Published in final edited form as:

Chem Res Toxicol. 2007 January ; 20(1): 11–19. doi:10.1021/tx060188l.

Structural Determinants of Binding of Aromates to Extracellular Matrix:

A Multi-Species Multi-Mode CoMFA Study

Yufen Zhang, Viera Lukacova[†], Vladimir Bartus, and Stefan Balaz^{*}

Department of Pharmaceutical Sciences, College of Pharmacy, North Dakota State University, Fargo, North Dakota 58105, USA

Abstract

For small molecules acting in tissues, including signaling peptides, effectors, inhibitors, and other drug candidates, nonspecific binding to the extracellular matrix (ECM) is a critical phenomenon affecting their disposition, toxicity, and other effects. A commercially available ECM mimic, forming a solidified layer at the bottom of the vials, was used to measure the association constants of 25 simple aromatic compounds to two forms of ECM proteins, solidified (s-ECM) and dissolved (d-ECM) in the buffer during the incubation. Except for small homologous series, the binding data did not correlate with the lipophilicity and acidity of the compounds, contrary to a common expectation for nonspecific binding. To elucidate the putative structures of averaged binding sites of s-ECM and d-ECM, the Comparative Molecular Field Analysis (CoMFA) was applied in a modified version taking into consideration that multiple modes and multiple species may be involved. The method shapes a receptor site model from a set of grid points, in which the interaction energies between a probe atom and superimposed ligands are calculated. Electrostatic and steric energies in the grid points are characterized by regression coefficients. The forward-selection, nonlinear regression analysis was used to optimize the coefficients in the novel multi-species, multi-mode CoMFA models. These models showed satisfactory descriptive and predictive abilities for both s-ECM and d-ECM binding data, which were better than those obtained with the standard, one-mode CoMFA analysis. The calibrated models, defining the electrostatic and van der Waals regions of putative binding sites, are suitable for the prediction of ECM binding for untested chemicals.

Keywords

multi-species, multi-mode binding; Comparative Molecular Field Analysis - CoMFA; extracellular matrix - ECM binding; protein binding; disposition

Introduction

For reactive chemicals and metabolites, noncovalent binding to proteins precedes the covalent bond formation. An understanding of this phenomenon in terms of the structures and properties of chemicals would help in the identification of vulnerable proteins and their parts. The binding of chemicals to nonreceptor proteins was shown to depend on lipophilicity (1-4) and acidity (5). Is this a general model or rather a local relationship that breaks down for an expanded set of chemicals? The answer has important implications for predictive toxicology, indicating

[†]Current address: Simulations Plus, Inc., Lancaster, CA 93534, USA

^{*}To whom correspondence should be addressed. Tel. +1-701-231-7749, Fax +1-701-231-8333, E-mail: stefan.balaz@ndsu.edu

whether the covalently modified protein sites are selected more or less randomly or determined by the primary, noncovalent interaction.

Noncovalent protein binding affects the disposition kinetics of nonreactive chemicals and needs to be considered in modeling toxicokinetics in risk assessment procedures. Reversible binding of chemicals to proteins other than the receptors in cellular or extracellular spaces reduces the free concentration that elicits toxic and other effects. Simultaneously, the process creates depots releasing the chemical over an extended time period.

Historically, human serum albumin was the first object of interest among nontarget proteins, also known as silent receptors (6). The fraction unbound in plasma that is mainly affected by binding to albumin was one of the considered parameters in modeling the distribution volumes of drugs (7). The development of prediction approaches to albumin binding affinities of chemicals follows the general trend in the area of quantitative structure-activity relationships - QSARs. Qualitative identification of binding structural features (8) was developed into the dependencies, within the homologous series, on global physicochemical properties, such as lipophilicity (2-4) and ionization (5), followed by fragment-based approaches for more diverse chemicals (9-12), and culminating in linear models with numerous global (13) and 3-D variables (14). The most advanced predictions, based on the estimation of free binding energies from the atom-level simulations using the experimental albumin structures (15-23), are yet to appear.

The extracellular matrix (ECM), as the most abundant protein component in the extracellular space, is one of the next candidates for which the prediction of binding affinities would be useful for predictive toxicology. In a quantitative study of the distribution volumes of drugs, binding in the interstitial space was approximated using albumin binding (7). Laminin and collagen IV account for almost 90% of the dry ECM mass. We developed a method for the quantification of the binding of chemicals to an ECM surrogate, Matrigel, using a reconstituted solidified ECM layer incubated with the medium containing studied chemicals (24). During the time necessary for a proper hydration of matrix proteins, protein dissolution from the solidified layer could not be avoided. The binding constants to dissolved Matrigel proteins were determined as well. The simple dependencies of the association constants on lipophilicity hold only for a rather small homologous series. To obtain predictions of ECM binding for larger and more diverse sets, the receptor site modeling methods, also known as 3D-QSARs (25) or ligand-based QSARs may be useful. Using the binding affinities and structures of ligands, the methods infer averaged shapes and properties of the binding sites in the absence of structural information on the protein. All 3D-QSAR methods work for ligands, which interact with a common binding site. This condition seems more likely to be satisfied for binding to purified proteins than in a mixture of proteins like Matrigel. In both cases, but especially in the latter one, corroborating data would be useful. A strong indication that the studied compounds bind to one major site, which can be different in solidified and dissolved Matrigel, stems from the very satisfactory fits of the binding data by an equation valid for the 1:1 binding to one binding site (24). The existence of a 3D-QSAR model, as shown below, also supports this notion.

In this study, we used one of the popular 3D-QSAR methods, Comparative Molecular Field Analysis, CoMFA (26). The method was extended in our laboratory (27) to deal with multiple binding modes (orientations and conformations) of each ligand in the same binding site. In this study, we also added a rigorous consideration of multiple species of bound ligands, which may originate by ionization, tautomerism, or the formation of hydrates of aldehydes, ketones, and other groups. Other approaches (28,29) pursue similar goals. The multi-species, multi mode (MSMM) CoMFA procedure was necessary to process the ECM binding data of 25 simple

organic compounds and create putative binding site models, which can be used to predict binding affinities of untested chemicals.

Experimental Procedures

Chemicals

The Matrigel from one batch was purchased from BD Biosciences, Labware Discovery Bedford, MA;. All tested chemicals, Bradford reagent, borate buffer species, and all solvents (analytical or spectroscopy grade) were purchased from Sigma, St. Louis, MO.

