

Rigorous Treatment of Multispecies Multimode Ligand–Receptor Interactions in 3D-QSAR: CoMFA Analysis of Thyroxine Analogs Binding to Transthyretin

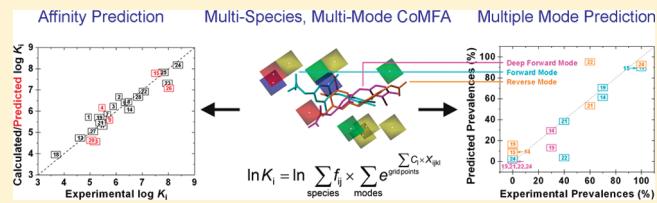
Senthil Natesan,^{†,‡} Tiansheng Wang,[‡] Viera Lukacova,[‡] Vladimir Bartus,[‡] Akash Khandelwal,[‡] and Stefan Balaz^{*,†}

[†]Albany College of Pharmacy and Health Sciences, Vermont Campus, Colchester, Vermont 05446, United States

[‡]North Dakota State University, College of Pharmacy, Nursing and Allied Sciences, Fargo, ND 58108, United States

Supporting Information

ABSTRACT: For a rigorous analysis of the receptor–ligand binding, speciation of the ligands caused by ionization, tautomerism, covalent hydration, and dynamic stereoisomerism needs to be considered. Each species may bind in several orientations or conformations (modes), especially for flexible ligands and receptors. A thermodynamic description of the multispecies (MS), multimode (MM) binding events shows that the overall association constant is equal to the weighted sum of the sums of microscopic association constants of individual modes for each species, with the weights given by the unbound fractions of individual species. This expression is a prerequisite for a precise quantitative characterization of the ligand–receptor interactions in both structure-based and ligand-based structure–activity analyses. We have implemented the MS-MM correlation expression into the comparative molecular field analysis (CoMFA), which deduces a map of the binding site from structures and binding affinities of a ligand set, in the absence of experimental structural information on the receptor. The MS-MM CoMFA approach was applied to published data for binding to transthyretin of 28 thyroxine analogs, each forming up to four ionization species under physiological conditions. The published X-ray structures of several analogs, exhibiting multiple binding modes, served as templates for the MS-MM superposition of thyroxine analogs. Additional modes were generated for compounds with flexible alkyl substituents, to identify bound conformations. The results demonstrate that the MS-MM modification improved predictive abilities of the CoMFA models, even for the standard procedure with MS-MM selected species and modes. The predicted prevalences of individual modes and the generated receptor site model are in reasonable agreement with the available X-ray data. The calibrated model can help in the design of inhibitors of transthyretin amyloid fibril formation.



INTRODUCTION

The majority of physiological compounds, drugs, and other man-made bioactive chemicals contain ionizable substructures. While electrostatic interactions of the ionized groups play important roles in receptor binding, their combination with the interactions of other ligand parts with the receptor may lead to the situations when several ligand species bind to the same binding site, with the bound species distribution different from that in solution, and varying among the ligands with similar skeletons. This scenario also applies to the species originating in aqueous solutions because of tautomerism and covalent hydration of aldehydes, ketones, and other groups,¹ as well as to stereoisomers, which can sometimes interconvert under physiologic conditions within the time intervals that are relevant for their effects.²

Whereas a unique binding mode is often considered a hallmark of ligand specificity, a growing body of experimental evidence has documented multiple binding modes, frequently for very similar ligands or, noticeably, even for one ligand.³ This

phenomenon is particularly pronounced for ligands having flexible substructures and interacting with flexible receptors, which are very common thanks to conformational freedom of the side chains of the receptor amino acids and larger-scale macromolecular motions. The multitude of interactions involved in each binding event, atom-level details, complex environment, and other factors make experimental elucidation of the nature of bound species and the conformation and orientation (mode) of each species a formidable task.^{4–6}

A quantitative description of the binding affinities of a set of ionizable (or otherwise speciated) and flexible ligands to the receptor in both structure-based and ligand-based quantitative structure–activity relationships (QSAR) requires a proper treatment of the multispecies (MS), multimode (MM) situations. The need for a generalized treatment of the MS-MM scenarios, as presented here, has been widely recognized.^{7,8}

Received: February 6, 2011

Published: April 08, 2011

Ligand-based receptor site models, also known as 3D-QSARs or pseudoreceptors, continue to serve as important tools in drug design. In this area, multiple binding modes were first considered in the qualitative active analog approach⁹ of Marshall, and quantitatively evaluated in the distance geometry-based and similar approaches by Crippen and his co-workers. Quadratic programming and other sophisticated optimization techniques were used to generate a close-fit receptor site model that simultaneously incorporated the best binding mode for each ligand.^{10–14} The minimum topological distance approach, although limited by a hypermolecule concept, also has a potential to consider multiple bound conformations.^{15–17}

The most frequently used 3D-QSAR methods, comparative molecular field analysis (CoMFA)¹⁸ and related approaches,¹⁹ characterize the hypothetical binding site by the regression coefficients that are assigned to the probe/ligand interaction energies (electrostatic, steric, sometimes hydrophobic fields) in individual intersection points of a grid encompassing the superimposed ligands. A treatment of multimode ligand binding was attempted by using the interaction energies that were obtained as a weighted average of the energies for individual modes, with either identical^{20,31} or different weights²² corresponding to the expected mode prevalences. The form of the standard CoMFA regression equation,¹⁸ however, differs from that obtained by a rigorous treatment of the multimode binding event (see eq 6 below and the accompanying text). Moreover, the mode prevalences depend on the optimized regression coefficients in a nonlinear way (eq 6), and, therefore, cannot be assigned a priori.²³

The distance geometry and CoMFA models, along with the majority of the receptor site models developed later, estimate the binding affinity as the sum of the pairwise interactions of individual ligand parts with some representation of the putative receptor site. Hopfinger, Dunn, and co-workers used different paradigms. In the molecular shape analysis (MSA) studies, they assumed that the binding affinities are proportional to the overlap volumes of the ligands with the chosen template in an assumed active conformation. To describe the multimode situation, the binding affinity was represented by a linear combination of the MSA, field, and other ligand descriptors, which were arranged in tensors resulting from different conformations and superpositions.^{24–26} In the more recent 4D-QSAR analysis, the binding affinities were assumed to depend on the grid cell occupancies, which were created by molecular dynamics simulations of the superimposed ligands.^{27–30} The studies demonstrated the significance of considering conformational sampling in the quality of 3D-QSAR models.^{31,32}

The majority of the aforementioned approaches considers several modes but, in the end, selects one mode for each ligand, which may not correspond to the real situation in many cases. With the advent of more efficient methods of structural analysis, it is becoming apparent that multiple binding modes occur more frequently than expected.^{33,34} To overcome the possible limitation of one binding mode in QSAR analysis, we have developed multimode models characterizing the occurrence of individual binding modes using the Boltzmann prevalences, which are not known a priori but are objectively quantified as a part of the overall model calibration.^{35,36} For some ligands, the resulting spectrum of prevalences may point to one dominant mode; for the remaining ligands, however, several modes can be suggested by the model. In addition to increased realism, this description allows a straightforward calibration of the model. The pertinent regression equation is based on the rigorous

relationship describing the observed association constant for the multimode binding as the sum of the microscopic association constants of individual modes. Nonlinear regression analysis needs to be used because the equation is nonlinear in optimized coefficients. The approach was later implemented²³ in CoMFA.¹⁸

Other approaches aspiring to treat the MS-MM situations were developed by Vedani and co-workers. The QUASAR and Raptor approaches sculpt the receptor site model out of a large set of atom-sized particles located in an envelope surrounding superimposed ligands.³⁷ A genetic algorithm is used to find a proper distribution of the particle types in the envelope, which provides the sum of interaction energies correlating in a linear way with the experimental binding free energies. The use of a single theory-to-experiment proportionality coefficient to scale the crude ligand-particle interaction energies and other contributions to the total free energy of binding is unusual. For instance, the methods for prediction of binding affinities utilizing known receptor structures, both ensemble-based (linear response method,³⁸ as well as its extended³⁹ and QM/MM⁴⁰ versions) and single-structure-based (VALIDATE,⁴¹ the free energy force field approach,^{42,43} COMBINE analysis,⁴⁴ and a single-structure linear response technique⁴⁵), routinely use more detailed scaling. In the QUASAR approach, the reduced ability to weigh the energy contributions may be compensated by huge statistical flexibility of the model resulting from the use of several probe types in hundreds of positions on the envelope. The number of degrees of freedom is further increased by the adaptation of the envelope to individual ligand shapes. Although physically corresponding to the induced fit, this step further increases the already high number of degrees of freedom in the model. Considering that even averaged receptor site models, as generated by all other methods, are usually heavily underdetermined, the induced fit that is specific for individual ligands makes a rigorous calibration of the models an extremely tedious task. The final result is represented by hundreds of often dissimilar models, which provide predictions of affinities but are less suitable for the mapping of the binding site.

The multimode situation is handled by expressing the total binding energy ΔG as a weighted sum of the contributions of individual modes, with the weights corresponding to the prevalences of the modes. The authors used the term ‘fractions’ instead of ‘prevalences’. We reserve the term ‘fractions’ for the species in the solution around the receptor, and use the term ‘prevalence’ to denote the percentage of the bound ligand that is found in the given binding mode. To avoid confusion, we will use our terminology throughout. Regrettably, the QUASAR methodology seems to be flawed in two aspects: the form of the ΔG expression and the calculation of the mode prevalences.

The overall binding free energy ΔG cannot be calculated as a weighted sum of the mode contributions. The correct relationship is shown in eq 6 below (speciation was not considered at this point, so the j -summation has only one term with the fraction $f=1$), with each $\ln K$ replaced by the corresponding $\Delta G/RT$ (K is the observed or microscopic association constant with the subscript i or ij , respectively, R is the gas constant, and T is absolute temperature). In addition to the form, eq 6 differs from the QUASAR correlation equation in the variables: the mode prevalences are superfluous in the latter equation.

Although called Boltzmann fractions, the QUASAR mode prevalences are actually calculated using a function (eq 2 in ref 37) that differs from the Boltzmann formula. This function was probably chosen to provide for the cancellation of the

theory-to-experiment proportionality factor in the expression so that the mode prevalences become independent of optimization and can be calculated before optimization. As discussed above, this is not a feasible endeavor because the prevalences are determined by the final, optimized receptor-site model. Also, unfortunately enough, the prevalences defined in this way do not even add up to 100% for all modes of a ligand.

The QUASAR approach was later applied to a multiple-species situation, without any apparent change in the correlation equation (described above). A comparison with the rigorous eq 6 shows that the QUASAR correlation equation lacks the fractions, f_{ij} , of individual species in the receptor surroundings. The absence of the fractions of individual species in the correlation equation is puzzling because these quantities clearly affect the overall binding affinity of the ligand. To illustrate the issue, let us consider two species (S1 and S2) of the same ligand, with S1 binding 100 times stronger than S2. The same amount of each species will interact with the receptor, if S2 is present in 100-times higher concentration, provided that the binding isotherms are linear (i.e., far from saturation). In a more recent approach called Raptor, the authors neglect electrostatics altogether, using only hydrophobic and H-bond interactions, adding another step that further compromises the analysis of multispecies problems. In Raptor, the induced-fit adjustments were made even more flexible than in QUASAR. The form of the regression equation containing several hundred parameters is unclear: whereas the contributions to the binding free energy of each species/mode were individually scaled initially,^{46,47} the scaling is not described in more recent papers.^{48,49}

In the present study, we examined the importance of the fractions of individual species in receptor binding and, consequently, extended our multimode approach by a rigorous inclusion of multiple species. The resulting correlation equation (eq 6 below) is suitable for both ligand-based and receptor-based 3D-QSAR approaches. In this study we validated the MS-MM method using the computationally cheaper ligand-based approach, although for the selected case, the binding of thyroxine analogs to transthyretin (TTR), information on the structures of the receptor and ligand–receptor complexes is available. The structural information was used to assess the quality of the ligand-based model. An advanced MS-MM receptor-based study^{50,51} would have required a much longer time frame and more computational resources. The published binding data of thyroxine analogs to TTR were chosen for the validation of the MS-MM approach because (1) the analogs are present in the solution as multiple species under the used experimental conditions; (2) several X-ray structures of the complexes, some exhibiting multiple binding modes, are available for a comparison with model predictions; (3) the experimentally quantified prevalences of individual modes in multimode complexes can serve as a rigorous test of model predictions; (4) published affinities of 28 analogs span almost 4 orders of magnitude, and (5) the case is of medicinal significance. The TTR ligands inhibit the first step in amyloid fibril formation, the dissociation of TTR tetramers.^{52,53} Elucidation of structural determinants of TTR binding helps in the development of potential treatments for amyloid diseases.^{54,55}

METHODS

Data Set. Published data on binding of thyroid hormone analogs to transthyretin, as determined at 37 °C using a dialysis method,⁵⁶ are summarized in Table 1. Diiodothyronine (T2) was

added to the original data set as compound 21, because experimentally determined prevalences of individual modes were available for this compound and this attribute was important for a more detailed comparison of the model with experimental data.⁵⁷ The binding affinity of 21 was determined at 23 °C by fluorescence quenching.⁵⁸ Fluorescence quenching and dialysis method provide similar results, as documented for thyroxine and triiodothyronine.⁵⁸ The binding affinity of 21 at 37 °C ($\log K = 5.4$) was recalculated using the van't Hoff equation using the published ΔH° (−13.4 kcal/mol) and ΔS° (−18.5 cal mol^{−1} deg^{−1}) values.

The training set consisted of 22 compounds and the test set is composed of 6 compounds: 4, 9, 13, 15, 20 and 26 (Table 1). The test set was selected to cover a substantial part of the activity range. Compound 10 was excluded from the training set because of singularity caused by the large benzyl substituent, which occupied the grid intersections unreachable by other studied compounds.

