Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10[†]

Jan Kmuníček,[‡] Santos Luengo,[§] Federico Gago,[§] Angel Ramirez Ortiz,[||] Rebecca C. Wade,[⊥] and Jiří Damborský*[,‡]

National Centre for Biomolecular Research, Masaryk University, Kotlarska 2, 611 37 Brno, Czech Republic, Department of Pharmacology, University of Alcala, 28871 Alcalá de Henares, Madrid, Spain, Department of Physiology and Biophysics, Mount Sinai School of Medicine, One Gustave Levy Place, Box 1218, New York, New York 10029, and European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Received March 7, 2001; Revised Manuscript Received May 21, 2001

ABSTRACT: Comparative binding energy (COMBINE) analysis was conducted for 18 substrates of the haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10 (DhlA): 1-chlorobutane, 1-chlorohexane, dichloromethane, 1,2-dichloroethane, 1,2-dichloropropane, 2-chloroethanol, epichlorohydrine, 2-chloroacetonitrile, 2-chloroacetamide, and their brominated analogues. The purpose of the COMBINE analysis was to identify the amino acid residues determining the substrate specificity of the haloalkane dehalogenase. This knowledge is essential for the tailoring of this enzyme for biotechnological applications. Complexes of the enzyme with these substrates were modeled and then refined by molecular mechanics energy minimization. The intermolecular enzyme−substrate energy was decomposed into residue-wise van der Waals and electrostatic contributions and complemented by surface area dependent and electrostatic desolvation terms. Partial least-squares projection to latent structures analysis was then used to establish relationships between the energy contributions and the experimental apparent dissociation constants. A model containing van der Waals and electrostatic intermolecular interaction energy contributions calculated using the AMBER force field explained 91% (73% cross-validated) of the quantitative variance in the apparent dissociation constants. A model based on van der Waals intermolecular contributions from AMBER and electrostatic interactions derived from the Poisson−Boltzmann equation explained 93% (74% cross-validated) of the quantitative variance. COMBINE models predicted correctly the change in apparent dissociation constants upon single-point mutation of DhlA for six enzyme−substrate complexes. The amino acid residues contributing most significantly to the substrate specificity of DhlA were identified; they include Asp124, Trp125, Phe164, Phe172, Trp175, Phe222, Pro223, and Leu263. These residues are suitable targets for modification by site-directed mutagenesis.

Haloalkane dehalogenases are microbial enzymes that catalyze dehalogenation reactions (*1*−*3*), which are important for the degradation of environmental pollutants (*4*−*6*). Halogenated aliphatic compounds are among the most frequently occurring pollutants. Large quantities of these compounds are widely used as pesticides, solvents, fire retardants, hydraulic and heat transfer fluids, and cleaning agents. They are environmentally dangerous and are hazardous to humans due their toxic, genotoxic, teratogenic, and irritating effects. Unfortunately, wild type enzymes often do not acquire sufficiently high activity or specificity for degradation of environmental pollutants. Protein design can be used to improve the catalytic properties of such enzymes. To tailor the enzyme for improved substrate specificity, the amino acid residues that participate in substrate binding must be identified so that they can be modified by site-directed mutagenesis. Comparative binding energy (COMBINE)[1] analysis has been shown to be a useful technique for deriving quantitative structure−activity relationships from a set of three-dimensional structures of enzyme−ligand complexes (*7*−*12*). Here, we use COMBINE analysis to derive a predictive model for substrate binding specificity in which

[†] Supported by NATO Linkage Grant MTECH.LG.974701.

* To whom correspondence should be addressed. Telephone: 420-5-41129377. Fax: 420-5-41129506. E-mail: jiri@chemi.muni.cz.

[‡] Masaryk University.

[§] University of Alcala.

[||] Mount Sinai School of Medicine.

[⊥] European Molecular Biology Laboratory. Current address: European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany.

[1] Abbreviations: BUW, block unscaled weights; COMBINE, comparative binding energy; DhlA, haloalkane dehalogenase from *X. autotrophicus* GJ10; E_{ele}^{ES}, enzyme−substrate electrostatic interaction energy in the presence of the surrounding solvent derived from the Poisson−Boltzmann equation; $E_{desolv-sur}$, surface area dependent term of the desolvation energy; E_{desolv}^{S}, desolvation energy of a substrate; E_{desolv}^{E}, desolvation energy of an enzyme; FFD, fractional factorial design; ΔG_{ele}, overall electrostatic free energy change upon binding; G_{ele}^{ES}, electrostatic energy for all atoms in the enzyme−substrate complex; G_{ele}^{S}, electrostatic energy for substrate atoms; G_{ele}^{E}, electrostatic energy for enzyme atoms; ΔG_{desolv}^{S}, change in desolvation energy of the substrate upon binding; ΔG_{desolv}^{E}, change in desolvation energy of the enzyme upon binding; ΔH_{vap}, enthalpy of vaporization; $INTC_{Q^2}$, intercept of the permutation plot for Q^2; K_m, enzyme−substrate dissociation constant; PLS, partial least-squares; Q^2, cross-validated correlation coefficient; R^2, correlation coefficient; SDEC, standard deviation of error of calculation for the working or training set; $SDEP_{int}$, standard deviation of error of (internal) predictions; $SDEP_{ext}$, standard deviation of error of (external) predictions; UV, scaling to unit variance; ΔU, total binding energy.

10.1021/bi010464p CCC: $20.00
Published on Web 07/06/2001

important interactions for binding are highlighted so that the model can be used to guide mutagenesis experiments to modify the enzyme's substrate specificity.

The haloalkane dehalogenase isolated from the soil bacterium *Xanthobacter autotrophicus* GJ10 (DhlA) is a soluble globular enzyme (*13*). DhlA is composed of 310 residues and has a molecular mass of ∼36 kDa. It consists of two different domains: the α/β-fold domain (main domain) which is conserved for all α/β-hydrolases (*14*, *15*) and the so-called cap domain. The main domain is composed of eight β-sheets surrounded by six α-helices, whereas the cap domain is composed of five additional α-helices. The active site of the enzyme is located between these two domains in an internal, predominantly hydrophobic cavity that can be reached from the solvent through a tunnel. The catalytic residues form a catalytic triad (nucleophile, base, and acid) that is highly conserved among all of the α/β-hydrolases known to date. The mechanism of dehalogenation is hydrolytic and requires the substrate and a water molecule in the active site; no other cofactor is necessary. During the hydrolytic dehalogenation, a carbon−halogen bond in the substrate is cleaved and the corresponding alcohol is formed. Details of the reaction mechanism have been investigated by crystallography (*16*−*18*), kinetic measurements (*19*−*23*), site-directed mutagenesis (*24*−*31*), and molecular modeling (*32*−*38*).

Previous theoretical studies were focused on the reaction mechanism of haloalkane dehalogenases (quantum mechanical calculations) and their conformational behavior (molecular dynamic simulations). The study presented here, on the other hand, deals with the substrate specificity of DhlA, and its aim is to construct a predictive model for estimation of the binding affinities for mutant proteins. To this end, a COMBINE analysis was carried out to identify the protein residues responsible for the differences in binding affinities of 18 chlorinated and brominated aliphatic substrates of DhlA. The effects of different scaling and variable selection procedures on the quality of the models were studied. The best model explained 93% (74% cross-validated) of the quantitative variance in binding constants and enabled identification of the residues that contribute most to the binding specificity; these are candidates for site-directed mutagenesis aimed at altering the substrate specificity of DhlA.

METHODS

Experimental Data. Apparent dissociation constants (K_m) were used as the measure of binding affinities for a set of 18 substrates. The binding affinities of these compounds vary over 4 orders of magnitude. The K_m values determined by Schanstra et al. (*39*) were logarithmically transformed (Table 1). Experimental activities were measured using steady-state kinetic analysis with purified DhlA. The K_m values for dichloromethane, 2-chloroethanol, and 2-chloroacetamide were fixed at the highest measured concentrations since the exact dissociation constants were not reported (*39*).

Overview of COMBINE Analysis. Binding energies are calculated for the set of enzyme−substrate complexes using a molecular mechanics force field. The total binding energy, ΔU, may be assumed to be given by the sum of five terms: (i) the sum of intermolecular interaction energies (Δu_i) between the substrate and each enzyme residue, $E_{\text{inter}}{}^{\text{ES}}$, (ii) the change in the intramolecular energy of the substrate upon binding to the enzyme, ΔE^{S}, (iii) the change in the intramolecular energy of the enzyme upon substrate binding, ΔE^{E}, (iv) the desolvation energy of a substrate, $E_{\text{desolv}}{}^{\text{S}}$, and (v) the desolvation energy of the enzyme, $E_{\text{desolv}}{}^{\text{E}}$.

