J Mol Model (2007) 13:1099–1108
DOI 10.1007/s00894-007-0234-3

ORIGINAL PAPER

CoMFA study of distamycin analogs binding to the minor-groove of DNA: a unified model for broad-spectrum activity

Santosh A. Khedkar · Alpeshkumar K. Malde · Evans C. Coutinho

Received: 30 December 2006 / Accepted: 18 July 2007 / Published online: 10 August 2007

Abstract A 3D-QSAR analysis has been carried out by comparative molecular field analysis (CoMFA) on a series of distamycin analogs that bind to the DNA of drug-resistant bacterial strains MRSA, PRSP and VSEF. The structures of the molecules were derived from the X-ray structure of distamycin bound to DNA and were aligned using the *Database alignment* method in *Sybyl*. Statistically significant CoMFA models for each activity were generated. The CoMFA contours throw light on the structure activity relationship (SAR) and help to identify novel features that can be incorporated into the distamycin framework to improve the activity. Common contours have been gleaned from the three models to construct a unified model that explains the steric and electrostatic requirements for antimicrobial activity against the three resistant strains.

Keywords Broad-spectrum antibacterials · Distamycin analogs · DNA minor-groove binders · Drug-resistant bacterial strains (MRSA, VSEF, PRSP) · 3D-QSAR · Unified CoMFA model

Introduction

The emergence of bacterial resistance to commonly used antibiotics is a severe health care problem and has instigated the search for newer classes of antibacterial agents [1, 2]. In particular, antibiotics with a novel mode of action are needed to overcome cross-resistance to known drugs. Only *one* novel antibiotic class, the oxazolidinones, has been introduced in the last 40 years [3]. In recent times, the diaryl quinolines as ATP synthase inhibitors have arisen as a second novel anti-tubercular class of drugs [4].

In the last few years, DNA has been the new target for development of antibiotics. A number of natural products like distamycin A [5] and actinomycin D [6] are known to interact with bacterial DNA, affecting DNA replication and RNA transcription. Analogs of distamycin and netropsin are able to bind with sequence specificity to the minor-groove of DNA [7]. DNA minor-groove binding molecules structurally related to distamycin A, have shown potent in vitro antibacterial activity against clinically relevant drug-resistant pathogens.

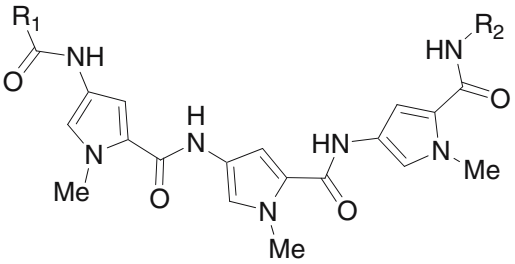
A 3D-QSAR by comparative molecular field analysis (CoMFA) [8] of a series of distamycin analogs that bind to the minor-groove of DNA of drug-resistant bacterial strains such as methicillin-resistant *Staphlococcus aureus* (MRSA), penicillin-resistant *Streptococcus pneumoniae* (PRSP) and vancomycin-susceptible *Enterococcus faecalis* (VSEF), that was undertaken to help understand the essential structural requirements for anti-bacterial activity against these pathogens is reported in this paper. A unified model that represents the structural requirements for broad-spectrum activity against the above-mentioned microbial strains has been formulated.

Biological data

A training set of 33 molecules (molecules **1** to **33** in Table 1) was designed that maintains the structural

S. A. Khedkar · A. K. Malde · E. C. Coutinho (✉)
Department of Pharmaceutical Chemistry,
Bombay College of Pharmacy,
Kalina, Santacruz (E),
Mumbai 400 098, India
e-mail: evans@bcpindia.org