Fractions of Individual Species. The binding affinities of the studied 28 ligands⁵⁶ were determined at pH 8.0. Under these conditions, the studied compounds will be present in the aqueous solution as several species, which are listed in Table 2. The carboxyl is regarded as ionized in all species. In the data set, 22 compounds with three ionization sites (OH, NH₂, COOH) are present as four species; compound 27 (Table 1) with ionization sites (OH, NH₂) forms similar four species, just without the carboxyl; and five compounds (23, 24, 25, 26, and 28 in Table 1) with two ionization sites (OH, COOH) have two species with ionized carboxyl and either OH or O[−]. The microconstants need to be used for calculation of the fractions. The SPARC online calculator⁵⁹ was used to calculate, at pH 8.0 and 37 °C, the respective fractions, which are shown in Table 1.

Superposition of Ligands. The template structures for multiple modes were selected from the PDB files for complexes of human transthyretin, 2ROX,⁶⁰ 1Z7J,⁶¹ and 1THA.⁶² Two other TTR PDB files were not used: 1F86 contains the complex without a typical thyroxine analog ligand and 1ICT has a low resolution. Three dimer crystal structures for the TTR complexes (2ROX, 1Z7J, 1THA) were superimposed by homology using Sybyl,⁶³ to generate the coordinates for the superposition of the three basic orientations (F, DF, and R), which are shown in Figure 6a and 6b and in Table 4. The reverse mode template from the PDB file 1Z7J is T4-AA without amino side chain. To obtain a full-size template for a straightforward superposition of tested ligands, an L-T4 molecule with the same orientation and conformation (Φ_1 , Φ_2 and Φ_3) as T4-AA was docked into the binding site using the Atom Fit function of Sybyl, and its side chain (Φ_4 and Φ_5), and the inner pocket were optimized. In this way, the side chain conformation with $\Phi_4 = 165.1^\circ$ and $\Phi_5 = 346.2^\circ$ was obtained.

In the MS-MM analyses, each compound was considered in four or two species (Table 1), and each species in twelve modes. All molecules with specified dihedrals were built in Sybyl⁶³ and optimized in Jaguar⁶⁴ using DFT with B3LYP functional and LAV3P** basis set. Atomic Mulliken charges for individual conformations were calculated by fitting electrostatic potential to atom centers.

For the single-species, multimode CoMFA analysis, several variants were used, in twelve modes: the single species with the highest fraction calculated from SPARC for each ligand, as well as all ligands as one species, whereby all four species were examined.

For standard CoMFA analyses, three setups were selected: (1) the most preferred modes and species from the MS-MM CoMFA analysis; (2) the prevalent species in solution and the forward (F)

Table 1. Studied Thyroxine Analogs: Stereochemistry of the α -Carbon C8, Structures (Figure 1), Binding Affinities to Transthyretin (K in L/mol), and the Fractions of Individual Species (Defined in Table 2) in the Medium Used in the Binding Experiments (pH 8.0; 37 °C), Estimated by SPARC^a

ligand no.	D/L	R ₁	R ₃	R ₅	R _{3'}	R _{5'}	log K	fractions of individual species			
								1	2	3	4
1	D	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	I	H	4.94	0.208	0.603	0.050	0.138
2	D	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	I	I	6.11	0.010	0.799	0.003	0.187
3	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	i-Bu	H	5.89	0.801	0.011	0.185	0.002
4	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	i-Pr	H	5.46	0.802	0.011	0.185	0.002
5	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	n-Pr	H	5.38	0.798	0.015	0.184	0.003
6	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	s-Bu	H	6.32	0.807	0.005	0.186	0.001
7	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	t-Bu	H	5.66	0.807	0.005	0.187	0.001
8	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	Br	Br	6.48	0.011	0.800	0.003	0.186
9	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	Br	H	5.75	0.213	0.599	0.051	0.136
10	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	Benzyl	H	5.59	0.792	0.018	0.186	0.004
11	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	Cl	H	5.42	0.202	0.610	0.049	0.139
12	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	Me	Me	4.62	0.803	0.014	0.180	0.003
13	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	Me	H	5.19	0.783	0.030	0.180	0.007
14	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	I	H	6.51	0.208	0.603	0.050	0.138
15	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	I	I	7.54	0.010	0.799	0.003	0.187
16	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	Me	Me	I	I	5.44	0.018	0.916	0.064	0.001
17	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	Me	Me	I	H	5.47	0.324	0.611	0.023	0.042
18	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	H	H	H	3.69	0.798	0.085	0.106	0.011
19	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	H	I	H	5.35	0.210	0.675	0.029	0.087
20	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	I	H	H	5.05	0.738	0.071	0.174	0.016
21	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	H	H	I	I	5.40	0.012	0.907	0.001	0.08
22	L	CH ₂ CH(NH ₃ ⁺)COO ⁻	I	H	I	I	7.06	0.010	0.873	0.001	0.115
23		(CH ₂) ₃ COO ⁻	I	I	I	I	7.97	0.014	0.986	n/a	n/a
24		CH ₂ COO ⁻	I	I	I	I	8.37	0.014	0.986	n/a	n/a
25		COO ⁻	I	I	I	I	7.81	0.012	0.988	n/a	n/a
26		(CH ₂) ₂ COO ⁻	I	I	I	I	8.02	0.014	0.986	n/a	n/a
27		(CH ₂) ₂ NH ₃ ⁺	I	I	I	I	5.11	0.012	0.950	0.001	0.038
28		CH ₂ COO ⁻	I	I	I	H	6.84	0.272	0.727	n/a	n/a

^a Fractions of species representing more than 10% of the ligand in the solution are shown in bold.

mode, which was used in most studies on thyroxine-protein binding;^{12,65–67} and (3) the prevalent species in solution and average energies for all modes.

Multi-Species Multi-Mode CoMFA Studies. The extended CoMFA models were created in Sybyl,⁶³ with specific routines written in C program and integrated via a script language (SPL). Steric and electrostatic interaction energies X_{ijk} were calculated for each compound in each species and each binding mode at each lattice intersection of a 3-D cubic lattice (2 Å in each coordinate direction). An sp³ carbon probe with a charge of +1 and an energy cutoff 30 kcal/mol was used. The grid extended 4.0 Å beyond aligned ligands in all directions with the following coordinates: 10–32 Å along the x-axis, 32–54 Å along the y-axis, and -10–20 Å along the z-axis. The energies X_{ijk} in each grid point were organized in the energy table. The number of rows (1752) is determined by the product of the numbers of ligands (28), species (4 or 2), and modes (12 or more). The number of columns is given by the number of used grid points (1980) multiplied by the number of used fields (2), steric and electrostatic.

Equation 6 is nonlinear in coefficients C , which characterize electrostatic or steric interactions in individual grid points. The choice of one of the two implemented optimization methods

depends on the total number (N) of coefficients C in the calibrated model and on the number of compounds (n). Nonlinear regression analysis is applied when $N \leq n/2$. Otherwise, the developed software automatically switches to the partial least-squares (PLS) analysis with a linearized form of eq 6.²³ The second option was never used in this study because acceptable models were found with $N < n/2$. Both optimization methods are sensitive to good initial estimates. Therefore, a forward-selection procedure, conceptually similar to that we used before for multi-mode analysis,²³ was a natural choice.

Before analysis, the variables (energy columns) were sorted using the following criteria: (1) high and sustained variability, (2) the minimal number of colinear columns with similar information, and (3) an even distribution of selected grid points around the ligands. Sustained variability means that the energy values cover the whole interval more or less evenly and are not clustered at one end. An extreme case is the situation; when all energies in a given column are equal (usually zero) and only one or two modes have different energies. These singularities originate when the structures in only one or two modes protrude into a spatial region and other modes do not affect the energies in that region. These modes need to be eliminated from the analysis because they do

Table 2. Individual Species of Thyroxine Analogs (Structures in Table 1) Formed by Ionization and Sorted by the Charge of Ionizable Groups^a

ligands	species							
	1		2		3		4	
	group ionization status	overall charge	group ionization status	overall charge	group ionization status	overall charge	group ionization status	overall charge
1–22	NH ₃ ⁺	0	NH ₃ ⁺	-1	NH ₂	-1	NH ₂	-2
	COO ⁻		COO ⁻		COO ⁻		COO ⁻	
	OH		O ⁻		OH		O ⁻	
27	NH ₃ ⁺	+1	NH ₃ ⁺	0	NH ₂	0	NH ₂	-1
	OH		O ⁻		OH		O ⁻	
23–26, 28	COO ⁻	-1	COO ⁻	-2	n/a		n/a	
	OH		O ⁻		n/a		n/a	

^a Overall charges of individual molecules are also shown. Only the species with ionized carboxyls were considered because of their vast dominance in the binding experiments (pH 8.0).

not provide statistically significant information about the poorly covered region.

The model calibration can be steered by user-defined inputs, for which the default values are shown in parentheses. The process starts with a quick scan through the models consisting of a few (4) variables, which are randomly selected from the top variables (10%). The initial values of coefficients C (± 0.1) are systematically evaluated by a grid search. In this step, no optimization is applied and eq 6 is evaluated for the input values of the coefficients. For the best models (top 10% r^2), coefficients are optimized. At each of the following steps, a reduction of the number of coefficients is attempted: the coefficients are eliminated, for which the process does not lead to a decrease in the r^2 value. The best models (top 5% r^2) are developed by a gradual addition of groups of a few (3) variables, until all variables from the selected set (top 10%) are used. The procedure is finalized by fine-tuning of the best models (top 5% r^2) by addition of single variables. The best model is selected on the basis of the following criteria: N , r^2 for the full set and for the leave-one-out cross-validation on the training set, and the standard errors of the parameters.

The description statistics, calculated for the training set, consists of the sum of squares of errors (SSE), that is, differences between calculated and experimental values, and the squared correlation coefficient $r^2 = 1 - \text{SSE}/\text{SYY}$, where SYY is the sum of squares of deviations of the experimental $\ln K$ values from their average. The prediction statistics contains similar indices calculated for the test set that was not used in the development of the model: predictive sum of squares of deviations between predicted and experimental values (PRESS), and the squared correlation coefficient $q^2 = 1 - \text{PRESS}/\text{SYY}$.

RESULTS AND DISCUSSION

Thyroxine hormones exist in human body in two forms, which are interconverted in liver: 3,5,3',5'-tetraiodo-L-thyronine (thyroxine, T4, 15 in Table 1) and 3,5,3'-triiodo-L-thyronine (T3, 14 in Table 1).⁶⁸ The active hormone, T3, has several functions in human body, including stimulation of metabolic rates and increasing the heart function.⁶⁹

Studied Data Set. Structures of the studied thyroxine analogs and their properties are summarized in Table 1. The molecules are

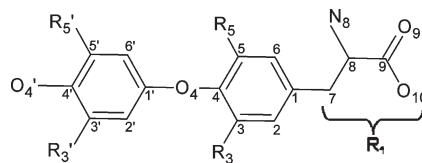


Figure 1. Heavy-atom connectivity and numbering of the skeleton of studied thyroxine analogs. The conformation of the diphenyl ether skeleton is defined by dihedrals Φ_1 (C3–C4–O4–Cl') and Φ_2 (C4–O4–Cl'–C2'). The amino acid side chain conformation is given by dihedrals Φ_3 (C2–C1–C7–C8), Φ_4 (C1–C7–C8–N8), and Φ_5 (N8–C8–C9–O10).

moderately flexible, with the main conformations defined by five dihedrals (Figure 1).⁷⁰ The side chain corresponding to C7–C8–(N8)–C9(O9)–O10 in Figure 1 is referred to as R₁ in Table 1.

Published binding affinities of thyroid hormone analogs to transthyretin,⁵⁶ as determined by a dialysis method at 37 °C, were used (Table 1), along with the data for 3',5'-diiodo-L-thyronine⁵⁷ (T2, 21 in Table 1). The binding affinities of the 28 ligands were determined⁵⁶ at pH 8.0.

Under these conditions, the carboxyls of all studied analogs completely ionize, while the NH₂ and OH groups undergo partial ionization. The pK_a values of the NH₂ and OH groups are quite similar, and two species with the same overall charge (NH₂ and OH, as well as NH₃⁺ and O⁻) need to be taken into account. All considered species are listed in Table 2. The species classification is based on the charges of ionizable groups. As a result, the species with the same number may have different overall charge, as indicated in Table 2. This species numbering was chosen because of the belief that the group charges are more important for specific ligand–receptor interactions than the overall charge of the molecule. However, it should be noted that the results are independent of species and mode numbering because the species- and mode-related terms are occurring in summations, and the only arrangement that matters is the correct association between the fraction of given species and its probe energies (eq 6 below).

The studied ligands exhibit rich protonation equilibria resulting in most cases from ionization of NH₂ and OH groups, while

the carboxyl is ionized in all cases where applicable. The amino acid derivatives **1–22** (Table 1) form four species with the charges -2 , -1 (two species), and 0 (Table 2). Compound **27** lacks the carboxyl and the charges of individual species are higher by one. Analogs **23–26** and **28**, lacking the amino group, form only two species differing in the ionization of the hydroxyl group.

Taking into account the importance of electrostatic and H-bond interactions in the ligand–receptor binding, the fractions of individual species need to be carefully characterized. The SPARC online calculator⁵⁹ was used to calculate, at pH 8.0 and $37\text{ }^{\circ}\text{C}$, the respective species fractions, which are summarized in Table 1. The ionization patterns are determined by the ionization state of phenolic hydroxyl group, which is promoted by the presence of halogen substituents in the phenolic ring.

Compounds containing halogen atoms in one ($3'$) or two ($3'$ and $5'$, Figure 1) positions of the phenolic ring undergo extensive deprotonation of the hydroxyl group leading to the dominance of species **2**, where the ionized hydroxyl is combined with the ammonium group. Species **2** can be accompanied by significant fractions of either species **1** with protonated hydroxyl and amino groups (compounds **17**, **19**, and **28**) or species **4**, where both protons were removed (compounds **2**, **8**, **15**, and **22**) or both species **1** and **4**, decreasing in this order (compounds **1**, **9**, **11**, and **14**). Species **2** is the only significant species for compounds **16**, **21**, and **23–27**, with the last group lacking NH_2 group in the amino acid side chain. The overall charges of individual species differ as shown in Table 2.

