

Volsurf analysis of carbapenem antibiotics

Munikumar Reddy Doddareddy, Joo Hwan Cha, Yong Seo Cho, Hun Yeong Koh,
Kyung Ho Yoo, Dong Jin Kim and Ae Nim Pae*

Life Science Division, Korea Institute of Science and Technology, PO Box 131, Cheongryang, Seoul 130–650, Korea

Received 16 February 2005; revised 10 March 2005; accepted 10 March 2005

Available online 6 April 2005

Abstract—Classical Volsurf approach was applied to a set of 70 carbapenem compounds acting as antibiotics. Antibacterial activity of *Staphylococcus aureus* SG 511 and *Escherichia coli* 078 representing Gram positive and Gram negative bacteria, respectively, was used for the analysis. The score plots obtained from principal component analysis showed clustering of compounds according to the activity and their loading plots explained the Volsurf descriptors responsible for the separation or peculiar behaviour of these compounds. Partial Least Square analysis yielded a seven component model for *S. aureus* with a cross-validated r^2 (q^2) value of 0.684 and conventional r^2 value of 0.883 and for *E. coli* it is a six component model with cross-validated r^2 (q^2) value of 0.514 and conventional r^2 value of 0.756. Both the PCA and PLS models were validated by an external test set of 15 compounds. All the compounds of the test set were fairly predicted with residual values less than one log unit. Comparatively activity data of *S. aureus* (Gram positive) was better explained than *E. coli* (Gram negative) by these models.

© 2005 Elsevier Ltd. All rights reserved.

1. Introduction

Carbapenems have attracted considerable attention as the most promising β -lactam antibiotics due to their chemical and metabolic stability as well as their potent antibacterial activities.¹ Research and development of carbapenem derivatives have been energetically conducted by many pharmaceutical companies worldwide, and imipenem² (IPM), panipenem³ (PAPM) and meropenem⁴ (MEPM) are already in clinical use. In order to enhance their resistance to renal dehydropeptidase-1 (DHP-1) and to reduce nephrotoxicity, IPM and PAPM are used as mixtures with cilastatin and betamipron, respectively.⁵ MEPM was the first derivative to be used as a single active agent in preparations, because its resistance to the enzyme was improved by the introduction of a 1 β -methyl group into the carbapenem nucleus. Nonetheless, the biological stability of MEPM is still not ideal. Furthermore, the marketed carbapenems are insufficiently active against methicillin-resistant *Staphylococcus aureus* (MRSA) and drug-resistant *Pseudomonas aeruginosa*. Thus, despite many advances in the carbapenem antibiotics, there exists continuous need

for novel carbapenems to overcome the limitations of existing drugs. The present study is dedicated to derive chemometric models for a dataset of carbapenems via classical Volsurf approach⁶ to understand the effect of Volsurf descriptors⁷ on the activity of these compounds.

Molecular descriptors calculated by the Volsurf program have been extensively used to model pharmacokinetic properties, for example, passive permeability through the gastrointestinal tract or through the blood–brain barrier. These descriptors quantify steric, hydrophobic and hydrogen bond interactions between model compounds and different environments. Since these interactions are the same as those involved in the ligand–receptor binding, Volsurf descriptors could potentially be relevant in modeling this process as well. This kind of approach was also successfully applied to quinolone antibiotics.⁸

Volsurf is an automatic procedure to convert 3D molecular fields into physico-chemically relevant molecular descriptors. In the standard procedure, the interaction fields with a water probe and a hydrophobic probe are calculated for all molecules in the dataset. However, grid maps produced by other probes (ionic probes, etc.) or by various molecular mechanics or semiempirical approaches (e.g., electrostatic potential) can also be used.

Keywords: Volsurf; Carbapenem; PCA; PLS.

* Corresponding author. Tel.: +82 2 958 5185; fax: +82 2 958 5189;
e-mail: anpae@kist.re.kr

The basic concept of Volsurf is to extract the information present in 3D molecular field maps into few quantitative numerical descriptors, which are easy to understand and interpret. Molecular recognition is achieved using image analysis software coupled with external chemical knowledge. Within this context Volsurf selects the most appropriate descriptors and parameterization according to the type of 3D maps under study.

The molecular descriptors obtained refer to molecular size and shape, to size and shape of both hydrophilic and hydrophobic regions and to the balance between them. Hydrogen bonding, amphiphilic moments, critical packing parameters are other useful descriptors. The Volsurf descriptors have been presented and explained in detail.⁷ The originality of Volsurf resides in the fact that surfaces, volumes and other related descriptors can be directly obtained from 3D molecular fields with simple computational algorithms. Moreover, Volsurf descriptors can be easily obtained for small, medium and large molecules, as well as for biopolymers such as DNA sequences, peptides and proteins. No parameterization is required. Furthermore, it was already demonstrated that the Volsurf descriptors are often hardly influenced by conformational sampling and averaging.^{6,7,9} This is probably due to the peculiarity of the Grid force field that allows for the conformational flexibility of external groups, hydrogens and lone pairs. In general, a protocol consisting in a simple 2D-to-3D structure conversion followed by energy minimization produces good results, without any MD sampling and Boltzmann averaging. These properties make Volsurf descriptors computationally efficient and well suited for fast quantitative structure–property relationship studies, especially when dealing with a large number of compounds.