Table 1 Dataset used for the CoMFA study



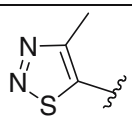
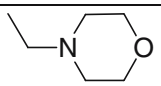
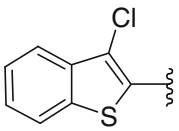
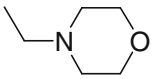
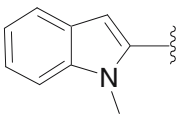
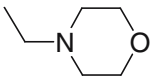
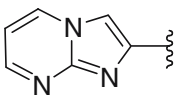
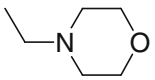
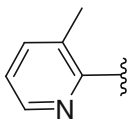
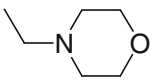
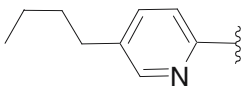
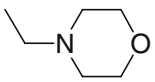
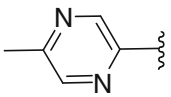
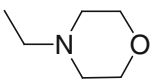
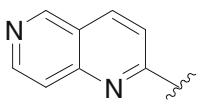
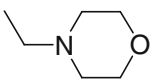
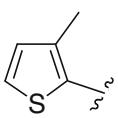
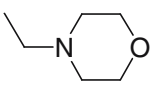
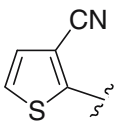
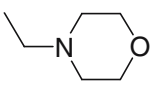
Mol ID[a]	R_1	R_2	Antibacterial activity (pMIC) against		
			MRSA[b]	PRSP[c]	VSEF[d]
1			4.91	5.49	4.89
2			6.44	6.74	5.97
3			5.21	4.91	4.31
4			4.91	6.11	4.90
5			4.53	5.79	4.89
6			4.92	5.52	4.31
7			5.49	6.39	6.09
8			4.91	6.42	5.21
9			4.59	5.19	4.59
10			5.20	6.10	5.50

Table 1 (continued)

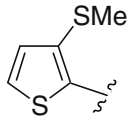
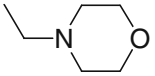
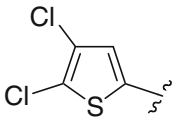
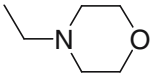
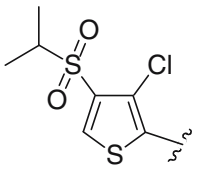
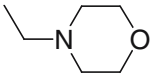
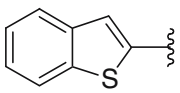
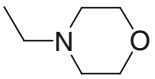
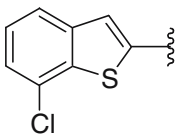
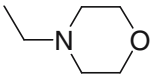
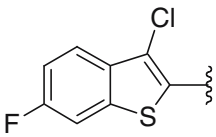
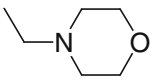
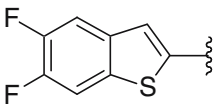
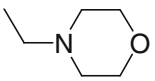
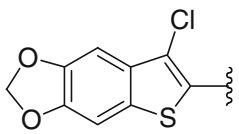
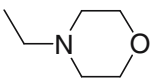
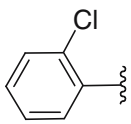
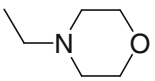
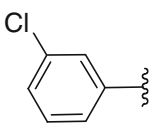
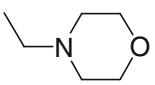
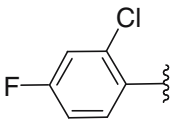
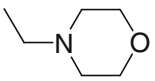
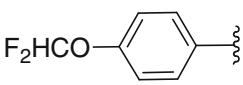
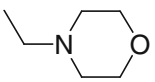
11			5.21	6.72	4.91
12			4.93	5.23	4.93
13			5.27	6.00	5.40
14			5.22	5.82	4.31
15			4.94	5.54	4.64
16			6.15	6.75	6.45
17			5.86	6.16	4.96
18			7.38	7.38	7.38
19			4.30	4.30	4.60
20			4.90	4.90	4.90
21			4.61	5.21	4.91
22			5.52	7.03	5.22

Table 1 (continued)

