation of the respective physicochemical (PC) property to the observed protection time. However, the quantitative contribution of any PC property to the protection time can only be judged from both the QSAR equation coefficient and the descriptor value quantifying it. We computed the mean descriptor values for this purpose as the arithmetic average of the descriptor values of all the training set compounds (i.e. MVD = { Σ descriptor value of all training set compounds} / 30). The product of the QSAR equation coefficient (QEC) and the mean descriptor value (MVD) would now provide the contribution of that PC property (CtoBA) to the protection time. (i.e. CtoBA = QSAR coefficient * MVD) Further, the significance of any PC property *vis-à-vis* all of the other PC properties appearing in the QSAR equation can be computed as the ratio of CtoBA to the sum total of all CtoBA. The percentage value of this quotient, is what we have termed as the 'Descriptor Significance Percentage' DSP (i.e. DSP = CtoBA * 100 / Σ abs(CtoBA)). Thus, the DSP values provide a better insight into the quantitative contribution of each of the descriptors to the protection times. The list of descriptors and their QEC, MVD, CtoBA and DSP is shown in Table–2.

The top five descriptors Jurs-RASA, Jurs-FPSA-3, JX, ADME-Solubility level and Shadow-Xlength contribute to 62% of the bioactivity. The largest contribution to the bioactivity is from Jurs-RASA with a positive 25% contribution. Jurs-RASA is defined as the ratio between the total hydrophobic surface area (Jurs-TASA) and the total solvent accessible surface area (Jurs-SASA). This observation is consistent with the first step of the MOA where the odorant molecule binds to the OBP and hydrophobicity or lipophilicity play a key role. ADME-Solubility level with negative 8.5%, Atype_H_47 with positive 3.5% and Fh2o with negative 1.1% also support the role of hydrophobicity in the MOA. This observation is in agreement with the earlier reports (McIver 1981: Suryanarayana; Pandey et al. 1991) that lipophillicity is directly related to repellency. The next largest contribution to bioactivity is from Jurs-FPSA-3 with a positive 11% value. Jurs-FPSA-3 is the quotient of Jurs-PPSA-3 and Jurs-SASA, where Jurs-PPSA-3 is the summation of the products of solvent accessible surface area and partial charge of all positively charged atoms. Thus, the 3D-QSAR model suggests that larger partial positive surface areas and larger partial positive charge along with smaller total solvent accessible surface area would correlate with higher repellency activity. This probably alludes to the second step of the mechanism of action where the odorant-OBP complex binds the neuronal GPCR peptide residues. The diffused or soft positively charged moiety's correlation with increased repellency activity is also corroborated by Jurs-FPSA-1 (Jurs-Fractional Positive Surface Area-1) defined as the

Table 2 Computation of DSP: Descriptor Significance Percentage

Descriptor	QEC	MVD	CtoBA	DSP
Jurs-RASA	12.508	0.882	11.028	25.038
Jurs-FPSA-3	66.492	0.074	4.916	11.161
JX	1.515	2.527	3.828	8.690
ADME_Solubility_Level	-1.210	3.100	-3.750	-8.514
Shadow-Xlength	-0.285	11.888	-3.393	-7.703
Kappa-3-AM	1.299	2.568	3.337	7.577
Energy	-0.042	51.357	-2.164	-4.912
Atype_H_47	0.196	7.933	1.555	3.530
DIPOLE_MOPAC	0.433	3.553	1.537	3.490
ADME_Absorption_T2_2D	0.538	2.736	1.473	3.344
Shadow-nu	-0.531	1.909	-1.014	-2.301
Jurs-FPSA-1	1.288	0.770	0.992	2.252
ADME_BBB_Level_2D	-0.689	1.300	-0.896	-2.034
Jurs - DPSA-3	-0.014	50.052	-0.717	-1.628
Shadow-XY	-0.008	59.513	-0.488	-1.109
Fh2o	0.097	-4.866	-0.473	-1.073
PMI-mag	-0.001	324.800	-0.438	-0.993
Shadow-Zlength	-0.057	6.306	-0.359	-0.815
Density	0.312	1.004	0.313	0.711
Jurs-RPCS	0.536	0.488	0.262	0.595
S_ssNH	-0.367	0.642	-0.236	-0.536
Jurs-FNSA-2	0.531	-0.438	-0.232	-0.528
S_aasC	-0.192	1.148	-0.221	-0.501
Jurs - DPSA-2	0.000	808.217	-0.131	-0.298
S_ssO	0.054	1.394	0.076	0.172
ADME_BBB_2D	-0.682	0.103	-0.071	-0.160
S_dssC	-0.531	0.114	-0.061	-0.138
HF_MOPAC	0.004	-11.390	-0.047	-0.107
Atype_C_5	-0.074	0.533	-0.040	-0.090
LUMO_MOPAC	0.001	0.094	0.000	0.000

