QSRR analysis of β-lactam antibiotics on a penicillin G targeted MIP stationary phase.

Kempe, Henrik; Kempe, Maria

Published in:
Analytical and Bioanalytical Chemistry

DOI:
10.1007/s00216-010-4254-y

2010

Henrik Kempe & Maria Kempe

"QSRR analysis of β-lactam antibiotics on a penicillin G targeted MIP stationary phase."

Analytical and Bioanalytical Chemistry 2010
Online October 08, 2010

http://dx.doi.org/10.1007/s00216-010-4254-y

QSRR analysis of β-lactam antibiotics on a penicillin G targeted MIP stationary phase

Henrik Kempe and Maria Kempe*

Biomedical Polymer Technology, Department of Experimental Medical Science, Biomedical Center, D11, Lund University, SE–221 84 Lund, Sweden

* Corresponding author

Phone: +46 46 2220857

Fax: +46 46 2221410

Email: Maria.Kempe@med.lu.se

URL: www.biomedicalpolymers.bmc.lu.se

Abstract

The imprinting factors of the β-lactam antibiotics penicillin V, methicillin, nafcillin, oxacillin, cloxacillin, dicloxacillin, and piperacillin on a poly(methacrylic acid–co–trimethylolpropane trimethacrylate) molecularly imprinted stationary phase targeted for penicillin G were correlated with molecular descriptors obtained by molecular computation. One-parameter linear regression and multivariate data analysis by principal component analysis (PCA) and partial least square (PLS) regression indicated that descriptors associated with molecular topology, shape, size, and volume were highly correlated with the imprinting factor and influential on the derived models.

Keywords: β-Lactam antibiotics; Molecular imprinting; Molecular computation; Multivariate data analysis; Principal component analysis; Partial least square regression

Introduction

Molecular imprinting is a method for the preparation of binding sites in synthetic polymers. Two distinct versions of the method have evolved based on the type of interactions between the template (the guest molecule) and the binding sites of the molecularly imprinted polymer (MIP). In the covalent approach, the template interacts with the polymer through one or several covalent interactions. Non-covalent molecular imprinting implies that template binding relies exclusively on non-covalent interactions. The latter approach, schematically visualized in Fig. 1, involves non-covalent self-assembly of functionalized and cross-linking monomers around templates, followed by fixation of the assemblies by polymerization. Upon extraction of the templates from the MIP, recognition sites capable of rebinding the template are uncovered in the rigid polymer network. The general understanding is that the templates function as molds, creating cavities in the polymer that are complementary to the templates in the positioning of functional groups as well as in shape.

MIPs are promising alternatives to biological recognition elements in that they are easy and inexpensive to produce, can be regenerated, and show high chemical, mechanical, and thermal stability. In addition, selectivity can be generated for substances for which a biological recognition element is non-existing or difficult to obtain. MIPs have many potential applications and those already explored include the use as stationary phases in chromatographic separations and solid-phase extractions, antibody mimetics in solid-phase binding assays, recognition elements in sensors, and catalysts of chemical reactions [1].

This study focuses on the MIP recognition of β-lactam antibiotics. Our interest in this class of compounds originates from the need of simple and

efficient methods for their detection in foodstuff. In animal husbandry, antibiotics are applied in therapeutic and prophylactic treatment of mastitis and other bacterial diseases and as food additives for growth promotion [2,3]. Due to the risk of the development of antibiotic resistant bacterial strains, the inhibition of starter cultures in the dairy industry, and the potential risk of allergic reactions in hypersensitive human individuals [4–6], legislative authorities have prohibited antibiotics in foodstuffs above certain MRLs (maximum residue levels). We have previously reported on the synthesis of a MIP library targeted for penicillin G for the purpose of identifying MIP candidates to serve as recognition elements in the detection of antibiotic residues [7] and recently described a method for solid-phase extraction (SPE) of penicillin G in milk using one of the formulations from the MIP library [8]. In the present study, the molecular recognition characteristics of this MIP formulation were further focused upon. The binding of penicillin G and a range of other β-lactam antibiotics was evaluated by determining the retention of the compounds on an HPLC column packed with particles of the MIP. Comparison was made with a column packed with a control polymer (CP) imprinted with an unrelated template (i.e., Boc-phenylalanine).

Quantitative structure-retention relationship (QSRR) analysis has proven to be a valuable method for relating chromatographic retention of compounds to their chemical structure [9]. To this end, molecular descriptors, either determined experimentally or calculated by computational methods, are correlated to the retention by regression analysis. QSRR models give insights into separation mechanisms as well as provide means of predicting the retention of untested compounds. The objective of the present study was to investigate if valid models of the retention of β-lactam antibiotics on a MIP stationary phase targeted for penicillin G can be generated from molecular descriptors of the antibiotics. From

the chromatographic retention data, the imprinting factor (IF) of each antibiotic were calculated as the ratio of the retention factor obtained on the MIP to that on the CP. 3D molecular models and descriptors of the antibiotics were computed by semiempirical quantum mechanical methods. The IFs were then correlated to the descriptors first by one-parameter models and then by multivariate data analysis using PCA (*principal component analysis*) and PLS (*partial least square* or *projection to latent structures*) regression. The statistical models were evaluated to identify descriptors with highest influence on the selective retention arising from the molecular imprinting procedure. The study provides an insight into what is required of a template analog to bind to a MIP. Such knowledge is valuable since (i) MIPs intended for quantitative analyses and solid-phase extractions are commonly imprinted with a structural analog of the target molecule in order to prevent interference from bleeding of remaining non-extracted template [10–15] and (ii) MIPs have been suggested as mimetics of biological receptors in the screening of lead compounds [16,17].

Experimental

Reagents

Penicillin G sodium salt, penicillin V, methicillin sodium salt, nafcillin sodium salt monohydrate, cloxacillin sodium salt monohydrate, dicloxacillin sodium salt monohydrate, piperacillin sodium salt, and oxacillin sodium salt monohydrate were purchased from Sigma (St. Louis, MO, USA). Methacrylic acid (MAA) and trimethylolpropane trimethacrylate (TRIM) were obtained from Aldrich (Milwaukee, WI, USA). 2,2'-Azobisisobutyronitrile (AIBN) was purchased from Acros (Geel, Belgium). Boc-L-Phe-OH was obtained from Advanced ChemTech (Louisville, KY, USA). P.a. grades of acetonitrile (CH_3CN) and methanol

(MeOH), used for the synthesis and extraction of the polymers, were purchased from Merck (Darmstadt, Germany). Acetonitrile of grade “HPLC far UV”, used for the chromatographic evaluation, was obtained from Labscan (Dublin, Ireland). MAA was distilled and TRIM was purified with inhibitor remover obtained from Aldrich (Milwaukee, WI, USA).

Synthesis of molecularly imprinted polymer (MIP) and control polymer (CP)

The MIP was synthesized as previously reported [7] by dissolving penicillin G (1.07 g, 3 mmol), MAA (2.58 g, 30 mmol), TRIM (15.23 g, 45 mmol), and AIBN (0.30 g, 1.8 mmol) in acetonitrile (22.5 mL). The pre-polymerization mixture was cooled on ice and purged with a stream of nitrogen gas for 10 min. The polymerization was performed at 350 nm for 24 h at 4 °C in a Rayonet photochemical mini-reactor model RMR-600 (Branford, CT, USA). The bulk polymer was ground in a Retsch Ultra Centrifugal Mill model ZM 100 (Haan, Germany). The ground particles were wet sieved in water using sieves from Retsch. The fraction containing particles <25 µm was sedimented in acetone and fine particles were removed by decantation (3 times). Penicillin G was extracted from the polymer network by microwave-assisted extraction with MeOH–HOAc (1:1), MeOH, and acetonitrile using a MARS 5 microwave-accelerated reaction system (CEM Corp., Matthews, NC) equipped with HP-500 Plus vessels. The samples were subjected to 300 W microwaves with a ramp over 10 min up to 120 °C and then for 15 min at constant temperature (120 °C). The particles were extracted three times with each extraction solvent. Between each run, the particles were collected in SPE columns, washed with MeOH, and dried. The particles were finally dried *in vacuo* overnight.

A control polymer (CP) was prepared following the same procedure as the one described for the MIP but substituting Boc-L-Phe-OH (0.76 g, 3 mmol) for penicillin G as the template.

Generation of chromatographic retention data

The polymer particles were suspended in chloroform–acetonitrile (17:3, v/v) and packed into stainless steel columns (150 × 3.0 mm) with acetone at 300 bar using an Alltech model 1666 slurry packer (Deerfield, IL, USA). The chromatography was performed with acetonitrile containing 1% of 10 mM Tris-HCl pH 7 as the mobile phase. Equilibration was done for 18 h prior to the injection of the analytes. The flow was 1.0 mL/min and detection was made at 220 nm. The HPLC system consisted of a Waters 600 Gradient Pump, a Waters 717 Plus Autosampler, and a Waters 2487 Dual Wavelength Absorbance Detector. Data were acquired via a PC using Chromatography Station for Windows (DataApex, Ltd., Prague, Czech Republic). 4 µg each of penicillin G, penicillin V, methicillin, nafcillin, oxacillin, cloxacillin, dicloxacillin, and piperacillin were injected and eluted in separate runs on the MIP and the CP columns, respectively. The retention factors (k_{MIP} and k_{CP}) were calculated as $k_{MIP\ alt.\ CP} = (t–t_0)/t_0$ where t and t_0 (the void) are the retention times of the analyte and acetone, respectively. The imprinting factors (IFs) were calculated as $IF = k_{MIP}/k_{CP}$.

Molecular computations

Chemical structures of the antibiotics, drawn with ChemDraw Ultra 10.0 of the ChemOffice 2006 software package (CambridgeSoft, Cambridge, MA, USA), were transferred into Chem 3D Ultra 10.0 to obtain 3D structures. The energy of the structures were minimized first by a modified version of Allinger's MM2 force field and then by the semiempirical molecular orbital MOPAC algorithm

with the Austin model 1 (AM1) Hamiltonian approximation. Molecular descriptors were calculated with the MOPAC application and the ChemProp interface. The Connolly molecular area (CMA), the Connolly solvent excluded volume (CSEV), and the ovality were calculated with a probe molecule of radius 0.001 Å (i.e., the minimum radius allowed by the software) [18].