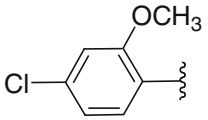
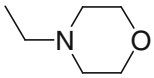
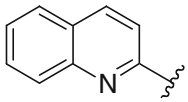
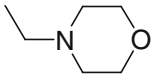
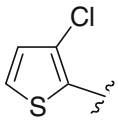
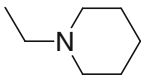
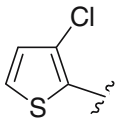
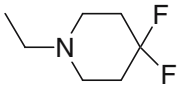
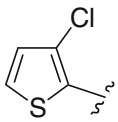
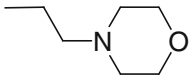
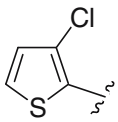
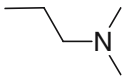
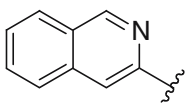
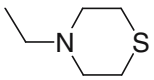
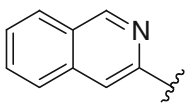
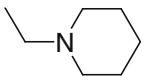
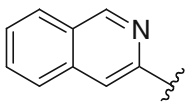
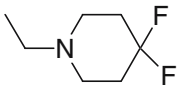
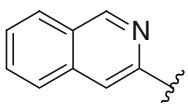
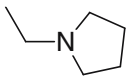
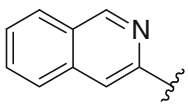
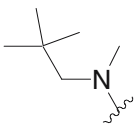
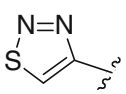
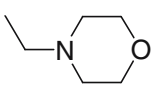
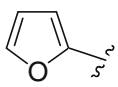
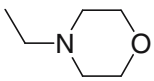
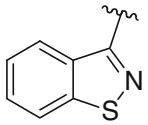
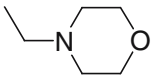
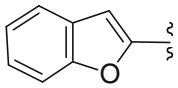
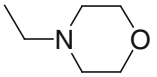
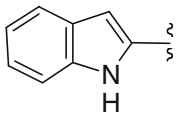
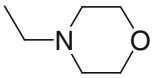
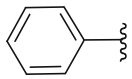
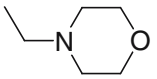
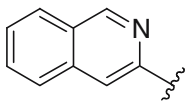
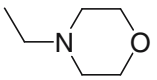
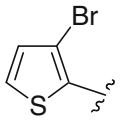
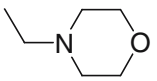
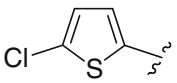
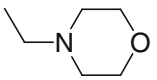
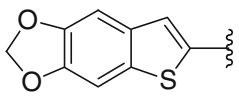
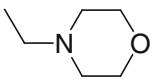
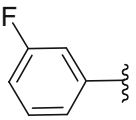
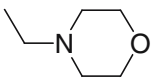
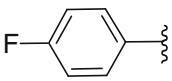
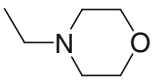
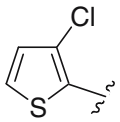
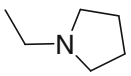
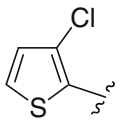
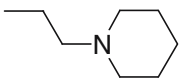
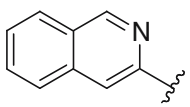
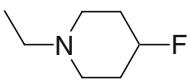
23			4.62	6.12	4.92
24			5.51	5.81	4.31
25			6.11	6.41	6.11
26			5.53	6.13	5.53
27			4.91	5.52	5.21
28			4.89	5.19	4.58
29			6.73	7.33	6.43
30			6.42	7.32	7.02
31			6.14	6.74	6.44
32			6.10	7.01	5.50
33			5.81	5.81	5.81
34			4.28	4.51	5.18
35			4.27	5.77	4.27

Table 1 (continued)

36			5.82	6.72	5.82
37			5.51	6.11	5.21
38			4.90	5.51	4.90
39			4.27	4.58	4.27
40			6.12	7.03	5.84
41			5.23	6.44	6.42
42			5.21	5.51	5.51
43			6.15	6.45	5.85
44			4.29	5.49	4.29
45			4.59	4.59	4.89
46			4.89	5.5	4.59
47			4.91	5.21	5.21
48			6.43	7.33	6.73

[[a] molecules **1** – **33**: training set and molecules **34** – **48**: test set; [b] methicillin-resistant *Staphlococcus aureus*; [c] penicillin-resistant *Streptococcus pneumoniae*, [d] vancomycin-susceptible *Enterococcus faecalis*]

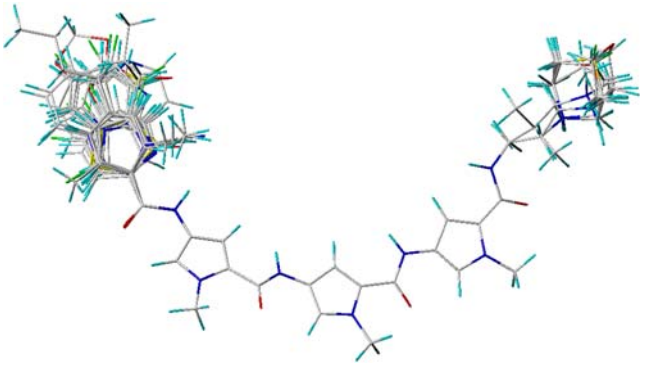


Fig. 1 Alignment of the training set molecules

diversity as well as the activity distribution against all the three types of microorganisms, on the basis of the Tanimoto coefficient [9]. Biological response [7] is the in vitro antimicrobial activity against MRSA, PRSP, and VSEF. The MIC values reported as μg/mL were converted to the negative logarithm of MIC in moles/L (pMIC). The test set of 15 molecules (molecules **34** to **48** in Table 1) was designed on the same basis as the training set.