2. Materials and methods

2.1. Dataset

A dataset of 85 carbapenem derivatives, 70 of which were used as training set and 15 as test set were obtained from various papers published by our synthesis groups.¹⁰ Table 1 contains the structures and pMIC (–logMIC) values for *S. aureus* SG 511 and *Escherichia coli* 078 representing Gram positive and Gram negative bacteria, respectively. The training set covers a spectrum of 2.64 log units (pMIC) in case of *S. aureus* with 8.15 for the most potent compound **70** and 5.51 for the weakest compound **42**, whereas it is 2.34 in case of *E. coli* with 8.15 for the most potent compound **56** and 5.81 for the weakest compound **42**. Three dimensional structures of the compounds were generated and subsequently refined by energy minimization using Tripos force field¹¹ and Gasteiger Huckel charge with distance dependent dielectric and conjugate gradient method with convergence criterion of 0.01 kcal/mol as implemented in SYBYL 6.9.¹²

2.2. Calculation of Volsurf descriptors

A set of 72 Volsurf descriptors⁷ was automatically generated from 3D molecular fields by using Volsurf 3.0

Table 1. Compounds used for the analysis

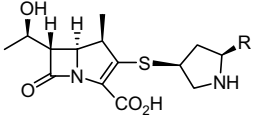
			
No	R	pMIC ^a	pMIC ^b
<i>Training set</i>			
1		6.71	7.60
2		7.89	7.60
3		7.60	7.60
4		7.89	7.60
5		7.89	7.60
6		7.60	7.60
7		7.31	7.01
8		7.01	7.89
9		7.31	7.60
10		7.31	7.60
11		7.01	7.60
12		7.01	7.60
13		7.01	7.06
14		6.71	7.89
15		6.41	7.60
16		6.71	7.60

Table 1 (continued)

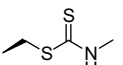
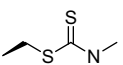
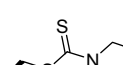
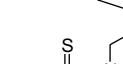
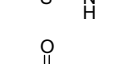
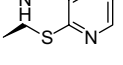
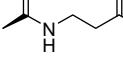
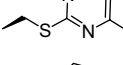
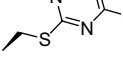
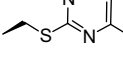
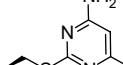
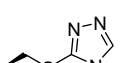

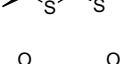
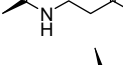
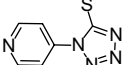
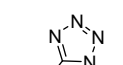
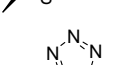
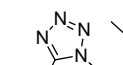
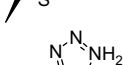
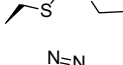
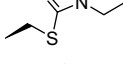
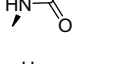
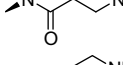
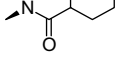
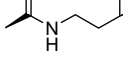
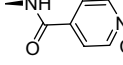
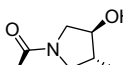
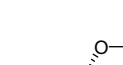
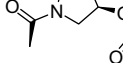
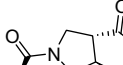
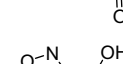
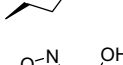
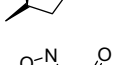
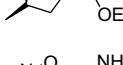
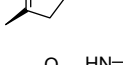
No	R	pMIC ^a	pMIC ^b
17		7.31	7.31
18		7.89	7.60
19		7.60	7.31
20		7.60	7.60
21		7.31	7.31
22		6.71	7.60
23		7.31	7.31
24		7.01	7.01
25		7.60	7.31
26		7.60	7.31
27		7.31	7.60
28		7.60	7.89
29		6.71	7.60
30		7.31	7.01
31		7.01	7.60
32		7.31	7.60
33		7.01	7.31
34		6.71	7.60

Table 1 (continued)

No	R	pMIC ^a	pMIC ^b
35		6.71	7.60
36		7.01	7.60
37		7.01	7.60
38		6.71	7.01
39		6.41	6.11
40		6.41	6.41
41		6.71	7.60
42		5.51	5.81
43		6.71	7.89
44		6.71	7.60
45		6.71	7.60
46		7.01	7.60
47		7.01	7.31
48		6.71	7.89
49		7.31	7.60
50		7.31	7.31
51		7.31	7.60
52		7.01	7.89

(continued on next page)

Table 1 (continued)

No	R	pMIC ^a	pMIC ^b
53		7.01	7.89
54		6.71	7.60
55		7.01	7.60
56		6.41	8.15
57		7.31	7.60
58		7.31	7.31
59		6.71	6.71
60		7.01	7.60
61		7.01	7.60
62		8.15	7.01
63		7.89	7.01
64		7.89	7.31
65		7.89	7.31
66		7.60	7.60
67		7.89	7.01
68		7.60	7.31
69		7.60	7.60
70		8.15	7.01

Table 1 (continued)

No	R	pMIC ^a	pMIC ^b
<i>Test set</i>			
71		7.01	7.89
72		7.60	7.31
73		7.31	7.60
74		6.71	7.60
75 ^c		7.89	7.01
76		7.60	7.60
77		7.89	7.60
78		7.60	7.31
79		7.60	7.60
80		6.41	6.41
81		6.11	6.11
82		5.81	5.81
83		7.01	6.71
84		6.41	8.15
85		7.89	7.01

Data set was obtained from Ref. 10. Compounds 1–70 were used as training set and 71–85 were used as test set.

^a pMIC (–log MIC) value of *S. aureus*.

^b pMIC value of *E. coli*.

^c The entire substituent group of compound 75 (imipenem) was shown as R.

program.¹³ The water probe (OH2) was used to simulate solvation–desolvation processes, while the hydrophobic probe (DRY) and the carbonyl probe (O) were used to simulate drug–membrane interactions. The DRY probe is a specific probe to compute the hydrophobic energy; R the overall energy of the hydrophobic probe is computed at each gridpoint as $E_{\text{entropy}} + E_{\text{LJ}} - E_{\text{HB}}$ where E_{entropy} is the ideal entropic component of the hydrophobic effect in an aqueous environment, E_{LJ} measures the induction and dispersion interactions occurring between any pair of molecules, and E_{HB} measures the H-bonding interactions between water molecules and polar groups on the target surface.

2.3. Statistical analysis

Principal component analysis (PCA)¹⁴ and partial least squares analysis (PLS)¹⁵ are the most common chemometric tools for extracting and rationalizing the information from any multivariate description of a biological system. Complexity reduction and data simplification are two of the most important features of such tools. PCA and PLS condense the overall information into two smaller matrixes, namely the score plot and the loading plot.¹⁶ The score plots represent the relative position of the objects in the space (two dimensional or three dimensional) of the principal components. They are useful to identify clusters of objects and single objects that behave in a peculiar way. Moreover, the position of the objects in the plots may serve to interpret the principal components (PCs). The first PCs try to explain the maximum amount of variation and therefore when there are clusters of objects, to distinguish among them. In this context, the PC can be interpreted as a compendium of distinctive features of the objects in these clusters. The loading plots represent the original variables in the space (two dimensional or three dimensional) of the principal components. The loading of a single variable indicates how much this variable participates in defining the PC (the squares of the loadings indicate their percentage in the PC). Variables contributing very little to the PCs have small loading values and are plotted around the centre of the plot. On the other hand, the variables that contribute most are plotted around the borders of the plot. Because the chemical interpretation of score and loading plots is simple and straight forward, PCA and PLS are usually preferred to other nonlinear methods, especially when the noise is relatively high.