Compounds **3–7**, **10**, **12**, **13**, **18**, and **20** exhibiting nonhalogen substitutions in the phenolic ring have the phenolic hydroxyl group mostly in the neutral, protonated form. This pattern leads to the dominance of species **1** with overall charge 0 , and significant fractions of species **3** with overall charge -1 .

For two or more ionizable groups, the dissociation macro-constants are equal to the microconstants only if the pK_a values differ by 3 – 4 units. The pK_a values of the phenolic hydroxyl and the amino group of the studied ligands are closer than the limit. In such situations, the acidities of the ionizable groups are not sufficiently different and lead, at a given pH value of the medium, to the presence of the pairs of equally charged species, in which proton binds alternatively to one of the two ionizable groups. As follows from Table 2, such pairs are formed by species **2** and **3** for compounds **1–22** and **27**. The less abundant species typically represents 5 – 10% of the prevalent species in the pair (Table 1), which is a significant portion that needs to be considered in the MS-MM analysis.

The contributions of the species with nonionized carboxyls to the protonation equilibria were neglected because the respective fractions would be smaller than 0.01% in the binding experiments (pH 8.0). Hypothetically, nonionized carboxyl could participate in H-bond, while ionized carboxyl could act as H-bond acceptor or form a salt bridge with a positively charged receptor atom. The difference in energies of these interactions should not lead to an increase in the binding affinity by 3 – 4 orders of magnitude that would be necessary for the species with fractions below 0.01% to significantly contribute to overall binding.

Binding of Thyroxine Analogs to Transthyretin. Thyroid hormones bind in the blood to transporting proteins, thyroxine-binding globulin, transthyretin (TTR), and albumin.⁷¹ TTR, exhibiting the highest binding capacity for T3 (**14** in Table 1) among the shown proteins, is composed of four identical monomers, each consisting of 127 amino acids⁷² and having an extensive β -sheet structure.⁷³ A channel running through the center of the compact tetramer contains two nearly symmetrical binding sites.^{73,57} The

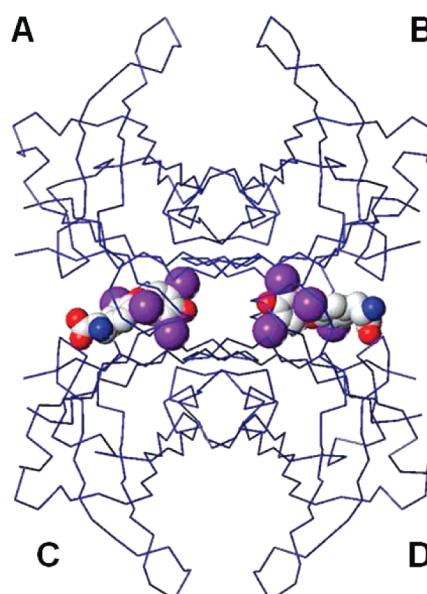


Figure 2. Two thyroxine molecules bound in the deep forward orientation (see below) at two opposite binding sites of transthyretin (PDB code 1ICT⁸¹). The transthyretin tetramer is shown as α -carbon wire with the subunits labeled and thyroxine molecules are shown in space-fill rendering.⁸² The affinities of the two binding sites are widely different due to negative cooperativity.

sites differ widely (up to 2 orders of magnitude) in the binding affinity^{74,75} because of negative cooperativity.⁷⁶ Monomers, released by the dissociation of the tetramer, may misfold and aggregate into protofilaments, filaments, and, under appropriate conditions, into amyloid fibrils.⁵² TTR ligands stabilize the tetramers and inhibit the process of TTR fibril formation in vitro.⁷⁷ Several classes of amyloidogenesis inhibitors were developed utilizing this principle.^{54,55,78–80} An X-ray structure of transthyretin with one thyroxine molecule in each binding site is shown in Figure 2.

Each binding site consists of three pairs of binding pockets:⁶⁵ inner pocket defined by Leu110, Ser117, Thr119; middle pocket defined by Leu17, Thr106, Ala108, Ala109, Val121; and outer pocket defined by Lys15 and Glu54 (Figure 3). Three binding orientations were observed in the X-ray structures of the transthyretin complexes with thyroxine analogs (PDB files 2ROX,⁸³ 1Z7J,⁶¹ and 1ICT⁸¹ or 1THA⁸²). They are characterized by the iodine positions in the three pockets,⁵⁷ as summarized for thyroxine in Table 3 and Figure 3. In the forward (F) orientation, $5'$ -iodine binds in the inner pocket, $3'$ -iodine binds in the middle pocket, and 3 - and 5 -iodines both bind in the outer pockets.⁸³ The interaction of the ionized carboxyl group of thyroxine with Lys15 of the C chain pushed the molecule deeper into the binding pocket. The resulting deep forward (DF) orientation is characterized by the shift of the $3'$ -iodine into the inner pocket and the binding of 3 - and 5 -iodines in the middle pockets.^{81,82} The thyroxine molecule is completely flipped in the reverse (R) orientation: both 3 - and 5 -iodines bind in the middle pockets and $3'$ - and $5'$ -iodines protrude into the outer pockets⁶¹ (Table 3, Figure 3).

Conformations of Ligands. Thyroxine derivatives bind to transthyretin in multiple binding modes (Table 3, Figure 3). The ligands extracted from the PDB files 1ICT, 2ROX, and 1Z7J served as the templates for the deep forward (DF), forward (F), and reverse (R) orientations, respectively. The ligand from the PDB file 1Z7J, T4-acetic acid (T4-AA; ligand **23** in Table 1),

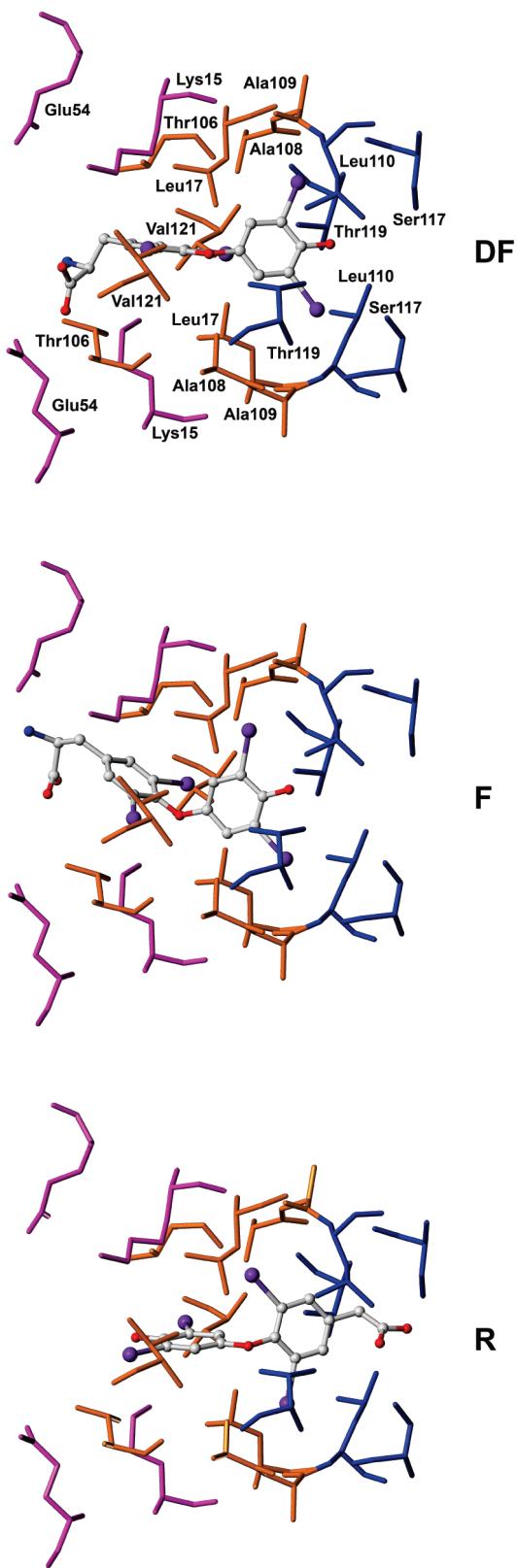


Figure 3. Deep forward (DF), forward (F), and reversed (R) binding orientations of thyroxine to transthyretin as observed in the PDB files 1ICT, 2ROX, and 1Z7J, respectively. The colors indicate the binding pockets: inner (blue), middle (orange), and outer (magenta). For illustration the residues of 1ICT are shown in all cases (capped sticks). Thyroxine is depicted as atom-colored balls and sticks.

Table 3. Deep Forward, Forward, and Reverse Binding Orientations of Thyroxine to Transthyretin As Documented by X-ray Data and Characterized by the Placement of Iodine Atoms in Different Binding Pockets

binding orientation	PDB file	binding pocket		
		inner	middle	outer
deep forward	1ICT	3'-I, 5'-I	3-I, 5-I	
forward	2ROX	5'-I	3'-I	3-I, 5-I
reverse	1Z7J		3-I, 5-I	3'-I, 5'-I

lacks the amino acid chain that is present in ligands 1–22 (Table 1). To obtain a suitable template for the reverse orientation, thyroxine (**15**, Table 1) molecule with the same orientation and conformation (dihedrals Φ_1 , Φ_2 , and Φ_3 , Figure 1) as T4-AA was docked into the binding site and its side chain was optimized to obtain dihedrals Φ_4 and Φ_5 . The magnitudes of dihedrals for all templates are summarized in Table 4.

The X-ray structures indicate that the ring substituents form predominantly hydrophobic bonds within the binding pockets and do not participate in other interactions. Since hydrophobic interactions do not exhibit strong directional preferences, we decided to examine, in addition to the three binding orientations shown in Table 3 and Figure 3, the conformations resulting from flipping the two benzene rings of ligands by 180°. Similar approaches to the generation of binding modes have been used previously.^{23,35,36,84–87} As illustrated in Figure 4, mode A is the experimentally observed conformation, from which other modes were derived by flipping the phenolic ring by 180° (conformation B), the tyrosine ring by 180° (conformation C), and both phenolic and tyrosine rings by 180° (conformation D).

In summary, twelve binding modes were generated, as listed in Table 4, representing four different skeleton conformations for each of the three orientations (DF, F, and R). It should be noted that species and modes have equivalent representation in the correlation equation (eq 6), and the results are independent from their numbering.

Additional Modes for Compounds with Flexible Substituents. Compounds **3**, **5**, **6**, and **10** (Table 1) contain, in position 3', flexible alkyl substituents with several rotatable bonds (isobutyl, *n*-propyl, *s*-butyl, and benzyl, respectively). Estimating exact conformations of alkyl chains in both ligands and receptors in the complexes is a challenge because of the inherent flexibility. In the X-ray structures, alkyls are often found in the regions of undefined electron density, indicating multiple binding modes.

To account for the increased potential to form multiple binding modes for compounds **3**, **5**, **6** and **10**, each of the four skeleton conformations (A, B, C, and D) obtained by flipping both phenolic and tyrosine rings, were expanded by utilization of several substituent conformations. A systematic conformational search was used to find low energy conformations of the alkyl chains, focusing mainly on the extended (protruding outward and occupying more space) and folded (occupying minimal space without steric clash) conformations. Four alkyl conformations, two extended conformations (e1 and e2) and two folded conformations (f1 and f2) were selected for compounds **3**, **5**, and **6**, and three alkyl conformations (e1, e2, and f1) for compound **10** (Figure 5). Eventually, compounds **3**, **5**, and **6** were represented by 192 total structures each (4 species × 3 orientations × 4 skeleton conformations × 4 alkyl conformations), and compound **10** was represented by 144 total structures (4 species × 3 orientations × 4 skeleton conformations × 3 alkyl conformations). In this context,

Table 4. Characteristics of the Twelve Binding Modes That Were Used for All Compounds: In Each Binding Orientation (DF, F, R; Figure 3), a Compound Can Bind in Four Skeleton Conformations (A, B, C, and D; Figure 4)^a

mode no.	orientation	conformation	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5	template ligand	PDB file
1	deep forward	A	93.8	19.2	115.8	339.4	350.0	19	1THA
2		B		199.2					
3		C	273.8	19.2	295.8				
4		D		199.2					
5	forward	A	102.9	346.9	97.9	215.3	273.7	15	2ROX
6		B		166.9					
7		C	282.9	346.9	277.9				
8		D		166.9					
9	reverse	A	73.3	6.4	141.1	165.1 ^b	346.2 ^b	24	1Z7J
10		B		186.4					
11		C	253.3	6.4	321.1				
12		D		186.4					

^a Torsions Φ_1 – Φ_5 are defined in Figure 1. ^b Optimized values for template 15 (Table 1) with the same orientation and conformation (Φ_1 , Φ_2 , and Φ_3) as compound 24.