Computational details

All molecules in the training and test sets were constructed in silico using the X-ray structure of distamycin (PDB code: 2DND) as the template, with the sketcher module in *SYBYL 7.1* [10] molecular modeling package installed on a Pentium PC with the Linux OS (RedHat Enterprise WS 2.3). The molecules were assigned the MMFF94 charges and minimized by Powell's method using the MMFF94 force field [11], terminating with a maximum value of the gradient as 0.5 kcal mol^{-1} $Å^{-1}$. A distance dependent dielectric was used and the three common pyrrole rings were constrained during the minimization. The low energy structures obtained were used for subsequent alignment. The molecules were aligned (Fig. 1) with reference to the highest active molecule using the amide-NMP-amide-NMP-amide-NMP-amide-CH_2-CH_2 substructure (NMP-*N*-methylpyrrole) which is common to all molecules, using the *Database Alignment* method in *Sybyl*.

The steric and electrostatic CoMFA fields were calculated at each lattice intersection of a regularly spaced grid of 2.0 Å in all three dimensions. The van der Waals potential and Coulombic energy were calculated with the Tripos force field. A distance-dependent dielectric constant of 1.0 was used for the electrostatics. A sp^3 carbon atom with +1.0 charge was used as the probe. The steric field was truncated at points where the value exceeded ±30.0 kcal mol^{-1}, and the electrostatic fields were ignored at the lattice points with very high steric interactions.

Partial least square (PLS) analysis

PLS method was used to correlate the CoMFA fields with the activity values. Cross-validation analysis was performed using samples-distance partial least squares (SAMPLS) [12]. The optimum number of components was selected for the model for which the q^2 was the highest and the standard error of prediction (SDEP) was lowest. Equal weights were assigned to the steric and electrostatic fields by the standard CoMFA scaling option. To speed up the analysis and reduce noise, columns were filtered whose value did not vary by at least 2.0 kcal mol^{-1}. Final analysis was performed to calculate the conventional r^2 using the optimum number of components. To further assess the robustness and statistical confidence of the derived models, bootstrapping [13] analysis for 100 runs was performed. Models with cross-validation (q^2) value above 0.3 were sought since at this value the probability of chance correlation is less than 5% [14].

CoMFA contour maps

Contour maps were generated as a scalar product of coefficients and standard deviation (StDev*Coeff) associated with each column. Favored and disfavored levels, fixed at 80% and 20%, respectively, were used to display the steric and the electrostatic fields. The contours for steric fields are shown in green (more bulk favored) and yellow (less bulk favored), while the electrostatic field contours are shown in red (more electronegative substituents favored) and blue (more electropositive substituents favored).

Table 2 Summary of CoMFA models built for antimicrobial activity against MRSA, PRSP and VSEF

	MRSA	PRSP	VSEF
N	6	6	6
r^2	0.97	0.98	0.97
q^2	0.68	0.66	0.61
SEE	0.14	0.15	0.15
F-value	137	218	156
r^2_{bs}	0.98	0.98	0.99
r^2_{pred}	0.85	0.86	0.76
SD	0.007	0.008	0.003
Contributions (%)			
Steric	0.71	0.72	0.70
Electrostatic	0.29	0.28	0.30

[q^2=cross-validated correlation coefficient (r^2_{cv}) using *SAMPLS*; *N*=optimum number of components; r^2=conventional (non cross-validated) correlation coefficient; *SEE*=standard error of estimate, r^2_{pred}=predictive (test molecules) correlation coefficient; r^2_{bs}=correlation coefficient after 100 runs of bootstrapping analysis; SD=standard deviation from 100 bootstrapping runs]