Table 1: Steady-State Dissociation Constants of Haloalkane Dehalogenase[a]

	compound	log K_m (mM)		compound	log K_m (mM)
1	1-chlorobutane	0.34	**10**	1,2-dibromopropane	0.11
2	1-chlorohexane	0.15	**11**	2-chloroethanol	2.60
3	1-bromobutane	−1.22	**12**	2-bromoethanol	1.04
4	1-bromohexane	−0.52	**13**	epichlorohydrine	1.68
5	dichloromethane	2.00	**14**	epibromohydrine	0.34
6	1,2-dichloroethane	−0.28	**15**	2-chloroacetonitrile	0.80
7	dibromomethane	0.38	**16**	2-bromoacetonitrile	−0.31
8	1,2-dibromoethane	−2.00	**17**	2-chloroacetamide	2.00
9	1,2-dichloropropane	1.11	**18**	2-bromoacetamide	1.30

[a] From ref *39*.

$$\Delta U = E_{\text{inter}}{}^{\text{ES}} + \Delta E^{\text{S}} + \Delta E^{\text{E}} + E_{\text{desolv}}{}^{\text{S}} + E_{\text{desolv}}{}^{\text{E}} \quad (1)$$

The second and third terms, describing changes in intramolecular energies upon binding, were neglected in the study presented here because the DhlA substrates are rather small molecules and there is no evidence for large differences in the structure of DhlA when different substrates are bound. Intermolecular energy contributions were decomposed into van der Waals and electrostatic interactions.

In the first step of COMBINE analysis, a set of structures of enzyme−substrate complexes is prepared and the total binding energy is calculated for each of these complexes. The following step is the decomposition of the enzyme−substrate interaction energy on a per residue basis for each of the complexes. A matrix is then constructed in which the rows represent the different compounds studied and the columns contain the residue-based energy information, which is separated into two blocks (van der Waals and electrostatic), plus an additional column containing the experimental binding affinities. Further columns can contain additional energy terms such as the substrate desolvation energy terms. This matrix is then projected onto a small number of orthogonal "latent variables" using partial least-squares (PLS) analysis (*40*, *41*), and the original energy terms are given weights, w_i, according to their importance in the model, in the form of PLS pseudocoefficients. The higher these coefficients are, the more significant they are for explaining the variance in the experimental data. Thus, in the simplest form, the COMBINE model for binding affinity, ΔG, is of the following form (C is a constant term):

$$\Delta G = \sum w_i \Delta u_i + C \quad (2)$$

Parametrization of Halogenated Substrates. The all-atom AMBER molecular mechanics force field (*42*) was used throughout, and consistent parameters for the haloalkanes were derived to describe the bonded and nonbonded interactions. For each molecule, molecular electrostatic potentials (MEPs) were calculated from the corresponding ab initio wave functions (RHF MP2//6-31G*) using Gaussian94 (*43*) following full energy minimization. Partial atomic charges were then derived by fitting each MEP to a monopole–monopole expression using the RESP methodology (*44*, *45*). One conformation of each molecule (trans) was employed in the fit, except for 1,2-dichoroethane for which both the gauche and trans conformations were considered. Atom types for carbon atoms in the haloalkanes (CT) were taken from the AMBER database. Equilibrium bond lengths and angles for chlorinated and brominated hydrocarbons were obtained by averaging equivalent terms from the ab initio 6-31G(d) energy-minimized structures (Table 1). Dihedral parameters involving halogens were adjusted so as to reproduce in the molecular mechanics force field the torsional barriers calculated ab initio. For this purpose, the SPASMS module in AMBER (*46*) was employed. Nonbonded parameters for halogen atoms were developed and tested following a previously reported procedure (*47*) with some modifications. In brief, periodic cubic boxes (27 Å × 27 Å × 27 Å) containing 149 solvent molecules of 1,2-dichloroethane, bromoethane, and acetonitrile were constructed to reproduce the density and enthalpy of vaporization (ΔH_{vap}) of these liquids at 300 K. The compressibility values (in 10^{-6} bar^{-1}) that were used were 84.6, 142.3, and 107.0, respectively (*48*). Molecular dynamics simulations were carried out at 300 K using the SANDER module in AMBER. Both the temperature and the pressure were coupled to thermal and pressure baths with relaxation times of 0.2 and 0.6 ps, respectively. In a 20 ps heating phase, the temperature was gradually increased under constant-volume conditions, and the velocities were reassigned at each new temperature according to a Maxwell–Boltzmann distribution. This was followed by an equilibration phase of 200 ps at 300 K, and by a 300 ps sampling period at constant pressure during which system coordinates were saved every 50 ps. All bonds involving hydrogens were constrained to their equilibrium values by means of the SHAKE algorithm (*49*), which allowed an integration time step of 2 fs to be used. A nonbonded cutoff of 10 Å was employed, and the lists of nonbonded pairs were updated every 25 steps. Density values were provided directly by the SANDER module. ΔH_{vap} values were calculated according to the equation

$$\Delta H_{\mathrm{vap}} = RT - E_{\mathrm{inter}} \tag{3}$$

where E_{inter} is the interaction energy of the system, which encompasses both the electrostatic and van der Waals components obtained directly from the SANDER output, divided by the number of molecules in each box.

Construction of Enzyme–Substrate Complexes and Energy Analysis. The complexes were modeled with AMBER 5.0 (*50*) using the structure (*16*) of DhlA complexed with the substrate 1,2-dichloroethane (DCE) (PDB entry 2DHC) as a template. The WHATIF 5.0 program (*51*) was used for adding the polar hydrogen atoms. His289 was singly protonated in the δ-position in accordance with its catalytic function. Nonpolar hydrogen atoms were added using the AMBER 5.0 graphic interface xLEaP. The substrates were manually docked in the enzyme active site and aligned so that the X–C$_1$–C$_2$ angle (where X is a halogen atom) of each substrate could be superimposed on that of the DCE molecule. These initial structures of the complexes were refined using the molecular mechanics force field of Cornell et al. (1994) implemented in AMBER 5.0. One hundred steps of steepest descent were followed by conjugate gradient energy minimization until the root-mean-square value of the potential energy gradient was less than 0.1 kcal mol^{-1} Å^{-1}. A nonbonded cutoff of 10.0 Å and a distance-dependent dielectric constant ($\epsilon = 4r_{ij}$) were used. The ANAL module of AMBER 5.0 was used for energy decomposition of the refined complexes.

Estimation of Surface Desolvation Energy. The surface desolvation energy ($E_{\mathrm{desolv-sur}}$) of a substrate was calculated as a sum of atomic surface accessibilities multiplied by hydrophobicity coefficients for specific atom types. Atomic surface accessibility was calculated using the NACCESS 2.1.1 program (*52*). This program is an implementation of the method of Lee and Richards. Appropriate hydrophobicity coefficients were taken from the literature (*53*, *54*) as follows: carbon-containing group, 18; neutral oxygen or nitrogen, −9; sulfur, −5; charged nitrogen, −38; and charged oxygen, −37. In this context, we assigned a value of 1 to the hydrophobicity coefficient of halogen atoms.

Estimation of the Electrostatic Contributions to the Free Energies of Binding. Continuum Electrostatics Calculations. The overall electrostatic free energy change upon binding (ΔG_{ele}) can be calculated from the total electrostatic energy of the system by running three consecutive calculations on the same grid (*55*, *56*): one for all the atoms in the complex ($G_{\mathrm{ele}}{}^{\mathrm{ES}}$), one for the substrate atoms alone ($G_{\mathrm{ele}}{}^{\mathrm{S}}$), and a third one for the enzyme atoms alone ($G_{\mathrm{ele}}{}^{\mathrm{E}}$). Since the grid definition is the same in the three calculations, the grid energy artifact cancels out when the electrostatic contribution to the binding free energy is expressed as the difference in energy between the bound and the unbound molecule:

$$\Delta G_{\mathrm{ele}} = G_{\mathrm{ele}}{}^{\mathrm{ES}} - (G_{\mathrm{ele}}{}^{\mathrm{S}} + G_{\mathrm{ele}}{}^{\mathrm{E}}) \tag{4}$$

An alternative method, which allows partitioning at the residue level, considers a different description of the binding process. This consists of first desolvating the apposing surfaces of both the substrate and enzyme and then letting the charges of the two molecules interact. It is then possible to separate the change in electrostatic free energy on molecular association (ΔG_{ele}) into three components (*55*–*57*): (i) the enzyme–substrate interaction energy in the presence of the surrounding solvent ($E_{\mathrm{ele}}{}^{\mathrm{ES}}$), (ii) the change in desolvation energy of the substrate upon binding ($\Delta G_{\mathrm{desolv}}{}^{\mathrm{S}}$), and (iii) the change in desolvation energy of the enzyme upon binding ($\Delta G_{\mathrm{desolv}}{}^{\mathrm{E}}$):

$$\Delta G_{\mathrm{ele}} = E_{\mathrm{ele}}{}^{\mathrm{ES}} + (\Delta G_{\mathrm{desolv}}{}^{\mathrm{S}} + \Delta G_{\mathrm{desolv}}{}^{\mathrm{E}}) \tag{5}$$