QSRR computations

To correlate the IFs (the response variables) to the molecular descriptors (the predictor variables), one-parameter regression analyses were performed using Prism-4.0b (Graphpad Software, Inc. San Diego, CA, USA) and multivariate data analysis was carried out by principal components analysis (PCA) and partial least squares (PLS; also referred to as projections to latent structures) regression using the Simca-P 8.0 software (Umetrics AB, Umeå, Sweden) [19,20]. The software automatically mean-centered and scaled the data to unit variance prior to modeling and determined the number of significant principal components (PCs) by cross-validation. Cross-validation was carried out by calculating and comparing the PRESS (i.e., the sum of the squared differences between predicted and observed values) for each model dimension with the RSS (i.e., the residual sum of squares) of the previous dimension. A component was considered significant when PRESS was significantly smaller than RSS. The model validity was estimated with the *validate* diagnostic tool of SIMCA-P by PLS model fitting to 100 data sets obtained by random permutation of the response (y) data of the original data set. The computed r^2y- and q^2-values of the derived models were plotted against the correlation coefficients between original and permuted y-variables. After linear regression, the r^2y- and q^2-intercepts of the regression lines were assessed. Intercepts below 0.3–0.4 and 0.05, respectively, were considered

indicative of a valid original model [19]. Theoretical background on PCA and PLS regression analysis can be found in Supplementary Materials.

Results and discussion

Particles of a *poly*(methacrylic acid–*co*–trimethylolpropane trimethacrylate) MIP, targeted for penicillin G (structure 1 in Scheme 1), were prepared as previously described [7]. Control polymer (CP) particles were prepared under identical conditions as the MIP but with Boc-L-Phe-OH (structure 2 in Scheme 1) as the template instead of penicillin G. MIP and CP particles were packed into HPLC columns and retention data of penicillin G and a range of β-lactam analogs were obtained by elution with acetonitrile containing 1–5% of 10 mM Tris-HCl pH 7 (Table 1). The retention obtained on a MIP stationary phase generally originates from both specific interactions and non-specific ones. The specific interactions take place in the recognition sites created during the molecular imprinting procedure while the non-specific binding is due to random interactions with the polymer. The non-specific binding to the MIP was here estimated from the binding to the CP stationary phase.

The analogs injected on the columns included penicillin V (structure 3 in Scheme 1), methicillin (structure 4), nafcillin (structure 5), oxacillin (structure 6), cloxacillin (structure 7), dicloxacillin (structure 8), and piperacillin (structure 9). The molecules are derivatives of 6-aminopenicillanic acid (6-APA, structure 10); they hence share the same backbone but each molecule has a unique substituent at the 6-amino position. All of the antibiotics were retained longer on the MIP column than on the CP column as reflected in their retention factors (k_{MIP} and k_{CP}) and imprinting factors (IFs) (Table 1), indicating that the penicillin G recognition sites created during the imprinting were able to bind not only penicillin G but also

the analogs. The IF is a measure of the imprinting effect and the selectivity of the MIP as it compares the retention on the MIP to that on the CP (IF = k_{MIP}/k_{CP}). The retention times decreased with increasing amounts of buffer used as additive in the mobile phase. The MIP was able to retain and discriminate between the different antibiotics at conditions with 1% and 2.5% of buffer additive in the eluent. At higher amounts of buffer, the specificity of the MIP was lost. These findings are in line with our previous studies on penicillin G imprinted *poly*(methacrylic acid–*co*–trimethylolpropane trimethacrylate); water weakens hydrogen bonding and polar interactions between the template and the polymer [7,8,21].

As discussed in the introduction, size and shape of a molecule have been recognized as important determinants in MIP recognition [22–24]. Among molecules having the same spatial positioning of functional groups as the template, the highest selective binding can be expected with those that provide the best shape complementarity to the MIP cavity. The best fit in this respect is obviously obtained with the template. Molecules with substituents that are too large to fit into the cavity will be excluded from the recognition sites due to steric hindrance and molecules with substituents smaller than those of the template will not provide optimal van der Waals contacts and hydrophobic interactions with the walls of the cavities. Both of these scenarios will result in binding affinities that are lower than that for the template.

To investigate how the structure and the properties of the antibiotics affect their retention on the MIP stationary phase in relation to the CP stationary phase, a set of molecular descriptors were computed (Table S1). The energy of the antibiotics was first minimized by semiempirical quantum mechanical methods to obtain stable 3D molecular conformations (Fig. S1 in Supplementary Materials)

and the molecular descriptors were thereafter computed. The IFs, obtained at elution with 1% of Tris-buffer pH 7 in acetonitrile, were plotted against each of the descriptors and linear regression was carried out to derive linear one-parameter models (Fig. S2 and Table S1 in Supplementary Materials). The data points of penicillin G were excluded from the regression analyses. Visual inspection of the plots as well as assessment of the statistical parameters of the regression lines indicated clear linear relationships between the IF and each of the following descriptors: the molar refractivity (MR), the critical volume (CV), the Connolly solvent-excluded volume (CSEV), the molecular weight (MW), the total connectivity (TC), and the COSMO volume (COSMO V). The descriptors are all related to molecular size and shape and thereby support the previous notion that these properties are important factors in the recognition process. The one-parameter models predicted an IF in the size range of 6.03–6.69 for penicillin G (the observed IF was 7.62). Hence, Penicillin G deviated from the linear one-parameter models derived from the data of the analogs. The remaining descriptors showed moderate or no linear correlation to the IFs of the antibiotic analogs (Fig. S2 and Table S1 in Supplementary Materials).

To obtain a model that correlated the IFs of the antibiotic analogs to all of the descriptors simultaneously, multivariate data analysis was carried out as outlined in Fig. 2. A PCA of the complete data set [i.e., a data matrix consisting of seven observations and thirty variables, the latter consisting of one response (IF) and twenty-nine descriptors] was first performed in order to detect outliers and to evaluate relations, deviations, or groupings within the data (Table 2). The t_2/t_1 score plot in Fig. 3a shows that the analogs were well within the 95% confidence ellipse based on Hotelling's T^2, i.e., no outliers were detected (the two remaining score plots are found in Fig. S3 in Supplementary Materials). The grouping of

nafcillin, oxacillin, cloxacillin, and dicloxacillin in Fig. 3a shows that these antibiotics have similar characteristics. Likewise, the proximity of penicillin V and methicillin indicates that these two antibiotics are more closely related to each other than to any of the other antibiotics. The p_2/p_1 loading plot in Fig. 3b shows that all variables to some extent contribute to the PCA-model (the two remaining loading plots are found in Fig. S4 in Supplementary Materials). Fig. 3b shows that the IF is correlated directly to the total connectivity (TC) and to a lower degree also to the electronic energy (EE), the total energy (TE), and the dipole moment (Dipole). The Connolly solvent excluded volume (CSEV), the molecular weight (MW), the molecular connectivity (MR), the sum of degrees (SD), the critical volume (CV), the COSMO volume (COSMO V), the shape attribute (SA), the core-core repulsion (CCR), the cluster count (CC), and the principal moment of inertia in the x-direction (I_x) are correlated with each other and with the IF inversely. It is noteworthy that many of these descriptors also gave the best one-parameter models.

The PLS regression method can be used to analyze numerous x-variables that are strongly correlated, noisy, and/or incomplete [19,20]. The method hence overcomes the shortcomings of traditional MLR (multiple linear regression). Since PCA showed that several of the descriptors were correlated, PLS regression appeared to be a suitable method for analyzing the data in the present case. Two different strategies were investigated (Fig. 2): In the first one, PLS regression was carried out on all descriptors. In the second strategy, only the descriptors that were indicated by PCA to be directly or inversely correlated with the IF were included. PLS regression on the full data set resulted in a three-component model (referred to as PLS model 1a in Fig. 2 and Table 2) that was found to be non valid by a response permutation test (the methodology of the test is detailed in the methods

section). The VIP plot indicated thirteen descriptors with a VIP value >1 and six descriptors with a VIP value >1.1 (Fig. S5a in Supporting Materials). A second PLS regression analysis, based on the six most influential descriptors from the first PLS model (i.e., the dimensions of the **X** and **Y** matrices were (7 × 6) and (7 × 1), respectively), resulted in a two-component model (PLS model 1b in Fig. 2 and Table 2). The response permutation test indicated that the model was now valid. The VIP plot showed that log P was the least influential descriptor with a VIP value of 0.46 (Fig. S5b in Supporting Material). A third PLS regression analysis was therefore carried out by omitting this descriptor. The result was a one-component model (PLS model 1c in Fig. 2 and Table 2) that used 94.2% of the x-variation for describing 91.1% and predicting 87.1% of the y-variance. The response permutation test indicated that the model was valid. The VIP plot in Fig. 4a shows satisfactory VIP values (>0.95) for the remaining descriptors, i.e., the Connolly solvent excluded volume (CSEV), the molar refractivity (MR), the molecular weight (MW), the critical volume (CV), and the total connectivity (TC). The plot in Fig. 4b shows the observed IFs vs. the IFs predicted by the model. The model predicted an IF for penicillin G of 6.28, which should be compared with the experimentally determined value of 7.62. Hence, the model derived from the antibiotic analogs underestimated the IF of the template penicillin G.

The second strategy, i.e., PLS regression on descriptors selected by PCA, started with an **X** matrix of dimension (7 × 14). The result of the analysis was a one-component model (PLS model 2a in Fig. 2 and Table 2) that was valid according to the response permutation test. The model predicted an IF of 6.16 for penicillin G. All descriptors except for the principal moment of inertia in the x-direction (I_x) and the dipole moment (Dipole) showed VIP values >1.0 (Fig. S6a

in Supporting Materials). A second PLS regression analysis was carried out by omitting the two least influential descriptors. The resultant two-component model (PLS model 2b in Fig. 2 and Table 2) used 98.7% of the x-variation for describing 92.4% and predicting 78.0% of the y-variance. The model was valid according to the response permutation test, although the validity of PLS model 2a appeared to be higher. The VIP plot shows that all descriptors had VIP values >0.95 (Fig. 5a). The observed vs. predicted IF plot is shown in Fig. 5b. The model predicted an IF of 6.45 for penicillin G. The five most influential descriptors according to the VIP plot [i.e., the total connectivity (TC), the Connolly solvent–excluded volume (CSEV), the molar refractivity (MR), the molecular weight (MR), and the critical volume(CV)] are the descriptors found to be most influential in the first modeling strategy detailed above. These descriptors also provided the best one-parameter models as discussed earlier. All of these descriptors are related to topology, shape, size, and volume: the total connectivity is a mathematically derived index of molecular branching, the Connolly solvent-excluded volume is the volume contained within the contact molecular surface formed by rolling a probe sphere over the molecule, the molar refractivity is a measure of the polarizability and is proportional to the molar volume, the molecular weight is the average molecular mass of the molecule, and the critical volume is the volume occupied per mole of the compound at the critical temperature and pressure.