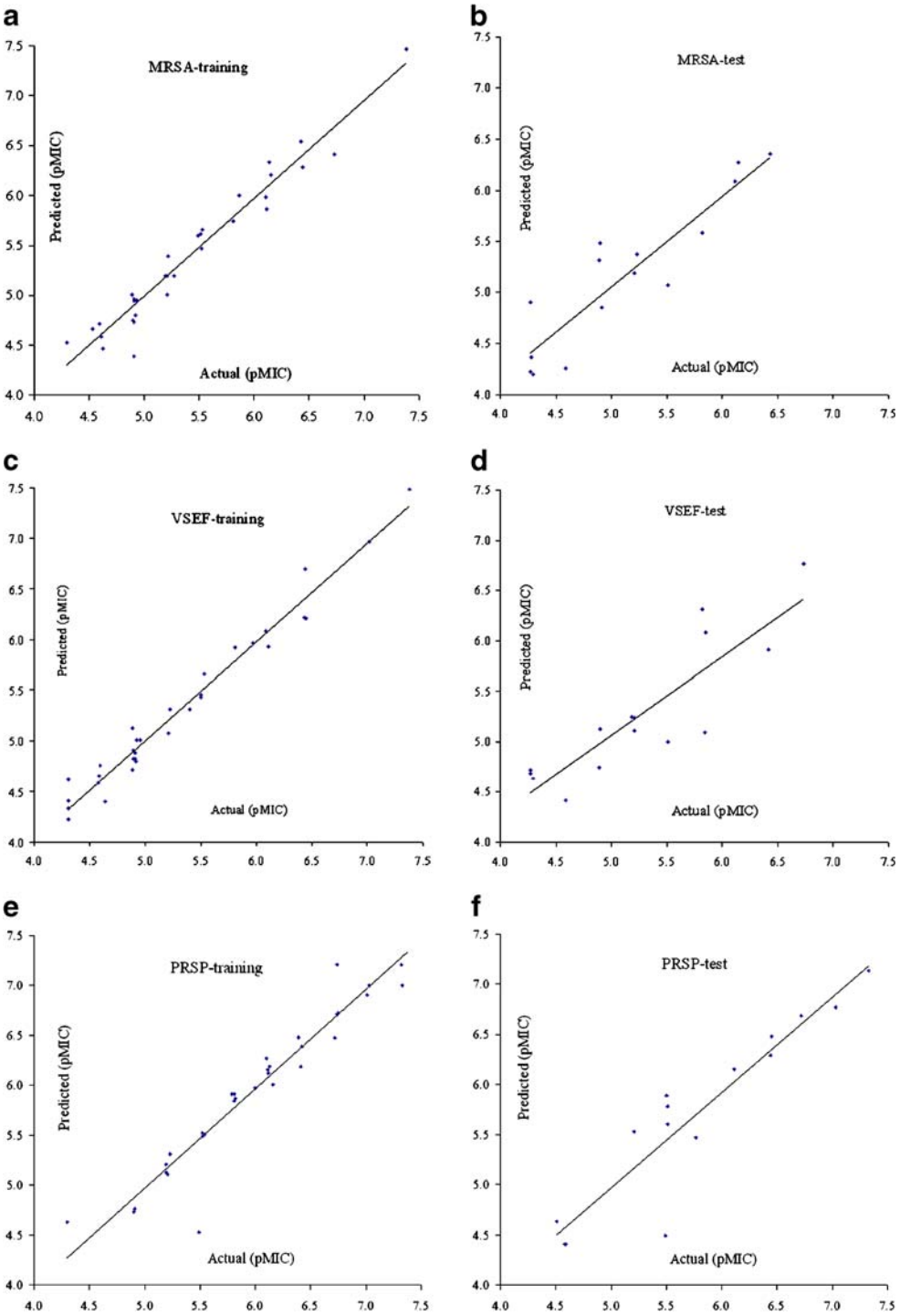


Fig. 2 Plots of actual versus predicted activities (pMIC) of training and test set molecules for MRSA, PRSP and VSEF

Results and discussion

The dataset was divided so as to include 33 molecules in the training set and 15 molecules in the test set. The training set encompasses molecules whose biological activity ranges from 4.3–7.4 log units and the biological activity of the test set molecules has a similar range. The alignment of the molecules in the training set used to generate the CoMFA model is shown in Fig. 1. The statistics of the CoMFA models is given in Table 2. Models were developed independently for each type of antimicrobial activity. The CoMFA models show a good predictive q^2 of 0.68, 0.66 and 0.61, respectively, for molecules with antimicrobial activity against MRSA, PRSP, and VSEF (Table 2). The conventional correlation coefficient (r^2) of 0.96 or more indicates a good linear correlation between the biological activity and the CoMFA descriptors. The predictive power of the models is indicated by the r^2_{pred}

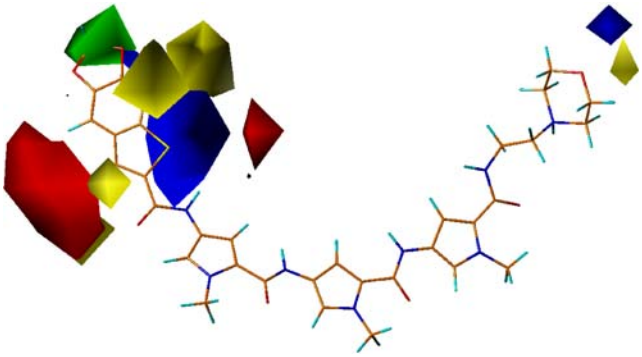


Fig. 3 CoMFA contour map for steric and electrostatic fields around molecule **18** with antimicrobial activity against MRSA. The green colored contour favors steric bulk while sites where steric bulk is disfavored are shown in yellow. The red contour shows regions where electronegative substituents are favored, while the blue contour is associated with positions where electropositive substituents improve activity