ECM binding

The determination of binding constants of small molecules to ECM was described previously (24), therefore only a brief synopsis is provided here. Matrigel (500 μ L, 5.76 mg/mL) was carefully loaded to the bottoms of vials and allowed to solidify at 37 $^{\circ}$ C. The borate buffer (2 mL; pH 7.4) was incubated with Matrigel for four hours to establish the protein dissolution equilibrium. Different concentrations of compounds in DMSO (20 μ L) were added to the buffer and incubated with Matrigel for another 2 h to achieve the binding equilibrium. In the separated liquid, UV-Vis spectrophotometry (Shimadzu 1601) was used to measure the absorbances at several wavelengths. The association constants K to solidified Matrigel (K_S) and dissolved Matrigel (K_D -subscript P was used in the original article) were determined using the fitting of the dependence of the absorbances on the ligand concentration c_L according to eq 1.

$$A_L - A = [(\epsilon_L + \epsilon_{LR} \times K_D \times c_R) \times c_L] / (1 + H \times V_G/V_A + K_D \times c_R + K_S) \quad (1)$$

Here, A_L and A denote the absorbances with and without the ligand added; c_R refers to the total concentration of the receptor (ECM); and ϵ_L and ϵ_{LR} are the absorption coefficients of the chemical and ligand-receptor complexes, respectively, multiplied by the light path. V_G and V_A are the volumes of gas phase and aqueous phase, respectively, and H is the Henry's law constant. The magnitudes of ϵ_L , ϵ_{LR} , and H were determined in separate experiments. Equation 1 is valid for a 1:1 binding to one binding site in both dissolved and solidified Matrigel. The SD values of the fitted K_D parameters and most K_S parameters were in the range of 3-20% of the optimized values on the logarithmic scale. Larger SD values, which were still satisfactory for nonlinear fits, were observed for 4-bromophenol, 4-chloro-3-methylphenol, 2,2-diphenylethanol, 2-nitrotoluene, and 2,4,5-trichloroaniline (compounds **9**, **10**, **14**, **18**, and **25**, respectively, in Table 1 below).

Computational Procedures

Structures of Compounds

The standard and MSMM CoMFA procedures were applied to measured binding data, separately to solidified and dissolved ECM surrogates (s-ECM and d-ECM, respectively), of a set of 25 simple aromatic molecules, which are summarized in Table 1 below. All compounds are substituted benzenes with groups differing in size, electronegativity, and ability to ionize. In the following text, the term 'molecule' means a given conformation of a compound in the most prevalent species for the standard CoMFA analysis, and the given orientation/conformation (mode) of a species of a compound for the MSMM CoMFA. In the latter case, a compound will be represented by a number of molecules that is given as the sum of all modes for all species.

All molecules were sketched de novo in Sybyl modeling suite (version 7.1, Tripos, St. Louis, MO), and initially assigned Gasteiger and Hückel atomic charges (30,31). The geometries were

initially optimized using the Tripos force field by the Powell's method, available in the Sybyl's Maximin2 procedure, until the energy gradient was smaller than 0.001 kcal/(mol·Å).

Ligand Superposition

N-(2-benzylphenyl)-2,2-dimethylpropanamide (BDPA, compound 4 in Table 1 and Chart 1 below) in the optimum conformation was chosen as the template because of its size and a high association constant toward solidified Matrigel. All molecules in all species with Gasteiger and Hückel charges were superimposed to BPDA using the Flexible Superposition Option in the FlexS (32) module of Sybyl with the minimum overlap volume set to 0.6, to generate good starting conformations. The structures were briefly minimized with the standard Tripos force field and then by the density functional approach with the B3LYP hybrid functional and LAV3P** basis set in the Jaguar quantum chemistry suite by Schrodinger, Portland, OR. As an output, Mulliken charges were calculated by fitting the electrostatic potential to the atomic center for all atoms. Flexible superposition by FlexS with the minimum overlap volume set to 0.6 was repeated for all molecules, now with more realistic charges. Six conformations/modes were generated for each species; the top-scoring pose was used in the standard CoMFA procedure. Mulliken charges were recalculated for each molecule/species/conformation.

CoMFA Interaction Energy Calculations

The CoMFA studies (33) were performed with the QSAR module of Sybyl. The superimposed molecules were first placed in a rectangular box with a size of $20 \times 18 \times 22 \text{ Å}^3$, with the following coordinate ranges: from -11 to 9 Å (x), from -8 to 10 Å (y), and from -10 to 12 Å (z), respectively. The grid points were set 2 Å apart in each direction. For each ligand/species in each binding mode, two types of interaction energies, steric and electrostatic, were calculated in each grid point with an sp^3 -carbon probe with the a +1 charge and the distance-dependent dielectric constant. The maximum allowable steric and electrostatic energy values were set to 30 kcal/mol. The electrostatic energy term was not calculated in the grid points where the steric energy term reached the limit value (30 kcal/mol). The energies are organized in a table, where each row holds the data for one molecule. Each column represents the steric or electrostatic energy X_{ijkl} of all molecules in one grid point. The dependent variables, bioactivity or affinity, are only available for a compound. Therefore, in the table for the MSMM CoMFA procedure, only one row per compound, the row associated with the first mode of the first species, contains the dependent variable.

Coverage of the 3D-Space by Superimposed Ligands

This aspect of 3D-QSAR analysis is crucial for the soundness of the final model; yet, it is often neglected in published studies. Singularity is defined as the situation when the electrostatic or steric energy X_{ijkl} values in a particular grid point, which are normally collected in one column, are identical for all but one or two molecules. Typically, this situation arises when a subspace of the grid is occupied by only one or two molecules, whereas other molecules exhibit zero interaction energies in the pertinent grid points. The sparsely populated subspace cannot be appropriately characterized by the CoMFA analysis. The regression coefficients are associated with the energies X_{ijkl} in the grid points and a correct optimization requires that there are several different energy values available for each grid point. The molecules causing the singularities cannot be included in the correlation and need to be removed from the data set. After the removal, the columns initially containing singularities have zero variance, and are eliminated from the data set. The subspace must be marked as such and the CoMFA model cannot be used for the prediction of activities of the compounds occupying the singular space (27).