The k_{MIP} was higher than the k_{CP} for all of the antibiotics. Hence, the imprinting of penicillin G created an increased affinity for β-lactam antibiotics as a group. The derived models showed linear relationship between the ratio of the retention factors (i.e., the IF) for the antibiotic analogs and a selection of the molecular descriptors. It is noteworthy that all of the derived models underestimated the IF for the template penicillin G. This finding supports the

hypothesis that the template functions as a mold that generates complementary recognition sites in the polymer network. If the imprinting effect had been a product of random structural changes (e.g., altered porosity, surface area, and exposition or shielding of functional groups) of the polymer network without formation of selective recognition sites, one would expect that the derived linear models should be valid for the template as well. Instead, the IF for the template was higher than predicted by the models. It remains to be investigated if molecular descriptors quantifying the spatial overlap of the template and the analogs can generate linear models for both the template and the analogs.

Models that relate the IF to the molecular descriptors can be used to predict the ability of a MIP to recognize an untested compound. A potential use of the models derived in the present study is for in-silico screening of virtual libraries of compounds for the purpose of identifying novel antibiotic leads. Another potential application of the models is to assess the feasibility of applying the MIP for solid-phase extraction of novel β-lactam antibiotics. As mentioned in the introduction section, molecules structurally related to the target molecules are often used as the template instead of the target molecule itself for the generation of molecularly imprinted solid-phase extraction media.

CONCLUSIONS

This study has demonstrated that molecular descriptors associated with topology, shape, size, and volume of a range of β-lactam antibiotics are correlated with their IFs on a MIP stationary phase imprinted with the template penicillin G. In one-parameter models, linear relationships between the IF and each of the following descriptors were observed: the molar refractivity, the critical volume, the

Connolly solvent-excluded volume, the molecular weight, the total connectivity, and the COSMO volume. These descriptors were influential also in models obtained by multivariate data analysis methods (i.e., PCA and PLS regression). The template penicillin G diverged from linearity with an IF that was higher than predicted from the models derived from the data of the analogs. This indicates that the imprinting procedure indeed generated molecular recognition sites for the template. Good correlation between the observed and the predicted IFs propose that the models can be used for predictions of untested antibiotics.

ACKNOWLEDGMENTS *This work was supported by the Swedish Foundation for Strategic Research through an Individual Grant for the Advancement of Research Leaders (INGVAR) to MK.*

References

1. Kempe H, Kempe M (2008) In: Albericio F, Pulla-Tuche J (eds) The Power of Functional Resins in Organic Synthesis, Wiley-VCH, Weinheim, 2008, pp 15–44
2. Johnston AM (2001) Int J Antimicrob Agents 18: 291–294
3. Boerlin P, Wissing A, Aarestrup FM, Frey J, Nicolet J (2001) J Clin Microbiol 39:4193–4195
4. Patriarca G, D'Ambrosio C, Schiavino D, Nucera E, Sun JY, DelNinno M, Misuraca C, Buonomo A (1999) Int Arch Allergy Immunol 118: 247–250
5. Threlfall EJ, Ward LR, Frost JA, Willshaw GA (2000) Int J Food Microbiol 62, 1–5
6. Smith DL, Harris AD, Johnson JA, Silbergeld EK, Morris Jr JG, (2002) Proc Natl Acad Sci USA 99:6434–6439
7. Cederfur J, Pei Y, Zihui M, Kempe M (2003) J Comb Chem 5:67–72
8. Kempe H, Kempe M (2010) Anal Bioanal Chem 396:1599–1606
9. David V, Medvedovici A (2007) J Liq Chromatogr Rel Technol 30:768–789
10. Andersson LI, Paprica A, Arvidsson T (1997) Chromatographia 46: 57–62
11. Matsui J, Fujiwara K, Takeuchi T (2000) Anal Chem 72: 1810–1813
12. Martin P, Wilson ID, Jones GR (2000) J Chromatogr A 889:143–147
13. Dirion B, Lanza F, Sellergren B, Chassaing C, Venn R, Berggren C (2002) Chromatographia 56:237–241
14. Theodoridis G, Kantifes A, Manesiotis P, Raikos N, Tsoukali-Papadopoulou H (2003) J Chromatogr A 987:103–109
15. Manesiotis P, Hall AJ, Courtois J, Irgum K, Sellergren B (2005) Angew Chem Int Ed Engl 44:3902 –3906
16. Ye L, Yu Y, Mosbach K (2001) Analyst 126: 760–765
17. O'Connor NA, Paisner DA, Huryn D, Shea KJ (2007) J Am Chem Soc 129:1680–1689
18. Connolly ML (1985) J Am Chem Soc 107:1118–1124
19. Eriksson L, Johansson E, Kettaneh-Wold N, Wold S (2001) Multi- and megavariate data analysis. Principles and applications. Umetrics, Umeå
20. Wold S, Sjöström M, Eriksson L (2001) Chemom Intell Lab Syst 58: 109–130

21. Benito-Peña E, Moreno-Bondi MC, Aparicio S, Orellana G, Cederfur J, Kempe M (2006) Anal Chem 78: 2019–2027
22. Shea KJ, Sasaki DY (1989) J Am Chem Soc 111: 3442–3444
23. Kempe M, Mosbach K (1994) Int J Peptide Protein Res 44: 603–606
24. Spivak DA, Simon R, Campbell J (2004) Anal Chim Acta 504: 23–30

Legends

Scheme 1. Structure of penicillin G (**1**); Boc-L-Phe-OH (**2**); penicillin V (**3**); methicillin (**4**); nafcillin (**5**); oxacillin (**6**); cloxacillin (**7**); dicloxacillin (**8**); piperacillin (**9**); and 6-aminopenicillanic acid (**10**)

Fig. 1 Schematic representation of the concept of non-covalent molecular imprinting. Monomers and template self-assemble by non-covalent interactions (I). A rigid polymer is formed after co-polymerization with a cross-linker (II). The template is extracted from the polymer (III), leaving a recognition site that can rebind the template (IV)

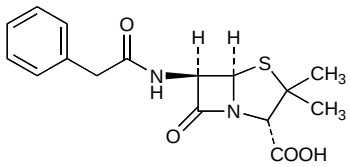
Fig. 2 Schematic flowchart of the generation of response and predictor variables (i.e., imprinting factors (IFs) and molecular descriptors) and QSRR models by multivariate data analysis. (a) A PCA model was derived by PCA on the full data set. The most influential descriptors were identified from the PCA loading plot; (b) PLS models were derived by PLS regression analysis starting with the full data set to generate PLS model 1a, followed by successive reduction of predictor variables to generate PLS models 1b and 1c; (c) PLS models were derived by PLS regression analysis starting with a reduced data set, containing predictor variables identified to be influential by PCA, to generate PLS model 2a followed by reduction of predictor variables to generate PLS model 2b. The derived models are marked with grey boxes

Fig. 3 (a) PCA t_2/t_1 score plot; (b) PCA p_2/p_1 loading plot

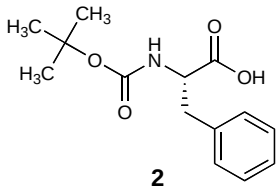
Fig. 4 (a) VIP plot of PLS model 1c; and (b) Plot of observed vs. predicted imprinting factor (IF) of PLS model 1c

Fig. 5 (a) VIP plot of PLS model 2b; and (b) Plot of observed vs. predicted imprinting factor (IF) of PLS model 2b