values. The overall statistics of the model reflect good correlation and predictive power. The plots of experimental *vs* predicted activities (pMIC) of molecules in the training and test sets against MRSA, VSEF, and PRSP are shown in Fig. 2. The CoMFA model, with its hundred or thousand of the terms, is generally represented as a 3D 'coefficient contour'. Colored contours in the map represent those areas in 3D space where changes in the steric and electrostatic fields of a compound correlate strongly with concomitant change in its biological activity. The CoMFA steric and electrostatic contour plots for CoMFA models developed using antimicrobial activity against MRSA, PRSP, and VSEF are shown in Figs. 3, 4 and 5, respectively.

The contour maps are devoid of any features at the R_2 substitution end of the crescent shape molecules, as the morpholine substituent is common in most molecules. Many features are seen at the R_1 substitution end of the molecule. For the CoMFA model correlating activity against MRSA (Fig. 3), a medium sized green colored contour is seen in the plane of the molecules away from the R_1 substitution end. Three medium sized and one small yellow colored contour are seen around the benzothiophene ring at the R_1 substitution end of the most active molecule **18**. Molecule **18** contains a dioxolane ring, attached to the benzothiophene ring at the R_1 substitution end of the molecule, which is buried in the sterically favorable green contour, thus contributing to the potency of this molecule. Activity seems to be modulated by the orientation of the aryl/heteroaromatic ring at the R_1 substitution end of the molecule. When this ring is coplanar with the average plane formed by the crescent shaped NMP units (as in molecules **2**, **16**, **25**, **29**, **30**, **31** and **32**), the molecule has high activity. A loss of coplanarity of the R_1 substitution with the average plane of NMP units places the substituent in a sterically disfavoured (yellow) region (as in molecules **14**, **17**, **22**, **24**

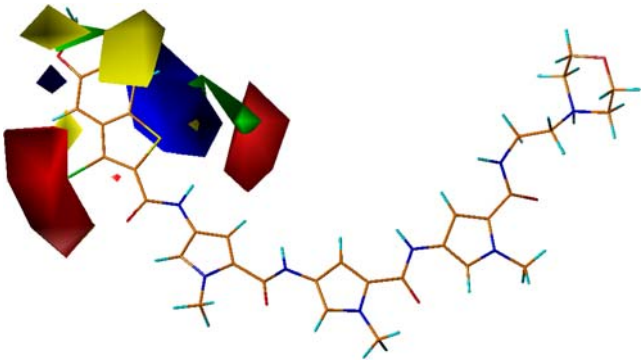


Fig. 4 CoMFA contour map for steric and electrostatic fields around molecule **18** with antimicrobial activity against PRSP

and **33**) and causes a decrease in the activity. This is in corroboration with the fact that these DNA minor groove binders are seen wrapped in between the DNA strands. The bifurcated hydrogen bonds that connect the distamycin molecule to the narrow minor groove are further stabilized by hydrophobic van der Waals contacts between the aromatic pyrrole rings of the distamycin analogs and the walls of the minor groove, leaving little room for substituents in this region as seen in Fig. 6. The *iso*-propyl moiety attached to the sulphonyl group in molecule **13** and the *n*-butyl substituent on the pyridyl ring of molecule **6** are buried in the unfavorable yellow colored contour, leading to lower activity. The indole ring in molecule **3** deviates from the plane of the NMP units, as a result its N-methyl substituent lies within the unfavorable yellow colored contours, which is associated with reduced activity. Inspection of the electrostatic contours (Fig. 3) reveals a small blue colored contour near the morpholine ring at the R_2 substitution end of the molecule. At the R_1 substitution end of molecule **18**, there is a large blue colored contour along with two (one large and the other small) red colored contours. The blue contour lies on the concave surface (that faces DNA upon binding, Fig. 6) of molecule **18** at the R_1