QEC - QSAR Model A Equation Coefficient values

MVD - Mean value of descriptors of all training cmpds = (Σdescriptor_value / 30)

CtoBA - Contribution to bioactivity = (QEC * MVD)

DSP - Descriptor Significance Percentage = (CtoBA * 100 / Σ abs(CtoBA))

sum of the solvent accessible surface area of all partial positively charged atoms with a positive 2.3% contribution and the positive 8.7% contribution from the Balaban index JX, which is inversely proportional to the electronegativities and covalent radii of the atoms in the repellent molecules. The fifth largest DSP contribution of negative 7.7% comes from the descriptor Shadow-Xlength, which is the measure of the projection of the molecule on the x-axis. The contribution of other shadow indices are shadow-Zlength (projection measure on the z-axis) of negative 0.8%, shadow-XY (the area of the shadow of the molecule in the XY plane) of negative 1.1% and shadow-nu (ratio of the largest to the smallest shadow measures) of negative 2.3%. This combination of shadow indices indicate that elongated rectangular box (parallelepiped) like molecular structure correlate with repellency activity. This alludes to the shape of the binding pocket of the OBP involved in the first step of the mechanism of action.

2.2 AMP based antibacterials 3D-QSAR

We selected 28 AMPs (Table-3) with diverse activity against *Staphylococcus aureus* ME/GM/TC resistant (ATCC 33592) (***SA***) and *Mycobacterium ranae* (ATCC 110) (***MR***) bacteria for this 3D-QSAR study. (Hicks; Bhonsle et al. 2007) Each peptide was constructed using the Biopolymer module of InsightII, energy minimized using the Steepest Descent Algorithm (Levitt; Lifson 1969) and subjected to a brief (1000 cycles) MD simulation followed by exhaustive minimization to give the local minimum conformation of the peptide. The conformational search was done using Monte Carlo Algorithm (Chang 1989). The conformations were clustered using Root Mean Squares (RMS) difference of torsion angles of the peptides (Accelrys 2005). We selected sets of cluster nuclei that gave the best 3D spatial representations, which were 20-30 conformers for some and 30-40 conformers for the rest peptides. All of the conformers of all the peptides were aligned and added to a study table for descriptor computation with default settings. The correlation matrix was computed for all the descriptor values of all the conformers of all the peptides to obtain the cross correlation coefficients and correlation with bioactivity. The descriptors that showed very poor correlation with bioactivity ($|r| < 0.01$) were removed. The cross correlation matrix showed that 33 descriptors exhibited very high cross correlation coefficient values ($|r| > \sim 0.9$). Removal of these highly cross correlated descriptors left behind the final 22 and 21 descriptors for ***SA*** and ***MR*** QSAR models. The list of these final descriptors for the two 3D-QSAR models is presented in Table-4. Our novel, gradual and stepwise bioactive conformer mining methodology mines the clustered conformations and identifies the bioactive conformers that most closely correlate with the observed bioactivity.