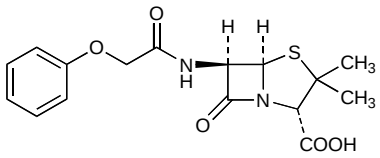
Scheme 1



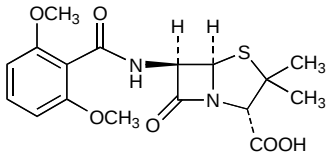
1



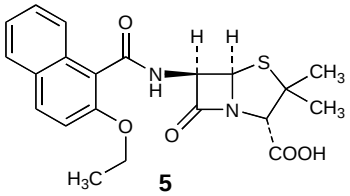
2



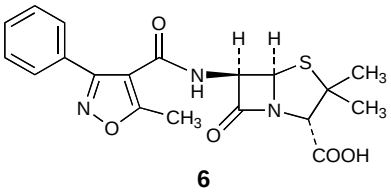
3



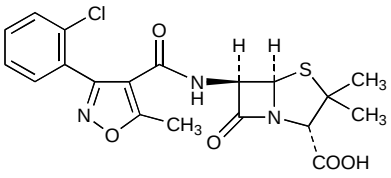
4



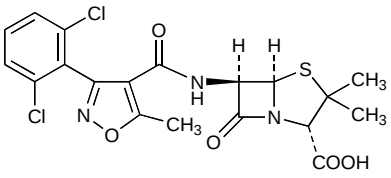
5



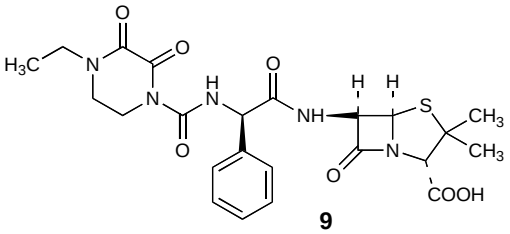
6



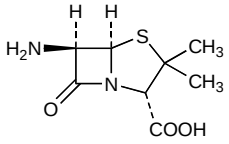
7



8



9



10

Fig. 1

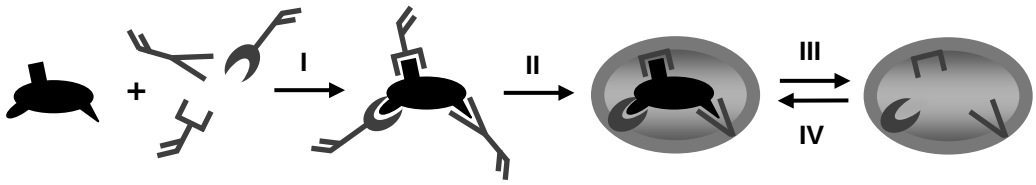


Fig. 2

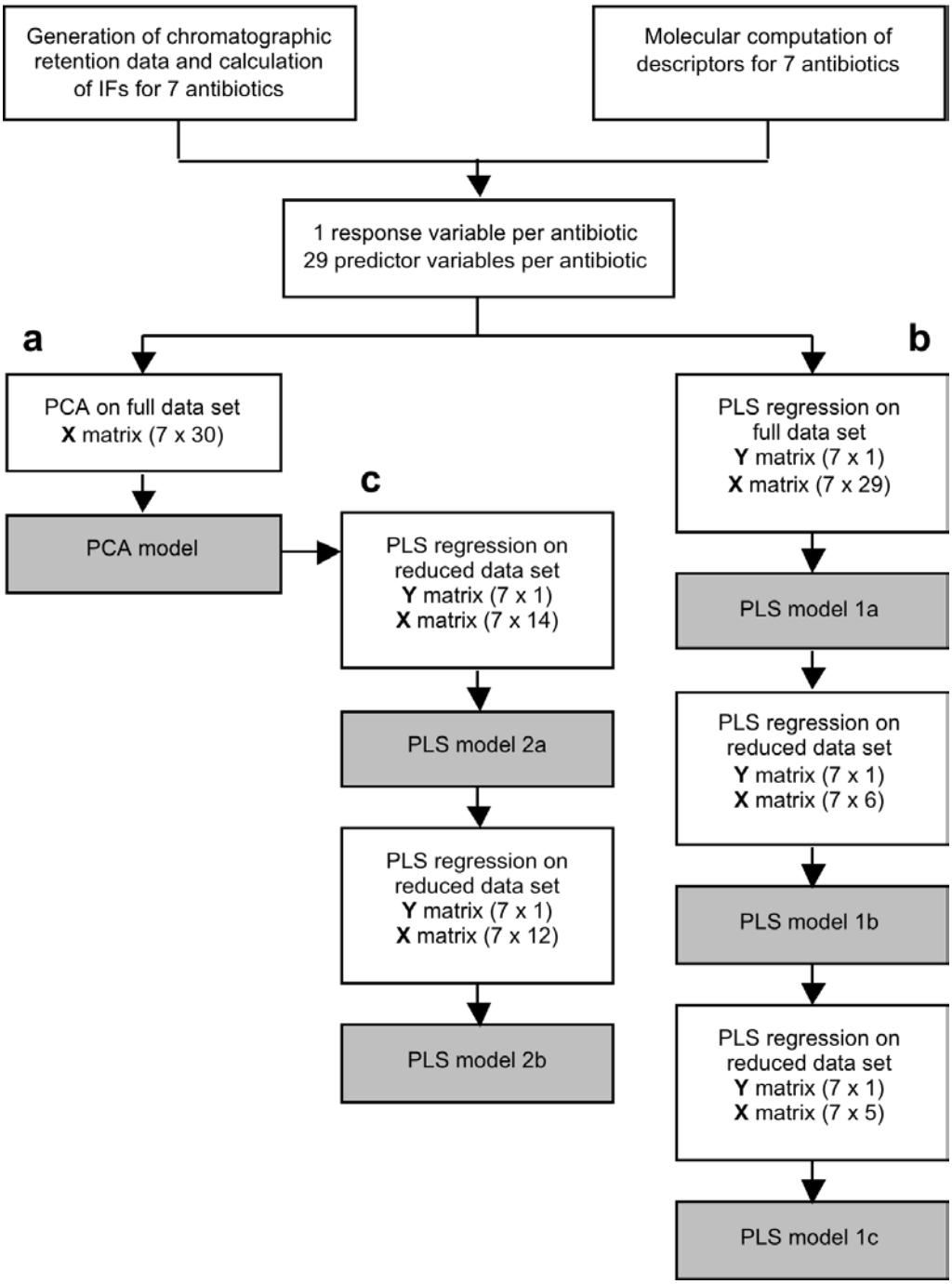

Generation of chromatographic retention data and calculation of IFs for 7 antibiotics
Molecular computation of descriptors for 7 antibiotics
1 response variable per antibiotic
29 predictor variables per antibiotic
a
b
c
PCA on full data set
X matrix (7 x 30)
PCA model
PLS regression on reduced data set
Y matrix (7 x 1)
X matrix (7 x 14)
PLS model 2a
PLS regression on reduced data set
Y matrix (7 x 1)
X matrix (7 x 12)
PLS model 2b
PLS regression on full data set
Y matrix (7 x 1)
X matrix (7 x 29)
PLS model 1a
PLS regression on reduced data set
Y matrix (7 x 1)
X matrix (7 x 6)
PLS model 1b
PLS regression on reduced data set
Y matrix (7 x 1)
X matrix (7 x 5)
PLS model 1c

Fig. 3a

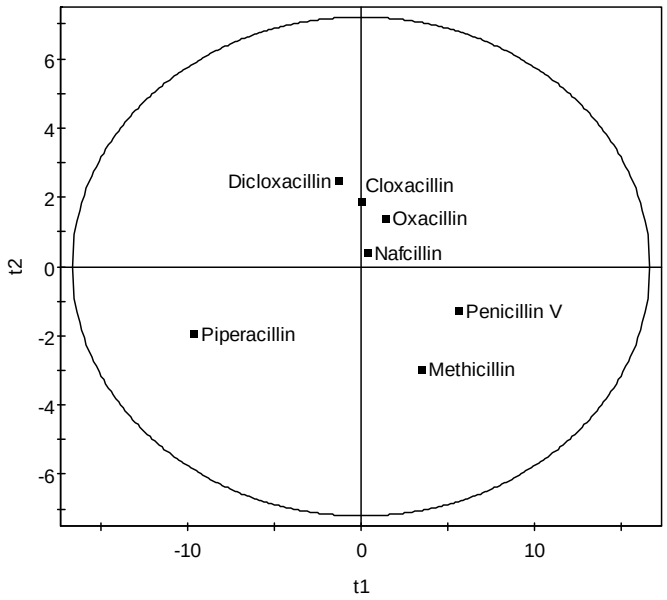

6
4
2
0
-2
-4
-6
t2
Dicloxacillin
Cloxacillin
Oxacillin
Nafcillin
Penicillin V
Piperacillin
Methicillin
-10
0
10
t1

Fig. 3b

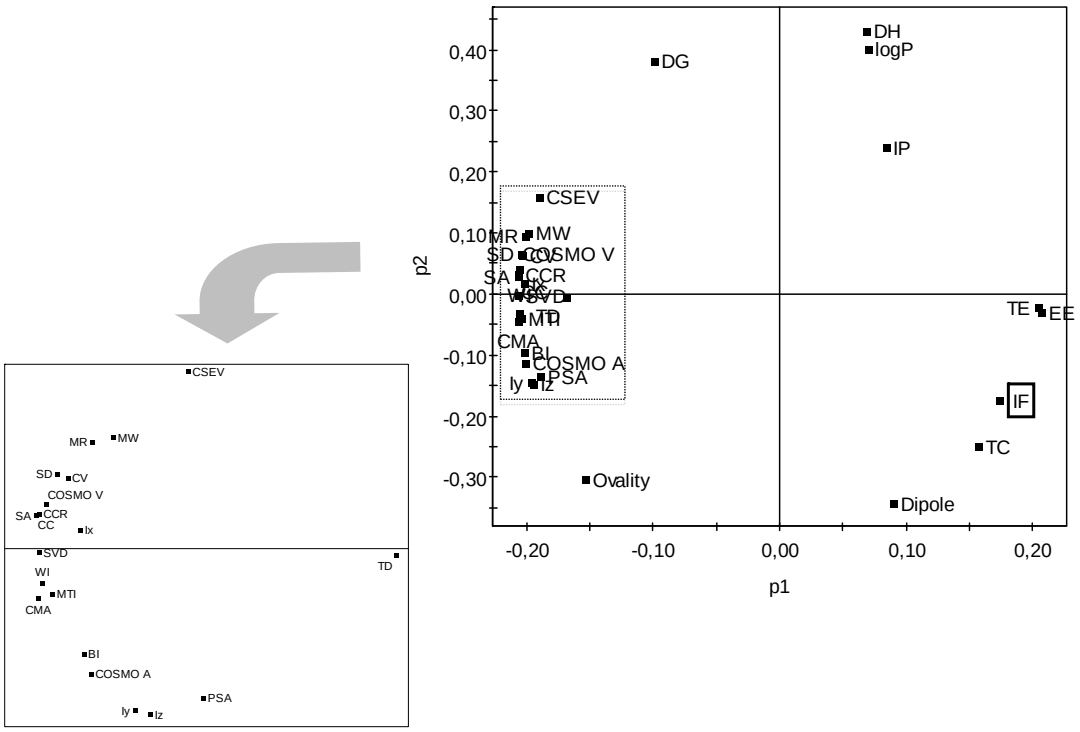

DH
logP
DG
IP
CSEV
MW
COSMO V
CCR
TE
EE
MTI
BI
COSMO A
PSA
IF
TC
Ovality
Dipole
0,40
0,30
0,20
0,10
0,00
-0,10
-0,20
-0,30
-0,20
-0,10
0,00
0,10
0,20
p1
p2
CSEV
MR
MW
SD
CV
COSMO V
SA
CCR
CC
Ix
SVD
TD
WI
MTI
CMA
BI
COSMO A
PSA
Iy
Iz

Fig 4a

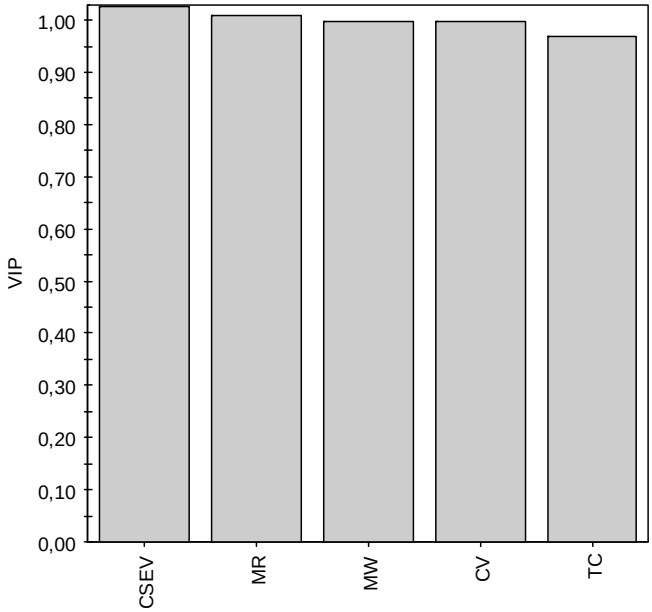


Fig. 4b

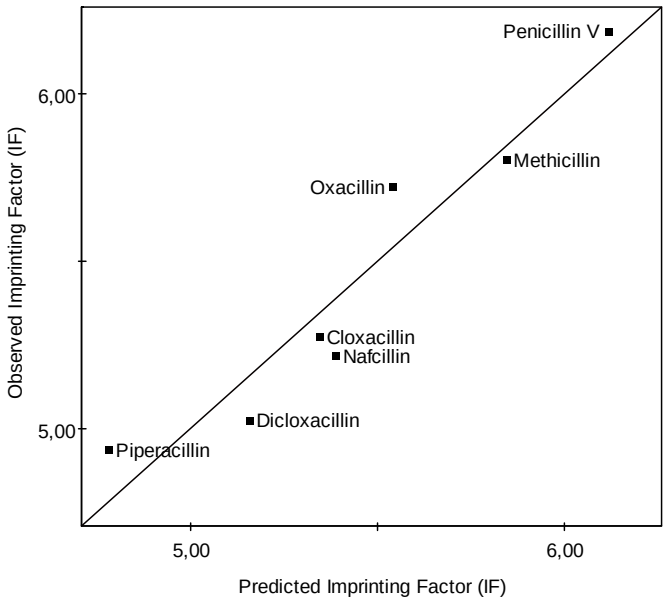

Penicillin V
6,00
Methicillin
Oxacillin
Cloxacillin
Nafcillin
Dicloxacillin
5,00
Piperacillin
Observed Imprinting Factor (IF)
5,00
6,00
Predicted Imprinting Factor (IF)