This decomposition is exact, contains all cross terms, and can be profitably used in COMBINE analysis. The first term in eq 5, that is, the electrostatic energy of interaction between the group of *E* atoms in the enzyme and the group of *S* atoms in the substrate, can be described (in kilocalories per mole) by

$$E_{\mathrm{ele}}{}^{\mathrm{ES}} = \sum_{i=1}^{E} q_i \phi_i \tag{6}$$

where *q* represents the atomic point charges of the *i*th atom of the enzyme and ϕ_i is the electrostatic potential at the *i*th atom of the enzyme created by the *S* atoms of the substrate. Equation 6 can be recast in the form of a sum of *N* residue-based contributions (e_N), where *N* is the number of residues in the enzyme, each of these containing *K* atoms:

$$E_{\mathrm{ele}}{}^{\mathrm{ES}} = \sum_{n=1}^{N}\sum_{k=1}^{K} q_{nk}\phi_{nk} = \sum_{n=1}^{N} e_n \tag{7}$$

This allows the total electrostatic free energy of substrate binding to be expressed as a sum of residue-based contributions plus two additional terms corresponding to the electrostatic components of the desolvation free energy of both the substrate and the enzyme:

$$\Delta G_{\mathrm{ele}} = \sum_{n=1}^{N} e_n + \Delta G_{\mathrm{desolv}}{}^{\mathrm{S}} + \Delta G_{\mathrm{desolv}}{}^{\mathrm{E}} \tag{8}$$

The electrostatic potentials used in eqs 6 and 7 can be calculated either as

$$\phi_i = 332\sum_{j=1}^{L}\frac{q_j}{\mathrm{e}r_{ij}} \tag{9}$$

where ϵ is the relative permittivity of the homogeneous dielectric medium and r_{ij} is the separation between every pair of atoms (as in the molecular mechanics force field) or to include the potential created by the response of the surrounding solvent to the substrate charges, by solving the linear form of the Poisson−Boltzmann equation:

$$\nabla[\epsilon(\vec{r})\nabla\phi(\vec{r})] = -4\pi\rho(\vec{r}) + \bar{\kappa}^2\phi(\vec{r}) \tag{10}$$

where ρ is the fixed solute charge distribution, κ is the modified Debye−Hückel constant that accounts for a Boltzmann distribution of the ions in solution, and $\epsilon(\vec{r})$ and $\phi(\vec{r})$ are the dielectric constant and the electrostatic potential, respectively, as a function of position. The solvent-corrected potential calculated with eq 10 can be either that generated by the charges on the enzyme at the positions of the uncharged substrate atoms or, alternatively, that created by the charges on the substrate at the location of each of the uncharged atoms of the enzyme. It is the latter that we have computed for the purpose of calculating the residue-based contributions to $E_{\mathrm{ele}}{}^{\mathrm{ES}}$ described in eq 7. Thus, the only effect that is missed by this approach is the solvent polarization created by the enzyme charges and its corresponding cross terms. However, the consequence of ignoring this effect is almost negligible when the complexes of a common receptor with a series of congeneric substrates are considered, as is demonstrated below by the similar $E_{\mathrm{ele}}{}^{\mathrm{ES}}$ values computed with eqs 4 and 8.

The latter two terms of eqs 5 and 8, corresponding to the differences in electrostatic free energies of desolvation of the substrate and enzyme upon complex formation, were calculated by considering the effects on the respective electrostatic free energies of replacing the high-dielectric medium of the solvent with the low-dielectric medium of the other molecule in those regions that are occupied by the binding partner in the complex.

Each of the components of eq 8, which describes the electrostatic effects of substrate binding, enters the energy matrix for COMBINE analysis as a different variable. From the previous derivation, and as demonstrated below in the Results, it is clear that the total electrostatic binding free energy is partitioned taking into account the cross terms in the reaction field and no double counting is done; i.e., the sum of electrostatic terms in the energy matrix yields the electrostatic free energy of binding for that particular compound. This is an important feature that allows easier interpretation of the regression models, avoiding convoluted effects in the energetic description of the variables. It is also worth noting that the formalism that is presented creates an electrostatic block with variance similar to that of the van der Waals block, making the direct use of PLS analysis possible without invocation of scaling procedures, which may produce spurious results in three-dimensional QSAR (*8*).

The Poisson−Boltzmann equation was solved using a finite difference method, as implemented in the DelPhi (*57*) module of Insight II. The atomic coordinates that were employed were those of the AMBER-optimized complexes. The interior of the enzyme, the substrates, and the complexes were considered a low-dielectric medium ($\epsilon = 4$), whereas the surrounding solvent was treated as a high-dielectric medium ($\epsilon = 80$) with an ionic strength of 0.145 M. Cubic grids with a resolution of 0.5 Å were centered on the molecular systems that were considered, and the charges were distributed onto the grid points (*55, 56*). Solvent-accessible surfaces (*58*), calculated with a spherical probe with a 1.4 Å radius (*59*), defined the solute boundaries, and a minimum separation of 10 Å was left between any solute atom and the borders of the box. The potentials at the grid points delimiting the box were calculated analytically by treating each atom with a partial atomic charge as a Debye−Hückel sphere (*55, 56*).

Chemometric Analysis. The program Q2 4.5.11 (Multivariate Infometric Analysis) was employed for data pretreatment, building of a model, and selection of variables by fractional factorial design (FFD) (*60*). The program SIMCA-P 8.0 (Umetri, Sweden) was used for permutation validation (*61*). The quality of models is described by the correlation coefficient (R^2), the cross-validated correlation coefficient (Q^2), the standard deviation of error of calculations (SDEC), the standard deviation of error of predictions (SDEP$_{\mathrm{int}}$ and SDEP$_{\mathrm{ext}}$), and the intercept of the permutation plot for Q^2 (INTC$_{Q^2}$). R^2 and SDEC are the descriptors of the quality of the fit and are given by eqs 11 and 12, respectively.

$$R^2 = 1 - \frac{\sum_i (y_{i\mathrm{calc}} - y_{i\mathrm{obs}})^2}{\sum_i (y_{i\mathrm{obs}} - y_{i\mathrm{mean}})^2} \tag{11}$$

$$\mathrm{SDEC} = \left[\sum_i \frac{(y_{i\mathrm{calc}} - y_{i\mathrm{obs}})^2}{N}\right]^{1/2} \tag{12}$$

R^2 takes values up to a maximum of 1, corresponding to a perfect fit. A value higher than 0.5 is generally considered statistically significant. Q^2 characterizes the predictive ability of a model and was computed using the Leave One Out/Leave Some Out cross-validation according to eq 13.

$$Q^2 = 1 - \frac{\sum_i (y_{i\text{pred}} - y_{i\text{obs}})^2}{\sum_i (y_{i\text{obs}} - y_{i\text{mean}})^2} \qquad (13)$$

A value higher than 0.4 is generally considered statistically significant. SDEP$_{\text{int}}$ and SDEP$_{\text{ext}}$ quantify the error in prediction for test and validation sets, respectively, and are standard deviations computed in a manner analogous to that of SDEC.

INTC$_{Q^2}$ is a measure of the background Q^2 obtained by model fitting with a randomized *y* variable (30 permutations with random seed). The *X* variable matrix contained 622 columns (620 energy contributions for 310 amino acid residues and two energy contributions for the catalytic water molecule) and 18 rows (enzyme–substrate complexes). The dependent *y* variable was represented by 18 logarithmic values of experimental binding constants K_m. Three different data pretreatment methods were applied to the energy interaction matrix during the PLS analysis: centering only (no scaling), block unscaled weights (BUW), and scaling to unit variance (UV). A sum of squares higher than 10^{-7} was the pretreatment threshold that was used for *X* variables to be considered active. This threshold served for elimination of variables with low-magnitude energies and variance. External validation was performed by splitting the data set of the complexes into two subsets. The compounds were ordered according to log K_m values and split into odd and even values to obtain homogeneous data sets. One of them served as a test set (compounds **2**, **4**, **6**, **8**, **9**, **13**–**15**, and **17**), while the other was used as a training and validation set (**1**, **3**, **5**, **7**, **10**–**12**, **16**, and **18**).

RESULTS

Parametrization of the Haloalkanes. The nonbonded parameters used in this work for the Cl and Br atoms and the cyano group present in the haloalkanes that were studied were derived from condensed phase molecular dynamics simulations of three relevant organic solvents. The good agreement found between the calculated and experimentally measured densities and enthalpies of vaporization (Figure 1) lends credence to the validity of these parameters. The derived parameters are provided in the Supporting Information.