Table 3: Peptide amino acid sequence and their anti-bacterial activity

C#	Amino Acid Sequence	*SA* μM†	*MR* μM†
1	NH_2KLTcOcKTcOcFTcOcKTcOcFTcOcKTcOcKRNH_2	10	30
2	AcGFTcOcGKTcOcGFTcOcGKTcKKKK-NH_2	3	10
3	NH_2GFTcOcGKTcOcGFTcOcGKTcKKKK-NH_2	10	10
4	NH_2KLTcOcGKTcOcGFTcOcGKTcKKKK-NH_2	30	3
5	AcFTcOcKTcOcFTcOcKTcKKKKNH_2	3	30
6	AcFTcOcKTcOcFTcOcKTcKKKKKKNH_2	3	3
7	AcGabaFTcOcGabaKTcOcGabaFTcOcGabaKTcKKKKKNH_2	100	10
8	AcβAlaFTcOcβAlaKTcOcβAlaFTcOcβAlaKTcKKKKNH_2	10	1
9	AcAhxFTcOcAhxKTcOcAhxFTcOcAhxKTcKKKKNH_2	10	3
10	AcGabaFTcOcGabaKTcOcGabaFTcOcGabaKTcKKKKKNH_2	30	3
11	AcGTcOcKTcOcGTcOcKTcKKKKNH_2	10	3
12	AcGFOcGKOcGFOcGKKKKKNH_2	10^5	100
13	AcGFGOcGKGOcGFGOcGKGKKKKNH_2	10^5	100
14	AcGFTcGKTcGFTcGKTcKKKKNH_2	10^5	30
15	AcGFTcGGKTcGGFTcGGKTcKKKKNH_2	10^5	30
16	AcGFFOcGKFOcGFFOcGKFKKKKNH_2	10	10
17	AcGFTcOcGKTcOcGFTcOcGKTcKKKKKNH_2	3	3
18	AcGFTcOcGKTcOcGFTcOcGKTcOOOONH_2	10	10
19	AcGFpaTcOcGKTcOcGFpaTcOcGKTcKKKKNH_2	10	3
20	AcGFTcOcGOTcOcGFTcOcGOTcOOOONH_2	3	10
21	AcGFTcOcGKTcOcGFTcOcGKTcKKKKCONH$CH_2CH_2NH_2$	3	10
22	AcGFTcOcGKTcOcGFTcOcGKTcKKKKCONH$CH_2CH_2CH_2NH_2$	10	10
23	NH_2ELMNSTcOcGLTcOcGKTcOcGLTcOcGKTcOcELMNSNH_2	10^5	10^5
24	NH_2GKGLTcOcGKTcOcGFTcOcGKTcOcGFTcOcGKTcOcGKRNH_2	10	NT
25	NH_2GKGLTcOcGRTcOcGFTcOcGRTcOcGFTcOcGRTcOcGKRNH_2	10	10^5
26	NH_2GKGLTcOcGLTcOcGKTcOcGLTcOcGKTcOcGLTcOcGLRNH_2	100	NT
27	NH_2GKGLTcOcGKTcOcGLTcOcGKTcOcGLTcOcGKTcOcGKRNH_2	10	NT
28	NH_2GKGLTcOcFKTcOcKFTcOcFKTcOcKFTcOcFKTcOcFKRNH_2	30	10^5

C# = Compound #; Tc = Tetrahydroisoquinolinecarboxylic acid; Oc = Octahydroindolecarboxylic acid; Fpa = 4Fluoro Phenylalanine; Gaba = γAminobutyric acid; Ahx = εAminohexanoic acid; Ac = Acetyl; NT = Not Tested; † Since all analogs were screened in the concentration range of 0.1 μM to 100 μM, compounds with MIC of $\leq$ 100 μM, were deemed to be active compounds. For QSAR purposes all inactive compounds were assigned an MIC of 1.0 M.