The singularities can be identified visually in the superposition of molecules or, more rigorously, by inspection of individual columns. To perform the latter task, the following procedure was coded in the C language and incorporated using the SPL scripting language into

the Sybyl software. The user specifies the value of SINGL (typically, SINGL = 1 or 2) that is equal to the number of energy values differing from the rest of values in the singular column. The range of interaction energies in each column is divided into the user-specified number of intervals (usually 10). The numbers of energies appearing in the intervals at the two ends of the range are counted. A singularity is detected in the following cases (n is the number of molecules as defined in the Structures of Compounds): (a) one interval contains SINGL energies and the other interval contains ($n - \text{SINGL}$) energies; the molecule(s) causing the singularity and the respective variable(s) are removed from the data set; (b) one interval contains more than one SINGL energy, but all of them come from one species of one compound and represent all modes for that species; the species and the column are eliminated from the set; (c) the same as b but valid for a compound instead of a species. Situations b and c reflect the fact that the association constant is available for the compounds, not for the species or modes. In situation c, a compound is excluded from the data set. In this case, additional information can be obtained about the singular space. After the calibration of the model without the singular compound in the data set, the activity of the singular compound can be predicted. If the prediction agrees with the experimental value, it means that the regression coefficients for the singular grid points are close to zero and, consequently, the singular subspace can be regarded as the solvent. However, a substantial difference between the predicted and experimental values means that the singular subspace is a part of the receptor that cannot be appropriately characterized by the set of compounds used.

Standard CoMFA Analysis

Ionizable compounds were analyzed as the species prevalent in the aqueous solution at pH 7.4, used in experiments. In the superposition, the top poses generated by the FlexS (32) module of Sybyl with the minimum overlap volume set to 0.6 were used. The partial least squares (PLS) analysis in the QSAR module of Sybyl was used to generate a model composed of latent variables representing orthogonal linear combinations of initial variables (33).

Multi-Species, Multi-Mode Binding Equilibria

All 3D-QSAR methods assume that the averaged binding site of the receptor (ECM in this case) is only slightly larger than the largest ligand in the studied series and accommodates only one ligand molecule at a time. The reversible ligand-receptor complexes usually form quickly via noncovalent interactions, and can be characterized by the association constant.

A ligand is present in the solution around the receptor in different conformations, which have similar energies close to the minimum energy under the given conditions. Upon contact with the binding site, the ligand molecules bind quickly and reversibly via noncovalent interactions. The probabilities of binding of each solution conformation are similar, because each conformation is quickly converted to a bound conformation that is, in general, different from the solution conformation. The differences in internal energies of free and bound molecules can be incorporated into the correlation equation. The probabilities of bound conformations (mode prevalencies) differ depending upon the free energy of binding. Some ligands will have only one binding mode with 100% prevalence, whereas other ligands may exhibit multiple modes, with prevalencies adding up to 100%.

Let us assume that the ligand is present in the solution in s species and all species bind in one binding mode. The experimentally observed association constant K_i for the binding of the i -th ligand (L_i) to the receptor (R) reflects the total concentration of the ligand/receptor complexes (LR_i) and does not distinguish between complexes differing in bound species:

$$K_i = \frac{[LR_i]}{[L_i] \times [R]} = \frac{\sum_{j=1}^s [LR_{ij}]}{[L_i] \times [R]} = \sum_{j=1}^s \frac{[L_{ij}]}{[L_i]} \times \frac{[LR_{ij}]}{[L_{ij}] \times [R]} = \sum_{j=1}^s f_{ij} \times K_{ij} \quad (2)$$

Here, LR_{ij} ($j=1,2,\dots, s$) is the 1:1 ligand/receptor complex, differing in the bound species. The square brackets denote the equilibrium concentrations. Thus, $[L_i]$ represents the total unbound concentration of all species of a ligand. In the fourth term, each summand was formally multiplied by $[L_{ij}]/[L_{ij}]$. The fraction of the j th species (f_{ij}) in the solution is given by the respective dissociation constant of the ligand or other equilibrium constant in the aqueous medium, and is independent of the receptor binding.

If each ligand species can bind in m binding modes, then the association constant of the j th species can be rewritten as follows:

$$K_{ij} = \frac{[LR_{ij}]}{[L_{ij}] \times [R]} = \frac{\sum_{k=1}^m [LR_{ijk}]}{[L_{ij}] \times [R]} = \sum_{k=1}^m K_{ijk} \quad (3)$$

Combining eqs 2 and 3, the observed association constant for multi-species, multi-mode binding can be expressed as follows:

$$K_i = \sum_{j=1}^s f_{ij} \times \sum_{k=1}^m K_{ijk} \quad (4)$$

Most ligand- and structure-based methods for the prediction of binding affinities describe the association constant K_i as a function of some sort of interaction energies with the putative binding site. Using eq 4, multi-species, multi-mode binding can be incorporated into any of these methods.

The simple eq 4 is in accordance with published analyses of formally analogous one-species situations: the statistical thermodynamic (34) and equilibrium (35,36) treatment of multi-mode binding in ligand/protein interactions and kinetic analyses of a reversible uni-molecular reaction leading to different products (37) or isomers (38). A similar treatment for one molecular species was used in the Mining Minima approach to structure-based prediction of binding affinities (39).

Multiple Species and Multiple Modes in CoMFA

In the CoMFA approach (33), the natural logarithm of the association constant for a ligand binding as j th species in k th binding mode ($\ln K_{ijk}$) is expressed as the weighed summation of the ligand/probe interaction energies X_{ijkl} . Placing this expression into eq. 4, the correlation equation for the MSMM CoMFA analysis is as follows:

$$K_i = \sum_{j=1}^s f_{ij} \times \sum_{k=1}^m e^{\sum_{l=1}^{f \times g} C_l \times X_{ijkl} + C_0} \quad (5)$$

The summation goes through up to $f \times g$ independent variables, where f is the number of fields used (steric, electrostatic, sometimes hydrophobic, polarizability, or hydrogen-bonding fields), and g is the number of the used grid points. The independent variables, X_{ijkl} are the energies of interaction between a probe placed in the l th grid point and i th ligand molecule in the j th species and k th binding mode. The regression coefficients C_l characterize the significance of the field contributions in each grid point for overall binding. The extension of eq 5 for other factors contributing to free binding energy (internal conformational energy, conformational entropy, and desolvation) is straightforward but was not applied in the present study, in order to minimize the number of optimized coefficients.