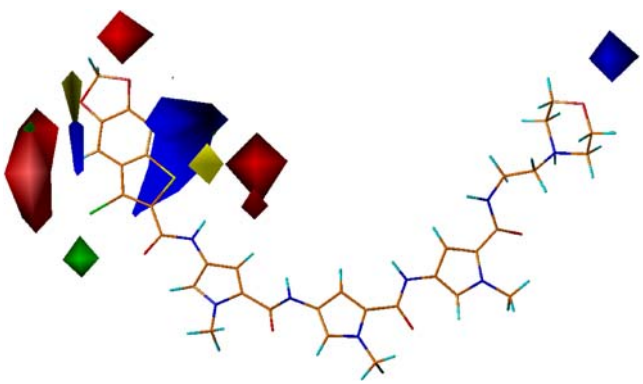


Fig. 5 CoMFA contour map for steric and electrostatic fields around molecule **18** with antimicrobial activity against VSEF

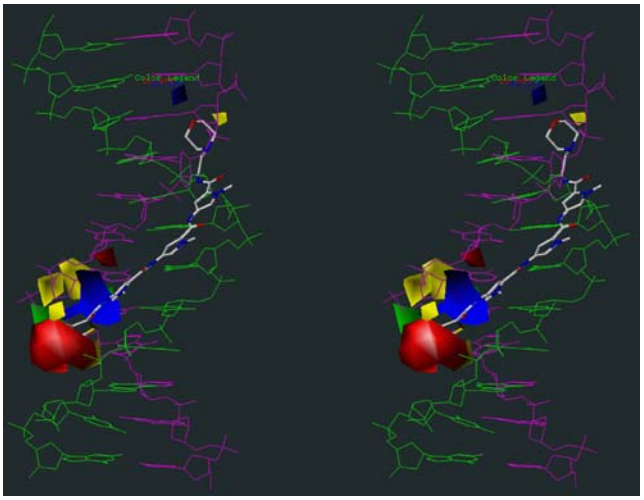


Fig. 6 Molecule **18** shown bound in the minor-groove of DNA. The CoMFA contours for antimicrobial activity against MRSA in relation to the DNA bases in the minor-groove have also been shown

substitution region, away from the plane of the molecule. The small red colored contour lies behind the blue contour. The large red colored contour covers the convex surface of the molecule, opposite the blue colored contour. Molecules with an electropositive substituent on the concave surface, which faces DNA on binding and an electronegative substituent on the opposite side (the convex surface), will exhibit good activity. Molecules with electronegative substituents on the convex surface of the R_1 substitution end (for example, Cl in **2**, **16** and **18**, the thiophene sulfur in **25** and **26**, and the benzothiophene sulfur in **17**) exhibit good activity. The electronegative substituent on the concave surface at the R_1 substitution end in for, e.g., molecules **41** (both N's of the thiadiazole), **3** (the N-methyl nitrogen), **4** (the two N's of pyrimidoimidazole), **6** (the pyridyl N), **8** (both N's of 1,6-naphthyridine), **10** (the CN group), **11** (the S-methyl sulphur), **12** (the 3-Cl of 2,3-dichlorothiophene), **15** (the Cl of 7-chlorobenzothiophene) and **23** (the methoxy O), lie within the unfavorable blue electrostatic blue contour causing a lowering of their activity.

The CoMFA model built using activity data for PRSP (Fig. 4) shows no contours at the R_2 substitution end of the molecules. The electrostatic contours are the same as that in the model developed for MRSA activity, except that the second red contour is bigger. An additional small green colored contour at the concave surface is present beside the blue electrostatic contour. The additional green steric contour might be the result of the methyl group of molecules **8** (-SMe) and **11** (-OMe), which exhibit poor activity against MRSA but good activity against PSIP strain. The second red contour, which is bigger than that in the CoMFA model built for molecules with activity against MRSA, is in proximity of the O and the S atoms in molecules **8** and **11** respectively.

The CoMFA models built using activity data against VSEF (Fig. 5) displays very few steric contours. The electrostatic contours are similar to those in the CoMFA model for molecules with activity against MRSA with an additional small red contour at the R_1 substitution end in the plane of the molecule.

The unified model (Fig. 7) shows that very few steric but a majority of the electrostatic contours are common to all the three models. The common electrostatic contours are at the R_1 substitution end of the molecule where there is a large blue colored contour along with two (one large and the other small) red colored contours. The blue contour lies on the concave surface that faces DNA upon binding at the R_1 substitution region (Fig. 6), away from the plane of molecule. The small red colored contour lies behind the blue contour. The large red colored contour covers the convex surface of the molecule, opposite the blue colored contour. Electrostatic interaction is a major requirement for DNA intercalation and the electrostatic contours are similar for molecules with activity against MRSA, PRSP and VSEF; it is therefore not surprising that the distamycin analogs have a broad spectrum of antibacterial activity.