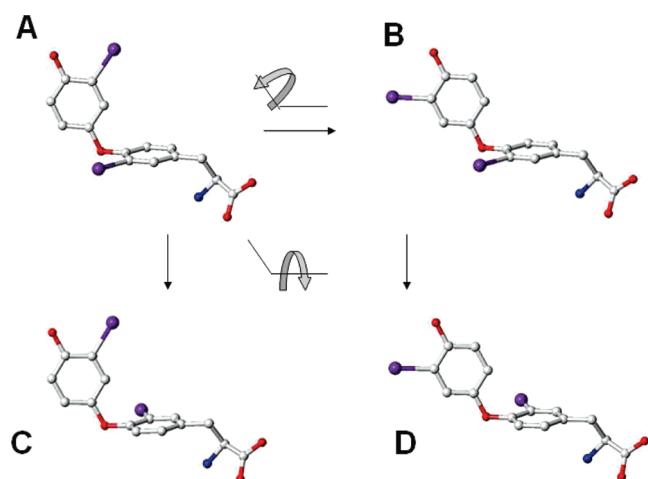


Figure 4. Four skeleton conformations of studied ligands, illustrated using 3,3'-diodo-L-thyronine (compound 19, Table 1), derived from crystallographic conformation (conformation A, PDB file 1THA). Conformation B was obtained by flipping the phenolic ring by 180°; conformation C results from flipping the tyrosine ring by 180°; and conformation D was generated by flipping both phenolic and tyrosine rings by 180°. Atom color coding: carbon, gray; nitrogen, blue; oxygen, red; iodine, purple.

a structure means the molecule characterized by ionization species, orientation, and conformation (the skeleton conformation for all compounds, plus the alkyl conformation for compounds 3, 5, 6, and 10). Geometry changes of symmetric alkyl substituents, i-Pr in compound 4 and t-Bu in compound 7 are less significant, and these compounds are not expected to form multiple binding modes based on the substituent conformations.

Superposition. Three crystal structures of the TTR dimer complexes (PDB files 1THA, 2ROX, 1Z7J) were superimposed by homology, to generate the coordinates for the superposition of the three orientations (DF, F, and R; Figure 6a). Four skeleton conformations (A, B, C, and D) or their multiple versions with flexible substituents of each species for each compound were superimposed on these templates (Figure 6c). For MS-MM analysis, 1720 structures were superimposed: 28 compounds, each comprising two or four species (Table 2), and either twelve,

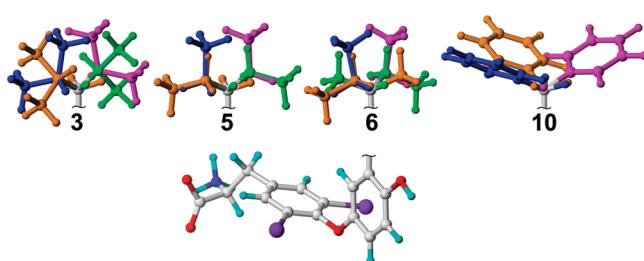


Figure 5. Multiple analyzed conformations of isobutyl, n-propyl, s-butyl, and benzyl substituents in compounds 3, 5, 6, and 10, respectively. Conformations are named as extended (e1, blue; e2, magenta) and folded (f1, orange; f2, green).

thirty six (compound 10) or forty eight (compounds 3, 5, and 6) modes for each species (Figure 6).

The single species, multimode CoMFA procedure was performed for each species (Table 2), as well as for the major species at pH 8.0 (Table 1). The pertinent species was superimposed in three different orientations (Figure 6b), each orientation containing four, twelve (compound 10) or sixteen (compounds 3, 5 and 6) conformations. In these cases, 468 structures were superimposed in total.

For the standard (single-species, single-mode) CoMFA procedure, three approaches were applied. In the first approach, the most prevalent species and modes for each compound from multispecies, multimode CoMFA analysis were included in standard CoMFA analysis. In the second approach, only forward modes of predominant species for all compounds were considered. In the third approach, for each compound, the average energies of all modes of predominant species (species with the highest fraction at pH 8.0) in each lattice point were considered.

Multi-Species, Multi-Mode Binding Equilibria. The studied compounds can be present in the solution as two (23–26, 28) or four species (1–22, 27) thanks to ionization (Table 2). No tautomers or hydrates are expected to exist for these compounds, according to the predictions by the SPARC web server,⁵⁹ but the conclusions in the following general discussion would apply to these species would they have been encountered.

The ligand species in any solution conformation can approach the receptor and bind in a conformation that is usually different from

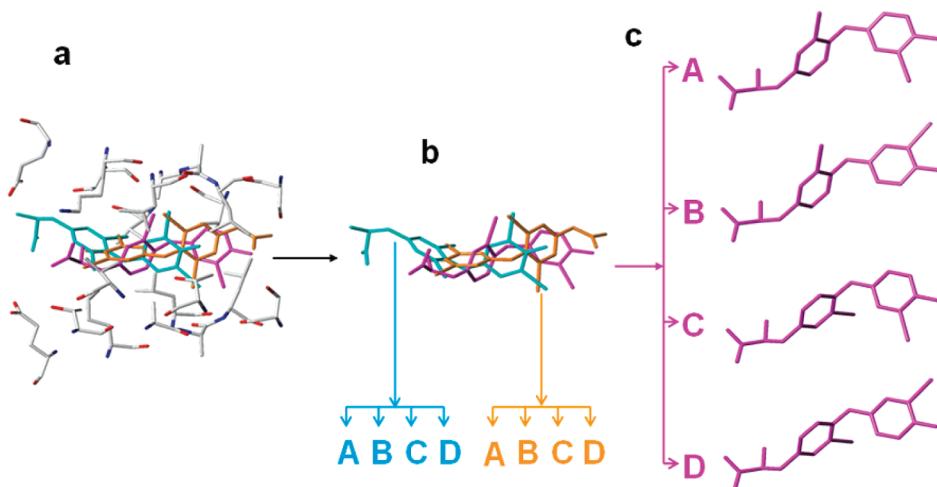


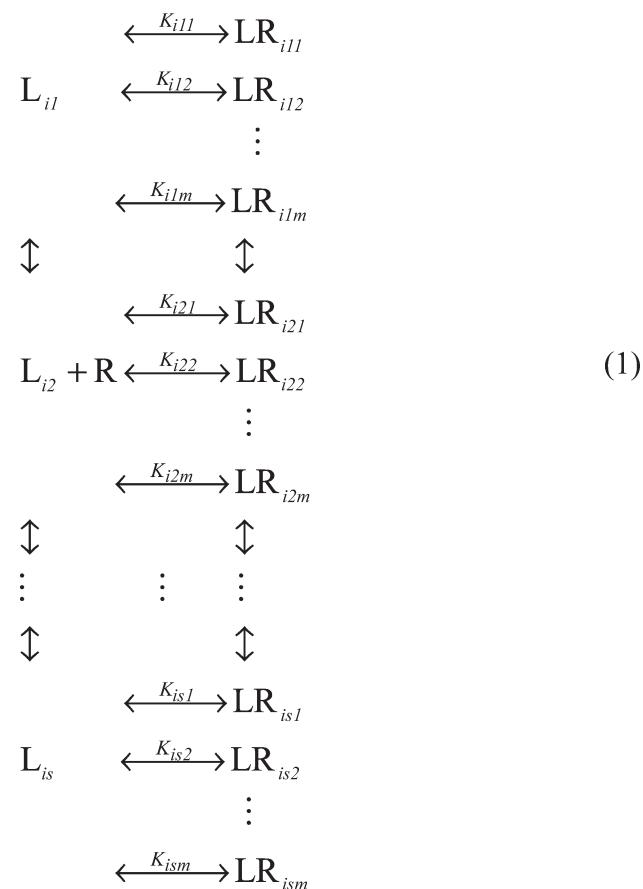
Figure 6. Generation of multiple modes: (a) superposition of the complex structures (PDB files 1THA, 2ROX, and 1Z7J), containing three binding orientation templates (DF orientation in magenta, F orientation in cyan, and R orientation in orange; the binding site amino acids in atom colors); (b) the three extracted templates; (c) the four skeleton conformations A, B, C, and D for the DF orientation, shown for 3,3'-diiodo-L-thyronine (**19** in Table 1). Similar skeleton conformations were constructed for F and R orientations. Additional conformations were generated for alkyl-substituted compounds **3**, **5**, **6**, and **10**, to accommodate the increased potential for multiple binding modes of alkyl chains.

the prevalent solution conformation. An extensive analysis of the structures of 150 complexes in the PDB showed⁸⁸ that ligands bind in the global and local minimum energy conformations with almost zero and less than 40% probabilities, respectively. The strain energies to the closest local minimum often reached significant levels, e.g. 9 kcal/mol for more than 10% bound ligands, with higher penalties associated with more flexible molecules. The vast majority of the ligand–receptor complexes are formed by noncovalent interactions. For these complexes, the change from the solution conformation to the bound conformation is usually fast, and can be treated as practically instantaneous. The energy loss, calculated as the energy difference of bound and free ligand conformations, can be incorporated directly into the correlation equation. For these reasons, the Boltzmann distribution of conformations in the solution need not be accounted for in a description of overall binding (an opposite view has been presented¹⁷).

Individual ligand species occurring in the solution differ in structures of some groups or fragments and, consequently, have different binding affinities. Changes in the ionization, tautomer, or hydrate state upon binding are much less probable than a change of conformation. Therefore, in contrast to the conformations, the fractions of individual species in the solution around the receptor play an important role in the description of binding. One could easily envision some exceptions to these expectations. For instance, a protomer could change the pK_a value upon binding, and convert to other species under the given conditions. The probability of the conversion depends on the overall binding energetics of each species, and on the structural predisposition of the receptor to accommodate/supply the released/bound protons for ionization conversion and water for hydrate conversion. The conversion would require a more complex description of binding equilibria. Since no experimental information is available on this complex phenomenon, we will treat the species binding as if no conversion is taking place upon binding. Since each species has a specific binding affinity, the species distribution of free and bound molecules is generally different.

Each ligand species can bind to the binding site of the receptor R in different orientations or conformations (binding modes). If the binding site is of similar size as the bound ligands, each

complex contains only one bound ligand molecule, that is, the modes are mutually exclusive. The multispecies, multimode (MS-MM) situation is outlined in eq 1



Each binding equilibrium represents the receptor binding of the i -th ligand in the j -th species and the k -th binding mode, and is characterized by the microscopic association constant K_{ijk} . In

principle, the number of species s and the number of binding modes m can differ for individual ligands. To avoid double indexing, s and m represent the maximum numbers of species and binding modes for the tested ligands. The missing species or modes of some ligands will have zero fractions f_{ij} or zero association constants K_{ijk} , respectively (see below).

The association constant for the i -th ligand bound as the j -th species in the k -th binding mode is defined as

$$K_{ijk} = \frac{[\text{LR}_{ijk}]}{[\text{L}_{ij}] \times [\text{R}]} \quad (2)$$

The microscopic association constants are the relevant quantities to be correlated with structure. Experimentally, however, the total association constant is usually determined. The measured equilibrium constant of the i -th ligand K_i reflects the total concentration of the ligand/receptor complexes, without distinguishing between complexes differing in species or in binding modes. To express the total association constant as a function of the microscopic association constants K_{ijk} , a series of rearrangements needs to be made, as shown in eq 3: (1) the total bound concentration $[\text{LR}_i]$ is given as the sum of s bound concentrations of individual species $[\text{LR}_{ij}]$; (2) each summand is formally multiplied by $[\text{L}_{ij}]/[\text{L}_{ij}]$ and, in this way, the fraction of each species, $f_{ij} = [\text{L}_{ij}]/[\text{L}_i]$, is introduced; and (3) the bound concentration of each species $[\text{LR}_{ij}]$ is given as the sum of m concentrations of individual binding modes $[\text{LR}_{ijk}]$

$$\begin{aligned} K_i &= \frac{[\text{LR}_i]}{[\text{L}_i][\text{R}]} = \sum_{j=1}^s \frac{[\text{LR}_{ij}]}{[\text{L}_i][\text{R}]} = \sum_{j=1}^s f_{ij} \frac{[\text{LR}_{ij}]}{[\text{L}_{ij}][\text{R}]} \\ &= \sum_{j=1}^s f_{ij} \sum_{k=1}^m \frac{[\text{LR}_{ijk}]}{[\text{L}_{ij}][\text{R}]} \end{aligned} \quad (3)$$

The fractions f_{ij} are assumed to be identical in the defined aqueous solution with and without the receptor. For speciation due to ionization, tautomerism, and hydration of aldehydes, ketones, and other groups, this is a plausible assumption, which is satisfied when the pH of the medium and the water activity are the same in bulk water and around the receptor.

Combining eqs 2 and 3, the expression relating the microscopic and total association constants reads

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

where f_{ij} is the fraction of the j -th molecular species in the i -th compound in the solution. For compounds present in the solution as one species, the simple eq 4 is in accordance with published analyses of formally analogous situations: the statistical thermodynamic⁸⁹ and equilibrium^{35,36} treatment of multimode binding in ligand/protein interactions and kinetic analyses of a reversible unimolecular reaction leading to different products⁹⁰ or isomers.⁹¹

For ionization equilibria, the expressions for the fractions f_{ij} of the j -th species can be derived from the definition of the ionization constants. In the case two or more ionization groups, attention must be paid to the actual macroscopic pK_a values. If the pK_a values are closer than 3–4 units, the ionization tendencies of the two ionizable groups are not clearly separated, and mixed species may occur. Consequently, such pK_a values are no longer equal to the microconstants, and multiple species can have

the same charge (e.g., neutral molecules and zwitterions as two species with the zero charge).

The prevalence of the k -th mode of the j -th bound species for the i -th ligand is

$$\begin{aligned} \frac{[\text{LR}_{ijk}]}{[\text{LR}_i]} &= \frac{K_{ijk} \times [\text{L}_{ij}] \times [\text{R}]}{K_i \times [\text{L}_i] \times [\text{R}]} = \frac{K_{ijk} \times f_{ij}}{K_i} \\ &\approx \frac{K_{ijk} \times f_{ij}}{\sum_{j=1}^s f_{ij} \times \sum_{k=1}^m K_{ijk}} \end{aligned} \quad (5)$$

The numerator and denominator come from eqs 2 and 3, respectively. The third quasi-equality ensures that the sum of all $s \times m$ prevalences equals unity. The prevalence of the j -th bound species in all modes and the prevalence of the k -th bound mode in all species can easily be summed up from the summands calculated by eq 5.