Table 4: A Rank ordering of the Physicochemical Properties defining anti-bacterial activity

Physico-chemical property	*Staphylococcus aureus* QSAR_DSP	Physico-chemical property	*Mycobacterium ranae* QSAR_DSP
Jurs-FPSA-1	29.347	Density	-30.784
Density	-16.01	Jurs-RASA	16.827
Jurs-TASA	-14.762	Jurs-PPSA-1	-15.494
Jurs-PNSA-1	10.54	Jurs-TPSA	10.218
Jurs-RASA	7.886	Jurs-RPSA	-5.444
Jurs-SASA	4.12	Hbond donor	-3.905
Jurs-DPSA-2	3.093	Hbond acceptor	3.729
Jurs-PNSA-2	-2.911	Jurs-FPSA-1	-3.409
Jurs-RPSA	-2.492	Fcharge	2.892
Rotlbonds	-2.164	Jurs-PNSA-1	-1.244
Hbond acceptor	1.91	RadOfGyration	1.164
Jurs-FPSA-3	1.709	Rotlbonds	-1.156
Fcharge	-0.742	Apol	1.148
Jurs-RPCG	-0.726	Jurs-PPSA-2	1.016
Jurs-PPSA-1	0.555	Jurs-PNSA-2	-0.632
Jurs-FNSA-3	-0.426	Jurs-RNCG	0.4
Dipole-mag	0.162	Dipole-mag	0.298
RadOfGyration	-0.127	Jurs-FNSA-3	-0.127
Jurs-RPCS	-0.126	AlogP	0.051
Hbond donor	0.113	Conformer Energy	0.037
Jurs-DPSA-3	0.053	Jurs-RPCG	-0.024
AlogP	-0.026	Jurs-DPSA-2	0

Thus, the bioactive conformer mining method, over seven iterative generations(Bhonsle; Bhattacharjee et al. 2007) resulted in two conformers each for the 12 peptides (C# **1, 2, 5, 6, 17, 19, 20, 21, 22, 24, 25 & 27** for ***SA*** and C# **4, 6, 8, 9, 10, 11, 17, 18, 19, 20, 21 & 22** for ***MR***) and one conformer for each the remaining 16 peptides. There are 4096 (2^{12}) ways to select the best set of 12 conformers, from the 24 conformers. The 4096 eighth generation models were computed employing a Tcl-based Cerius2 script. The final ***SA*** and ***MR*** 3D-QSAR models showed non-validated r^2 of 0.988 and 0.997, leave-one-out cross-validated r^2 of 0.839 and 0.997 with PRESS values of 22.92 and 29.19 respectively. The 3D-QSAR equations for predicting the activity against ***SA*** is given in equation 1 and that against ***MR*** is given in equation 2. The correlation plots of the predicted *vs.* the observed anti-bacterial activities of these two 3D-QSAR models are shown in Figure 3. Internal validation (cross-validation) tests of the final 3D-QSAR models were performed at two levels. Both of the models showed $q^2_{LOO} > 0.83$ for the leave-one-out (LOO) cross-validation tests. For the leave-10%-out or leave-three-out (L10O) cross-validation tests,

SA model showed q^2_{L10O} of 0.875, whereas ***MR*** model showed q^2_{L10O} value of 0.537. We performed randomization tests of ninety-nine trials each at 99% confidence level for ***SA*** and ***MR*** 3D-QSAR models. None of the random r values were found to be larger than the non-random r values for either the ***SA*** or the ***MR*** models. The mean random r value for the ***SA*** model was 0.572 (r^2 = 0.327), and for the ***MR*** model was 0.617 (r^2 = 0.380). This proved that the ***SA*** and ***MR*** QSAR models are not obtained by chance.