Forward-selection nonlinear regression analysis of logarithmized eq 5 optimizes the number and magnitudes of coefficients C . Because of the selection of significant variables, the final number of coefficients C is much smaller than the initial pool of the $f \times g$ values. Coefficients C define the putative partial association constants K_{ijk} for each binding mode of each species, which are equal to individual exponentials in eq 5. The K_{ijk} values are further used to calculate the prevalences, that is, Boltzmann probabilities, of individual species and modes as the ratios of the particular K_{ijk} divided by the weighted sum of all other K_{ijk} values as shown in the right-hand term of eq 4.

Forward-selection Nonlinear Regression Analysis

All available CoMFA columns were first sorted according to the decreasing standard deviations and used for the analyses in this order, starting at the top of the list. The analyses were performed in the spirit of the approach described in ref 27, with the following differences: (1) nonlinear regression analysis (NLR), based on the published algorithm (40), was integrated into the approach; (2) the last two phases in the original approach, called Shaping and Detailing, were fused into a single phase, called Shaping; and (3) individual fits were compared using $r^2_{\text{adj}} = 1 - (1 - r^2)(n - 1)/(n - k - 1)$, where k is the number of columns. The procedure was written in such a way that the NLR analysis automatically switches to the PLS analysis with linearized eq 5 (27), once the number of optimized coefficients exceeds 80% of the number of compounds. This switch was not activated in the analysis of the current data set.

In the first phase, called Outlining, NLR was carried out in an exhaustive-search mode, with the user-specified number of columns (STCOL = 3 to 7) having the highest standard deviations, and using the user-specified initial coefficient values. The best 5% fits were subjected to the one-by-one deletion of the columns with the lowest contribution, until r^2_{adj} did not improve anymore. A user-specified fraction of the worst models (usually 95%) was discarded.

The second phase, Shaping, started with an addition of a new group of columns. The number of columns in the group is user-specified, ADCOL = 3 to 6. If the fit improved, the columns were iteratively eliminated until r^2_{adj} stopped improving. The results were compared with the analysis before the addition of the ADCOL columns, and kept if the model improved. The process was repeated with the next group of ADCOL columns. This evolution of the model continues until r^2_{adj} stops improving.

For the best models of one run, characterized by the specific values of STCOL, ADCOL, and starting values of the coefficients, SD of the coefficients were calculated. The columns with the highest SD/coefficient ratios were iteratively dropped from the regression equation, if r^2 decreased by less than 5%.

Typically, several hundred runs were performed, many of them resulting in acceptable models. The best MSMM CoMFA model was selected on the basis of the statistical indices, number of included columns, and SD/coefficient ratios as the primary criteria, and the spatial distribution of the included grid points and used fields as secondary criteria. The whole process

took 3-5 h on a current SGI or Linux workstation, which is a much shorter time than in the original proof-of-the-concept study (27). The main reasons for the improved performance are the use of NLR instead of the exclusive use of linearized PLS analysis, and the coding of the routines in the C language instead of the SPL script language.

Results and Discussion

Binding of Chemicals to ECM

Monitoring the binding of chemicals to solidified ECM surrogate is complicated by the dissolution of ECM proteins into the buffer. After establishing the dissolution equilibrium, the association constants were measured to both solidified and dissolved ECM (K_S and K_D , respectively) as described before (24), and are summarized in Table 1. The pK_a values for ionizable compounds were taken from the Syracuse Research Corporation's PhysProp database (<http://esc.syrres.com/pointer/default.asp>), except for the values for compounds **23** and **24**, which were predicted (41) using the SPARC web server (<http://ibmlc2.chem.uga.edu/sparc/index.cfm>). Among the 25 simple aromatic compounds analyzed (Table 1; more complex structures shown in Chart 1), 11 compounds have the pK_a values within 3 logarithmic units of the pH 7.4 that was used in the experiment. For these compounds, both neutral and ionized species are present in the solution in the fractions $f > 0.1\%$ (Table 3), and both could participate in the binding. The measured association constants correlated neither with lipophilicity, parametrized by the 1-octanol/water partition coefficient P , nor with the pK_a values, as occasionally observed for nonspecific protein binding (1). The next choice to examine was the possibility of a structure-specific binding to an averaged binding site. To elucidate structural determinants of ECM binding, we used one of the popular receptor site modeling methods, CoMFA, first in the standard version and then in our modified version for multiple species and multiple binding modes (MSMM CoMFA).

Ligand Superposition and CoMFA Setup

In the absence of structural information on the binding site, the selection of the template for superposition was based on the following criteria: (a) high affinity toward solidified ECM surrogate because this form of ECM is of primary interest; (b) large size of the molecule; and (c) a comparatively low number of rotatable bonds. The criteria were best satisfied by one of the strongest binders to solidified ECM, *N*-(2-benzylphenyl)-2,2-dimethylpropanamide (BDPA), listed as compound **4** in Table 1 and shown in Chart 1.

The FlexS (32) module in Sybyl was used to generate the superposition. The method does not need predefined information on the pharmacophore atoms shared by the template and superimposed ligands, and explicitly takes into account molecular flexibility of the superimposed ligand. The algorithm maximizes the overlap of the groups with similar binding abilities and, simultaneously, the overlap volume of the reference and superimposed ligands.

The superimposed molecules were enclosed in a regular grid. The sp^3 -carbon probe with the +1 charge was placed into each intersection and the interaction energies between the probe and the superimposed ligands were calculated. The CoMFA analyses were performed according to described protocols.

To probe the predictive ability of the resulting CoMFA models, the data set was split into two subsets, the training set and test set. The compounds for the test set were selected in a way ensuring that (1) their binding affinities were evenly distributed within the range of binding affinities to solidified Matrigel and dissolved Matrigel, and (2) all substitution patterns and ionization states were included. The test set consisted of compounds **2**, **11**, **13**, and **17** (Table

1). These compounds were excluded from model calibration and were only used to test the predictivity.