Score and loading plots are interconnected so that any descriptor change in the loading plot is reflected by changes in the position of compounds in the score plot. Pairwise comparison can be made directly with interactive plots¹⁷ as developed in Volsurf program,¹³ and the relative contributions to the property are shown in the related descriptors space.

PCA is a least-squares method and for this reason its results depend on data scaling. The initial variance of a column variable partly determines its importance in the model. To avoid this problem, column variables were scaled to unit variance before analysis.¹⁸ The col-

umn average was then subtracted from each variable. From a statistical point of view, this corresponds to moving the multivariate system to the centre of data, which becomes the starting point of the mathematical analysis. The same autoscaling and centering procedures were applied to the PLS analysis.

Once the PCA model was developed, PCA predictions for new compounds or external test set compounds were made by projecting the compound descriptors into the PCA model. This was made by calculating the score vector **T** of descriptors *X* and average *x* for the new compounds, using the loading *P* of the PCA model, according to the following equation (Eq. 1)

$$\mathbf{T} = (\mathbf{X} - \mathbf{x})\mathbf{P}'(\mathbf{PP}')^{-1} \quad (1)$$

For the PLS discrimination, external predictions were made using the following equation¹⁹ (Eq. 2)

$$Y = y - x\mathbf{P}'(\mathbf{PP}')\mathbf{BQ} + \mathbf{XP}'(\mathbf{PP}')^{-1}\mathbf{BQ} \quad (2)$$

where *y* is the *Y* column average and **Q** is the loading vector for the *y* space and *B* the coefficient between the *X* and *Y* spaces.

3. Results and discussion

In our work we first conducted PCA to search for the relationship between the 3D structure and the activity of the training set containing 70 carbapenem compounds (Table 1) and 72 Volsurf descriptors. No biological input was given to the model. Five significant principal components were observed by cross-validation technique (Table 2). These components explained about 84% of the total variance of the matrix. The score plot for the first two PCs is shown in Figure 1. The plots were colour coded according to their activities with most active compounds coloured blue followed by green, cyan and then red for the least potent compounds. Table 3 shows the activity data of both *S. aureus* and *E. coli* in decreasing order of potency. The score plot in Figure 1a show that compounds are ‘roughly’ distributed in decreasing order of activity from right to left in case of *S. aureus*, whereas opposite trend was observed in case of *E. coli* data (Fig. 1b) where compounds are arranged from left to right. Visually better trend was observed in score plot of PC1 and PC3. Figure 2a shows the score plot of PC1 and PC3 in which the separation of active compounds (*S. aureus*) was clearer in the form of a cluster at the right side of the plot. It is striking to

Table 2. Summary of PCA analysis

Components	XVarEXP	XAccum
1	45.62	45.62
2	15.09	60.72
3	9.62	70.33
4	7.44	77.79
5	6.64	84.41

XVarExp. Percentage of *X*-matrix variance explained by that component. XAccum. Accumulative percentage of the *X*-matrix variance explained by the model.

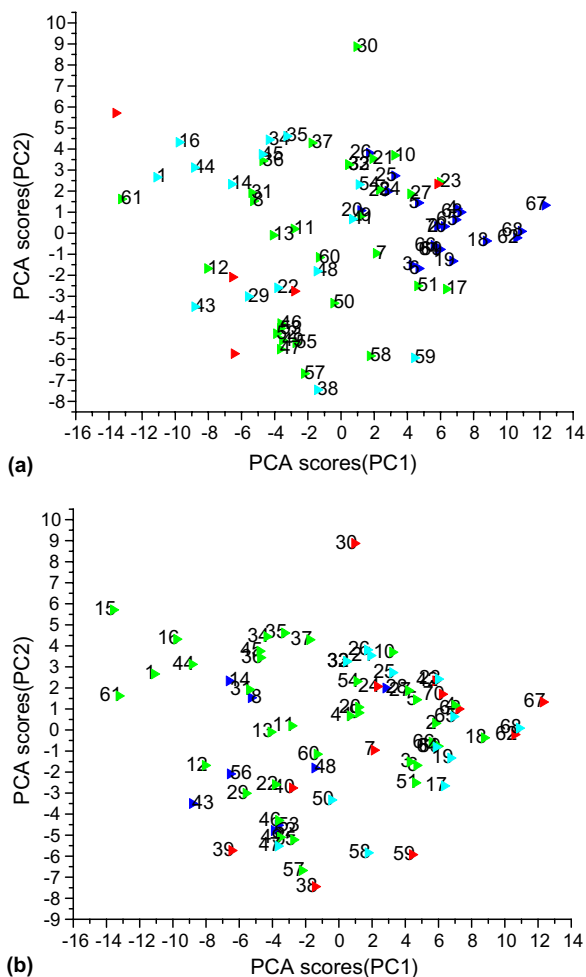


Figure 1. (a) Score plot of PC1 versus PC2 of PCA analysis. Training set compounds are coloured according to *S. aureus* activities (pMIC). Blue (8.15–7.60), green (7.31–7.01), cyan (6.71) and red (6.41–5.51). (b) Score plot of PC1 versus PC2 of PCA analysis. Training set compounds are coloured according to *E. coli* activities (pMIC). Blue (8.15–7.89), green (7.60), cyan (7.31) and red (7.01–5.81).

see that all top 20 compounds (blue) of *S. aureus* data lie in that cluster. In case of *E. coli* data, the trend was little different as shown in Figure 2b with most active compounds (blue) at the top left hand side of the plot and decreasing in the direction of arrow inside the ellipse, whereas all less potent compounds (red) were distributed around the ellipse. Compounds **1**, **15**, **16** and **65** that are located far away from the remaining can be considered as outliers whose behaviour cannot be explained by these models. These score plots are significant from the fact that in PCA no activity data was given as input and the plots obtained just from the structures and the generated Volsurf descriptors clearly restricted the most active compounds to a specific region.