Construction of COMBINE Models. The set of 18 enzyme–substrate complexes was modeled, and each complex was energy-minimized. The positions of the substrates inside the active site after energy minimization are shown in Figure 2. Different types of models were built using several scaling methods. The Q^2 value was used as the criterion to determine the optimal dimensionality of the PLS models. The FFD variable selection procedure was then applied to all models using two different techniques (retaining uncertain variables and not retaining uncertain variables). The complete set of PLS models with their statistical parameters is listed in Table 2.

Effect of Scaling on the Predictive Ability. The unscaled models have statistical criteria similar to those of the BUW-scaled models (Table 2). The autoscaled models have higher R^2 values but significantly lower Q^2 values, which is indicative of overfit. External validation was employed for testing the predictive power of the models (Table 3). External validation of the models without scaling and the BUW-scaled models produces similar statistics. The autoscaled models unequivocally provide the best SDEP values. This result, however, does not seem consistent with the fact that autoscaled models have lower Q^2 values and is presumably a chance effect. Indeed, *y*-value permutation tests indicated the presence of chance correlation in the autoscaled models. The chance correlation is also apparent from Table 4. Many of the best scoring energy contributions in the autoscaled models are not provided by the residues lining the active site, and some of them are provided by residues on the protein surface.

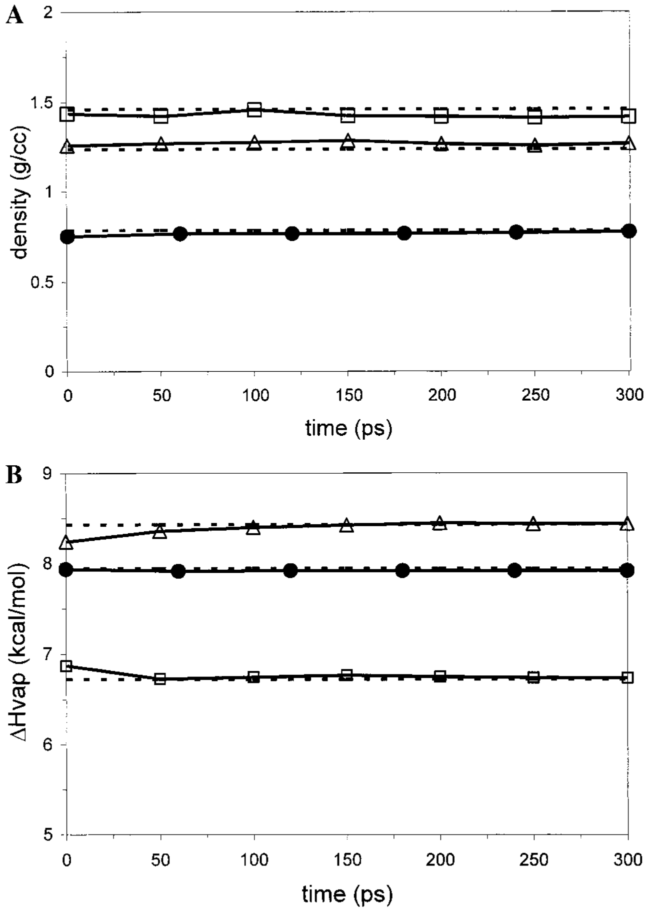

FIGURE 1: (A) Time evolution of the calculated density (grams per cubic centimeter) during the molecular dynamics simulations of the solvent boxes: bromoethane (□, exptl, 1.460), 1,2-dichloroethane (△, exptl, 1.235), and acetonitrile (●, exptl, 0.786). (B) Time evolution of the calculated enthalpy of vaporization (ΔH_{vap}, kilocalories per mole) during the molecular dynamics simulations of the solvent boxes: bromoethane (□, exptl, 6.72), 1,2-dichloroethane (△, exptl, 8.43), and acetonitrile (●, exptl, 7.94). The experimental values are displayed as horizontal dashed lines.

Effect of Variable Selection on the Predictive Ability. Two types of FFD variable selection procedures were performed. FFD resulted in higher Q^2 values, indicating improved internal predictive ability of the models (Table 2), but at the same time resulted in lower R^2 values, suggesting that some of the variables that are important for explaining K_m were excluded from the data set. The 24 variables retained in model 3 and the 23 variables in model 6 were sufficient for explaining 87% of the variance (77% cross-validated) in the K_m values. External validation, however, confirmed the

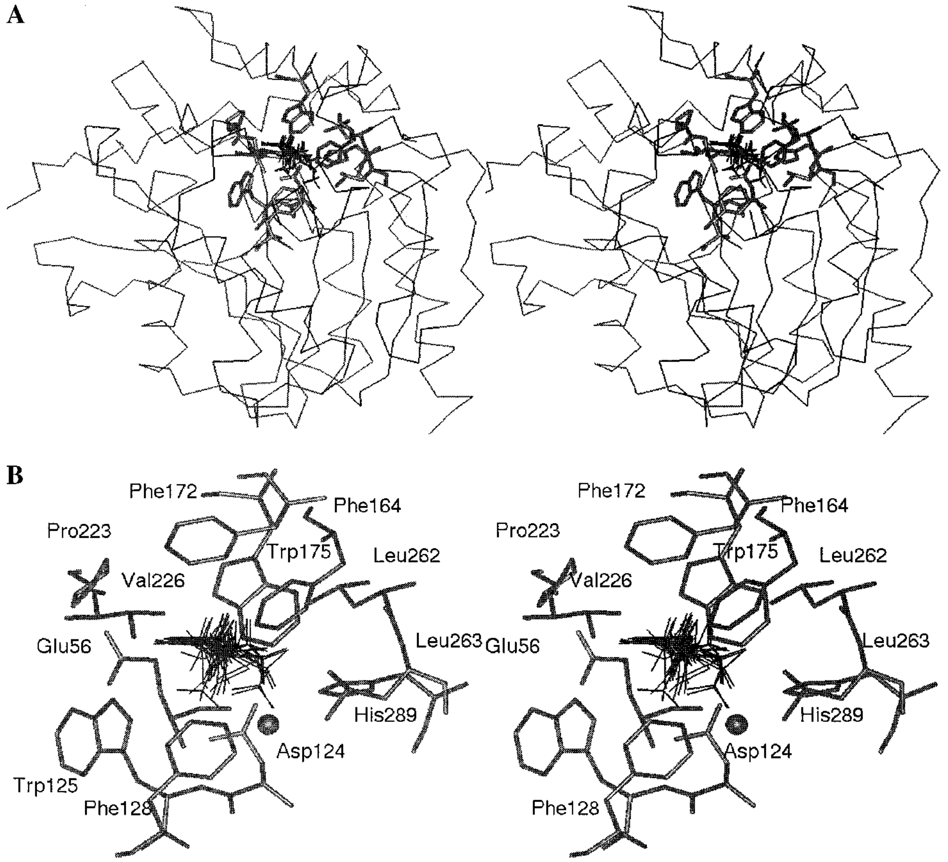

FIGURE 2: Stereoview of the DhlA haloalkane dehalogenase in complex with ligands. The 18 ligands are superimposed in the positions obtained from energy refinement. The protein backbone is represented by a Cα trace (A). The side chains of first-shell residues lining the active site that are identified as significant contributors to substrate specificity in COMBINE analysis are shown in stick representation, and in a magnified view in panel B where they are labeled according to the DhlA sequence. The active site water molecule that is included in the COMBINE analysis is shown as a sphere.

greater robustness of the models derived without variable selection (Table 3).

Effect of Surface Desolvation Energy on the Predictive Ability. Inclusion of the surface desolvation energy term ($E_{\mathrm{desolv-sur}}$) does not significantly improve the predictive ability of the models (models 10−18) in comparison to the ability of models without this term (models 1−9). Statistical criteria for the models with and without surface desolvation terms are very similar (Table 2). The variable $E_{\mathrm{desolv-sur}}$ showed low weighted regression coefficients and was eliminated from the models that employed variable selection. Addition of $E_{\mathrm{desolv-sur}}$ did, however, have a large influence on the number of variables retained after FFD variable selection.