Standard CoMFA Analysis

In the first attempt to create a 3D-QSAR model, the top poses for all compounds in their major species were used. Compounds **3** and **24** (Table 1 and Chart 1) were deleted because they caused singularities. The standard CoMFA models were built for the solidified ECM surrogate (s-ECM) and for the dissolved ECM surrogate (d-ECM). The best models correlated the binding affinities with 75 and 27 energies (the weights steric/electrostatic 0.500/0.500 and 0.250/0.750), arranged in 5 and 2 latent variables, respectively. The statistical indices are summarized in Table 2. The descriptive abilities of the CoMFA models were much better ($r^2 = 0.930$) for s-ECM than those for d-ECM ($r^2 = 0.630$). Unfortunately, the predictive abilities for both s-ECM and d-ECM were disappointing (Figure 2): the predictive correlation coefficients for the test set of four compounds were only $q^2 = 0.178$ and -0.455 , respectively. The weak predictivity was also indicated by the leave-one-out cross-validation in the training set (the data now shown in Table 2): the predictive squared correlation coefficients for the omitted compounds were $r^2_{\text{LOO}} = 0.105$ and 0.309 , respectively. The unsatisfactory performance of the standard CoMFA model prompted us to examine the fit of the data with the MSMM CoMFA model.

MSMM CoMFA Analysis

The FlexS superposition of all compounds is shown in Figure 1A. Six conformations, representing putative binding modes, were generated for each species of each compound except for compound **18** (Table 1), for which only five conformations were obtained.

The interaction energies between the probe (sp^3 -carbon with the charge +1) and the superimposed molecules were calculated and organized in a table. The columns containing singularities were rigorously eliminated, along with the modes causing the singularities. Complete compounds were excluded, if all modes in one species were declared as singular. Using these criteria, several modes of compounds **4**, **9**, **10**, **12**, and **23** (Table 1) were eliminated. The final superposition, used in the MSMM CoMFA analysis, is shown in Figure 1B. All singular molecules were eliminated, as follows from a comparison of Figures 1A and 1B. The aniline ring of sulfisomidine (**24**, Table 1 and Chart 1) is seen in several modes on the left side in Figure 1A. In the bottom part of Figure 1A, the flexible side chain of several modes of atenolol (**3**, Table 1 and Chart 1) can be identified. Although the spaces appear to be well covered, all aniline rings and all side chains represent several binding modes of the same compounds. Therefore, compounds **3** and **24** were classified as singular.

The coefficient optimization was performed by the forward-selection nonlinear regression analysis. Statistical indices for the final selected MSMM CoMFA models for s-ECM and d-ECM are summarized in Table 2, along with the results for the standard CoMFA models. The descriptive abilities significantly improved by the introduction of the MSMM binding: the values for $r^2 = 0.964$ and 0.849 for s-ECM and d-ECM, respectively. More importantly, the predictive abilities for the 4-member test set reached useful levels (Figure 2): $q^2 = 0.829$ and 0.806 , respectively. The use of scrambled affinities, aiming at the detection of possible chance correlations, did not lead to acceptable models ($r^2 = 0.167$ and 0.102 for s-ECM and d-ECM, respectively).

Prevalences of Binding Modes

The Boltzmann mode prevalences are calculated as $K_{ijk}/\sum K_{ijk}$, where K_{ijk} corresponds to individual exponentials in the summation in eq 5. All mode prevalences for the 2-species, 6-mode CoMFA analyses are summarized in Table 1S in the Supporting Information. Only

compounds **7** and **18** (Table 1) exclusively demonstrated one binding mode for s-ECM. The remaining ligands exhibit two or more modes, each having at least 5-% mode prevalence. For the binding to d-ECM, multiple modes were observed for all compounds. These facts illustrate why the standard, one-mode CoMFA procedure experienced problems with the current data sets and the MSMM approach performed better. The results of the one-mode CoMFA model using the most prevalent modes from the MSMM CoMFA analysis (the results not shown) were not acceptable either. Consequently, there seems to be a need for the use of multiple modes and species in the CoMFA modeling of studied data.

Binding of Individual Species

A significant portion (10 out of 25) of tested compounds ionize in the way that either neutral or ionized species dominate in the solution (Table 3). Compound **12**, 3,5-dichlorophenol, is an exception and has 14.2% of molecules ionized and 85.8% neutral at pH 7.4. The bound fraction of each species of each compound was obtained by summing up the contributions of all modes in that particular species (Table 1S in the Supporting Information). The species present in the solution in larger fractions contribute more to the binding for both s-ECM and d-ECM, except compounds **13**, **23**, and **24** (Table 1), which exhibit the opposite trend. Compound **13** exists almost exclusively in neutral form, which, however, represents less than a half of bound molecules. The neutral form of sulfacetamide (**23**, Table 1 and Chart 1), which is present in about 2.3% fraction in the solution, contributed 23.6% and 87.7% to the binding for s-ECM and d-ECM, respectively. The neutral form of sulfisomidine (**24**, Table 1 and Chart 1), a singular compound with acceptable predictions, has a small fraction, 0.4%, in the solution but accounts for 100 % of the binding to s-ECM and 3.9% for d-ECM.

Charged species seem to play important roles in the binding process. The studied set includes compounds with positively (**2**, **3**, and **13**) and negatively charged ions (**6**, **9**, **10**, **12**, **16**, **21**, **23**, and **24**; Table 1) originating from the ionization at pH 7.4. Cations contribute more to the binding than neutral species in all three compounds. This trend is most significant for compound **13**, containing a tertiary amine group: the cations represent 56.8% and 68.3% of the molecules bound to s-ECM and d-ECM, respectively, whereas their fraction in the solution is only about 0.5%. Aniline (**2**, Table 1) cations, representing about 0.2% of the molecules in the solution, account for 20% of the molecules bound to d-ECM and 3.9 % bound to s-ECM. Singular compound **3** was not satisfactorily predicted and its species distribution cannot be analyzed.

For anions binding to s-ECM, the contributions to the binding are only slightly higher than their fractions in the solution in three (compounds **6**, **9**, and **10**; Table 1) out of eight cases. Anions of compounds **12**, **23**, and **24** bind less than neutral species, especially to d-ECM.

CoMFA Maps of the Binding Site

Graphic representations of the best MSMM CoMFA models for s-ECM and d-ECM are given in Figure 3. The regions shown indicate the subspaces where variations of steric or electrostatic properties in the structures of compounds in the training set led to the most significant changes in binding affinities. The CoMFA contour plots show all eight steric columns and six electrostatic columns in the best model for s-ECM, and three steric columns and six electrostatic columns in the best model for d-ECM. The green polygons indicate the regions with positive coefficients associated with the van der Waals energies, where increased steric bulk is associated with enhanced activity. The yellow regions mark the opposite case. The regions where increased positive charge is favorable for the affinity are indicated by the blue polygons,

Supporting Information Available. Table 1S contains the bound fractions (prevalences) of all modes and species. This material is available free of charge via the Internet at <http://pubs.acs.org>.

whereas those where increased negative charge leads to increased affinity are shown as the red polygons.