Carbapenems along with penicillins, aminocillins, cephalosporins and monobactams all come under β -lactam antibiotics.²⁰ They exhibit their activity by inhibiting bacterial cell wall synthesis.²¹ Although several mechanisms have been explained for this inhibition, the most important is the inhibition of terminal peptidoglycan

Table 3. Compounds arranged in decreasing order of activity for both *S. aureus* and *E. coli* data

No	Compound	<i>S. aureus</i> SG 511	No	Compound	<i>E. coli</i> 078
1	70	8.15	1	56	8.15
2	62	8.15	2	28	7.89
3	4	7.89	3	48	7.89
4	5	7.89	4	43	7.89
5	2	7.89	5	53	7.89
6	18	7.89	6	8	7.89
7	67	7.89	7	52	7.89
8	63	7.89	8	14	7.89
9	65	7.89	9	10	7.60
10	64	7.89	10	11	7.60
11	20	7.60	11	12	7.60
12	69	7.60	12	13	7.60
13	3	7.60	13	9	7.60
14	66	7.60	14	15	7.60
15	6	7.60	15	16	7.60
16	68	7.60	16	45	7.60
17	28	7.60	17	18	7.60
18	19	7.60	18	46	7.60
19	25	7.60	19	20	7.60
20	26	7.60	20	54	7.60
21	9	7.31	21	22	7.60
22	7	7.31	22	51	7.60
23	23	7.31	23	49	7.60
24	49	7.31	24	6	7.60
25	30	7.31	25	5	7.60
26	32	7.31	26	27	7.60
27	10	7.31	27	4	7.60
28	27	7.31	28	29	7.60
29	21	7.31	29	3	7.60
30	17	7.31	30	31	7.60
31	58	7.31	31	32	7.60
32	57	7.31	32	2	7.60
33	51	7.31	33	34	7.60
34	50	7.31	34	35	7.60
35	8	7.01	35	1	7.60
36	36	7.01	36	36	7.60
37	24	7.01	37	69	7.60
38	37	7.01	38	37	7.60
39	61	7.01	39	66	7.60
40	60	7.01	40	61	7.60
41	31	7.01	41	60	7.60
42	13	7.01	42	41	7.60
43	55	7.01	43	57	7.60
44	53	7.01	44	55	7.60
45	52	7.01	45	44	7.60
46	33	7.01	46	19	7.31
47	46	7.01	47	21	7.31
48	47	7.01	48	23	7.31
49	12	7.01	49	68	7.31
50	11	7.01	50	50	7.31
51	45	6.71	51	65	7.31
52	44	6.71	52	25	7.31
53	54	6.71	53	26	7.31
54	43	6.71	54	47	7.31
55	59	6.71	55	64	7.31
56	1	6.71	56	33	7.31
57	48	6.71	57	17	7.31
58	41	6.71	58	58	7.31
59	34	6.71	59	24	7.01
60	35	6.71	60	30	7.01
61	38	6.71	61	62	7.01
62	29	6.71	62	7	7.01
63	22	6.71	63	63	7.01
64	16	6.71	64	67	7.01

Table 3 (continued)

No	Compound	<i>S. aureus</i> SG 511	No	Compound	<i>E. coli</i> 078
65	14	6.71	65	38	7.01
66	56	6.41	66	70	7.01
67	15	6.41	67	59	6.71
68	39	6.41	68	40	6.41
69	40	6.41	69	39	6.11
70	42	5.51	70	42	5.81

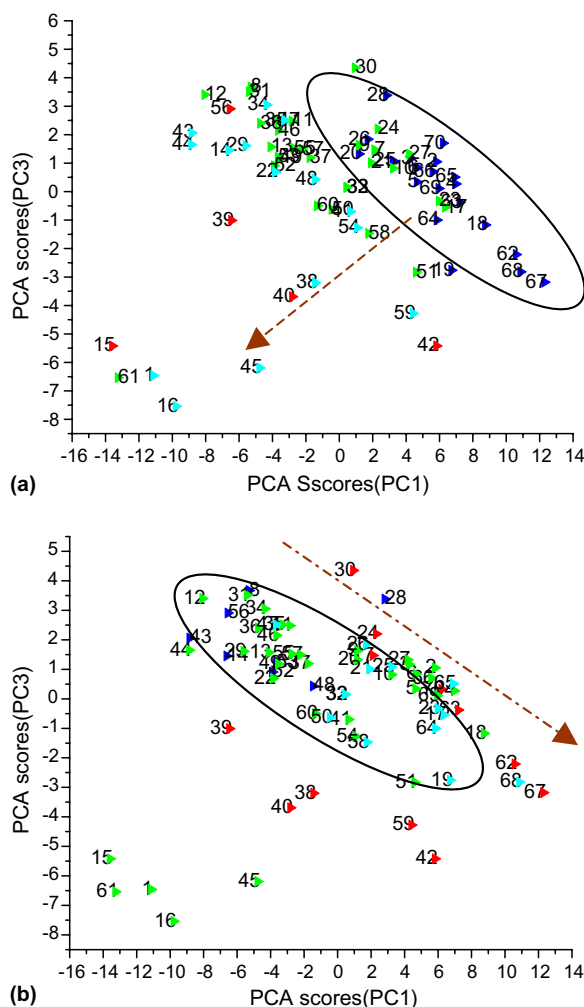


Figure 2. (a) Score plot of PC1 versus PC3 of PCA analysis. Compounds are coloured according to *S. aureus* activities (pMIC). Blue (8.15–7.60), green (7.31–7.01), cyan (6.71) and red (6.41–5.51). Arrow shows the direction of decreasing trend in activities. (b) Score plot of PC1 versus PC3. Compounds are coloured according to *E. coli* activities (pMIC). Blue (8.15–7.89), green (7.60), cyan (7.31) and red (7.01–5.81). Arrow shows the direction of decreasing trend in activities inside the ellipse. Lower active compounds (red) are dispersed around the ellipse.