EQUATIONS

The *SA* 3D-QSAR model is described by equation: 1

SA Predicted Activity = [(-1.49592 * Fcharge) + (0.0098147 * Dipole-mag) + (0.013993 * Jurs-SASA) + (0.00233 * Jurs-PPSA-1) + (0.187647 * Jurs-PNSA-1) + (0.0021686 * Jurs-PNSA-2) + (0.00036919 * Jurs-DPSA-2) + (0.0015025 * Jurs-DPSA-3) + (438.251 * Jurs-FPSA-1) + (267.258 * Jurs-FPSA-3) + (120.432 * Jurs-FNSA-3) - (715.316 * Jurs-RPCG) - (12.8649 * Jurs-RPCS) - (0.065752 * Jurs-TASA) - (125.513 * Jurs-RPSA) + (125.513 * Jurs-RASA) - (183.99 * Density) + (1.03397 * Hbond acceptor) + (0.039473 * Hbond donor) - (0.306856 * Rotlbonds) + (0.114808 * AlogP) - (0.10004 * RadOfGyration) - 225.589]

The *MR* QSAR model is described by equation: 2

MR Predicted Activity = [(-0.0083585 * Conformer Energy) + (2.05758 * Fcharge) + (5.3259e-05 * Apol) + (0.0061422 * Dipole-mag) - (0.023941 * Jurs-PPSA-1) - (0.008252 * Jurs-PNSA-1) + (5.5381e-05 * Jurs-PPSA-2) + (0.00018566 * Jurs-PNSA-2) - (18.282 * Jurs-FPSA-1) + (13.321 * Jurs-FNSA-3) - (8.46841 * Jurs-RPCG) + (66.6262 * Jurs-RNCG) + (0.052889 * Jurs-TPSA) - (96.9761 * Jurs-RPSA) + (96.9761 * Jurs-RASA) - (127.577 * Density) + (0.768698 * Hbond acceptor) - (0.498282 * Hbond donor) - (0.060764 * Rotlbonds) - (0.075759 * AlogP) + (0.337835 * RadOfGyration) + 110.841]

The seventeen physiochemical properties common to the ***SA*** and ***MR*** 3D-QSAR models are shown in Table 4. The five physicochemical properties specific to the ***SA*** QSAR model are Jurs–Fractional-Positive-Surface-Area–3 (Jurs-FPSA-3), Jurs-Relative-Positive-Charge-Surface-area (Jurs-RPCS), Jurs-Differential-Positively-charged-Surface-Area-3 (Jurs-DPSA-3), Jurs-total-Solvent-Accessible-Surface-Area (Jurs-SASA) and Jurs-TotAl-hydrophobic-Surface-Area (Jurs-TASA). While the five physicochemical properties specific to the ***MR*** QSAR model are sum-of-all-atomic-polarizabilities (Apol), Conformer Energy, Jurs-Partial-Positively-charged-Surface-Area-2 (Jurs-PPSA-2), Jurs-Relative-Negative-CharGe (Jurs-RNCG), and Jurs-Total-Polar-Surface-Area (Jurs-TPSA). The commonality of physicochemical properties shows the minimal requirement for activity