Compounds **3** and **24** (Table 1 and Chart 1) were larger than the superposition of the other molecules and showed singularities in all modes. They were excluded from the training set and their affinities were later predicted from the calibrated models and provided further insight. The MSMM CoMFA predictions for atenolol (**3**, Table 1 and Chart 1) are much higher than the experimental affinities. Therefore, the model is deemed inconclusive for the respective singular region (shown in gray in Figure 3), as defined by this compound. The affinities of the compounds, which penetrate the singular region after the FlexS superposition, cannot be reliably predicted by the MSMM CoMFA model presented. However, sulfisomidine (**24**, Table 1 and Chart 1) affinities are predicted well, especially for s-ECM. This fact leads to the conclusion that the regression coefficients in the singular regions defined by compound **24** have the magnitudes close to zero and the respective regions can be considered to be water (shown in cyan in Figure 3). The predictions for the two singular compounds allowed a complete characterization of one of the two singular regions and increased the size of the subspace that is completely characterized by the MSMM CoMFA model.

The contour map for the s-ECM is more complex than the d-ECM map. Although the numbers of columns are different, some similarities can be recognized by comparing the contour plots in Figure 3, despite slightly different viewing angles in the s-ECM and d-ECM models, which were used to provide unobstructed views of the regions. The singular region (gray) and a major water region (cyan) are in both cases positioned close to the flexible *tert*-butyl substituent of *N*-(2-benzylphenyl)-2,2-dimethylpropanamide (compound **4**, BDPA, Table 1 and Chart 1), which is shown to illustrate spatial relationships. Similar positions in the vicinity of the fixed benzene ring can also be seen in both models for the attractive (green) and repulsive (yellow) steric regions and a water region (cyan). However, the electrostatic regions exhibit quite a different behavior. Three electrostatic regions are located close to the moving benzene ring. The positions of the blue region are similar, in contrast to the patterns for the red regions. In summary, the s-ECM and d-ECM models are different, despite similarities in some parts. This fact may be associated with the different composition of both protein mixtures (**24**).

Concluding Remarks

This study was initiated by the inability to correlate the binding constants of studied aromates to ECM with their physicochemical properties, such as lipophilicity and acidity. One of the frequently used 3D-QSAR techniques, CoMFA, capable of revealing the structural determinants of the binding behavior, was deployed. Because the standard, one-mode CoMFA provided unsatisfactory results, we used our modification of the method for multi-species, multi-mode binding. The standard CoMFA procedure utilizes one binding mode of one species for each ligand in the superposition. Frequently, the outliers are repositioned, to generate a better fit. The presented MSMM CoMFA procedure enables a systematic examination of several binding species and modes for all ligands, not just for the outliers. This outcome does not require an increase in the number of optimized regression coefficients, only a different correlation equation and, consequently, a different fitting procedure.

The resulting models exhibit excellent descriptive and predictive abilities and provide a structure-based interpretation of the binding activity of aromates to ECM. A singular region, where the MSMM CoMFA models are inconclusive, is small. Therefore, the models can be used for the prediction of binding affinities of untested chemicals, which do not exceed the size of the model. For the predictions, all species of the compounds need to be superimposed by the same procedure as was used for the training set, which does not require a common skeleton. The top six modes of each species are checked for singularity. The interaction energies X_{ijkl} are calculated for all nonsingular modes, which have all parts of the molecules placed in

the regions completely characterized by the optimized regression coefficients C . The predicted affinity is obtained according to eq 5, using the optimized C values, with individual exponentials representing the partial association constants K_{ijk} of individual modes and species. The predictions for solidified ECM should be useful in the areas of pharmacokinetics, toxicology, and risk assessment.

Acknowledgments

This work was supported in part by the NIH NCRR grants 1 PP20 RR 15566 and 1 P20 RR 16471.

References

- (1). Hansch C, Kiehs K, Lawrence GL. The role of substituents in the hydrophobic bonding of phenols by serum and mitochondrial proteins. *J. Am. Chem. Soc* 1965;87:5770–5773. [PubMed: 5845425]
- (2). Hersey A, Hyde RM, Livingstone DJ, Rahr E. A quantitative structure-activity relationship approach to the minimization of albumin binding. *J. Pharm. Sci* 1991;80:333–337. [PubMed: 1865333]
- (3). Colmenarejo G, Alvarez-Pedraglio A, Lavandera JL. Cheminformatic models to predict binding affinities to human serum albumin. *J. Med. Chem* 2001;44:4370–4378. [PubMed: 11728183]
- (4). Valko K, Nunhuck S, Bevan C, Abraham MH, Reynolds DP. Fast gradient HPLC method to determine compounds binding to human serum albumin. Relationships with octanol/water and immobilized artificial membrane lipophilicity. *J. Pharm. Sci* 2003;92:2236–2248. [PubMed: 14603509]
- (5). Ermondi G, Lorenti M, Caron G. Contribution of ionization and lipophilicity to drug binding to albumin: A preliminary step toward biodistribution prediction. *J. Med. Chem* 2004;47:3949–3961. [PubMed: 15267234]
- (6). Davis BD. Binding of sulfonamides by plasma proteins. *Science* 1942;95:78. [PubMed: 17791075]
- (7). Poulin P, Theil FP. Prediction of pharmacokinetics prior to in vivo studies. 1. Mechanism-based prediction of volume of distribution. *J. Pharm. Sci* 2002;91:129–156. [PubMed: 11782904]
- (8). Wanwimolruk S, Birkett DJ, Brooks PM. Structural requirements for drug binding to site II on human serum albumin. *Mol. Pharmacol* 1983;24:458–463. [PubMed: 6195516]
- (9). Saiakhov RD, Stefan LR, Klopman G. Multiple computer-automated structure evaluation model of the plasma protein binding affinity of diverse drugs. *Perspect. Drug Discovery Des* 2000;19:133–155.
- (10). Kratochwil NA, Huber W, Muller F, Kansy M, Gerber PR. Predicting plasma protein binding of drugs: A new approach. *Biochem. Pharmacol* 2002;64:1355–1374. [PubMed: 12392818]
- (11). Hajduk PJ, Mendoza R, Petros AM, Huth JR, Bures M, Fesik SW, Martin YC. Ligand binding to domain-3 of human serum albumin: A chemometric analysis. *J. Comput.-Aided Mol. Des* 2003;17:93–102. [PubMed: 13677478]
- (12). Kratochwil NA, Huber W, Muller F, Kansy M, Gerber PR. Predicting plasma protein binding of drugs - revisited. *Curr. Opin. Drug Discovery Dev* 2004;7:507–512.
- (13). Colmenarejo G. In silico prediction of drug-binding strengths to human serum albumin. *Med. Res. Rev* 2003;23:275–301. [PubMed: 12647311]
- (14). Aureli L, Cruciani G, Cesta MC, Anacardio R, De Simone L, Moriconi A. Predicting human serum albumin affinity of interleukin-8 (cxcl8) inhibitors by 3D-QSAR approach. *J. Med. Chem* 2005;48:2469–2479. [PubMed: 15801837]
- (15). Carter DC, He XM. Structure of human serum albumin. *Science* 1990;249:302–303. [PubMed: 2374930]
- (16). Carter DC, He XM, Munson SH, Twigg PD, Gernert KM, Broom MB, Miller TY. Three-dimensional structure of human serum albumin. *Science* 1989;244:1195–1198. [PubMed: 2727704]
- (17). He XM, Carter DC. Atomic structure and chemistry of human serum albumin. *Nature* 1992;358:209–215. [PubMed: 1630489]
- (18). Curry S, Mandelkow H, Brick P, Franks N. Crystal structure of human serum albumin complexed with fatty acid reveals an asymmetric distribution of binding sites. *Nat. Struct. Biol* 1998;5:827–835. [PubMed: 9731778]