Effect of Continuum Electrostatic Energy Terms on the Predictive Ability. Replacement of AMBER electrostatic interactions with the enzyme−substrate interactions calculated by numerically solving the linearized Poisson−Boltzmann equation slightly improved both R^2 and Q^2 (models 37−42; compare with models 1−6). Only the unscaled and BUW-scaled models were taken into account in these comparisons since the chance correlation was detected earlier in the autoscaled models. When the electrostatic term calculated using AMBER was replaced with the overall electrostatic energy change upon binding calculated with DelPhi (ΔG_{ele}), a new set of global interaction energies was obtained. The residue-based electrostatic interaction energies computed with DelPhi, as depicted in eq 7, were used to replace the corresponding AMBER values in the COMBINE energy matrix. To account for the change in the electrostatic energy of desolvation of the substrate and the enzyme binding site upon complex formation, two new variables, $\Delta G_{\mathrm{desolv}}{}^{\mathrm{E}}$ and $\Delta G_{\mathrm{desolv}}{}^{\mathrm{S}}$, were incorporated in the analysis as additional terms. Addition of $\Delta G_{\mathrm{desolv}}{}^{\mathrm{E}}$ and $\Delta G_{\mathrm{desolv}}{}^{\mathrm{S}}$ into models with both AMBER electrostatic and van der Waals interactions did not lead to statistically better models (models 19−27). The replacement of AMBER electrostatic interactions with $\Delta G_{\mathrm{desolv}}{}^{\mathrm{E}}$ and $\Delta G_{\mathrm{desolv}}{}^{\mathrm{S}}$ terms led to models with improved predictive ability (models 28−33). Addition of these two terms to models with AMBER electrostatic interactions replaced with the corresponding values from Poisson−Boltzmann calculations slightly improves the predictive ability of the COMBINE models (models 46−54). The substrate desolvation energy term makes the most important contribution to the first principal component in these models. The desolvation energy of the enzyme is also among the five most significant energy contributions.

Chemometric Analysis of Model 4 (BUW-Scaled, without FFD and E_{desolv}). The most influential variables are almost the same in the unscaled and BUW-scaled models (Table 4). Model 4 was chosen for detailed description because it shows one of the best statistical parameters out of the models based on per residue van der Waals and electrostatic contributions. This model has four latent variables, yields an R^2 of 0.91, a Q^2 of 0.73, and an $\mathrm{SDEP}_{\mathrm{int}}$ of 0.59, and shows good external validation (Tables 4 and 5). Validation

Table 2: Summary of the COMBINE Models

model[a]	BUW[b]	UV[c]	FFD[d]	E_{vdw}[e]	E_{ele}[f]	$E_{\mathrm{desolv-sur}}$[g]	$E_{\mathrm{ele}}^{\mathrm{ES}}$[h]	$\Delta G_{\mathrm{desolv}}^{\mathrm{S}}$[i]	$\Delta G_{\mathrm{desolv}}^{\mathrm{E}}$[j]	obj[k]	var[l]	A[m]	R^2	SDEC	Q^2	$\mathrm{SDEP}_{\mathrm{int}}$
1	−	−	N	+	+	−	−	−	−	18	448	4	0.911	0.342	0.728	0.596
2	−	−	R	+	+	−	−	−	−	18	436	3	0.893	0.375	0.734	0.590
3	−	−	E	+	+	−	−	−	−	18	24	2	0.865	0.420	0.770	0.548
4	+	−	N	+	+	−	−	−	−	18	448	4	0.910	0.343	0.734	0.590
5	+	−	R	+	+	−	−	−	−	18	434	3	0.888	0.382	0.734	0.590
6	+	−	E	+	+	−	−	−	−	18	23	2	0.865	0.420	0.774	0.543
7	−	+	N	+	+	−	−	−	−	18	448	4	0.959	0.198	0.649	0.576
8	−	+	R	+	+	−	−	−	−	18	232	4	0.964	0.185	0.700	0.532
9	−	+	E	+	+	−	−	−	−	18	105	4	0.970	0.167	0.748	0.488
10	−	−	N	+	+	+	−	−	−	18	449	5	0.924	0.316	0.722	0.603
11	−	−	R	+	+	+	−	−	−	18	445	3	0.885	0.388	0.704	0.622
12	−	−	E	+	+	+	−	−	−	18	7	2	0.864	0.421	0.774	0.544
13	+	−	N	+	+	+	−	−	−	18	449	5	0.924	0.316	0.722	0.604
14	+	−	R	+	+	+	−	−	−	18	445	3	0.882	0.393	0.705	0.622
15	+	−	E	+	+	+	−	−	−	18	7	2	0.864	0.421	0.774	0.544
16	−	+	N	+	+	+	−	−	−	18	449	4	0.959	0.198	0.648	0.577
17	−	+	R	+	+	+	−	−	−	18	227	4	0.967	0.178	0.718	0.516
18	−	+	E	+	+	+	−	−	−	18	111	4	0.966	0.178	0.747	0.488
19	−	−	N	+	+	−	−	+	+	18	450	4	0.903	0.356	0.705	0.621
20	−	−	R	+	+	−	−	+	+	18	443	3	0.885	0.388	0.704	0.622
21	−	−	E	+	+	−	−	+	+	18	8	2	0.864	0.421	0.774	0.544
22	+	−	N	+	+	−	−	+	+	18	450	6	0.939	0.283	0.779	0.538
23	+	−	R	+	+	−	−	+	+	18	439	3	0.885	0.388	0.704	0.622
24	+	−	E	+	+	−	−	+	+	18	7	2	0.864	0.421	0.774	0.544
25	−	+	N	+	+	−	−	+	+	18	450	4	0.959	0.197	0.649	0.576
26	−	+	R	+	+	−	−	+	+	18	222	4	0.962	0.190	0.687	0.544
27	−	+	E	+	+	−	−	+	+	18	112	4	0.967	0.176	0.749	0.487
28	−	−	N	+	−	−	−	+	+	18	199	3	0.909	0.345	0.745	0.577
29	−	−	R	+	−	−	−	+	+	18	192	2	0.858	0.431	0.733	0.591
30	−	−	E	+	−	−	−	+	+	18	7	2	0.864	0.421	0.774	0.544
31	+	−	N	+	−	−	−	+	+	18	199	4	0.918	0.328	0.752	0.570
32	+	−	R	+	−	−	−	+	+	18	188	2	0.861	0.427	0.745	0.578
33	+	−	E	+	−	−	−	+	+	18	7	2	0.842	0.455	0.741	0.582
34	−	+	N	+	−	−	−	+	+	18	199	3	0.842	0.386	0.494	0.692
35	−	+	R	+	−	−	−	+	+	18	117	3	0.858	0.366	0.677	0.553
36	−	+	E	+	−	−	−	+	+	18	45	2	0.817	0.416	0.716	0.518
37	−	−	N	+	−	−	+	−	−	18	502	5	0.930	0.303	0.742	0.580
38	−	−	R	+	−	−	+	−	−	18	490	3	0.898	0.365	0.755	0.567
39	−	−	E	+	−	−	+	−	−	18	14	2	0.831	0.471	0.718	0.607
40	+	−	N	+	−	−	+	−	−	18	502	5	0.925	0.314	0.738	0.586
41	+	−	R	+	−	−	+	−	−	18	490	3	0.890	0.379	0.756	0.564
42	+	−	E	+	−	−	+	−	−	18	17	2	0.831	0.471	0.718	0.607
43	−	+	N	+	−	−	+	−	−	18	502	4	0.924	0.269	0.581	0.629
44	−	+	R	+	−	−	+	−	−	18	269	4	0.913	0.286	0.467	0.709
45	−	+	E	+	−	−	+	−	−	18	171	4	0.954	0.208	0.556	0.648
46	−	−	N	+	−	−	+	+	+	18	504	5	0.933	0.297	0.731	0.593
47	−	−	R	+	−	−	+	+	+	18	497	3	0.898	0.365	0.709	0.617
48	−	−	E	+	−	−	+	+	+	18	6	3	0.885	0.387	0.788	0.527
49	+	−	N	+	−	−	+	+	+	18	502	5	0.922	0.320	0.733	0.590
50	+	−	R	+	−	−	+	+	+	18	491	3	0.889	0.381	0.696	0.630
51	+	−	E	+	−	−	+	+	+	18	9	2	0.864	0.421	0.774	0.544
52	−	+	N	+	−	−	+	+	+	18	504	4	0.924	0.268	0.580	0.630
53	−	+	R	+	−	−	+	+	+	18	269	4	0.933	0.252	0.477	0.703
54	−	+	E	+	−	−	+	+	+	18	172	4	0.945	0.227	0.549	0.653

[a] Model identifier. [b] Block unscaled weights. [c] Scaling to unit variance. [d] Fractional factorial design (FFD): N, without FFD; R, FFD with retained uncertain variables; E, FFD with excluded uncertain variables. [e] van der Waals energy contributions from AMBER (*X* matrix). [f] Electrostatic energy contributions from AMBER (*X* matrix). [g] Surface term of desolvation energy in the *X* matrix. [h] Enzyme−substrate interaction energy in the presence of the surrounding solvent in the *X* matrix. [i] Change in desolvation energy of the substrate upon binding in the *X* matrix. [j] Change in desolvation energy of the enzyme upon binding in the *X* matrix. [k] Number of objects. [l] Number of variables *X* after threshold application. [m] Number of latent variables.