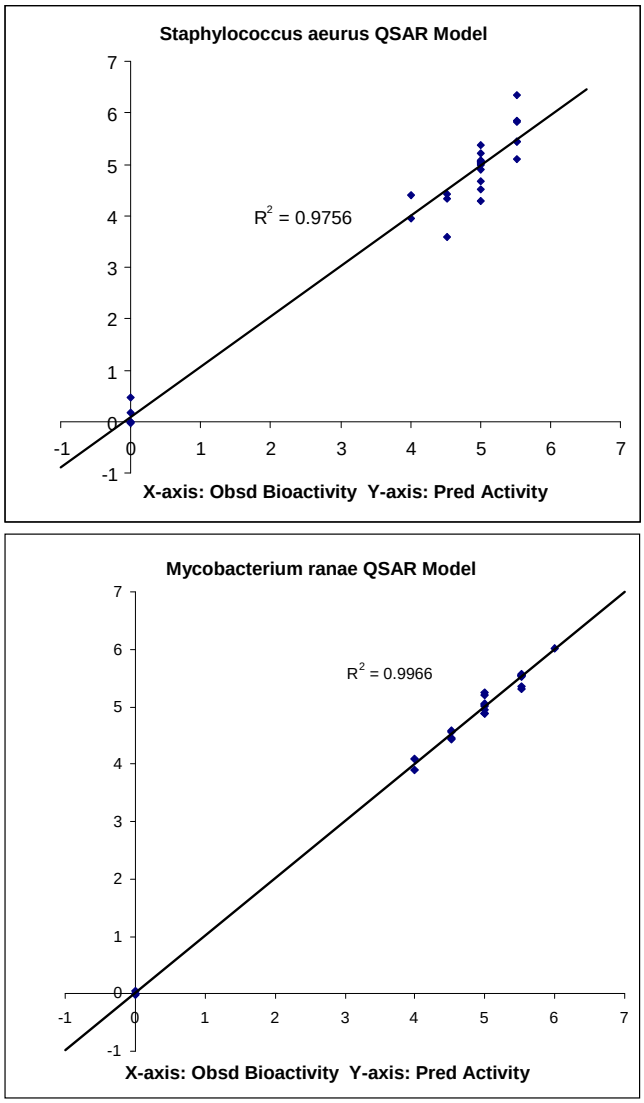


Fig 3 The correlation plot of predicted *vs.* observed anti-bacterial activities of the two 3D-QSAR models.

against ***SA*** and ***MR***. The importance of electrostatic potential for the AMP bioactivity can be seen from the physicochemical properties such as Dipole-magnitude (Dipole-mag), Formal charge (Fcharge), Jurs-Fractional-Negatively-charged-Surface-Area (Jurs-FNSA-3), Jurs-Relative-Polar-Surface-Area (Jurs-RPSA), Jurs–Fractional-Positive-Surface-Area–1 (Jurs-FPSA-1), Jurs–Fractional-Negative-Surface-Area–1 (Jurs-PNSA-1), Jurs–Fractional-Negative-Surface-Area–2 (Jurs-PNSA-2), Jurs–Partially-Positive-Surface-Area–1 (Jurs-PPSA-1), and Jurs-Relative-Positive-CharGe (Jurs-RPCG). While the significance of the AMP molecular shape for bioactivity is evident from the physicochemical properties such as molecular Density (Density), number-of-H-bond-acceptors (H-bond acceptor), Jurs-RelAtive-hydrophobic-Surface-Area (Jurs-RASA), number-of-H-bond-donor (H-bond donor), molecular-Radius-Of-Gyration

(RadOfGyration), and number-of-Rotatable-bonds (Rotlbonds). The importance of amphipathicity is alluded to by the physicochemical properties such as Jurs-RASA, Jurs-RPSA, and AlogP. The top six descriptors (DSP) *viz*. Jurs-FPSA-1 (29.35%), Density (-16.01%), Jurs-TASA (-14.76%), Jurs-PNSA-1 (10.54%), Jurs-RASA (7.89%), and Jurs-SASA (4.12%) account for 82% of the ***SA*** predicted activity. The correlation of non-polar surface area to bioactivity is evident from the descriptors such as Jurs-TASA with –14.76% DSP contribution and Jurs-RASA with 7.89% DSP contribution. The significant descriptors accounting for 82% of ***MR*** predicted activity are Density (-30.78%), Jurs-RASA (16.83%), Jurs-PPSA-1 (-15.49%), Jurs-TPSA (10.22%), Jurs-RPSA (-5.44%), and H-bond donor (-3.91%). The correlation of the polar surface area to the ***MR*** bioactivity is evident from the descriptors Jurs-PPSA-1 with –15.5% DSP, Jurs-TPSA with 10.5% DSP contribution & Jurs-RPSA with –5.44% DSP contribution. The hydrophobicity and hydrophilicity correlation with the ***MR*** bioactivity is shown by the descriptors Jurs-RASA with 16.8% DSP contribution, and H-bond donor with –3.9% DSP. The contribution of shape to ***MR*** predicted bioactivity comes from the descriptor Density with –30.78% DSP contribution.

3. CONCLUSION

The 3D-QSAR modeling efforts presented herein demonstrate the utility and advantages of the novel bioactive confirmation mining methodology in the quest of predictive 3D-QSAR models.