- (19). Sugio S, Kashima A, Mochizuki S, Noda M, Kobayashi K. Crystal structure of human serum albumin at 2.5 Å resolution. *Protein Eng* 1999;12:439–446. [PubMed: 10388840]
- (20). Bhattacharya AA, Curry S, Franks NP. Binding of the general anesthetics propofol and halothane to human serum albumin: High resolution crystal structures. *J. Biol. Chem* 2000;275:38731–38738. [PubMed: 10940303]
- (21). Bhattacharya AA, Grune T, Curry S. Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin. *J. Mol. Biol* 2000;303:721–732. [PubMed: 11061971]
- (22). Petitpas I, Bhattacharya AA, Twine S, East M, Curry S. Crystal structure analysis of warfarin binding to human serum albumin: anatomy of drug site I. *J. Biol. Chem* 2001;276:22804–22809. [PubMed: 11285262]
- (23). Petitpas I, Gruene T, Bhattacharya AA, Curry S. Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids. *J. Mol. Biol* 2001;314:955–960. [PubMed: 11743713]
- (24). Zhang Y, Lukacova V, Reindl K, Balaz S. Quantitative characterization of binding of small molecules to extracellular matrix. *J. Biochem. Biophys. Methods* 2006;67:107–122. [PubMed: 16516301]
- (25). Oprea, TI. 3D QSAR modeling in drug design. In: Bultinck, P., editor. *Computational Medicinal Chemistry for Drug Discovery*. Marcel Dekker; New York: 2004. p. 571–616.
- (26). Cramer RD III, Patterson DE, Bunce JD. Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc* 1988;110:5959–5967.
- (27). Lukacova V, Balaz S. Multimode ligand binding in receptor site modeling: Implementation in CoMFA. *J. Chem. Inf. Comput. Sci* 2003;43:2093–2105. [PubMed: 14632461]
- (28). Vedani A, Briem K, Dobler M, Dollinger H, McMasters DR. Multiple-conformation and protonation-state representation in 4D-QSAR: The neurokinin-1 receptor system. *J. Med. Chem* 2000;43:4416–4427. [PubMed: 11087566]
- (29). Doweyko, AM. Multiple conformer protocol: A new method for the identification of preferred ligand binding motifs using cross-validated 3D-QSAR models. In: Hoeltje, HD.; Sippl, W., editors. *Rational Approaches to Drug Design*. Prous Science; Barcelona, Spain: 2001. p. 307–315.
- (30). Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity: a rapid access to atomic charges. *Tetrahedron* 1980;36:3219–3228.
- (31). Streitwieser, A, Jr.. *Molecular Orbital Theory for Organic Chemists*. Wiley; New York: 1961.
- (32). Lemmen C, Lengauer T, Klebe G. FLEXS: A method for fast flexible ligand superposition. *J. Med. Chem* 1998;41:4502–4520. [PubMed: 9804690]
- (33). Cramer, RD., III; Wold, S. Comparative molecular field analysis (CoMFA). 1994. U.S. Patent 5,307,287
- (34). Wang J, Szewczuk Z, Yue SY, Tsuda Y, Konishi Y, Purisima EO. Calculation of relative binding free energies and configurational entropies: A structural and thermodynamic analysis of the nature of non-polar binding of thrombin inhibitors based on hirudin55–65. *J. Mol. Biol* 1995;253:473–492. [PubMed: 7473728]
- (35). Balaz S, Hornak V, Haluska L. Receptor mapping with multiple binding modes: Binding site of PCB- degrading dioxygenase. *Chemom. Intell. Lab. Syst* 1994;24:185–191.
- (36). Hornak V, Balaz S, Schaper KJ, Seydel JK. Multiple binding modes in 3D-QSAR: Microbial degradation of polychlorinated biphenyls. *Quant. Struct.-Act. Relat* 1998;17:427–436.
- (37). Jullien L, Proust A, LeMenn JC. How does the Gibbs free energy evolve in a system undergoing coupled competitive reactions? *J. Chem. Educ* 1998;75:194–199.
- (38). Smith, WR.; Missen, RW. *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*. John Wiley and Sons; New York: 1982. p. 1–364.
- (39). Head MS, Given JA, Gilson MK. Mining Minima: Direct computation of conformational free energy. *J. Phys. Chem. A* 1997;101:1609–1618.
- (40). Fletcher R, Powell MJD. A rapidly convergent descent method for minimization. *Comput. J* 1963;6:163–168.

- (41). Hilal SH, Karickhoff SW, Carreira LA. A rigorous test for SPARC's chemical reactivity models: Estimation of more than 4300 ionization pKas. *Quant. Struct.-Act. Relat* 1995;14:348–355.