by permutation confirms no chance correlation in the model. The robustness of this model is further supported by the fact that a very similar model is derived when COMBINE analysis is performed for the same structures with different software (the COMBINE program, A. R. Ortiz) with slightly different pretreatment and cross-validation procedures (data not shown). Score plots for model 4 are shown in Figure 3, while the loading plots are shown in Figure 4. The most important variable in the first principal component is the electrostatic interaction energy of Asp124, 124^{ele}. This component is most important for explaining the variance in the binding affinities of compounds **11**, **12**, and **18** (listed in order of decreasing significance). The most important variable in the second principal component is the van der Waals interaction energy of Trp125, 125^{vdw}. This component is particularly important for explaining the binding affinity of compounds **13** and **14**. In the third component (focused on compounds **2**, **5**, **4**, **8**, **10**, **14**, and **18**), the most significant

Table 3: External Validation and Permutation Test of COMBINE Models

model[a]	obj[b]	var[c]	A[d]	R^2	SDEC	Q^2	$\mathrm{SDEP_{int}}$	$\mathrm{SDEP_{ext}}$	INTC_{Q^2}
1*	9	444	4	0.985	0.143	0.742	0.587	0.561	−0.02
2*	9	425	3	0.980	0.165	0.879	0.401	1.008	−0.08
3*	9	16	3	0.982	0.157	0.924	0.318	0.973	−0.11
4*	9	444	4	0.981	0.160	0.738	0.590	0.629	−0.05
5*	9	426	3	0.981	0.160	0.890	0.383	1.003	−0.07
6*	9	15	3	0.982	0.157	0.924	0.318	0.973	0.01
7*	9	444	4	0.989	0.103	0.544	0.662	0.559	0.39
8*	9	204	4	0.994	0.078	0.729	0.510	0.537	0.29
9*	9	105	4	0.998	0.048	0.846	0.385	0.589	0.37

[a] Model identifier. The number corresponds to the models presented in Table 2. An asterisk indicates that only half of the objects were used in a model. [b] Number of objects in a model. Working set: **2**, **4**, **6**, **8**, **9**, **13**−**15**, and **17**. Validation set: **1**, **3**, **5**, **7**, **10**−**12**, **16**, and **18**. [c] Number of variables X after threshold application. [d] Number of latent variables.

variables are 263^{vdw}, 172^{vdw}, 226^{vdw}, 175^{vdw}, and 125^{vdw}. In the fourth component (which explains the variance in binding affinity of compounds **5**, **1**, **6**, **9**, **4**, **2**, **18**, and **17**), variables 226^{vdw} and 172^{vdw} are most influential. The overall importance of each variable in model 4, as quantified by the weighted regression coefficients, is presented in Figure 5. The goodness of fit and external predictive ability of the model are presented in Figure 6. A stereoview of the energy contributions selected by COMBINE analysis with the assigned type of energy contributions is shown in Figure 7. A picture illustrating unfavorable interactions between long substrate molecules (butanes and hexanes) and Trp175 together with positive interactions of these molecules with three amino acid residues (Leu263, Met152, and Cys150) on the opposite side of the active site is shown in Figure 8.

External Predictions of Binding Affinities for Mutant Proteins. The applicability of the COMBINE models for predictions was validated using two mutants of DhlA for which the crystal structures were determined (*19*, *20*). Four substrate molecules with available experimental binding constants were modeled in the active sites of the mutant proteins. The experimental binding constants were calculated only for models with the best $\mathrm{SDEP_{ext}}$ values (models 4 and 40). Although the internal predictive ability of model 49 was as good as that of model 40, model 49 led to significantly worse external predictions. The trends in the changes of binding affinity due to mutation are predicted correctly using both model 4 and model 40 without exception (Table 5). The largest error was obtained for the substrate 1-bromo-2-chloroethane, but this is not unexpected since the prediction is made for both a new substrate and a new enzyme. Generally, the predictions made using model 4 and model 40 are equivalent as shown by their $\mathrm{SDEP_{ext}}$ values (0.66 and 0.67, respectively).

DISCUSSION

Rational engineering of enzyme substrate specificity requires detailed knowledge of the interactions taking place between the enzyme and the substrates at atomic resolution. Interaction energies based on molecular mechanics calculations are employed for the study of enzyme−substrate interactions in comparative binding energy (COMBINE) analysis (*7*). In this study, the applicability of COMBINE analysis for protein engineering purposes has been investigated. COMBINE analysis was conducted for 18 substrates of the haloalkane dehalogenase DhlA. The effect of scaling, variable selection, and addition of desolvation energy terms on the predictive ability of the resulting models was investigated. No scaling and BUW-scaling procedures provided robust models with good predictive ability ($Q^2 \geq 0.72$ and $\mathrm{SDEP_{int}} \leq 0.59$), while autoscaling resulted in models with chance correlation.

Variable selection demonstrated that the energy contributions from only a limited number of amino acid residues (1%) are sufficient to explain a large proportion of the variance (91%) in the binding constants. The models retaining all energy contributions after applying pretreatment thresholds showed very good fitting properties and predictive ability, and it was concluded that variable selection procedures (i.e., exclusion of the nonsignificant variables from the data matrix) are not necessary for modeling the DhlA−substrate binding affinities. Apparently, PLS was effective enough to filter out nonsignificant interaction energy contributions by giving them small weighting coefficients. The inclusion of a surface area dependent desolvation energy term did not improve the predictive ability of the models. The inclusion of the electrostatic enzyme−substrate interactions computed by numerical solution of the Poisson−Boltzmann equation improved the quality of the models and resulted in a COMBINE model that achieved very good predictive ability ($Q^2 = 0.74$, $\mathrm{SDEP_{int}} = 0.59$, $\mathrm{SDEP_{ext}} = 0.67$). The incorporation of two additional terms, representing the electrostatic energy contributions to the partial desolvation of the substrates and the enzyme upon binding, resulted in a COMBINE model with good fitting properties ($Q^2 = 0.78$) that provided good internal predictions ($\mathrm{SDEP_{int}} = 0.54$) and slightly worse predictions for objects not included in model development ($\mathrm{SDEP_{ext}} = 0.82$).

A mechanistic interpretation of the models that were constructed provides a detailed understanding of the structure−affinity relationships of DhlA substrates. The BUW-scaled model containing van der Waals and electrostatic energy terms for every residue (model 4) was chosen for this purpose. Simultaneous examination of the score plots (Figure 3), loading plots (Figure 4), weighted regression coefficients plot (Figure 5), and the structures of enzyme−substrate complexes (Figure 7) enables identification of important interactions between the substrate molecules and amino acid residues that are key to understanding the differences in affinity. We expect that knowledge of these interactions can be used advantageously to propose mutant enzymes with modified specificities. It is apparent from COMBINE analysis that only a limited number of interactions are important for explaining most of the differences in binding among the substrates of DhlA. van der Waals interactions are considerably more important than electrostatic interactions. This result can be rationalized for the haloalkane dehalogenase DhlA because its active site is small, being evolutionarily optimized for the natural substrate, 1,2-dichloroethane, whereas most of the substrates analyzed in this study have a volume larger than that of 1,2-dichloroethane, resulting in a number of close contacts between the ligands and the enzyme active site. Furthermore, the substrates that were analyzed are simple, uncharged, and mainly hydrophobic molecules. Experience from studying different series of inhibitors interacting with different protein structures