cross-linking,²² which is a terminal event in the bacterial cell wall formation. These compounds act by inactivating the enzyme peptidoglycan transpeptidase, which catalyses the cross-linking of the peptidoglycan. The labile β -lactam ring directly involves in the inactivation by acylation of the transpeptidase. But these drugs have to

pass through the cell membrane in order to reach their active site. As all the compounds of the training set contain the β -lactam ring for the inactivation, one of the main roles of the side chain substitutions is to enhance the pharmacokinetic properties of the drug. Volsurf method is specifically designed to produce descriptors related to pharmacokinetic properties and membrane permeation. *S. aureus* is a Gram positive bacteria, whose cell wall is made of peptidoglycan through which the drug has to pass through by 'passive diffusion' to reach the active site. Therefore the decreasing trend of activity from right to left in first score plot (Fig. 1a) and the cluster of higher active compounds on the right hand side in the second score plot (Fig. 2a) in case of activity data of *S. aureus* (Table 3) gives an indication of the efficiency of Volsurf descriptors in predicting the passive diffusion. *E. coli* is a Gram negative bacteria, whose cell wall is made of high amount of lipopolysaccharide, which contains various pores through which hydrophilic substances pass to reach the active site. Even though the process is highly complex and rate dependent on both the compound and membrane through which it is passing, the score plot (Fig. 2b) showed all top eight compounds **56**, **28**, **48**, **43**, **8**, **52** and **14** (blue) lying in the same region and the less potent compounds (red) around the ellipse as shown in the figure. Further the significance of these score plots increases from the fact that they are obtained from PCA, where no external information or training was given to the dataset and so are highly unbiased. These score plots can be used as a projection maps to understand the behaviour of untested carbapenem antibiotics.

The PCA model was validated by using an external test set of 15 compounds (Table 1). In Volsurf it is possible to apply a PCA model to an external test set, in order to obtain 'predicted' scores that can be used to obtain scores plots representing both the original series and the external objects. This representation can be seen as a projection of the external series in the same dimensionally reduced space obtained for the original series, with a rotation defined by the original loading matrix. As score plot of PC1 and PC3 was clearer than the score plot of PC1 and PC2, we used the former one for validation. Figure 3 shows the score plot (PC1 vs PC3) of predicted scores of the test set. All seven compounds **72**, **75**, **76**, **77**, **78**, **79** and **85** (blue), whose pMIC values are more than 7.60 (*S. aureus*) are correctly predicted in the same region (Fig. 3a) as specified in the training set model (Fig. 2a). In case of *E. coli* except **71** and **76** all other compounds **73**, **74**, **77**, **79** and **84** (blue and green), whose pMIC values are more than 7.60 (*E. coli*) are correctly predicted in the area (Fig. 3b) specified for the *E. coli* in the original score plot (Fig. 2b). The above observations suggest that PCA model explains *S. aureus* data better than *E. coli* data. The two score plots and the successful prediction of the external test set by using PCA show that Volsurf descriptors are highly successful in explaining the molecular basis for the carbapenem antibiotics.

To understand which Volsurf descriptors are really responsible for the clustering of compounds in the score

Table 4. Summary of PLS analysis

	<i>S. aureus</i> SG 511	<i>E. coli</i> 078
q^2	0.684	0.514
r^2	0.883	0.756
N	7	6
SDEP	0.276	0.271
SDEC	0.168	0.192
% of Variance	79.13	81.22

q^2 —Cross-validated correlation coefficient, N —optimum number of components, r^2 —noncross-validated correlation coefficient, SDEP—standard deviation of error of predictions, SDEC—standard deviation of error of calculations.

was better explained for *S. aureus* (Gram positive) bacteria than *E. coli* (Gram negative) bacteria. Both the models were used to predict the activities of the external training set containing 15 compounds, which also include known compounds like meropenem (**71**) and imipenem (**75**) (Table 1). Figures 5 and 6 show the graphs

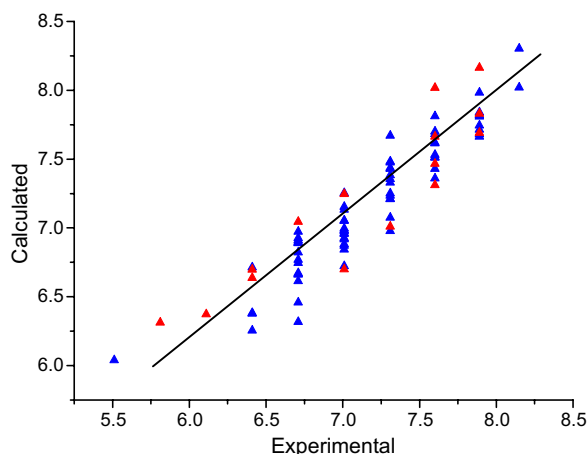


Figure 5. Plot of experimental versus calculated activities of *S. aureus* obtained from PLS analysis. Blue triangles show the predictions of training set and red triangles show the predictions of test set.

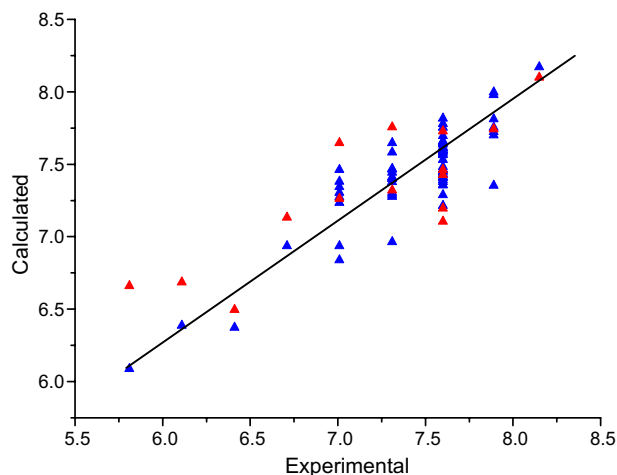


Figure 6. Plot of experimental versus calculated activities of *E. coli* obtained from PLS analysis. Blue triangles show the predictions of training set and red triangles show the predictions of test set.