Fig. 5a

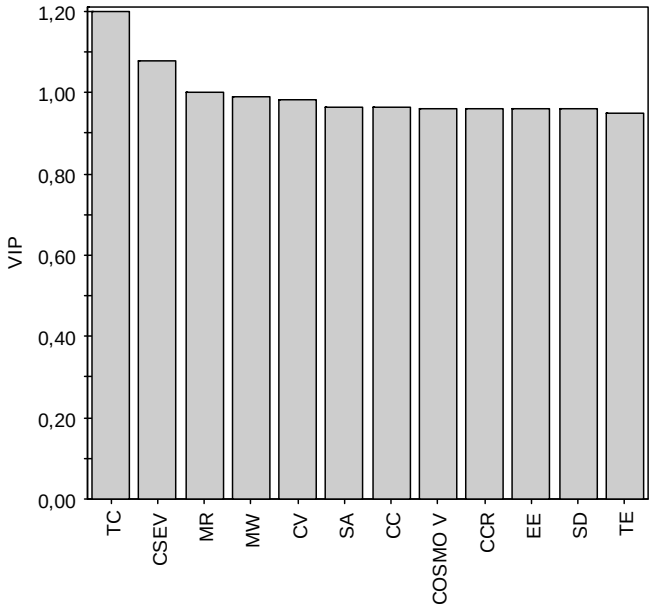

VIP
1,20
1,00
0,80
0,60
0,40
0,20
0,00
TC
CSEV
MR
MW
CV
SA
CC
COSMO V
CCR
EE
SD
TE

Fig. 5b

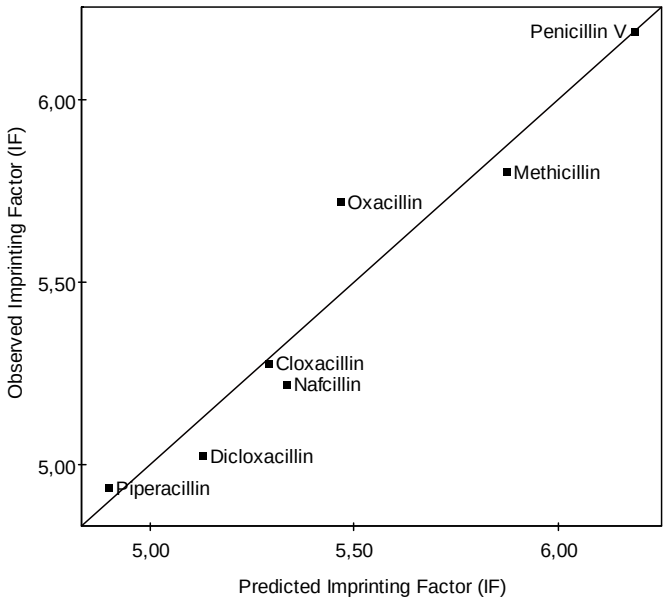

Penicillin V
Methicillin
Oxacillin
Cloxacillin
Nafcillin
Dicloxacillin
Piperacillin
6,00
5,50
5,00
Observed Imprinting Factor (IF)
5,00
5,50
6,00
Predicted Imprinting Factor (IF)

Table 1 Retention factors (k_{MIP} and k_{CP}) and imprinting factors (IFs) of antibiotics eluted on MIP and CP stationary phases with mobile phases consisting of acetonitrile containing the indicated amount of 10 mM Tris-HCl buffer pH 7 as additive

Antibiotic	Percentage of buffer additive in the mobile phase								
	1			2.5			5		
	k_{MIP}	k_{CP}	IF	k_{MIP}	k_{CP}	IF	k_{MIP}	k_{CP}	IF
Penicillin G (1 in Scheme 1)	16.14	2.12	7.62	6.55	1.19	5.49	2.21	0.33	6.63
Penicillin V (3 in Scheme 1)	12.11	1.96	6.19	5.01	1.11	4.54	1.75	0.32	5.54
Methicillin (4 in Scheme 1)	7.25	1.25	5.80	3.20	0.75	4.24	1.27	0.21	6.05
Nafcillin (5 in Scheme 1)	8.91	1.71	5.22	4.13	1.02	4.06	1.68	0.32	5.29
Oxacillin (6 in Scheme 1)	10.50	1.84	5.72	4.62	1.07	4.34	1.72	0.33	5.16
Cloxacillin (7 in Scheme 1)	8.75	1.66	5.28	4.05	0.98	4.12	1.54	0.30	5.15
Dicloxacillin (8 in Scheme 1)	7.95	1.58	5.03	3.72	0.95	3.93	1.46	0.30	4.90
Piperacillin (9 in Scheme 1)	8.32	1.69	4.94	3.87	0.97	3.99	1.44	0.25	5.88

Table 2 Multivariate data analysis

Model	Number of x variables	Number of y variables	Number of PCs	r^2x	$r^2x_{cumulative}$	r^2y	$r^2y_{cumulative}$	q^2	$q^2_{cumulative}$	PLS *validate* test
PCA model	30	–	1 2 3	0.774 0.145 0.059	0.774 0.920 0.979	– – –	– – –	0.500 0.457 0.527	0.500 0.728 0.872	–
PLS model 1a	29	1	1 2 3	0.772 0.134 0.072	0.772 0.906 0.978	0.746 0.180 0.049	0.746 0.926 0975	0.575 0.568 0.483	0.575 0.816 0.905	Intercepts: r^2 = (0, 0.61) q^2 = (0, -0.03)
PLS model 1b	6	1	1 2	0.783 0.190	0.783 0.974	0.926 0.026	0.926 0.952	0.915 0.247	0.915 0.936	Intercepts: r^2 = (0, 0.21) q^2 = (0, -0.23)
PLS model 1c	5	1	1	0.942	0.942	0.911	0.911	0.871	0.871	Intercepts: r^2 = (0, -0.11) q^2 = (0, -0.27)
PLS model 2a	14	1	1	0.906	0.906	0.831	0.831	0.698	0.698	Intercepts: r^2 = (0, -0.03) q^2 = (0, -0.20)
PLS model 2b	12	1	1 2	0.959 0.028	0.959 0.987	0.845 0.078	0.845 0.924	0.715 0.228	0.715 0.780	Intercepts: r^2 = (0, 0.32) q^2 = (0, -0.02)

SUPPLEMENTARY MATERIALS

QSRR analysis of β-lactam antibiotics on a penicillin G targeted MIP stationary phase

Henrik Kempe and Maria Kempe*

Biomedical Polymer Technology, Department of Experimental Medical Science, Biomedical Center, D11, Lund University, SE–221 84 Lund, Sweden

* Corresponding author

Phone: +46 46 2220857

Fax: +46 46 2221410

Email: Maria.Kempe@med.lu.se

URL: www.biomedicalpolymers.bmc.lu.se

PCA theory

PCA starts with a data matrix of N observations and K variables. Each observation of the matrix is first plotted in a K-dimensional variable space. After mean-centering and scaling to unit variance the data are approximated to a line passing through the average point in the K-dimensional space. The line is referred to as the first principal component (PC). Projection of the observations onto the PC-line gives the *score* value of each observation. A second PC, orthogonal to the first PC and again passing through the average point is then calculated. More PCs are calculated as needed in order to model the variation in the data adequately. The optimal dimensionality is determined by cross-validation, a procedure that compares the predictive residual sum of squares (PRESS, the sum of the squared differences between predicted and observed values) of the present dimension with the residual sum of squares (RSS) of the previous dimension. The model building continues as long as PRESS is significantly smaller than RSS. Projection of the observations onto the plane defined by two PCs gives a *score plot*. The score plot reveals strong outliers and shows the relation between the observations. The *loadings* are calculated from the angle between the PC and the original variable in the K-dimensional space. The *loading plot* shows the influence of the variables and how they are correlated. The most influential variables are those positioned furthest away from the loading plot origin. Variables that are highly correlated are grouped together while variables that are inversely correlated are positioned on opposite sides of the plot origin.

PLS regression analysis theory

In PLS regression, two data matrices **X** (the predictor variables) and **Y** (the response variables) are related to each other by a linear multivariate model. Prior

to PLS regression analysis, the data are mean-centered and scaled to unit variance. The first PLS component calculated is a line in the x-space that passes through the origin, approximates the x variables, and provides a good correlation with the y-vector. Projection of the observations onto the line forms the x score vector $\mathbf{t_1}$. If one component is insufficient to model the variation in the y data, a second component is calculated. The second component is orthogonal to the first component and passes through the origin of the x-space. Projection of the observations onto this second component produces the score vector $\mathbf{t_2}$, etc. The procedure is repeated until the variation in the y-data is modeled adequately according to the cross-validation method (giving a final number of components equal to A). The x score vectors $\mathbf{t_a}$ (a = 1,2, …, A) can be seen as new variables, often referred to as *latent variables*. The scores matrix (**T**) is related to the original x data matrix (**X**) as **T** = **XW*** where **W*** is the x-weights matrix containing the x-weight vectors $\mathbf{w_a^*}$ (a = 1,2, …, A). The x-weights show the influence of the original x-variables on the new latent variables, i.e., the score vectors $\mathbf{t_a}$. The VIP (variable influence on projection) parameter is a weighted sum of squares of **w*** that takes the y-variance in each dimension into account. Predictor variables with large VIP values (>0.7–0.8; preferably >1) are the most relevant for explaining the variation in the response variables.