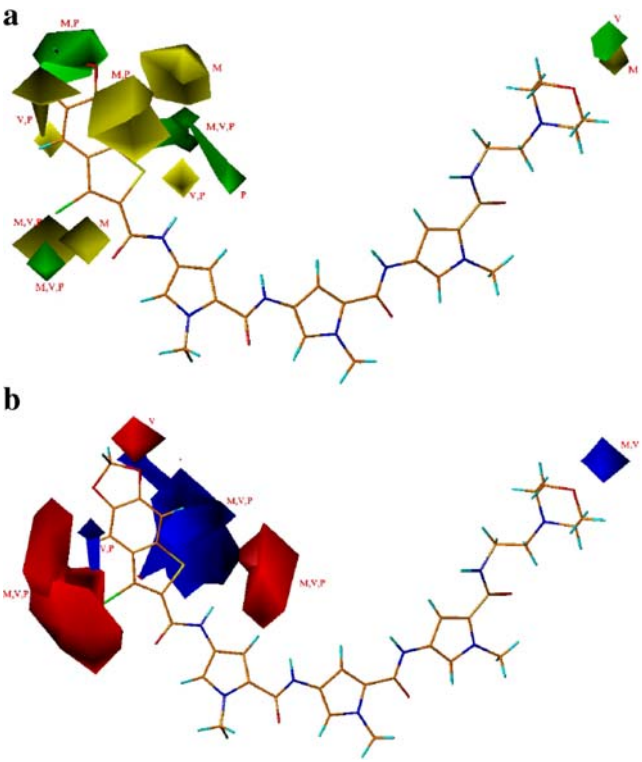


Fig. 7 The unified model depicting the common steric (**a**) and electrostatic (**b**) contours for models developed with antimicrobial activity against MRSA (M), VSEF (V) and PRSP (P)

Conclusions

To understand structure activity relationships of the distamycin analogs, a 3D-QSAR analysis was carried out using CoMFA to map the structural features contributing to the inhibitory activity of these molecules. Three CoMFA models, each using activity data against MRSA, PRSP and VSEF are all statistically robust with a good correlation and predictive power. The key feature in all three models is that molecules with an electropositive substituent on the concave surface, which faces DNA upon binding, and an electronegative substituent on the opposite side (the convex surface) at the R_1 substitution of the distamycin skeleton would exhibit good inhibitory activity and also contribute to their broad-spectrum activity. Analysis of the CoMFA contours shows details on the fine relationship linking structure and activity, and offer clues for structural modifications that can improve the activity.

Acknowledgements SAK and AKM thank the Council of Scientific and Industrial Research (CSIR), India for the Fellowship. The computational facilities at BCP were provided by the All India Council for Technical Education (AICTE), India (F. No. 8022/RID/NPROJ/RPS-5/2003-04).

References

1. Clough J (2002) Drug Discov Today 7:1036–1038
2. Yoneyama H, Katsumata R (2006) Biotechnol Biochem 70:1060
3. Ford C, Hamel J, Stapert D, Moerman J, Hutchinson D, Barbachyn M, Zurenko G (1999) Infect Med 16:435–445
4. Andries K, Verhasselt P, Guillemont J, Göhlmann HWH, Neefs J-M, Winkler H, Gestel JV, Timmerman P, Zhu M, Lee E, Williams P, Chaffoy D, Huitric E, Hoffner S, Cambau E, Truffot-Pernot C, Lounis N, Jarlier V (2005) Science 307:223–227
5. Arcamone F, Penco S, Orezzi P, Nicolella V, Pirelli A (1964) Nature 203:1064–1065
6. Choudhary AK, Brown JR, Longmore RB (1978) J Med Chem 21:607–612
7. Kaizerman JA, Gross MI, Ge Y, White S, Hu W, Duan J-X, Baird EE, Johnson KW, Tanaka RD, Moser HE, Burli RW (2003) J Med Chem 46:3914–3929
8. Cramer RD, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
9. Tanimoto TT, IBM Internal Reports 17th Nov, 1957
10. Sybyl 7.1, Tripos Associates Inc., 1699 S Hanley Rd, St Louis, MO 631444, USA
11. Halgren TJ (1990) Am Chem Soc 112:4710–4723
12. Bush BL, Nachbar RB (1993) J Comput-Aided Mol Des 7:587–619
13. Cramer RD, Bunce JD, Patterson DE, Frank IE (1988) Quant Struct-Act Relat 7:18–25
14. Clark M, Cramer RD, Jones DM, Patterson DE, Simeroth PE (1990) Tetrahedron Comput Methodol 3:47–59