Mulit-Species Multi-Mode CoMFA Studies. For the application in CoMFA, eq 4 needs to be logarithmized and the microscopic association constants must be expressed in terms of the ligand–probe interaction energies X for the probes, which are placed in individual cross sections (total number g) of a grid surrounding the superimposed ligands

$$\begin{aligned} \ln K_i &= \ln \sum_{j=1}^s f_{ij} \times \sum_{k=1}^m e^{\ln K_{ijk}} \\ &= \ln \sum_{j=1}^s f_{ij} \times \sum_{k=1}^m e^{1-\sum_{l=1}^{2 \times g} C_l \times X_{ijkl} + C_0} \end{aligned} \quad (6)$$

Here, C are the regression coefficients characterizing the weights of the energies X , one coefficient per grid point and energy type. Two types of interaction energies, steric and electrostatic, are usually used; hence, the number of optimized coefficients is formally $2 \times g$. The number of coefficients is minimized in the coefficient calibration procedure. The subscripts of interaction energies X and species fractions f indicate the relatedness to the i -th ligand, j -th species, k -th mode, and l -th grid point. Other contributions to the binding free energy (the differences upon binding in conformational energy, entropy, solvation) can be added to the summation in the exponents. In this case, we only used the CoMFA energies because the size of the used data set (28 compounds) did not allow a rigorous optimization of more than seven or eight coefficients C . The regression coefficients are associated with the grid points, and, consequently, are independent of the species and modes. Therefore, the inclusion of multiple species and multiple modes does not increase the number of optimized regression coefficients C . For the same reason, the final set of optimized coefficients is independent of the numbering of individual species and modes.

CoMFA models were created in SYBYL, with the multispecies, multimode routines written in C language and integrated into SYBYL using the Sybyl Programming Language (SPL) scripts. An sp^3 carbon probe with a charge of +1 was used, with an energy cutoff 30 kcal/mol. For a comparison, reduced models considering only single species and multiple modes, as well as standard (one-species, one-mode) CoMFA model were also generated. The respective superpositions are shown in Figure 7.

Model Calibration and Selection. Our goal was to find the minimal best-fitting model (eq 6) with acceptable statistical criteria, as common in many areas of science. In the calibration

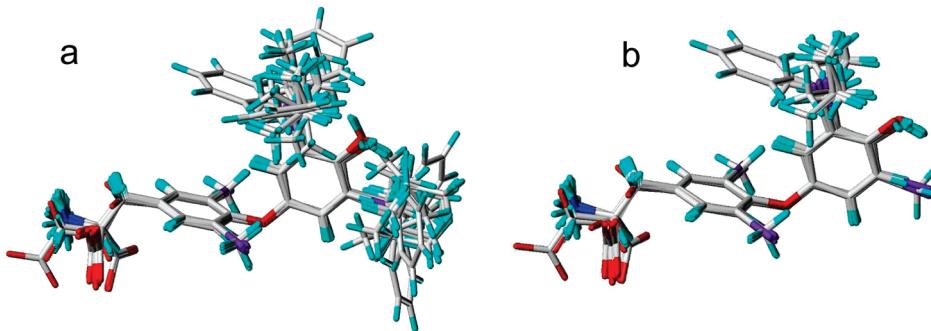


Figure 7. Superpositions for the forward orientation in (a) the multimode situation (156 structures arising from 24 ligands in 4 modes, 1 ligand in 12 modes, and 3 ligands in 16 modes) and (b) the one-mode situation (28 structures). For multiple species, the results are similar, because the species differ essentially only in the presence of hydrogens. To generate overall superpositions, similar setups for the deep-forward and reverse orientations (Figures 4 and 6) were added.

Table 5. Numbers of single-species and multi-species multi-mode models satisfying the shown criteria

$r^2 \geq$	$q^2 \geq$	standard deviations of coefficients \leq (%)	number of coefficients \leq	number of models satisfying the criteria	
				SSMM	MSMM
0.70				858	35 715
0.80				752	29 028
0.90				496	22 129
0.90	0.60			281	14 128
0.90	0.70			147	10 837
0.90	0.70	40		15	2 076
0.90	0.70	30		5	615
0.90	0.70	30	11	5	403
0.90	0.70	30	10	1	319
0.90	0.70	30	9	1	245
0.90	0.70	30	8	1	144

procedure, 21 compounds were used as the training set. The test set consisted of 6 compounds. Compound **10** was excluded from calibration because of singularity caused by the large benzyl substituent, which occupied the grid intersections unreachable by other studied compounds. Equation 6 is nonlinear in coefficients C , which need to be optimized by nonlinear regression analysis, requiring the initial estimates of C . The extensive search based on the selection of variables and a systematic variation of the initial estimates gave rise to more than 100 000 calibrated models. The final models were selected based on the subsequent application of the following criteria for the training set: r^2 , q^2 , the maximum relative standard deviation of optimized coefficients, and the number of used coefficients. The q^2 values were calculated for the leave-one-out procedure on the training set, starting with the selected variables and the initial estimates providing the respective r^2 value. The main goal of this exercise was the identification of instable models, because the usefulness for cross-validation was limited under these conditions. The numbers of multimode models, including both single-species and multispecies models, which satisfy the mentioned criteria are given in Table 5.

The best CoMFA models in individual categories are summarized in Table 6 and characterized by the description and prediction statistical indices. The numbers of adjustable coefficients for individual situations are also shown. Descriptive

abilities (SSE, r^2) were good in all shown cases, and especially for the classical PLS-based models. The consideration of multiple modes (lines 1–3) improved the predictive abilities (PRESS, q^2). The standard CoMFA model that was built using the most preferred species and modes from multispecies, multimode analysis shows similar predictive ability as is that of the MS-MM analysis itself, emphasizing the importance of multiple species and multiple modes consideration. We unsuccessfully attempted to increase the comparatively low descriptive ability of the most complex model (line 1) by using the selected variables and initial estimates from the single species model (line 2). The multispecies, multimode model (eq 6) and the single species, multimode model (eq 6 with one fraction f set to 1) contain the same numbers of nonlinear adjustable coefficients C and differ in the used equations, so the extrapolation from simpler to more complex models is not as straightforward as we are used to for models of the same functional forms, differing in the number of optimized coefficients. The addition of multiple species to the multimode CoMFA analysis led to a better predictive ability (PRESS, q^2) of the model in terms of the deviations from the experimental binding affinities, although the improvement is not impressive. However, the predictions for individual binding modes were in a much better agreement with experimental data for the multispecies CoMFA model (see Table 9 below). The standard CoMFA analysis with the averaged energies showed

poor predictive ability. The calculated and predicted binding affinities of the training set and the test set, respectively, as related to the experimental data, are shown in Figure 8.

Predicted Binding Orientations. The results for the multi-species, multimode model, calculated using eq 5, are shown in Table 7. The sum of the prevalences of the four conformations (A, B, C and D) is also given to emphasize the contributions of the three basic binding orientations (DF, F, and R). Twelve out of 27 compounds (compound 10 was excluded from the prediction because of singularity caused by the large benzyl substituent) bind in one major orientation with the prevalence $\geq 90\%$: five compounds (4, 6, 7, 20, and 23) bind in DF orientation, six compounds (1, 2, 22, 24, 26, and 27) bind in R orientation, and thyroxine (15) binds in its preferred F orientation. Fifteen compounds exhibit multiple binding orientations with prevalences $\geq 10\%$. Compounds 14, 17, and 18 bind mainly in similar DF and F orientations, while compounds 8, 12, 13, and 28 bind in very different DF and R orientations. Compounds 3, 5, 16, 21, and 25 bind in both F and R orientations. The remaining compounds 9, 11, and 19 bind in all three major orientations.

How are the binding preferences associated with the structural features of the binding site? The inner and middle pockets contain H-bonding serine and threonine residues, and the

nitrogen and oxygen atoms of the peptide backbone, while the outer pockets are formed by charged lysine and glutamate residues (Figure 3). The contacts made with the iodine atoms of the ligands are also tighter and more numerous as a direct result of the architecture of the binding site which flares out in a trumpet shape as the distance from the channel center increases. Thus the contacts for the outer pockets occur mainly through the side chains of amino acid residues, while those for the inner pockets often involve the peptide backbone itself.

The effects of ligand structure on binding orientations and affinities are difficult to predict, as seen in the following examples. A subtle change in 3' substituent from i-Pr (compound 4) to n-Pr (5) causes the shift in preference from DF orientation (4) to F and R orientations (compound 5), while keeping the overall binding affinities very similar. On the contrary, the difference between s-Bu and t-Bu in position 3' leads to a significant change in affinities of compounds 6 and 7, while their DF orientation remains unchanged. Compound pairs 1 and 14 and 2 and 15 differ in the D/L configuration of the chiral α -carbon of the amino acid chain. This change is sufficient to affect affinities by more than an order of magnitude and change the preferred orientations from R for the weak-binding D-enantiomers 1 and 2 to DF and F for stronger binders 14 and 15 with L configuration.

DF orientation is exhibited by 15 out of 27 of compounds either as a single dominant mode or along with similar F mode or dissimilar R mode. Significant ($>10\%$) binding in DF orientation is associated with the presence of the H substituent in position 5', although there are exceptions: compound 1 does not bind in DF mode despite having the 5'-H substituent and compounds 8, 12, and 23 exhibit DF mode binding although their 5'-substituents are Br, CH₃, and I, respectively. In addition, the interaction of the ionized carboxyl group of the compounds with Lys15 of the C chain seems to be important for pushing the molecules deeper into the binding pocket. The optimum chain length of compound 23 results in 100% binding in the DF mode, despite the bulky iodine in the 5' position, while the shorter chain lengths of 24, 25, and 26 do not exhibit this push. Compound 27 cannot exhibit it either, since it lacks carboxyl group.

F orientation is the least dominant orientation: it is adopted by 12 compounds, in all cases combined with DF and/or R orientations. This orientation is three times (compounds 14,

Table 6. Statistical Indices for the CoMFA Analyses Differing in the Number of Considered Species and Modes

CoMFA model		statistical indices				
species	modes	number of coefficients	SSE	r^2	PRESS	q^2
multi	multi	8	2.127	0.92	2.049	0.75
single ^a	multi	8	1.544	0.94	2.263	0.72
single ^b	single ^b	6 (695) ^e	1.275	0.95	2.319	0.72
single ^a	single ^c	6 (859)	0.941	0.96	3.05	0.63
single ^a	single ^d	6 (3859)	0.858	0.97	48.13	-4.87

^a The species with the highest aqueous fraction at pH 8.0 (Table 1) was used for each ligand. ^b The most preferred modes and species from the MS-MM approach were used for each ligand. ^c Only forward mode was used for all ligands. ^d Average energies of all plausible modes used for each ligand. ^e The number of latent variables, with the number of all used columns given in the brackets.

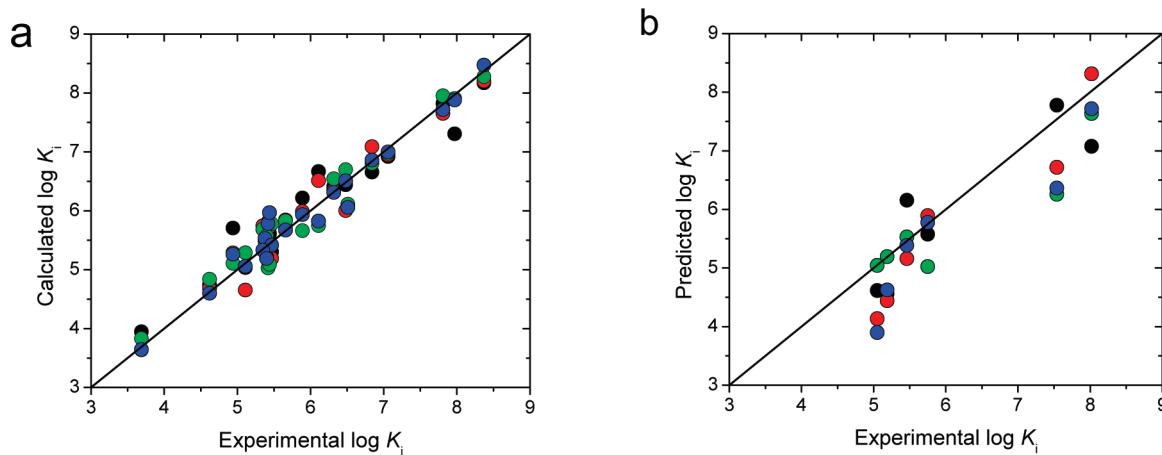


Figure 8. Experimental data vs calculated binding affinities for the training set (a) and predicted binding affinities for the test set (b). Multispecies, multimode model (black points), single-species, multimode model (red), single-species, single-mode (the most preferred modes from MS-MM model) (green), and single-species, single-mode (forward mode, blue).