Table 4: Most Important *X* Variables[a] in COMBINE Models

	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9
	$\mathbf{172^{vdw}}$	$\mathbf{124^{vdw}}$	$\mathbf{124^{vdw}}$	$\mathbf{125^{vdw}}$	$\mathbf{124^{vdw}}$	$\mathbf{124^{vdw}}$	$\mathbf{164^{vdw}}$	$\mathbf{125^{vdw}}$	$\mathbf{125^{vdw}}$
	$\mathbf{125^{vdw}}$	$\mathbf{125^{vdw}}$	$\mathbf{125^{vdw}}$	$\mathbf{172^{vdw}}$	$\mathbf{125^{vdw}}$	$\mathbf{125^{vdw}}$	$\mathbf{172^{vdw}}$	$\mathbf{164^{vdw}}$	120^{vdw}
	$\mathbf{164^{vdw}}$	$\mathbf{172^{vdw}}$	$\mathbf{164^{vdw}}$	$\mathbf{164^{vdw}}$	$\mathbf{172^{vdw}}$	$\mathbf{164^{vdw}}$	$\mathbf{125^{vdw}}$	120^{vdw}	126^{vdw}
	$\mathbf{124^{vdw}}$	$\mathbf{164^{vdw}}$	$\mathbf{172^{vdw}}$	$\mathbf{124^{vdw}}$	$\mathbf{164^{vdw}}$	$\mathbf{172^{vdw}}$	171^{vdw}	145^{vdw}	$\mathbf{164^{vdw}}$
	222^{vdw}	$\mathbf{175^{vdw}}$	222^{vdw}	222^{vdw}	$\mathbf{175^{vdw}}$	222^{vdw}	97^{ele}	126^{vdw}	97^{ele}
	$\mathbf{175^{vdw}}$	$\mathbf{289^{vdw}}$	$\mathbf{289^{vdw}}$	$\mathbf{175^{vdw}}$	$\mathbf{289^{vdw}}$	$\mathbf{289^{vdw}}$	261^{vdw}	97^{ele}	$\mathbf{172^{vdw}}$
	$\mathbf{128^{vdw}}$	222^{vdw}	$\mathbf{175^{vdw}}$	$\mathbf{128^{vdw}}$	222^{vdw}	$\mathbf{175^{vdw}}$	196^{vdw}	$\mathbf{172^{vdw}}$	171^{vdw}
	$\mathbf{289^{vdw}}$	$\mathbf{223^{vdw}}$	$\mathbf{262^{vdw}}$	$\mathbf{289^{vdw}}$	$\mathbf{223^{vdw}}$	176^{ele}	$\mathbf{128^{vdw}}$	171^{vdw}	196^{vdw}
	$\mathbf{223^{vdw}}$	$\mathbf{125^{ele}}$	176^{ele}	$\mathbf{223^{vdw}}$	125^{ele}	293^{ele}	173^{vdw}	127^{vdw}	121^{ele}
	$\mathbf{226^{vdw}}$	260^{ele}	293^{ele}	$\mathbf{226^{vdw}}$	260^{ele}	64^{ele}	$\mathbf{175^{vdw}}$	261^{vdw}	59^{ele}
	$\mathbf{263^{vdw}}$	176^{ele}	64^{ele}	$\mathbf{263^{vdw}}$	176^{ele}	168^{vdw}	120^{vdw}	130^{vdw}	261^{vdw}
	$\mathbf{262^{vdw}}$	$\mathbf{226^{ele}}$	168^{vdw}	$\mathbf{262^{vdw}}$	226^{ele}	167^{vdw}	145^{vdw}	196^{vdw}	189^{vdw}
	$\mathbf{124^{ele}}$	$\mathbf{172^{ele}}$	167^{vdw}	$\mathbf{124^{ele}}$	$\mathbf{262^{vdw}}$	271^{vdw}	189^{vdw}	$\mathbf{128^{vdw}}$	$\mathbf{128^{vdw}}$
	179^{vdw}	$\mathbf{175^{ele}}$	271^{vdw}	179^{vdw}	$\mathbf{172^{ele}}$	230^{ele}	168^{vdw}	121^{ele}	273^{ele}
	125^{ele}	$\mathbf{262^{vdw}}$	230^{ele}	$\mathbf{125^{ele}}$	$\mathbf{175^{ele}}$	291^{ele}	104^{vdw}	$\mathbf{175^{vdw}}$	104^{vdw}
	176^{vdw}	179^{vdw}	291^{ele}	176^{vdw}	179^{vdw}	116^{ele}	184^{vdw}	104^{vdw}	203^{vdw}
	$\mathbf{172^{ele}}$	$\mathbf{263^{ele}}$	116^{ele}	$\mathbf{172^{ele}}$	176^{vdw}	72^{ele}	203^{vdw}	184^{vdw}	184^{vdw}
	$\mathbf{226^{ele}}$	311^{ele}	72^{ele}	$\mathbf{226^{ele}}$	$\mathbf{263^{ele}}$	244^{ele}	170^{vdw}	203^{vdw}	163^{ele}
	260^{ele}	$\mathbf{164^{ele}}$	244^{ele}	$\mathbf{56^{vdw}}$	$\mathbf{164^{ele}}$	134^{vdw}	126^{vdw}	173^{vdw}	175^{vdw}
	176^{ele}	176^{vdw}	61^{ele}	260^{ele}	311^{ele}	61^{ele}	174^{vdw}	57^{ele}	173^{vdw}
	$\mathbf{56^{vdw}}$	$\mathbf{128^{vdw}}$	134^{vdw}	224^{vdw}	$\mathbf{128^{vdw}}$	274^{vdw}	167^{vdw}	122^{vdw}	57^{ele}
	224^{vdw}	$\mathbf{262^{ele}}$	274^{vdw}	176^{ele}	149^{vdw}	50^{vdw}	121^{ele}	189^{ele}	244^{ele}
	260^{vdw}	149^{vdw}	50^{vdw}	260^{vdw}	262^{ele}	104^{ele}	57^{ele}	189^{vdw}	281^{vdw}
	165^{vdw}	55^{ele}	104^{ele}	165^{vdw}	260^{vdw}	–	260^{vdw}	174^{vdw}	282^{vdw}
	57^{vdw}	224^{ele}	–	57^{vdw}	55^{ele}	–	189^{ele}	123^{vdw}	287^{vdw}
vdw first shell[b]	12	9	7	12	9	6	5	5	4
ele first shell[c]	3	7	0	4	4	0	0	0	0
vdw second shell[d]	7	4	7	7	5	7	16	16	14
ele second shell[e]	3	5	10	2	7	10	4	4	7

[a] Variables are sorted according to absolute values of weighted regression coefficients (only 25 top scoring variables are listed); first-shell residues are in bold. [b] Number of first-shell residues displaying van der Waals type interactions. [c] Number of first-shell residues displaying electrostatic type interactions. [d] Number of second-shell residues displaying van der Waals type interactions. [e] Number of second-shell residues displaying electrostatic interactions.

Table 5: External Predictions of Steady-State Dissociation Constants for Haloalkane Dehalogenase Mutants Using Models 4[d] and 40[e]

		wild type	Phe172Trp		
no.[a]	substrate	expt[b]	expt[b]	prediction[d]	prediction[e]
2′	1-chlorohexane	1.40	0.57	0.22	0.28
6′	1,2-dichloroethane	0.53	5.13	8.47	8.83
8′	1,2-dibromoethane	0.01	0.03	0.20	0.20
19′	1-bromo-2-chloroethane	0.07	0.10	1.28	1.38

		wild type	Trp175Tyr		
no.[a]	substrate	expt[c]	expt[c]	prediction[d]	prediction[e]
6″	1,2-dichloroethane	0.53	2.85	0.83	0.73
8″	1,2-dibromoethane	0.01	0.06	0.04	0.03

[a] A single prime corresponds to Phe172Trp, and a double prime corresponds to Trp175Tyr. [b] From ref *20*. [c] From ref *19*.

by COMBINE analysis (*7*−*12*) indicates that there is not a single trend across all systems, and that the dominating interactions depend on the physicochemical features of the variations in the ligand series and the characteristics of the protein binding site. Examination of the weighted regression coefficients plot reveals that most of the important van der Waals interactions show positive coefficients with only one exception, Trp125. Most of the electrostatic interactions show negative coefficients as a result of covariations in the behavior of the variables. In those cases where electrostatic desolvation energy is a penalty to binding and important for explaining differences in activity, there will be some electrostatic interactions in the binding site that will correlate with the desolvation energy. These interactions will be detected by COMBINE analysis as opposing binding, even though their individual contribution is favorable to binding. COMBINE analysis is therefore detecting an overall unfavorable electrostatic desolvation effect that the favorable electrostatic interactions within the binding site are unable to overcome. The coefficients in models including explicit desolvation terms corroborate this interpretation.

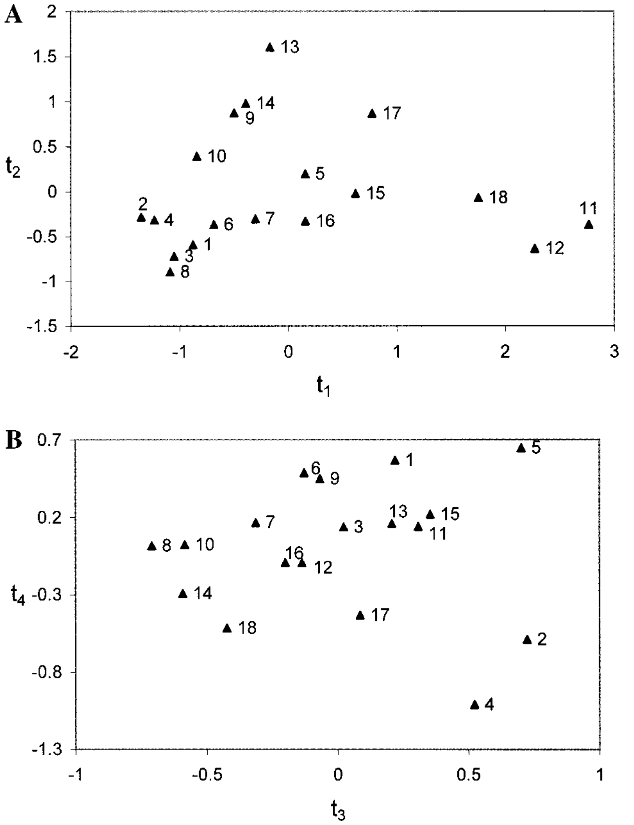

FIGURE 3: Score plots for latent variables t_1 vs t_2 (A) and t_3 vs t_4 (B) for model 4. The objects (compounds) are numbered according to Table 1.