of experimental versus calculated values for *S. aureus* and *E. coli* models, respectively. Blue triangles show the predictions of the training set and red triangles show the predictions of the test set. All the compounds were fairly predicted with residual values less than one log unit (Table 5). Figure 7 shows the PLS coefficients plot for the correlation of Volsurf descriptors with *S. aureus* activity. Activity particularly increases with high values of integrity moments (IWOH2) of the water probe, high values of hydrophobic regions (DDRY), integrity moments (IDDRY) and amphiphilic moment (A) of hydrophobic probe. Hydrophilic regions (WOH2), capacity factors (CwOH2), local interaction energy minima distances (DOH2) of water probe, hydrophilic–lipophilic balance (HL) of hydrophobic probe and volumes of interactions (WO), H-bond interaction energies (HBO) of carbonyl probe are inversely related to activity. Almost opposite correlation was observed between Volsurf descriptors and activity of *E. coli* (Fig. 8). This can be explained from the fact that hydrophobic drugs pass through membrane by passive diffusion in case of *S. aureus* (Gram positive), where as relatively hydrophilic drugs pass through pores to reach the active site in case of *E. coli* (Gram negative). Figures 9 and 10 show the visual comparison of Grid 3D molecular fields of the *S. aureus* active compound **70** and *E. coli* active compound **56** (Table 3) calculated with a water probe, respectively. The cyan zones around molecules represent the hydrophilic regions. The arrows represent the vectors of the integrity moments, which measure the unbalance between the centre of mass and the position of the hydrophilic regions around them. From the field maps it is clear that hydrophilic regions of **56** are larger than **70** and are distributed in different regions. The higher integrity moments in case of **70** indicates that there is a clear concentration of hydrated regions in only one part of the molecule, whereas the lower integrity moments in case of **56** indicates that the polar moieties are distributed throughout the molecule and the resultant barycentre is close to the centre of the molecule. Higher integrity moments and less hydrophilic regions in case of **70** (Fig. 9) are in complete agreement with the PLS coefficient plot for the correlation of Volsurf descriptors with *S. aureus* activity (Fig. 7). We can deduce from the plot that activity is directly proportional to the higher values of integrity moments (IWOH2) and inversely proportional to hydrophilic regions (WOH2) and capacity factors (CwOH2), which is a measure of amount of hydrophilic regions per surface unit. Whereas in case of *E. coli* PLS coefficient plot (Fig. 8) the activity is inversely proportional to integrity moments (IWOH2) and directly proportional to hydrophilic regions (WOH2) and Capacity factors (CwOH2), which is in agreement with grid map of higher active compound **56** (*E. coli*).

4. Conclusion

Classical Volsurf analysis was conducted on a dataset of 70 carbapenem compounds acting as antibiotics. The chemometric models obtained were highly significant giving useful information about the behaviour of the Volsurf descriptors in relation to the biological activity.

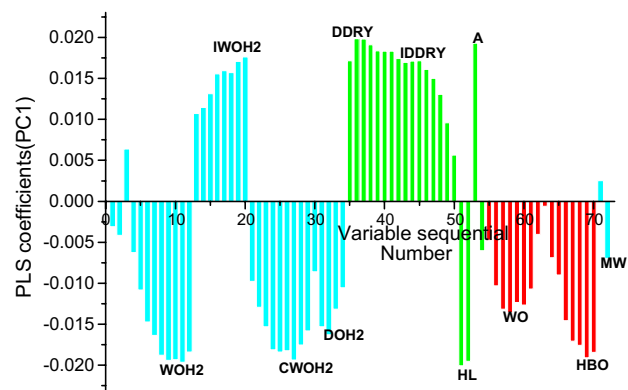
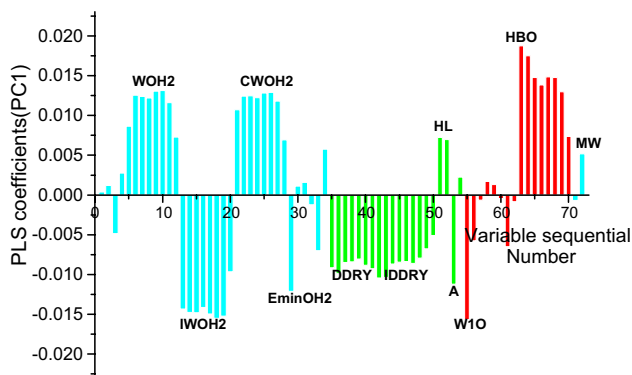
Table 5. Experimental versus calculated activities of training set and test sets of both *S. aureus* and *E. coli* as obtained from PLS analysis