Table 7. Prevalences (%) of the Twelve Binding Modes for Studied Ligands (eq 5)^a

no.	prevalence for different modes														
	deep forward				forward				reverse						
	1 A	2 B	3 C	4 D	sum	5 A	6 B	7 C	8 D	sum	9 A	10 B	11 C	12 D	sum
1	0	0	0	0	0	0	0	0	0	47	3	47	3	100	
2	0	0	0	0	0	0	0	0	0	25	25	25	25	100	
3	0	4	0	4	8	28	0	28	0	56	2	16	2	16	36
4	0	50	0	50	100	0	0	0	0	0	0	0	0	0	0
5	0	3	0	3	6	12	0	12	0	24	4	31	4	31	70
6	0	46	0	46	93	0	1	0	1	1	0	3	0	3	6
7	0	50	0	50	100	0	0	0	0	0	0	0	0	0	0
8	6	6	6	6	24	1	1	1	1	5	18	18	18	18	70
9	21	7	21	7	56	8	4	8	4	24	5	5	5	5	20
11	22	9	22	9	63	6	4	6	4	21	6	2	6	2	16
12	12	12	12	12	48	2	2	2	2	9	11	11	11	11	43
13	27	16	26	16	86	1	2	1	2	5	3	2	3	2	9
14	12	3	12	3	29	23	8	23	8	61	3	1	3	1	9
15	0	0	0	0	2	23	22	22	22	90	2	2	2	2	9
16	1	1	1	1	2	3	2	2	2	10	22	22	22	22	88
17	7	2	7	2	19	27	11	27	11	76	2	0	2	0	5
18	2	2	25	25	54	0	0	23	23	46	0	0	0	0	1
19	1	0	8	4	13	0	0	47	23	70	0	0	15	1	16
20	24	24	24	24	98	0	0	0	0	2	0	0	0	0	1
21	2	2	2	2	9	10	10	10	10	38	13	13	13	13	53
22	0	0	1	1	1	0	0	2	2	4	3	3	45	45	95
23	25	25	25	25	100	0	0	0	0	0	0	0	0	0	0
24	1	1	1	1	5	1	1	1	1	3	23	23	23	23	93
25	0	0	0	0	0	10	10	10	10	41	15	15	15	15	59
26	1	1	1	1	5	0	0	0	0	0	24	24	24	24	95
27	0	0	0	0	0	0	0	0	0	0	25	25	25	25	100
28	34	10	34	10	87	2	0	2	0	4	3	2	3	2	9

^a All species contributing to the given mode are included. To emphasize the differences among the deep forward, forward, and reverse orientations, the sums of prevalences for A, B, C and D conformations for each orientation are also shown. The significant (>10%) modes and orientations for individual compounds are printed in bold.

17, and 18) shared exclusively with the similar DF orientation, six times (3, 5, 15, 16, 21, and 25) with dissimilar R orientation, and three times (9, 11, and 19) with both DF and R orientations.

The most frequently occurring R orientation is observed with prevalence >10% for 20 of 27 compounds. Ten (2, 8, 12, 16, 21, 22, and 24–27) of twelve compounds having both 3' and 5' positions filled with non-H atoms have high prevalences of R orientation, except compounds 15 and 23. Compound 23 exhibits dominant DF orientation because its butyrate group is too large to allow all four halogens to make the interactions typical for R orientation. A small fraction of compound 15 is binding in R orientation, although F orientation dominates.

The tetra-iodo-substituted compounds 2, 15, and 23–27 prefer to bind in R orientation but this ability depends on the side chain. Too long (23) or too short side chains (25) lead to the preference for other orientations. For the alanyl side chains, the L configuration (15) is less supportive of R orientation than the D configuration (2).

Compound 21 exhibits a peculiar combination of two dissimilar binding orientations, R and F. The presence of two phenolic iodine substituents and the absence of tyrosine iodines explains the preference for the R binding orientation in which both phenolic iodines make stronger interactions with the entrance of the binding site, and F orientation in which both the phenolic iodines experience a forward penetration to make strong interaction with the binding pocket.

For compounds 3, 5, and 6 with flexible alkyl substituents in position R_{3'}, the breakdown of individual skeleton conformations into the contributions of the two extended (e1, e2) and two folded (f1, f2) substituent conformations (Figure 5) is shown in Table 8. Compound 3 is mostly bound with extended isobutyl substituent in the A and C skeleton conformations and F orientation. Compound 5 binds mostly in two modes: the A and C skeleton conformations with extended n-propyl substituent conformations in F orientation, as well as the B and D skeleton conformations with the folded n-propyl substituent conformations in R orientation. Compound 6 binds in the DF

orientation in B and D skeleton conformations, with about equal representation of the extended and folded s-butyl conformations.

Binding Orientations: Experiment versus Prediction. Double orientation prevalences for compound **14**, **15**, **19**, **21**, and **22** have been established by published X-ray diffraction studies based on the relative intensities of iodine peaks in inner, middle and outer pockets of transthyretin.⁵⁷ These experimentally determined orientations are as follows: for compound **14**, 70% F and 30% DF or, alternatively, 45% F and 55% R;⁵⁷ for compound **19**, 70% F or R, 30% DF; for compounds **21** and **22**, 60% R and 40% F, respectively. The orientation templates from two PDB files, compound **15** from 2ROX and compound **24** from 1Z7J, where no multiple modes were detected, can be considered as having 100% of the observed F and R orientations, respectively. The ligand in PDB file 1ICT that was used as the DF template, was not studied because no experimental affinity was available. The orientation prevalences predicted by the MS-MM model are in good agreement with the experimental data, except

Table 8. Prevalences (%) of the Sixteen Modes for Compounds **3, **5** and **6** (eq 5) Containing Flexible Alkyl Substituent in Position R_3' , Each Represented by a Skeleton Conformation (A–D) Combined with Either One of Two Extended (e1, e2) Conformations or One of Two Folded (f1, f2) Conformations (Figure 5)^a**

mode	prevalence for different modes								
	compound 3			compound 5			compound 6		
	DF	F	R	DF	F	R	DF	F	R
Ae2	0	27	1	0	12	1	0	0	0
Be2	1	0	12	1	0	0	23	0	0
Ce2	0	27	1	0	12	1	0	0	0
De2	1	0	12	1	0	0	23	0	0
Af2	0	1	1	0	0	1	0	0	0
Bf2	2	0	5	2	0	30	24	0	3
Cf2	0	1	1	0	0	1	0	0	0
Df2	2	0	5	2	0	29	24	0	3
Total	8	56	36	6	24	70	93	1	6

^a Eight modes based on conformations e1 and f1 were only occurring in R orientations of skeleton conformations A–D of **5** (1% or 2% each, altogether 13%), and are not shown. The modes representing more than 10% of bound molecules are printed in bold. All species contributing in the given mode are included. The sums of prevalences for all conformations of each orientation are shown, with the prevalent orientations (>20%) printed in bold.

for compound **22** as shown in Table 8. For compounds **14**, **15**, **21**, and **24**, the predictions are within 15–20% of experimental data. For compounds **19** and **22**, the major modes are identified. Consideration of multiple species seem to be necessary for prediction of binding orientations: the SS-MM CoMFA analysis did not pick the major binding modes for compounds **14**, **15**, and **19** exhibiting F orientation and compound **21** exhibiting R orientation as their major binding modes, although it correctly identified the orientations for compounds **22** and **24** (Table 9).

Binding Species. The binding prevalences for different species (Table 2) of each compound were calculated using an equation based on eq 5. For each species, the contributions of the twelve modes were calculated (see Supporting Information). The prevalence distribution for the bound molecules according to the multispecies, multimode model is shown in Table 10, summarized for the three binding orientations. Species 1 is practically absent for each bound compound. Species 2 is only dominant for compound **16** and then for the deaminated compounds **23–26** and **28**, which do not form species 3 and 4. Species 3 prevails in the bound molecules of compounds **4**, **7**, **12**, and **13**. The remaining 18 compounds prefer to bind as species 4.

There are no clear-cut correlations between the fractions bound in remaining three species and given orientation that would cover the entire set of compounds. Some qualitative patterns were seen for compounds bound preferably (>~50%) in given orientation and given species: F orientation is associated with species 4 (compounds **14**, **15**, **17**, and **19**); R orientation is associated with species 2 (compounds **16**, **24–27**) or species 4 (compounds **1**, **2**, **5**, **8**, **21**, **22**, and **27**); and DF orientation is associated with species 2 (compounds **23** and **28**) or species 3 (compounds **4** and **7**) or species 4 (compounds **6**, **9**, **11**, and **20**).

The published summaries of experimental observations state that the hydroxyl group contributes to binding more in the ionized form⁵⁸ and the amino group negatively affects the binding.⁵⁶ The first statements holds for **16** (**1–3**, **5**, **6**, **8**, **9**, **11**, **14**, **15**, **17–22**) of 27 compounds, for which species 4 is only marginally present in solution (Table 1) but predicted to be the most abundant (prevalences $\geq 60\%$) bound species (Table 10). The majority of these compounds (except **1**, **2**, and **22**) bind significantly (>20%) in F or DF orientations (Table 7) as species 4. These orientations maximize the electrostatic interaction of the ionized carboxyl of the ligands with the NH₃⁺ group of Lys15. This interaction is stronger in species 4 with nonionized amino group of the ligand as apposed to ionized amino group in species 1 and 2. In these cases, the prediction is consistent with the second published statement about the amino group

Table 9. Prevalences (%) of the Deep Forward, Forward, and Reversed Orientations (DF, F, and R, Respectively) from Experiments, the Multi-Species, Multi-Mode CoMFA Model and the Single-Species, Multi-Mode CoMFA Model

compounds	deep forward			forward			reverse		
	exp	MS-MM	SS-MM	exp	MS-MM	SS-MM	exp	MS-MM	SS-MM
14	30.0	29.3	81.7	70.0	61.3	0	0	9.4	18.3
15	0	1.6	33.5	100.0	89.7	0	0	8.7	66.5
19	30.0	13.1	73.5	70.0 ^a	70.5	0	0	16.5	26.5
21	0	8.6	95.4	40.0	38.2	0	60.0	53.2	4.6
22	0	1.2	14.1	40.0	3.6	0	60.0	95.2	85.9
24	0	4.6	3.2	0	2.6	36.0	100.0	92.8	60.9

^a This includes both forward and reverse modes that could not be distinguished.

Table 10. Predicted Binding Prevalences (%) for Four Species, Including Contribution for Each Orientations from Both Multi-Species, Multi-Mode CoMFA Model^a

compound	predicted binding prevalences for different species																
	species 1				species 2				species 3				species 4				
	NH ₃ ⁺ , OH, COO ⁻				NH ₃ ⁺ , O ⁻ , COO ⁻				NH ₂ , OH, COO ⁻				NH ₂ , O ⁻ , COO ⁻				
compound	DF	F	R	sum	DF	F	R	sum	DF	F	R	sum	DF	F	R	sum	
1	0	0	0	0	0	0	7	7	0	0	0	0	0	0	0	92	92
2	0	0	0	0	0	0	21	21	0	0	0	0	0	0	0	79	79
3	0	0	0	0	0	8	0	8	0	1	0	2	7	46	36	90	
4	0	0	0	0	0	0	0	0	78	0	0	78	21	0	0	21	
5	0	0	0	0	0	3	0	3	0	0	1	1	5	21	70	96	
6	0	0	0	0	0	0	0	0	4	0	1	4	89	1	6	96	
7	0	0	0	0	0	0	0	0	89	0	0	89	11	0	0	11	
8	0	0	0	0	0	0	14	14	0	0	0	0	24	5	56	86	
9	0	0	0	0	0	0	6	6	1	0	0	1	55	24	14	93	
11	0	0	0	0	0	0	4	4	1	0	0	1	62	21	12	95	
12	0	0	0	0	0	0	31	31	27	4	0	31	20	6	12	38	
13	0	0	0	0	0	0	6	6	43	0	0	44	42	5	2	50	
14	0	0	0	0	0	0	0	0	1	0	0	1	28	61	8	98	
15	0	0	0	0	0	0	1	1	0	1	0	1	2	89	8	98	
16	0	0	0	0	0	0	87	87	1	0	0	1	1	10	1	12	
17	0	0	0	0	0	0	4	4	0	0	0	1	18	76	1	95	
18	0	0	0	0	0	0	1	1	14	3	0	17	39	43	0	82	
19	0	0	0	0	0	0	2	2	1	0	0	1	13	70	14	97	
20	0	0	0	0	0	0	1	1	13	0	0	14	84	2	0	86	
21	0	0	0	0	0	0	2	2	0	0	0	0	8	38	51	98	
22	0	0	0	0	0	0	5	5	0	0	0	0	1	4	90	95	
27	0	0	0	0	0	0	49	49	0	0	0	0	0	0	51	51	

compound	species 1				species 2			
	OH, COO ⁻				O ⁻ , COO ⁻			
	DF	F	R	sum	DF	F	R	sum
23	0	0	0	0	100	0	0	100
24	0	0	0	0	4	3	93	100
25	0	1	0	1	0	40	59	99
26	0	0	0	0	5	0	95	100
28	3	0	0	3	84	4	9	97

^a The letters F, DF, and R denote forward, deep forward, and reverse orientation, respectively. Compound 27 has 4 species but without the COO⁻ group. The deamino compounds (23–26 and 28) with 2 species are listed separately. Species and orientations representing 10% and more of bound compounds are printed in bold.

negatively affects the binding.⁵⁶ For the five deamino compounds (23–26 and 28), species 2 (O⁻, COO⁻, Table 2) is the dominant species in solution and also the major bound species (almost 100%). The same behavior is predicted for compound 16.

Compounds 3–7 and 13 with alkyl substituents in position 3' and H in position 5' are present in solution mainly as species 1 (Table 1) and bind as species 3 and 4 (Table 2). Compounds 4 and 7 with s*i*-propyl and t-butyl respectively, in position 3' prefer to bind in DF orientation (Table 7) as species 3 (Table 7), with nonionized phenolic hydroxyl group possibly making H bond interactions with Ser 117 residues of inner binding pockets. These are the only compounds for which species 3 is dominant bound species. In contrast, their counterparts with i-butyl (3),

n-propyl (5), and s-butyl (6) groups in position 3', adopting several orientations, bind as species 4. Compound 27 with CH₂CH₂NH₃⁺ side chain group is predicted to bind as both species 2 and 4. Compound 13 with methyl group at 3' position prefers to bind as both species 3 and 4, whereas compound 12 with methyl groups at both 3' and 5' positions prefers to bind as species 2, 3, and 4. Compounds that are predicted to bind significantly in R orientation (1, 2, 8, 12, 16, 21, 22, 24–27), bind preferentially as species 2 and 4, with their phenolic ring pointing outside of the outer pocket where the ionized hydroxyl group forms an electrostatic interaction with Lys15.