4. REFERENCES

Accelrys, Inc., 2005: Performing a conformaitonal analysis. *Cerius2 Version 4.10L Manual*, 49.

Bhonsle, J. B., A. K. Bhattacharjee, and R. K. Gupta, 2007: Development of QSAR models for insect repellent amides. *J Mol Model*, **13,** 179-208.

Blondelle, S. E., K. Lohner, and M.-I. Aguilar, 1999: Lipid-induced conformaiton and lipid-binding properties of cytolytic and antimicrobial peptide: determination and biological specificity. *Biochimica et Biophysica Acta*, **1462,** 89-108.

Brewste, D., 2001: The story of mankind's deadliest foe. *BioMedical Journal*, **323,** 289.

Brogden, K. A., 2005: Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? *Nature Reviews Microbiology*, **3,** 238-250.

Chang, G., Guida, W.C., Still, W.C., 1989: An internal coordinate Monte Carlo method for searching conformational space. *J. Am. Chem. Soc.*, **111,** 4379-4386.

Dennison, S. R., J. Wallace, F. Harris, and D. A. Phoenix, 2005: Amphiphilic α-helical antimicrobial peptides and thier structure/function relationships. *Protein and peptide Letters*, **12,** 31-39.

Ghose, A. K. and G. M. Crippen, 1986: Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. I. Partition coefficients as a measure of hydrophobicity. *Journal of Computational Chemistry*, **7,** 565-77.

Hicks, R. P., J. B. Bhonsle, D. Venugopal, B. W. Koser, and A. J. Magill, 2007: De novo design of selective antibiotic peptides by incorporation of unnatural amino acids. *J Med Chem*, **50,** 3026-36.

Justice, R. W., H. Biessmann, M. F. Walter, S. D. Dimitratos, and D. F. Woods, 2003: Genomics spawns novel approaches to mosquito control. *BioEssays*, **25,** 1011-1020.

Kier, L. B., 1985: A shape index from molecular graphs. *Quan Structure-Activity Relationships*, **4,** 109-16.

Kier, L. B. and L. H. Hall, 1992: *The Electrotopological State Index: An Atom-Centered Index for QSAR*. Vol. 22, *Advances in Drug Research*, Academic Press, 205.

Levitt, M. and S. Lifson, 1969: Refinement of protein conformations using a macromolecular energy minimization procedure. *J Mol Biol*, **46,** 269-79.

Marsali, M. and J. Gasteiger, 1980: Charge distribution from molecular topology and orbital electronegativity. *Journal Croatica Chimica Acta*, **53,** 601-614.

Mayo, S. L., B. D. Olafson, and W. A. I. Goddard, 1990: DREIDING: A generic force field. *J. Phys. Chem.*, **94,** 8897-8909.

McIver, S. B., 1981: A model for the mechanism of action of the repellent DEET on Aedes aegypti (Diptera:Culicidae). *J Medical Entomology*, **18,** 357-61.

Rohrbaugh, R. H. and P. C. Jurs, 1987: Descriptions of molecular shape applied in studies of structure/activity and structure/property relationships. *Analytica Chimica Acta*, **199,** 99-109.

Stanton, D. T. and P. C. Jurs, 1990: Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure-property relationship studies. *Analytical Chemistry*, **62,** 2323-9.

Suryanarayana, M. V. S., K. S. Pandey, S. Prakash, C. D. Raghuveeran, R. S. Dangi, R. V. Swamy, and K. M. Rao, 1991: Structure-activity relationship studies with mosquito repellent amides. *Journal of Pharmaceutical Sciences*, **80,** 1055-7.

Yao, S. W., V. H. Lopes, F. Fernandez, X. Garcia-Mera, M. Morales, J. E. Rodriguez-Borges, and M. N. Cordeiro, 2003: Synthesis and QSAR study of the anticancer activity of some novel indane carbocyclic nucleosides. *Bioorg Med Chem*, **11,** 4999-5006.

DISCLAIMER