Figure 1.

FlexS superposition of the compounds studied on the BDPA (compound **4**, Table 1 and Chart 1) as a template: (A) all compounds and (B) - nonsingular compounds. All five or six modes generated by FlexS are included. Note the space in the left bottom part that was occupied by singular compounds **3** and **24** in part A and is empty in part B. Atoms are color-coded: carbon, gray; oxygen, red; nitrogen, blue; sulfur, yellow; and chlorine, green. Hydrogens are suppressed.

Figure 2.

Calculated/predicted vs experimental affinities for the CoMFA models for s-ECM (left) and d-ECM (right) binding, respectively. Predicted binding affinities are for compounds in the test set (compounds **2**, **11**, **13**, and **17** in Table 1) using the model developed for the training data set. Calculated and predicted affinities by the MSMM CoMFA models are shown as full (●) and open (○) circles, respectively, and those by one-mode CoMFA model are shown as compound numbers (Table 1). The identity line is plotted as a visual aid.

Figure 3.

Contour plots of steric and electrostatic fields for the best MSMM CoMFA models for solidified and dissolved ECM proteins (left and right, respectively). Compound **4** (Table 1 and Chart 1) is shown to illustrate spatial relations. Green and yellow contours indicate the regions where an increase and decrease, respectively, in steric bulk enhance binding affinities. Blue and red electrostatic contours correspond to regions where an increase in positive or negative charge, respectively, enhances affinity. The singular region, for which the model is inconclusive, is shown in gray. The regions representing water are shown in cyan.

3 **4** **11** **22** **23** **24**

Chart 1.
Structures of More Complex Tested Chemicals (other Data in Table 1).

Table 1
Measured and Calculated/Predicted Affinities of Tested Compounds to Solidified and Dissolved ECM Surrogate

No.	Compound	$\log K_s^a$			$\log K_o^a$		
		Exp.	Calculated/Predicted b by CoMFA		Exp.	Calculated/Predicted b by CoMFA	
			Standard	MSMM		Standard	MSMM
1	acetophenone	-0.30	-0.28	-0.43	2.68	2.61	2.68
2 ^c	aniline	-0.56	-0.39	-0.50	1.75	3.05	2.20
3 ^d	atenolol	-0.54	-0.82	33.79	3.20	3.53	10.79
4	BDPA ^e	0.03	-0.04	0.04	3.54	2.61	3.30
5	benzamide	-0.61	-0.69	-0.50	1.77	2.15	1.61
6	benzenesulfonamide	-0.92	-0.92	-0.96	2.34	2.13	2.25
7	benzyl alcohol	-0.82	-0.76	-0.89	3.44	3.19	2.94
8	BPCA ^f	-0.61	-0.63	-0.59	2.56	2.64	2.23
9	4-bromophenol	-0.16	-0.22	-0.16	3.40	2.97	3.39
10	4-chloro-3-methylphenol	-0.20	-0.29	-0.26	3.22	3.29	3.22
11 ^c	cinnamide	-0.12	-0.55	-0.22	3.64	2.58	3.50
12	3,5-dichlorophenol	-0.53	-0.54	-0.48	3.04	2.99	2.89
13 ^c	N,N-dimethylaniline	-0.26	-0.46	-0.34	2.68	2.65	2.90
14	2,2-diphenylethanol	-0.02	-0.04	0.00	3.41	3.25	3.41
15	3-phenylpropanal	0.05	0.07	0.01	2.61	2.51	2.81
16	2-hydroxybiphenyl	-0.60	-0.60	-0.63	2.36	2.62	2.79
17 ^c	4-nitroaniline	-0.95	-0.66	-0.73	2.10	2.52	2.46
18	2-nitrotoluene	-0.09	-0.06	-0.09	2.90	2.72	2.58
19	3-nitrotoluene	-0.68	-0.42	-0.63	2.84	2.45	2.87
20	4-nitrotoluene	-0.67	-0.70	-0.67	2.11	2.28	2.54
21	phenol	-0.48	-0.41	-0.39	2.45	2.90	2.69
22	phenylurea	-0.50	-0.50	-0.50	2.42	2.36	2.48
23	sulfacetamide	-0.64	-0.70	-0.62	1.03	1.81	1.37

No.	Compound	$\log K_S^a$			$\log K_D^a$		
		Exp.	Calculated/Predicted b by CoMFA		Exp.	Calculated/Predicted b by CoMFA	
			Standard	MSMM		Standard	MSMM
24 ^d	sulfisomidine	-0.61	-0.79	-0.42	2.18	2.50	3.15
25	2,4,5-trichloroaniline	-0.28	-0.31	-0.29	3.45	4.08	3.54

^aThe K_D values are given in L/mol, and K_S values have no units.

^bCalculated values for the training set, predicted values for the test set.

^cA test set member.

^dA singular compound.

^e*N*-(2-Benzylphenyl)-2,2-dimethylpropanamide.

^fBiphenyl-2-carbaldehyde.

Table 2

Statistical Indices for the Best CoMFA Models

	Solidified ECM		Dissolved ECM	
	standard	MSMM	standard	MSMM
r^2	0.930	0.964	0.630	0.849
SD	0.078	0.057	0.325	0.229
q^2	0.178	0.829	-0.455	0.806

Table 3

Fractions of Species in Solution and in the Bound State

no.	pK _a	fraction (%)					
		free (f)		bound to			
		neutral	ionized	solidified ECM		dissolved ECM	
				neutral	ionized	neutral	ionized
2	4.6	99.8	0.2	96.1	3.9	79.8	21.2
3	9.2	1.7	98.3	0.0	100.0	0.0	100.0
6	10.1	99.8	0.2	98.0	2.0	100	0.0
9	9.2	98.3	1.7	97.1	2.9	100	0.0
10	9.6	99.3	0.7	94.3	5.7	100	0.0
12	8.2	85.8	14.2	95.8	4.2	98.5	1.5
13	5.1	99.5	0.5	43.2	56.8	31.7	68.3
16	10.0	99.7	0.3	99.8	0.2	100	0.0
21	10.0	99.7	0.3	99.8	0.2	100	0.0
23	5.8	2.3	97.7	23.6	76.4	87.7	12.3
24	5.0	0.4	99.6	100	0.0	3.9	96.1