Compound	<i>S.a</i> ^a	<i>S.a</i> ^b	Res	<i>E.c</i> ^a	<i>E.c</i> ^b	Res
<i>Training set</i>						
1	6.71	6.82	0.11	7.60	7.58	−0.02
2	7.89	7.72	−0.17	7.60	7.48	−0.12
3	7.60	7.61	0.01	7.60	7.41	−0.19
4	7.89	7.82	−0.07	7.60	7.41	−0.19
5	7.89	7.66	−0.22	7.60	7.65	0.05
6	7.60	7.51	−0.09	7.60	7.43	−0.17
7	7.31	7.44	0.13	7.01	7.34	0.33
8	7.01	6.72	−0.29	7.89	7.98	0.09
9	7.31	7.24	−0.07	7.60	7.40	−0.20
10	7.31	7.21	−0.10	7.60	7.46	−0.14
11	7.01	6.88	−0.13	7.60	7.65	0.05
12	7.01	6.98	−0.02	7.60	7.56	−0.04
13	7.01	7.13	0.12	7.60	7.44	−0.16
14	6.71	6.89	0.18	7.89	8.00	0.11
15	6.41	6.38	−0.03	7.60	7.36	−0.24
16	6.71	6.75	0.04	7.60	7.61	0.01
17	7.31	7.67	0.36	7.31	7.31	−0.00
18	7.89	7.81	−0.08	7.60	7.29	−0.31
19	7.60	7.70	0.10	7.31	7.38	0.07
20	7.60	7.36	−0.24	7.60	7.21	−0.39
21	7.31	7.33	0.02	7.31	7.44	0.13
22	6.71	6.91	0.20	7.60	7.82	0.22
23	7.31	7.48	0.17	7.31	7.47	0.16
24	7.01	7.25	0.24	7.01	7.27	0.26
25	7.60	7.43	−0.17	7.31	7.41	0.10
26	7.60	7.52	−0.09	7.31	7.28	−0.03
27	7.31	7.43	0.12	7.60	7.43	−0.17
28	7.60	7.53	−0.07	7.89	7.35	−0.54
29	6.71	6.90	0.19	7.60	7.78	0.18
30	7.31	7.35	0.04	7.01	7.30	0.29
31	7.01	6.84	−0.17	7.60	7.61	0.01
32	7.31	6.98	−0.33	7.60	7.47	−0.13
33	7.01	6.98	−0.03	7.31	7.47	0.16
34	6.71	6.77	0.06	7.60	7.57	−0.03
35	6.71	6.66	−0.05	7.60	7.53	−0.07
36	7.01	6.87	−0.14	7.60	7.64	0.04
37	7.01	6.96	−0.05	7.60	7.60	−0.00
38	6.71	6.32	−0.39	7.01	6.84	−0.17
39	6.41	6.38	−0.03	6.11	6.39	0.28
40	6.41	6.26	−0.15	6.41	6.37	−0.04
41	6.71	6.93	0.22	7.60	7.75	0.15
42	5.51	6.04	0.53	5.81	6.09	0.28
43	6.71	6.67	−0.04	7.89	7.81	−0.08
44	6.71	6.46	−0.25	7.60	7.66	0.06
45	6.71	6.66	−0.05	7.60	7.45	−0.15
46	7.01	7.00	−0.01	7.60	7.59	−0.01
47	7.01	6.96	−0.05	7.31	7.58	0.27
48	6.71	6.89	0.18	7.89	7.70	−0.19
49	7.31	7.25	−0.06	7.60	7.62	0.02
50	7.31	7.08	−0.23	7.31	7.65	0.34
51	7.31	7.21	−0.10	7.60	7.66	0.06
52	7.01	7.05	0.04	7.89	7.75	−0.14
53	7.01	7.05	0.05	7.89	7.72	−0.17
54	6.71	6.97	0.26	7.60	7.70	0.10
55	7.01	7.15	0.15	7.60	7.59	−0.01
56	6.41	6.71	0.31	8.15	8.17	0.02
57	7.31	7.48	0.17	7.60	7.38	−0.22
58	7.31	7.38	0.07	7.31	7.41	0.10
59	6.71	6.61	−0.10	6.71	6.94	0.23
60	7.01	6.92	−0.09	7.60	7.47	−0.13
61	7.01	6.93	−0.08	7.60	7.60	−0.00
62	8.15	8.02	−0.13	7.01	7.38	0.37
63	7.89	7.98	0.10	7.01	7.23	0.22
64	7.89	7.68	−0.21	7.31	7.29	−0.02

Table 5 (continued)

Compound	<i>S.a</i> ^a	<i>S.a</i> ^b	Res	<i>E.c</i> ^a	<i>E.c</i> ^b	Res
65	7.89	7.84	−0.05	7.31	7.40	0.09
66	7.60	7.81	0.21	7.60	7.35	−0.25
67	7.89	7.75	−0.14	7.01	6.94	−0.07
68	7.60	7.68	0.08	7.31	6.96	−0.35
69	7.60	7.62	0.02	7.60	7.39	−0.21
70	8.15	8.30	0.15	7.01	7.46	0.45
<i>Test set</i>						
71	7.01	7.25	0.24	7.89	7.74	−0.15
72	7.60	7.66	0.06	7.31	7.76	0.45
73	7.31	7.01	−0.30	7.60	7.46	−0.14
74	6.71	7.04	0.33	7.60	7.73	0.13
75	7.89	7.69	−0.19	7.01	7.65	0.64
76	7.60	8.02	0.41	7.60	7.20	−0.40
77	7.89	7.83	−0.06	7.60	7.11	−0.49
78	7.60	7.47	−0.13	7.31	7.32	0.01
79	7.60	7.31	−0.29	7.60	7.43	−0.17
80	6.41	6.70	0.29	6.41	6.50	0.09
81	6.11	6.37	0.27	6.11	6.69	0.58
82	5.81	6.31	0.51	5.81	6.66	0.85
83	7.01	6.70	−0.31	6.71	7.13	0.42
84	6.41	6.64	0.23	8.15	8.10	−0.05
85	7.89	8.16	0.28	7.01	7.26	0.25

Res—residual values.

^a Experimental.^b Calculated.**Figure 7.** PLS coefficients plot for the correlation of Volsurf descriptors with *S. aureus* activity.**Figure 8.** PLS coefficients plot for the correlation of Volsurf descriptors with *E. coli* activity.

These models were validated by an external test set of 15 compounds whose activities are fairly predicted. The

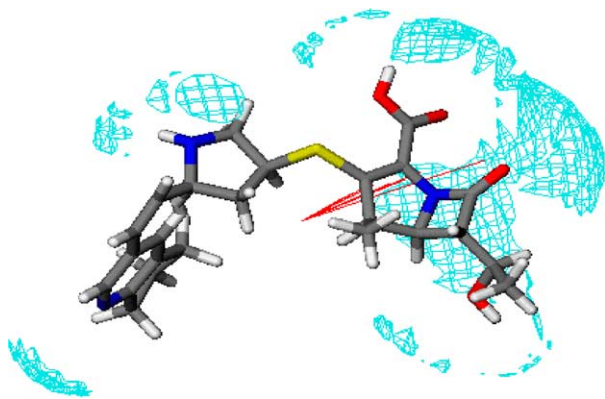


Figure 9. Grid 3D molecular fields of highest active compound (**70**) in case of *S. aureus* calculated with a water probe. The arrows represent the integrity moment's pattern.