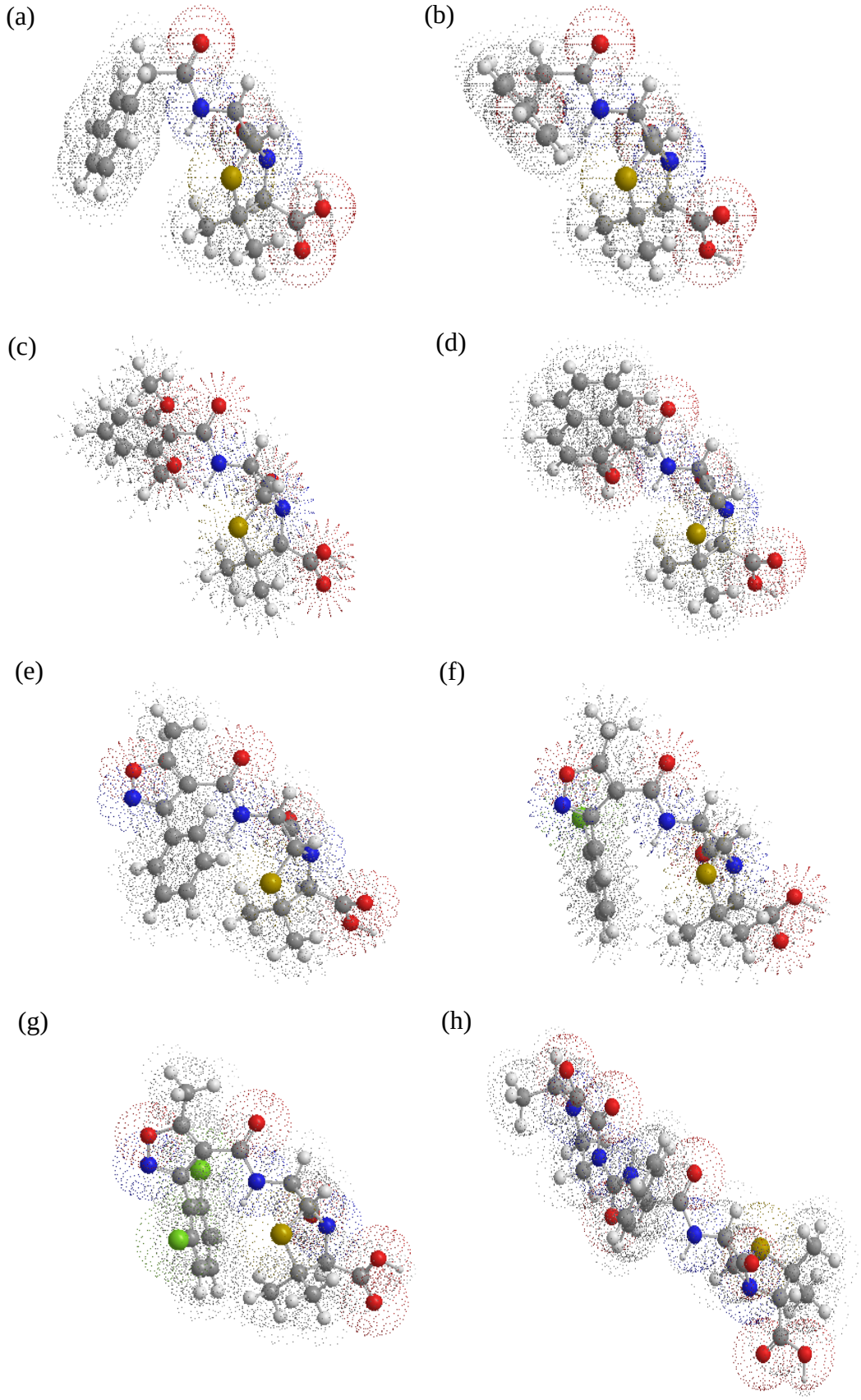


Fig. S1 3D structures of (a) penicillin G; (b) penicillin V; (c) methicillin; (d) nafcillin; (e) oxacillin; (f) cloxacillin; (g) dicloxacillin; and (h) piperacillin

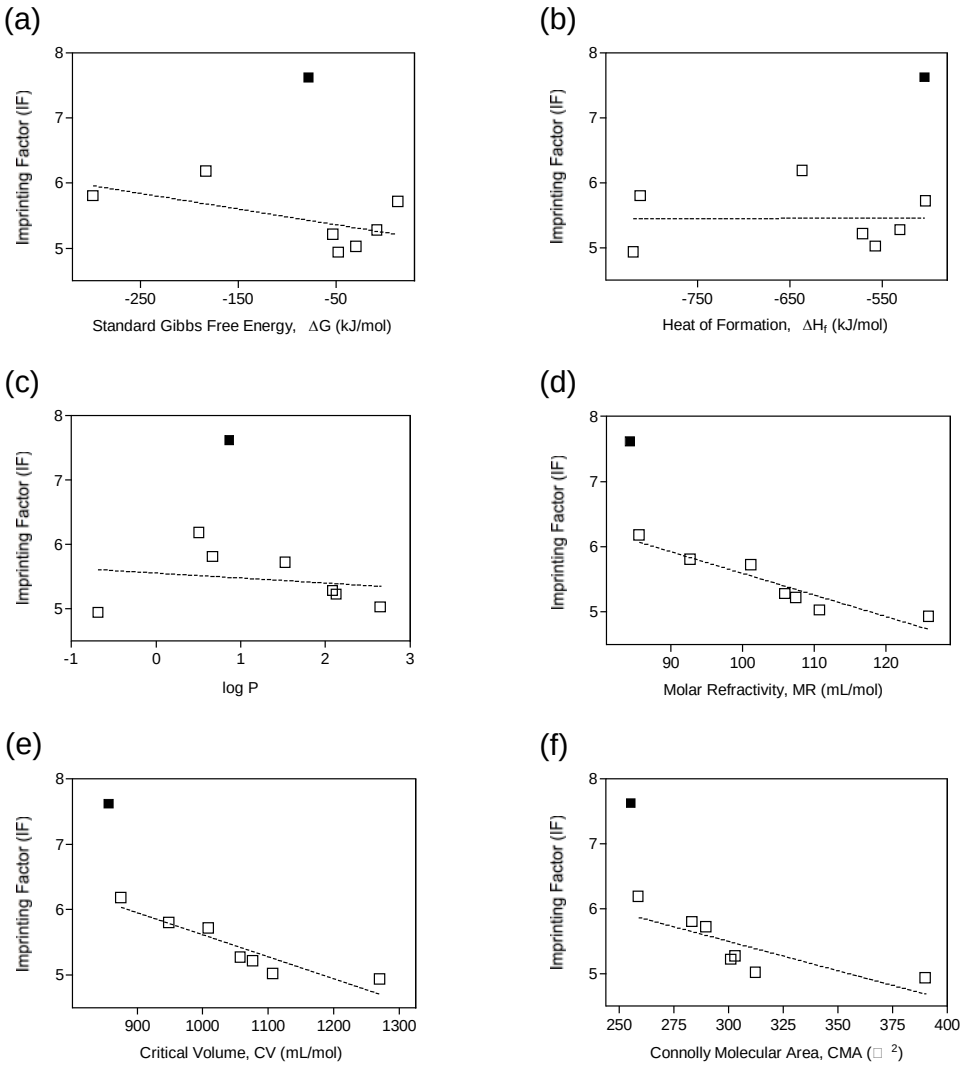
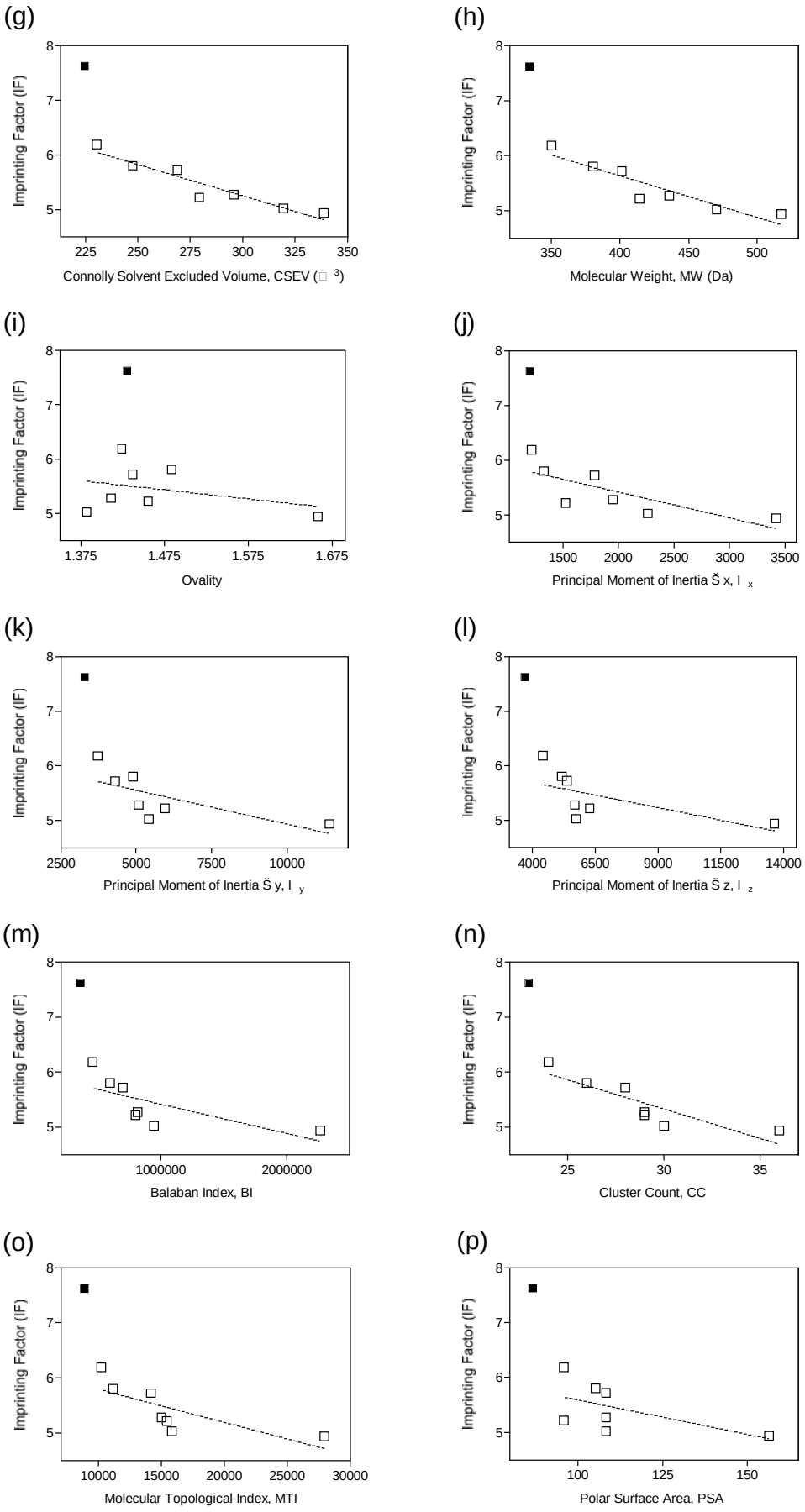
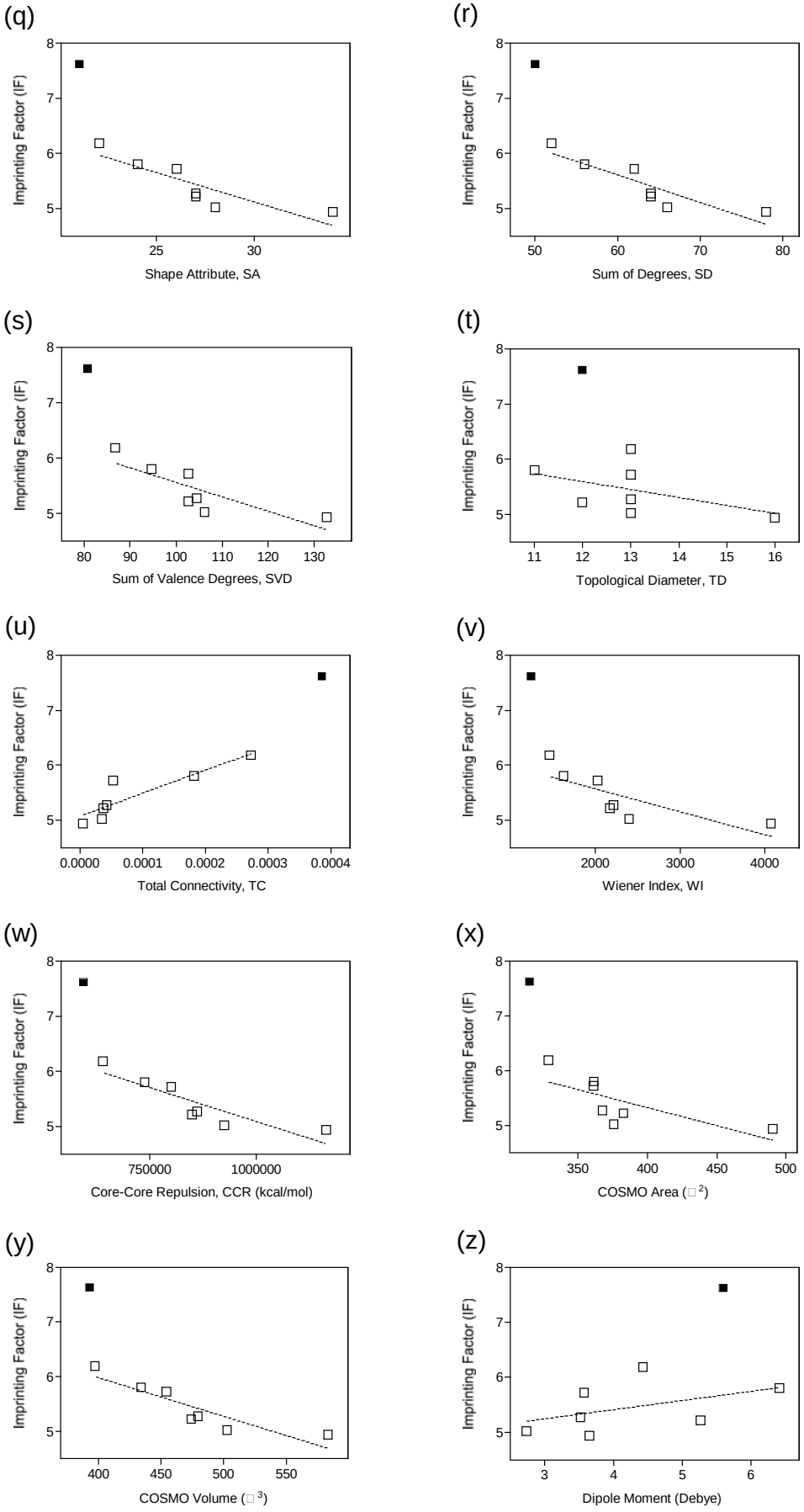