Species Presence in Solution versus Bound State. The correlation of the species fractions in aqueous solution and

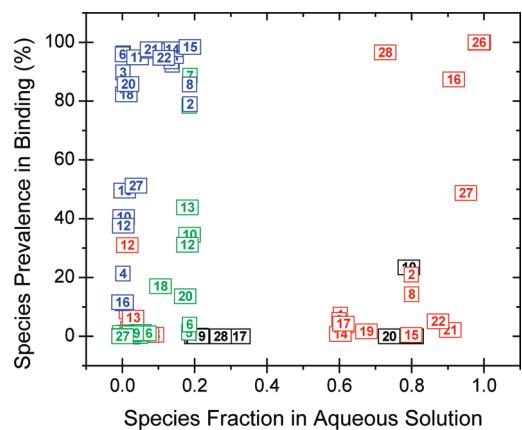


Figure 9. Correlation of bound species prevalence and species fraction in the aqueous solution for studied ligands (the numbers correspond to those in Table 1). Species 1, 2, 3, and 4 (defined in Table 2) are referred to by black, red, green and blue colors, respectively.

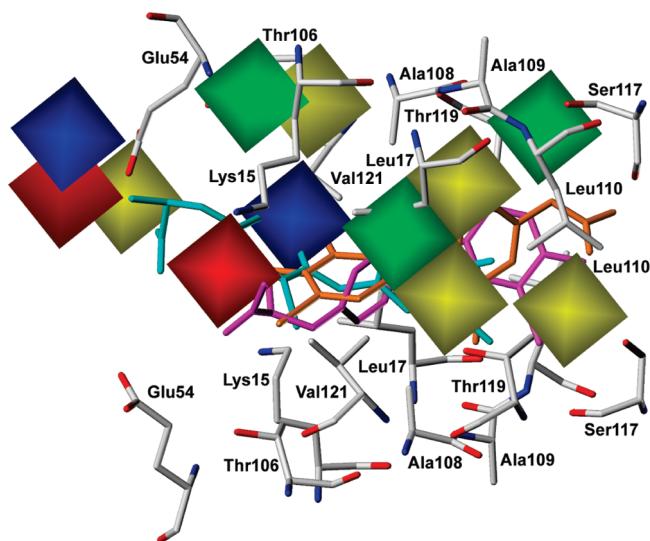


Figure 10. Contour map of the multispecies, multimode model. Green and yellow contours mark sterically favorable and unfavorable regions, respectively. In red and blue regions, positive and negative charges are favored, respectively. Key residues of the binding site are colored by atom type.

species prevalences in binding is shown in Figure 9. The differences between ionization tendencies of the free and bound molecules originate from the differences in the interaction energies of individual species. For instance, species 2 of compounds 2, 8, 15, 19, 21, and 22 (shown in the lower right corner of Figure 9) have low binding prevalences, although they are well represented in the aqueous solution. The same compounds exhibit the opposite preferences for species 4, as can be seen in the upper left corner of Figure 9. Compound 20, which exists in solution as species 1 and 3 with nonionized hydroxyl group in the phenol ring, undergoes deprotonation to preferentially bind as species 4 in DF orientation. For the descriptive and predictive ability of the model, the inclusion of multiple species and multiple modes was crucial (Table 7). Unfortunately, no experimental data (proton inventory of binding using calorimetry in buffers differing in solvation

enthalpies of protons^{92,93}) are currently available to test these interesting predictions.

Contour Maps. The steric and electrostatic contours for the MSMM CoMFA model are shown in Figure 10, along with three binding mode templates and the key residues of the receptor from the X-ray structure with PDB code 1ICT.

There are two sterically favorable green regions in the middle and inner pockets of the binding site accommodating lengthy and bulky alkyl groups occupying 3' position in compounds binding in DF modes. Three yellow steric hindrance regions define the middle and inner pockets.

The outer pocket shows a comparatively complex pattern of the red and blue regions, favoring negative and positive potentials, respectively. This pattern characterizes the locations of flexible Glu54 and Lys15 residues in both chains of both carboxyl and amino groups in the outer pocket while binding in F and DF modes, and the presence of deprotonated phenolic hydroxyl group while binding in R mode.

The receptor site map for the multimode CoMFA model (not shown) bears some similarities to the multispecies, multimode model: steric hindrance regions in similar locations in the middle and inner pockets; and a contour favoring positive and negative charges in the entrance of the pocket. The contours for the standard CoMFA model (not shown) are rather imbalanced: they predict regions favoring positive and negative charges around the inner pocket as well as a sterically favorable region in the middle pocket.

CONCLUSIONS

The multispecies, multimode (MS-MM) CoMFA model was rigorously developed and used to describe structural determinants of the binding of thyroxine analogs to transthyretin. The superposition was based on three experimentally determined binding orientations of the thyroxine ligands. Compared to the MM and standard CoMFA analyses, the MS-MM CoMFA model has shown a better predictive ability. The predicted prevalencies of individual binding modes are in good agreement with published experimental X-ray studies documenting multiple modes. The created receptor site models are consistent with structures of the binding site and of the ligands, and provide clues for elucidation of structural determinants of binding affinities. Consideration of multiple modes allows examination of different binding conformations of alkyls and other flexible substituents. Inclusion of multiple modes and multiple species was crucial for obtaining models with good predictive ability that was documented even for standard CoMFA models with the species and modes selected by the MS-MM procedure. If further analyses confirm the applicability, the MS-MM CoMFA model has a potential to become an important tool for prediction of binding affinities in the absence of experimental data on the receptor structure. Such a model is long overdue because the majority of physiological compounds and man-made bioactive chemicals ionize under physiological conditions. Ionized molecules are more soluble in water, exhibit more selective transport and tend to bind less to nontarget proteins and membranes, as compared to nonionized molecules. These traits make them better candidates for specific "bullets" targeting selected functions in the cell. Conceptual structure-based and ligand-based models for prediction of their binding affinities are prerequisites for the utilization of this potential.

■ ASSOCIATED CONTENT

S Supporting Information. Additional material as described in the text (the prevalences of all modes and species). This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*Address: Albany College of Pharmacy and Health Sciences, Vermont Campus, Department of Pharmaceutical Sciences, 261 Mountain View Drive, Colchester, VT 05446. E-mail: stefan.balaz@acphs.edu. Phone: 802-735-2615.

■ ACKNOWLEDGMENT

This work was supported in part by the NIH NIGMS grant R01 GM80508, NIH NCRR grants 1 P20 RR 15566 and 1 P20 RR 16471, and the NSF EPSCoR EPS-0814442 program.

■ REFERENCES

- (1) Hilal, S. H.; Bornander, L. L.; Carreira, L. A. Hydration equilibrium constants of aldehydes, ketones and quinazolines. *QSAR Comb. Sci.* **2005**, *24*, 631–637.
- (2) Davies, N. M. Chiral inversion. *Chirality Drug Des. Dev.* **2004**, *351*–392.
- (3) Mattos, C.; Ringe, D. Multiple binding modes. In *3D QSAR in Drug Design: Theory, Methods, and Applications*; Kubinyi, H., Ed.; Escom: Leiden, the Netherlands, 1993, 226–254.
- (4) Case, D. A.; Karplus, M. Dynamics of ligand binding to heme protein. *J. Mol. Biol.* **1979**, *132*, 343–368.
- (5) Nelson, R. D.; Gottlieb, D. I.; Balasubramanian, T. M.; Marshall, G. R. Opioid peptides: analysis of specificity and multiple binding modes through computer-aided drug design and structure-activity studies. *NIDA Res. Monogr.* **1986**, *69*, 204–230.
- (6) Dean, P. M. Molecular similarity. In *3D QSAR in Drug Design: Theory, Methods, and Applications*; Kubinyi, H., Ed.; Escom: Leiden, the Netherlands, 1993, 150–172.
- (7) Martin, Y. C.; Hackbarth, J. J. Theoretical model-based equations for the linear free energy relationships of the biological activity of ionizable substances. 1. Equilibrium-controlled potency. *J. Med. Chem.* **1976**, *19*, 1033–1039.
- (8) Kubinyi, H. Opinion: Drug research: myths, hype and reality. *Nat. Rev. Drug Discovery* **2003**, *2*, 665–668.
- (9) Marshall, G. R.; Barry, C. D.; Bosshard, H. E.; Dammkoehler, R. A.; and Dunn, D. A. The conformational parameter in drug design: The active analog approach. In *Computer-Assisted Drug Design*; Olson, E. C., Christoffersen, R. E., Eds.; ACS: Washington, DC, 1979; pp 205–226.
- (10) Crippen, G. M. Distance geometry approach to rationalizing binding data. *J. Med. Chem.* **1979**, *22*, 988–997.
- (11) Crippen, G. M. Quantitative structure–activity relationships by distance geometry: Systematic analysis of dihydrofolate reductase inhibitors. *J. Med. Chem.* **1980**, *23*, 599–606.
- (12) Crippen, G. M. Quantitative structure–activity relationships by distance geometry: Thyroxine binding site. *J. Med. Chem.* **1981**, *24*, 198–203.
- (13) Ghose, A. K.; Crippen, G. M. Quantitative structure–activity relationship by distance geometry: Quinazolines as dihydrofolate reductase inhibitors. *J. Med. Chem.* **1982**, *25*, 892–899.
- (14) Boulu, L. G.; Crippen, G. M. Voronoi binding site models: Calculation of binding modes and influence of drug binding data accuracy. *J. Comput. Chem.* **1989**, *10*, 673–682.
- (15) Ciubotariu, D.; Deretey, E.; Oprea, T. I.; Sulea, T.; Simon, Z.; Kurunczi, L.; Chiriac, A. Multiconformational Minimal Steric Difference. Structure-acetylcholinesterase hydrolysis rates relations for acetic acid esters. *Quant. Struct.–Act. Relat.* **1993**, *12*, 367–372.
- (16) Sulea, T.; Kurunczi, L.; Simon, Z. Dioxin-type activity for polyhalogenated aryllic derivatives: A QSAR model based on MTD-method. *SAR QSAR Environ. Sci.* **1995**, *3*, 37–61.
- (17) Sulea, T.; Kurunczi, L.; Oprea, T. I.; Simon, Z. MTD-ADJ: A multiconformational minimal topologic difference for determining bioactive conformers using adjusted biological activities. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 133–146.
- (18) Cramer, R. D. I.; Patterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- (19) Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.
- (20) Kim, K. H.; Martin, Y. C. Direct prediction of linear free energy substituent effects from 3D structures using comparative molecular field analysis. 1. Electronic effects of substituted benzoic acids. *J. Org. Chem.* **1991**, *56*, 2723–2729.
- (21) Nicklaus, M. C.; Milne, G. W.; Burke-TR, J. QSAR of conformationally flexible molecules: Comparative molecular field analysis of protein-tyrosine kinase inhibitors. *J. Comput.-Aid. Mol. Des.* **1992**, *6*, 487–504.
- (22) Cramer, R. D. I.; Wold, S. B. Comparative Molecular Field Analysis (CoMFA). U.S. Patent 5,307,287 1994.
- (23) Lukacova, V.; Balaz, S. Multimode ligand binding in receptor site modeling: Implementation in CoMFA. *J. Chem. Inf. Comp. Sci.* **2003**, *43*, 2093–2105.
- (24) Hopfinger, A. J.; Burke, B. J.; Dunn, W. J. I. A generalized formalism of three-dimensional quantitative structure–property relationship analysis for flexible molecules using tensor representation. *J. Med. Chem.* **1994**, *37*, 3768–3774.
- (25) Burke, B. J.; Dunn, W. J.; Hopfinger, A. J. Construction of a molecular shape analysis–three-dimensional quantitative structure–activity relationship for an analog series of pyridobenzodiazepinone inhibitors of muscarinic 2 and 3 receptors. *J. Med. Chem.* **1994**, *37*, 3775–3788.
- (26) Dunn, W. J.; Hopfinger, A. J.; Catana, C.; Duraiswami, C. Solution of the conformation and alignment tensors for the binding of trimethoprim and its analogs to dihydrofolate reductase: 3D-quantitative structure–activity relationship study using molecular shape analysis, 3-way partial least-squares regression, and 3-way factor analysis. *J. Med. Chem.* **1996**, *39*, 4825–4832.
- (27) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B. Q.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.* **1997**, *119*, 10509–10524.
- (28) Albuquerque, M. G.; Hopfinger, A. J.; Barreiro, E. J.; deAlencastro, R. B. Four-dimensional quantitative structure–activity relationship analysis of a series of interphenylene 7-oxabicycloheptane oxazole thromboxane A(2) receptor antagonists. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 925–938.
- (29) Hopfinger, A. J.; Reaka, A.; Venkatarangan, P.; Duca, J. S.; Wang, S. Construction of a virtual high throughput screen by 4D-QSAR analysis: Application to a combinatorial library of glucose inhibitors of glycogen phosphorylase b. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1151–1160.
- (30) Venkatarangan, P.; Hopfinger, A. J. Prediction of ligand–receptor binding free energy by 4D-QSAR analysis: Application to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 1141–1150.
- (31) Liu, J.; Pan, D.; Tseng, Y.; Hopfinger, A. J. 4D-QSAR analysis of a series of antifungal P450 inhibitors and 3D-pharmacophore comparisons as a function of alignment. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 2170–2179.
- (32) Pan, D.; Liu, J.; Senese, C.; Hopfinger, A. J.; Tseng, Y. Characterization of a ligand–receptor binding event using receptor-dependent four-dimensional quantitative structure–activity relationship analysis. *J. Med. Chem.* **2004**, *47*, 3075–3088.