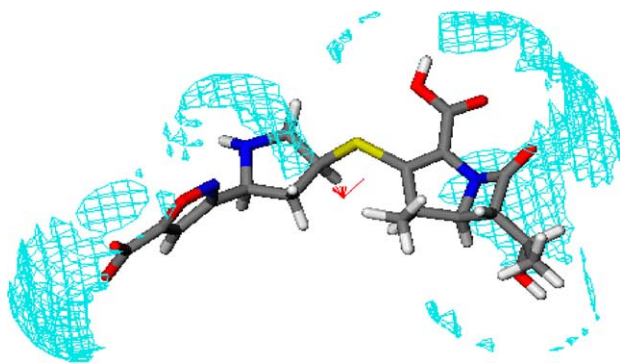


Figure 10. Grid 3D molecular fields of highest active compound (**56**) in case of *E. coli* calculated with a water probe. The arrows represent the integrity moment's pattern.

score plots obtained from PCA can be used as projection maps where unknown compounds can be projected to understand their behaviour even before actual biological screening. The results show that Volsurf approach is highly efficient in predicting the biological activities and pharmacokinetic behaviour of these carbapenem antibiotics.

Acknowledgements

This work was supported by the Korea Ministry of Science and Technology and Center for Bioactive Molecular Hybrids.

References and notes

- (a) Shih, D. H.; Baker, F.; Cama, L.; Christensen, B. G. *Heterocycles* **1984**, *21*, 29; (b) Shih, D. H.; Cama, L.; Christensen, B. G. Abstract 33, 23rd Interscience Conference on Antimicrobial Agents and Chemotherapy, Las Vegas, NV, 1983.
- Leanza, W. J.; Wildonger, K. J.; Miller, T. W.; Christensen, B. G. *J. Med. Chem.* **1979**, *22*, 1435.
- Miyadera, T.; Sugimura, T.; Hasimoto, T.; Tanaka, T.; Iino, K.; Sugawara, S. *J. Antibiot.* **1983**, *36*, 1034.
- (a) Fukasawa, Y.; Okuda, T. *J. Antibiot.* **1990**, *43*, 314; (b) Sunagawa, M.; Matsumure, H.; Inoue, T.; Fukasawa, M.; Kato, M. *J. Antibiot.* **1990**, *43*, 519.

- (a) Kahan, F. M.; Kropp, H.; Sundelof, J. G. Thienamycin: development of imipenem–cilastatin. *J. Antimicrob. Chemother.* **1983**, *12*, 1; (b) Shimada, J.; Kawahara, Y. Overview of a new carbapenem, panipenem/betamipron. *Drugs Exp. Clin. Res.* **1994**, *20*, 241.
- Cruciani, G.; Pastor, M.; Guba, W. *Eur. J. Pharm. Sci.* **2000**, *11*–2, S29.
- Cruciani, G.; Crivori, P.; Carrupt, P. A.; Testa, B. Molecular fields in quantitative structure–permeation relationships: the VolSurf approach. *THEOCHEM* **2000**, *503*, 17.
- Cianchetta, G.; Mannhold, R.; Cruciani, G.; Baroni, M.; Cecchetti, V. *J. Med. Chem.* **2004**, *47*, 3193.
- Crivori, P.; Cruciani, G.; Carrupt, P.-A.; Testa, B. *J. Med. Chem.* **2000**, *43*, 2204.
- (a) Kang, Y. K.; Shin, K. J.; Yoo, K. H.; Seo, K. J.; Park, S. Y.; Kim, D. J.; Park, S. W. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2385; (b) Kim, D. J.; Seo, K. J.; Lee, K. S.; Shin, K. J.; Yoo, K. H.; Kim, D. C.; Park, S. W. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 2799; (c) Shin, K. J.; Koo, K. D.; Yoo, K. H.; Kim, D. C.; Kim, D. J.; Park, S. W. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1421; (d) Kang, Y. K.; Shin, K. J.; Yoo, K. H.; Seo, K. J.; Hong, C. Y.; Lee, K. S.; Park, S. Y.; Kim, D. J.; Park, S. W. *Bioorg. Med. Chem. Lett.* **2000**, *10*, 95; (e) Lee, J. H.; Lee, K. S.; Kang, Y. K.; Yoo, K. H.; Shin, K. J.; Kim, D. C.; Kong, J. Y.; Lee, Y.; Lee, S. J.; Kim, D. J. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 4399.
- Clark, M.; Cramer, R. D., III; Van Opdenbosch, N. The tripos force field. *J. Comput. Chem.* **1989**, *10*, 982.
- SYBYL 6.9. Tripos Inc., 1699 Hanley Road, St. Louis, MO 63144.
- Volsurf version 3.0 software by Molecular Discovery Ltd, 2000–2004.
- Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52.
- Dunn, W. J.; Wold, S. Pattern Recognition Techniques in Drug Design. In *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon: Oxford, 1990; Vol. 4, pp 691–714.
- Cruciani, G.; Clementi, S. GOLPE: Philosophy and Applications in 3D-QSAR. In *Advanced Computer-Assisted Techniques in Drug Discovery*; van de Waterbeemd, H., Ed.; VCH: Weinheim, 1994; pp 61–88.
- Cruciani, G.; Clementi, S.; Baroni, M.; Pastor, M. Recent Development in 3D-QSAR Methodologies. In *Rational Molecular Design in Drug Research In Alfred Benzon Symposium 42*; Liljefors, T., Jorgensen, F. S., Krosgaard-Larsen, P., Eds.; Munksgaard: Copenhagen, 1998; pp 87–97.
- Wold, S.; Albano, C.; Dunn, W. J., III; Edlund, U.; Esbensen, K.; Geladi, P.; Helberg, S.; Johansson, E.; Lindberg, W.; Sjostrom, M. Multivariate Data Analysis in Chemistry. In *Chemometrics Mathematics and Statistics in Chemistry*; Kowalsky, B. R., Ed.; Holland: Dordrecht, 1983; pp 17–96.
- Clementi, S.; Cruciani, G.; Curti, G.; Skagerberg, B. PLS response surface optimisation: the CARSO procedure. *J. Chemom.* **1989**, *3*, 499.
- David, G. In *Antibiotics of the Beta-Lactam group*; Gruneberg, R., Ed.; John Wiley and Sons: London, 1982.
- Chambers, H. F. In *Beta-Lactam Antibiotics and Other Inhibitors of Cell Wall Synthesis*; Katzung, B. G., Ed.; McGraw-Hill: New York, 2001.
- Hardman, J. G.; Linbird, L. E.; Molinoff, P. B.; Ruddon, R. W.; Gilman, A. G. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 9th ed.; New York: McGraw Hill, 1996.