Fig. S2 Correlation of the retention factor and (a) the standard Gibbs free energy, (b) the heat of formation, (c) the log P, (d) the molar refractivity, (e) the critical volume, (f) the Connolly molecular area, (g) the Connolly Solvent-Excluded Volume, (h) the molecular weight, (i) the ovality, (j) the principal moment of inertia – x, (k) the principal moment of inertia – y, (l) the principal moment of inertia – z, (m) the Balaban index, (n) the cluster count, (o) the molecular topological index, (p) the polar surface area, (q) the shape attribute, (r) the sum of degrees, (s) the sum of valence degrees, (t) the topological diameter, (u) the total connectivity, (v) the Wiener index, (w) the core-core repulsion, (x) the COSMO area, (y) the COSMO volume (z) the dipole moment, (å) the electronic energy, (ä) the ionization potential, and (ö) the total energy

***(Fig. S2** Continued)*

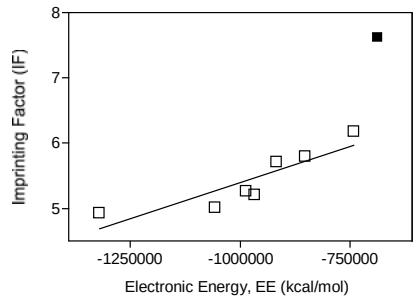


***(Fig. S2** Continued)*

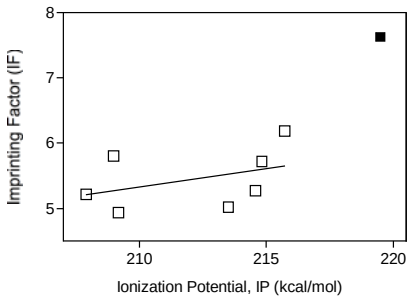


***(Fig. S2** Continued)*

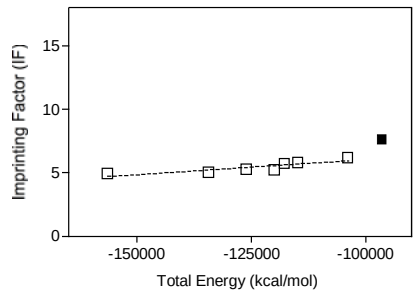
(□)



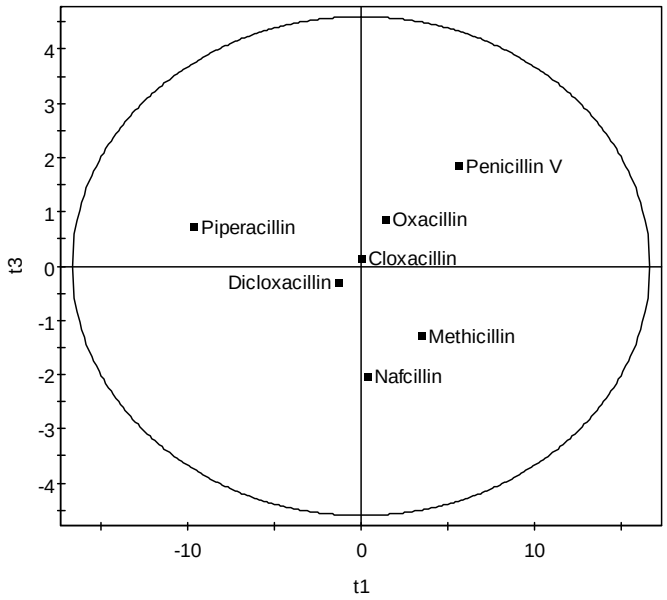
(□)



(□)



(a)



(b)

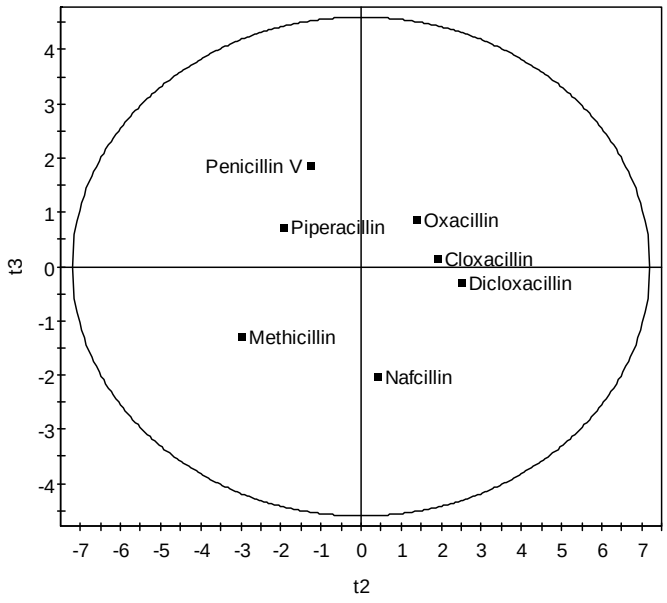
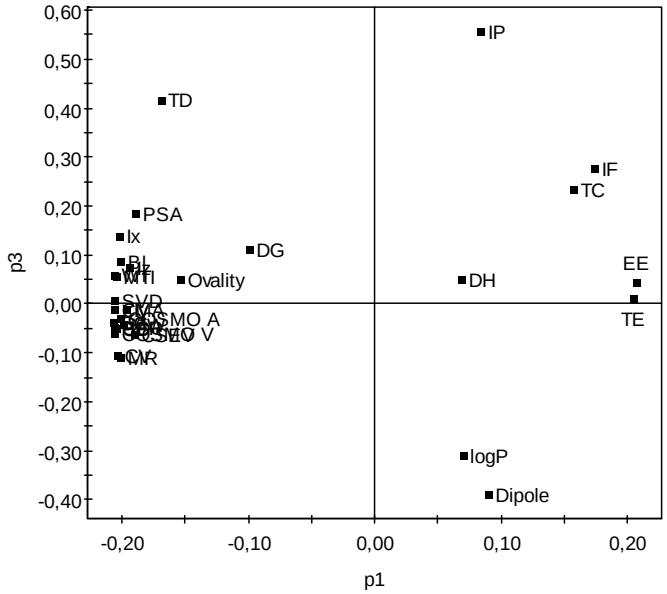


Fig. S3 (a) PCA t_3/t_1 score plot and PCA t_3/t_2 score plot

(a)



(b)