- (33) Marshall, G. R. Molecular modeling in drug design. In *Burger's Medicinal Chemistry*; Wolff, M. E., Ed.; John Wiley & Sons: New York, 1995; pp 573–659.
- (34) Lewi, P. J.; de Jonge, M.; Daeyaert, F.; Koymans, L.; Vinkers, M.; Heeres, J.; Janssen, P. A. J.; Arnold, E.; Das, K.; Clark, A. D., Jr.; Hughes, S. H.; Boyer, P. B.; de Bethune, M. P.; Pauwels, R.; Andries, K.; Kukla, M.; Ludovici, D.; de Corte, B.; Kavash, R.; Ho, C. On the detection of multiple-binding modes of ligands to proteins, from biological, structural, and modeling data. *J. Comput.-Aid. Mol. Des.* **2003**, *17*, 129–134.
- (35) Balaz, S.; Hornak, V.; Haluska, L. Receptor mapping with multiple binding modes: Binding site of PCB-degrading dioxygenase. *Chemometr. Intell. Lab.* **1994**, *24*, 185–191.
- (36) Hornak, V.; Balaz, S.; Schaper, K. J.; Seydel, J. K. Multiple binding modes in 3D-QSAR: Microbial degradation of polychlorinated biphenyls. *Quant. Struct.–Act. Relat.* **1998**, *17*, 427–436.
- (37) Vedani, A.; McMasters, D. R.; Dobler, M. Multi-conformational ligand representation in 4D-QSAR: Reducing the bias associated with ligand alignment. *Quant. Struct.–Act. Relat.* **2000**, *19*, 149–161.
- (38) Aqvist, J.; Medina, C.; Samuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.
- (39) Carlson, H. A.; Jorgensen, W. L. An extended linear response method for determining free energies of hydration. *J. Phys. Chem.* **1995**, *99*, 10667–10673.
- (40) Khandelwal, A.; Lukacova, V.; Kroll, D. M.; Comez, D.; Raha, S.; Balaz, S. Simulation-based predictions of binding affinities of matrix metalloproteinase inhibitors. *QSAR Comb. Sci.* **2004**, *23*, 754–766.
- (41) Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959–3969.
- (42) Tokarski, J. S.; Hopfinger, A. J. Prediction of ligand-receptor binding thermodynamics by free energy force field (FEFF) 3D-QSAR analysis: Application to a set of peptidomeric renin inhibitors. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 792–811.
- (43) Venkatarangan, P.; Hopfinger, A. J. Prediction of ligand-receptor binding thermodynamics by free energy force field three-dimensional quantitative structure-activity relationship analysis: Applications to a set of glucose analogue inhibitors of glycogen phosphorylase. *J. Med. Chem.* **1999**, *42*, 2169–2179.
- (44) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.
- (45) Huang, D.; Caflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.* **2004**, *47*, 5791–5797.
- (46) Lill, M. A.; Vedani, A.; Dobler, M. Raptor: Combining dual-shell representation, induced-fit simulation, and hydrophobicity scoring in receptor modeling: Application toward the simulation of structurally diverse ligand sets. *J. Med. Chem.* **2004**, *47*, 6174–6186.
- (47) Lill, M. A.; Dobler, M.; Vedani, A. In silico prediction of receptor-mediated environmental toxic phenomena—application to endocrine disruption. *SAR QSAR Environ. Res.* **2005**, *16*, 149–169.
- (48) Lill, M. A.; Winiger, F.; Vedani, A.; Ernst, B. Impact of induced fit on ligand binding to the androgen receptor: A multidimensional QSAR study to predict endocrine-disrupting effects of environmental chemicals. *J. Med. Chem.* **2005**, *48*, 5666–5674.
- (49) Lill, M. A.; Dobler, M.; Vedani, A. Prediction of small-molecule binding to cytochrome P450 3A4: Flexible docking combined with multidimensional QSAR. *ChemMedChem* **2006**, *1*, 73–81.
- (50) Khandelwal, A.; Lukacova, V.; Comez, D.; Kroll, D. M.; Raha, S.; Balaz, S. A combination of docking, QM/MM Methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* **2005**, *48*, 5437–5447.
- (51) Khandelwal, A.; Balaz, S. QM/MM linear response method distinguishes ligand affinities for closely related metalloproteins. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 326–339.
- (52) Colon, W.; Kelly, J. W. Partial denaturation of transthyretin is sufficient for amyloid fibril formation in vitro. *Biochemistry* **1992**, *31*, 8654–8660.
- (53) Uemichi, T.; Murrell, J. R.; Zeldenrust, S.; Benson, M. D. A new mutant transthyretin (Arg 10) associated with familial amyloid polyneuropathy. *J. Med. Genet.* **1992**, *29*, 888–891.
- (54) Baures, P. W.; Oza, V. B.; Peterson, S. A.; Kelly, J. W. Syntesis and evaluation of inhibitors of transthyretin amyloid formation based on the non-steroidal anti-inflammatory drug, flufenamic acid. *Bioorg. Med. Chem.* **1999**, *7*, 1339–1347.
- (55) Oza, V. B.; Petrassi, H. M.; Purkey, H. E.; Kelly, J. W. Synthesis and evaluation of anthranilic acid-based transthyretin amyloid fibril inhibitors. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1–6.
- (56) Andrea, T. A.; Cavalieri, R. R.; Goldfine, I. D.; Jorgensen, E. C. Binding of thyroid hormones and analogues to the human plasma protein prealbumin. *Biochemistry* **1980**, *19*, 55–63.
- (57) De La Paz, P.; Burridge, J. M.; Oatley, S. J.; Blake, C. C. F. Multiple modes of binding of thyroid hormones and other iodothyronines to human plasma transthyretin. In *The Design of Drugs to Macromolecular Targets*; Beddell, C. R., Ed.; John Wiley and Sons: Chichester, U.K., 1992, 119–172.
- (58) Nilsson, S. F.; Peterson, P. A. Evidence for multiple thyroxine-binding sites in human prealbumin. *J. Biol. Chem.* **1971**, *246*, 6098–6105.
- (59) Hilal, S. H.; Karickhoff, S. W.; Carreira, L. A. Estimation of microscopic, zwitterionic ionization constants, isoelectric point and molecular speciation of organic compounds. *Talanta* **1999**, *50*, 827–840.
- (60) Wojtczak, A.; Cody, V.; Luft, J. R.; Pangborn, W. Structures of human transthyretin complexed with thyroxine at 2.0 angstrom resolution and 3',5' dinitron acetyl L thyronine at 2.2 angstrom resolution. *Acta Crystallogr. D.* **1996**, *52* (Part 4), 758–765.
- (61) Neumann, P.; Cody, V.; Wojtczak, A. Ligand binding at the transthyretin dimer-dimer interface: Structure of the transthyretin-T4Ac complex at 2.2 angstrom resolution. *Acta Crystallogr. D.* **2005**, *D61*, 1313–1319.
- (62) Wojtczak, A.; Luft, J.; Cody, V. Mechanism of molecular recognition. Structural aspects of 3,3'- diiodo-L-thyronine binding to human serum transthyretin. *J. Biol. Chem.* **1992**, *267*, 353–357.
- (63) Sybyl 8.1; Tripos Inc.: St. Louis, Missouri, 2009.
- (64) Jaguar; Schrödinger LLC: Portland, OR, 2003.
- (65) Blaney, J. M.; Jorgensen, E. C.; Connolly, M. L.; Ferrin, T. E.; Langridge, R.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. Computer graphics in drug design: Molecular modeling of thyroid hormone-prealbumin interactions. *J. Med. Chem.* **1982**, *25*, 785–790.
- (66) Blaney, J. M.; Weiner, P. K.; Dearing, A.; Kollman, P. A.; Jorgensen, E. C.; Oatley, S. J.; Burridge, J. M.; Blake, C. C. F. Molecular mechanics simulation of protein-ligand interactions: Binding of thyroid hormone analogues to prealbumin. *J. Am. Chem. Soc.* **1982**, *104*, 6424–6434.
- (67) Somack, R.; Andrea, T. A.; Jorgensen, E. C. Thyroid hormone binding to human serum prealbumin and rat liver nuclear receptor: Kinetics, contribution of the hormone phenolic hydroxyl group, and accommodation of hormone side-chain bulk. *Biochemistry* **1982**, *21*, 163–170.
- (68) Pitt-Rivers, R.; Stanbury, J. B.; Rapp, B. Conversion of thyroxine to 3,5,5'-triiodothyronine in vivo. *J. Clin. Endocr. Metab.* **1955**, *15*, 616–620.
- (69) Baxter, J. D.; Eberhardt, N. L.; Apriletti, J. W.; Johnson, L. K.; Ivarie, R. D.; Morris, J. A.; Seeburg, P. H.; Goodman, H. M. Thyroid hormone receptors and responses. *Recent Prog. Horm. Res.* **1979**, *35*, 97–153.
- (70) Cody, V. Role of iodine in thyroid hormones: molecular conformation of a halogen-free hormone analogue. *J. Med. Chem.* **1980**, *23*, 584–587.
- (71) Bartalena, L.; Robbins, J. Thyroid hormone transport proteins. *Clin. Lab. Med.* **1993**, *13*, 583–598.

- (72) Kanda, Y.; Goodman, D. S.; Canfield, R. E.; Morgan, F. J. Amino acid sequence of human plasma prealbumin. *J. Biol. Chem.* **1974**, *249*, 6796–6805.
- (73) Balke, C. C. F.; Swan, I. D. A.; Rerat, C.; Berthou, J.; Laurent, A.; Rerat, B. X-ray study of the subunit structure of prealbumin. *J. Mol. Biol.* **1971**, *61*, 217–224.
- (74) Cheng, S. Y.; Pages, R. A.; Saroff, H. A.; Edelhoch, H.; Robbins, J. Analysis of thyroid hormone binding to human serum prealbumin by 8-anilinonaphthalene-1-sulfonate fluorescence. *Biochemistry* **1977**, *16*, 3707–3713.
- (75) Cheng, S. Y.; Wilchek, M.; Cahnmann, H. J.; Robbins, J. Affinity labeling of human serum prealbumin with N-bromoacetyl-L-thyroxine. *J. Biol. Chem.* **1977**, *252*, 6076–6081.
- (76) Ferguson, R. N.; Edelhoch, H.; Saroff, H. A.; Robbins, J.; Cahnmann, H. J. Negative cooperativity in the binding of thyroxine to human serum prealbumin. *Biochemistry* **1975**, *14*, 282–289.
- (77) Miroy, G. J.; Lai, Z.; Lashuel, H. A.; Peterson, S. A.; Strang, C.; Kelly, J. W. Inhibiting transthyretin amyloid fibril formation via protein stabilization. *P. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 15051–15056.
- (78) Klabunde, T.; Petrassi, H. M.; Oza, V. B.; Raman, P.; Kelly, F. W.; Sacchettini, J. C. Rational design of potent human transthyretin amyloid disease inhibitors. *Nat. Struct. Biol.* **2000**, *7*, 312–320.
- (79) Petrassi, H. M.; Klabunde, T.; Sacchettini, J.; Kelly, J. W. Structure-based design of N-phenyl phenoxazine transthyretin amyloid fibril inhibitors. *J. Am. Chem. Soc.* **2000**, *122*, 2178–2192.
- (80) Green, N. S.; Foss, T. R.; Kelly, J. W. Genistein, a natural product from soy, is a potent inhibitor of transthyretin amyloidosis. *P. Natl. Acad. Sci. USA* **2005**, *102*, 14545–14550.
- (81) Wojtczak, A.; Neumann, P.; Cody, V. Structure of a new polymorphic monoclinic form of human transthyretin at 3 Å resolution reveals a mixed complex between unliganded and T4-bound tetramers of TTR. *Acta Crystallogr. D* **2001**, *D57*, 957–967.
- (82) Wojtczak, A.; Luft, J.; Cody, V. Mechanism of molecular recognition. Structural aspects of 3,3'-diiodo-L-thyronine binding to human serum transthyretin. *J. Biol. Chem.* **1992**, *267*, 353–357.
- (83) Wojtczak, A.; Cody, V.; Luft, J. R.; Pangborn, W. Structures of human transthyretin complexed with thyroxine at 2.0 Å resolution and 3',5'-dinitro-N-acetyl-L-thyronine at 2.2 Å resolution. *Acta Crystallogr. D* **1996**, *52*, 758–765.
- (84) Clare, B. W. QSAR of benzene derivatives: comparison of classical descriptors, quantum theoretic parameters and flip regression, exemplified by phenylalkylamine hallucinogens. *J. Comput.-Aid. Mol. Des.* **2002**, *16*, 611–633.
- (85) Supuran, C. T.; Clare, B. W. Quantum theoretic QSAR of benzene derivatives: Some enzyme inhibitors. *J. Enzyme Inhib. Med. Chem.* **2004**, *19*, 237–248.
- (86) Clare, B. W.; Supuran, C. T. A physically interpretable quantum-theoretic QSAR for some carbonic anhydrase inhibitors with diverse aromatic rings, obtained by a new QSAR procedure. *Bioorgan. Med. Chem.* **2005**, *13*, 2197–2211.
- (87) Clare, B. W.; Supuran, C. T. Predictive flip regression: A technique for QSAR of derivatives of symmetric molecules. *J. Chem. Inf. Model.* **2005**, *45*, 1385–1391.
- (88) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499–2510.
- (89) Wang, J.; Szewczuk, Z.; Yue, S. Y.; Tsuda, Y.; Konishi, Y.; Purisima, E. O. 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 hirudin55–6S. *J. Mol. Biol.* **1995**, *253*, 473–492.
- (90) Jullien, L.; Proust, A.; LeMenn, J. C. How does the Gibbs free energy evolve in a system undergoing coupled competitive reactions? *J. Chem. Educ.* **1998**, *75*, 194–199.
- (91) Smith, W. R. Missen, R. W. *Chemical Reaction Equilibrium Analysis: Theory and Algorithms*. New York: John Wiley and Sons, 1982; pp 1–364.
- (92) Parker, M. H.; Lunney, E. A.; Ortwine, D. F.; Pavlovsky, A. G.; Humblet, C.; Brouillette, C. G. Analysis of the binding of hydroxamic acid and carboxylic acid inhibitors to the stromelysin-1 (matrix metalloproteinase-3) catalytic domain by isothermal titration calorimetry. *Biochemistry* **1999**, *38*, 13592–13601.
- (93) Venkatasubban, K. S.; Schowen, R. L. The proton inventory technique. *CRC Crit. Rev. Biochem. Mol.* **1984**, *17*, 1–44.