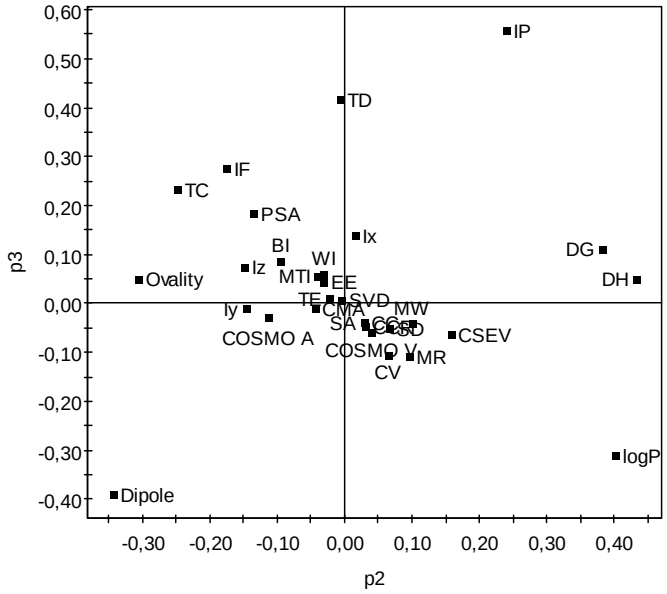
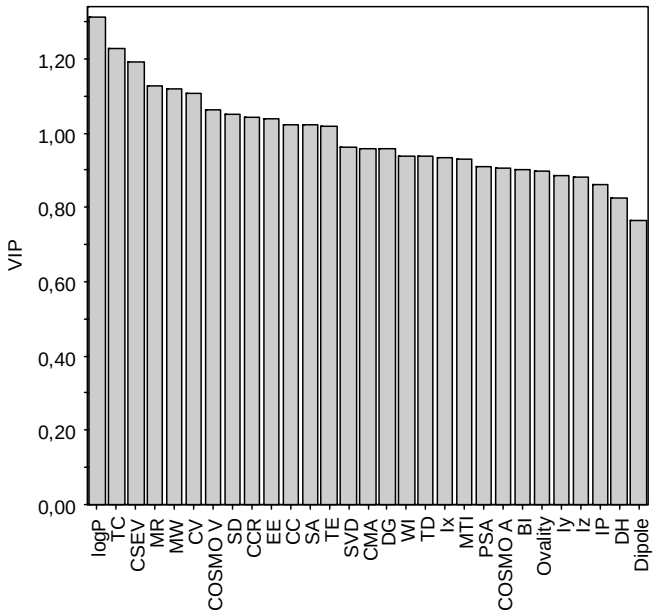


Fig. S4 (a) PCA p_3/p_1 loading plot and (b) PCA p_3/p_2 loading plot

(a)



(b)

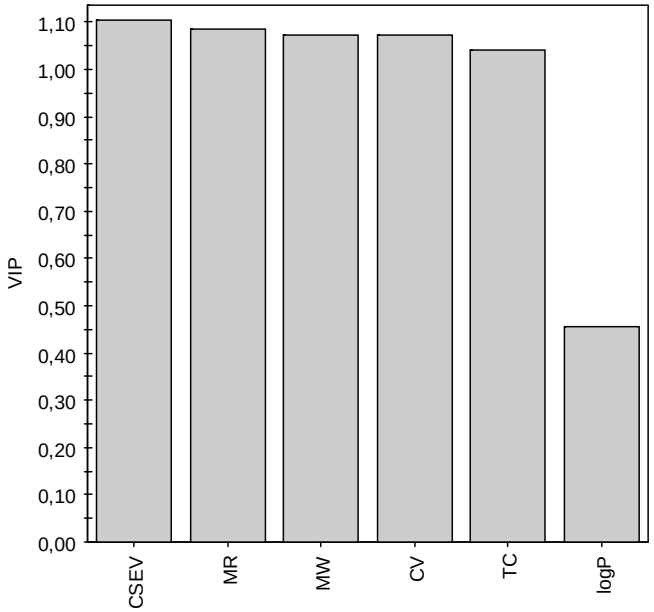
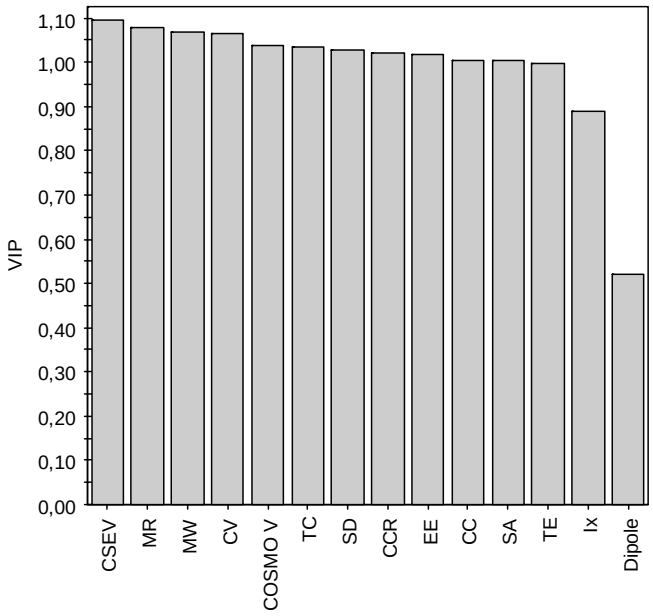


Fig. S5 (a) VIP plot of PLS model 1a; (b) VIP plot of PLS model 1b

(a)



(b)

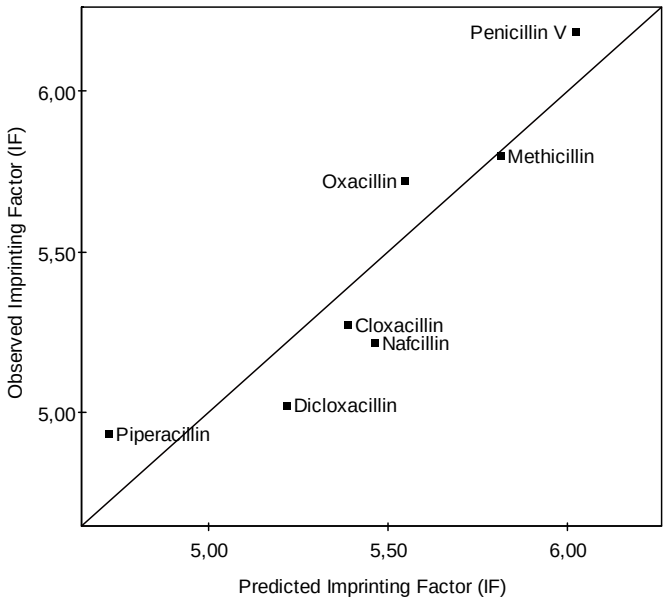


Fig. S6 (a) VIP plot of PLS model 2a; (b) Plot of observed vs. predicted imprinting factor (IF) of PLS model 2a

Table S1 One-parameter models and statistical data[a]

Type of Descriptor	**Descriptor**	**Abbre-viation**	**Unit of Descriptor**	**Model**	**r^2**	**$S_{y.x}$**	**F**	**P**
Thermo-dynamic	Standard Gibbs Free Energy	ΔG (DG)	kJ/mol	k_{MIP} = -0.002408*ΔG + 5.244	0.349	0.406	2.680	0.163
	Heat of Formation	ΔH_f (DH)	kJ/mol	k_{MIP} = 0.00001516*ΔH_f + 5.463	0	0.504	0	0.993
	log P	logP		k_{MIP} = -0.07676*log P + 5.551	0.038	0.494	0.197	0.676
	Molar Refractivity	MR	mL/mol	k_{MIP} = -0.03293*MR + 8.885	0.874	0.179	34.58	0.002
Steric/ Topology	Critical Volume	CV	mL/mol	k_{MIP} = -0.003366*CV + 8.984	0.851	0.194	28.66	0.003
	Connolly Molecular Area	CMA	$Å^2$	k_{MIP} = -0.009032*CMA + 8.211	0.649	0.298	9.258	0.029
	Connolly Solvent-Excluded Volume	CSEV	$Å^3$	k_{MIP} = -0.01140*CSEV + 8.679	0.903	0.157	46.46	0.001
	Molecular Weight	MW	Da	k_{MIP} = -0.007563*MW + 8.662	0.854	0.192	29.31	0.003
	Ovality	Ovality	–	k_{MIP} = -1.665*Ovality + 7.892	0.110	0.475	0.616	0.468
	Principal Moment of Inertia, x-axis	I_x	g/mol$Å^2$	k_{MIP} = -0.0004705*I_x + 6.360	0.591	0.322	7.229	0.043
	Principal Moment of Inertia, y-axis	I_y	g/mol$Å^2$	k_{MIP} = -0.0001242*I_y + 6.177	0.478	0.364	4.585	0.085
	Principal Moment of Inertia, z-axis	I_z	g/mol$Å^2$	k_{MIP} = -0.00009186*Iz + 6.061	0.395	0.392	3.262	0.131
	Balaban Index	BI	–	k_{MIP} = -0.0000005313*BI + 5.951	0.493	0.358	4.873	0.078
	Cluster Count	CC	–	k_{MIP} = -0.1064*CC + 8,525	0.757	0.248	15.62	0.011
	Molecular Topological Index	MTI	–	k_{MIP} = -0.00005995*MTI + 6.394	0.578	0.327	6.862	0.047
	Polar Surface Area	PSA	$Å^2$	k_{MIP} = -0.01249*PSA + 6.842	0.316	0.417	2.314	0.189
	Shape Attribute	SA	–	k_{MIP} = -0.1066*SA + 8.318	0.757	0.248	15.61	0.011

	Sum of Degrees	SD	–	k_{MIP} = -0.04973*SD + 8.593	0.793	0.229	19.12	0.007
	Sum of Valence Degrees	SVD	–	k_{MIP} = -0.02618*SVD + 8.183	0.657	0.295	9.596	0.027
	Topological Diameter	TD	–	k_{MIP} = -0.1437*TD + 7.322	0.228	0.443	1.476	0.279
	Total Connectivity	TC	–	k_{MIP} = 4177*TC + 5.078	0.805	0.222	20.71	0.006
	Wiener Index	WI	–	k_{MIP} = -0.0004169*WI + 6.405	0.604	0.317	7.626	0.040
Electronic	Core-Core repulsion	CCR	kcal/mol	k_{MIP} = -0.000002458*CCR + 7.552	0.781	0.236	17.82	0.008
	COSMO Area	COSMO A	$Å^2$	k_{MIP} = -0.006569*COSMO A + 7.957	0.533	0.344	5.695	0.063
	COSMO Volume	COSMO V	$Å^3$	k_{MIP} = -0.007040*COSMO V + 8.799	0.806	0.222	20.81	0.006
	Dipole Moment	Dipole	Debye	k_{MIP} = 0.1660*Dipole + 4.752	0.203	0.450	1.277	0.310
	Electronic Energy	EE	kcal/mol	k_{MIP} = 0.000002230*EE + 7.635	0.779	0.234	17.63	0.009
	Ionization Potential	IP	kcal/mol	k_{MIP} = 0.05805*IP -6.860	0.172	0.458	1.037	0.355
	Total Energy	TE	kcal/mol	k_{MIP} = 0.00002358*TE + 8.396	0.749	0.253	14.88	0.012

[a] r^2, the coefficient of determination; $S_{y.x}$, the standard deviation of residuals; F, the F-test; and P, the P value