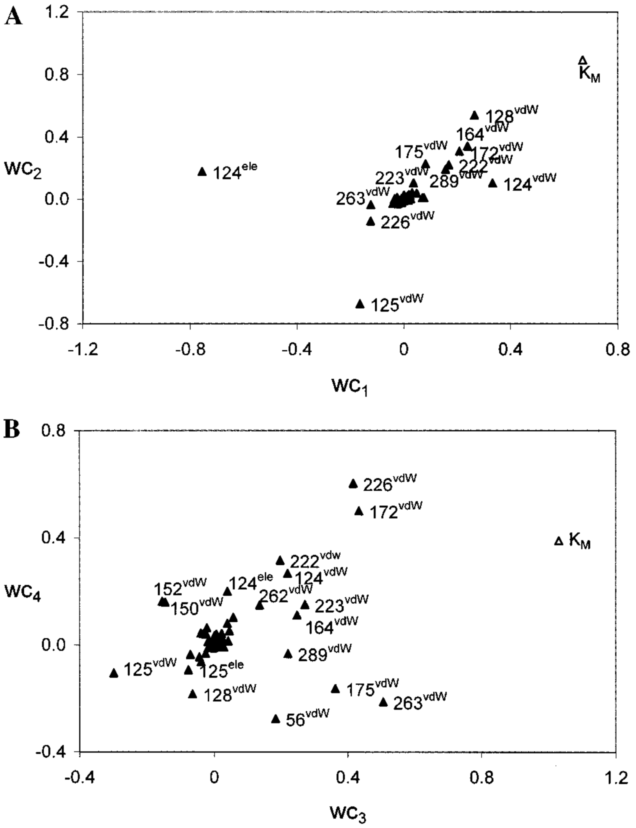

FIGURE 4: Loading plots of wc_1 vs wc_2 (A) and wc_3 vs wc_4 (B) for model 4. Selected variables (energy contributions) are numbered according to the DhlA sequence.

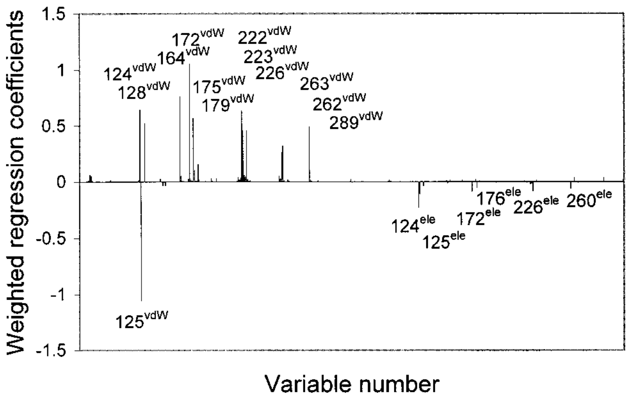

FIGURE 5: Plot of weighted regression coefficients for model 4. Selected variables (energy contributions) are numbered according to the DhlA sequence.

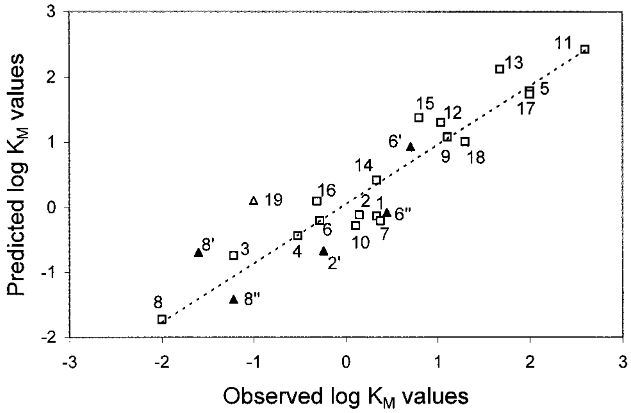

FIGURE 6: Plot of observed vs predicted K_m values for model 4. The objects (compounds) are numbered according to Tables 1 and 5. Compounds from the working (training) set are represented as white squares, and compounds from the prediction set are represented as triangles (the white triangle represents a prediction made for a complex of a new substrate *and* a new mutant enzyme). A single prime corresponds to the Phe172Trp mutant, and a double prime corresponds to the Trp175Tyr mutant.

The van der Waals interactions with positive coefficients can also be explained on a physical basis. Better van der Waals interactions result in better binding affinity. Most of the residues with positive van der Waals contributions line the active site cavity (Table 4 and Figure 7). Those further away may be explained by general improvement in packing. These residues explain most of the third and fourth principal components but also contribute to the first and second components.

A favorable van der Waals interaction of Asp124 is observed for the substrates 1-chlorohexane, 1-bromohexane, 1,2-dibromoethane, and 1,2-dibromopropane, while unfavorable interactions are observed with 2-chloroethanol, 2-bromoethanol, and 2-bromoacetamide in the first component. Asp124 is a nucleophile that initiates the dehalogenation reaction by nucleophilic attack on the carbon atom bonded to halogen in a substrate molecule (*16*, *25*). This attack leads to formation of a covalent alkyl−enzyme ester and a halide ion. Asp124 is positioned on a nucleophile elbow (*14*) and points toward the active site cavity.

The aromatic ring of Phe128 displays steric hindrance with epichlorohydrine, epibromohydrine, 1,2-dichloropropane, 2-chloroacetamide, and 2-bromoacetamide in the second component. We noted that the much smaller Ala is present in the equivalent position of dehalogenases LinB and DhaA (*37*). Both enzymes exhibit better activity with β-substituted haloalkanes than DhlA. Substitution of Phe128 with a smaller amino acid may result in enzymes with improved affinity for β-substituted substrates.

van der Waals energies of Leu263, Phe172, Val226, Trp175, Trp125, Cys150, and Met152 are among the most significant interactions in the third component. Leu263 makes unfavorable contacts with the two largest substrates in the data set, namely, 1-bromohexane and 1-chlorohexane. A smaller amino acid in position 263 may improve the affinity for long-chain substrates. Both Cys150 and Met152 have the opposite effect on binding long-chain substrates. The van der Waals interaction energy of Cys150 with substrates **2** and **4** is 1 order of magnitude lower than with other substrates. Trp175 makes direct van der Waals contact with the halogen substituent of all substrates and provides stabilization in a manner similar to that of Trp125 (*28*). Most of the substrates interact with Trp175 favorably, but 1-bromohexane, 1-chlorohexane, 1-bromobutane, and 1-chlorobutane make unfavorable van der Waals contacts with Trp175 $HN_{\epsilon1}$. Mutagenesis of Trp175 results in proteins with low activity (*19*, *28*); therefore, improved binding for these substrates can only be achieved by mutations in neighboring residues (helix 5). Nine of 12 in vivo mutants of DhlA with improved activity toward 1-chlorohexane (*30*) carried modifications in helix 5 or its close surroundings. Priest and coworkers suggested that this region is critical for the specificity of DhlA. This observation is in line with the COMBINE model which localizes seven highly significant interactions in helix 5 (Table 4 and Figure 7). The importance of Val226 could not be directly attributed to the binding of specific

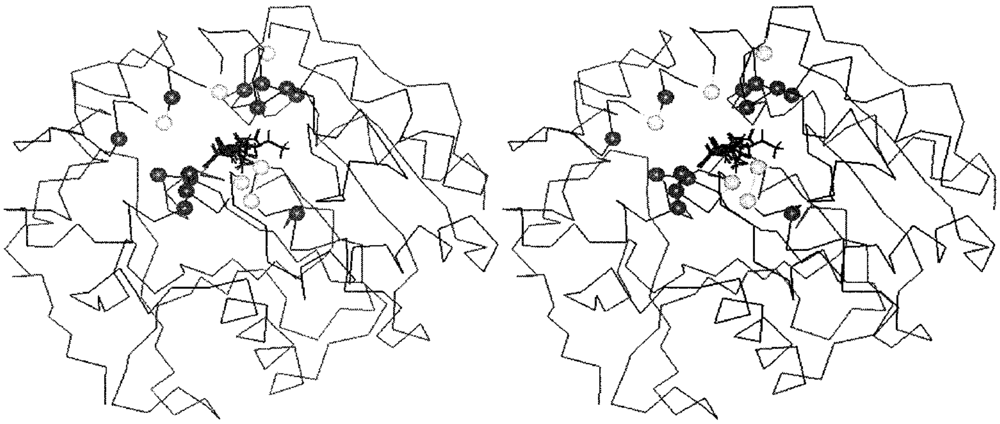

FIGURE 7: Stereoview of DhlA with assigned energy contributions. The protein backbone is represented by the Cα trace. The Cα atoms of the residues showing the most important energy contributions in model 4 are shown as balls (see Table 4). Dark-colored residues have only van der Waals energy contributions, while light-colored residues have both van der Waals and electrostatic energy contributions. The substrate molecules are in the positions obtained from energy minimization.

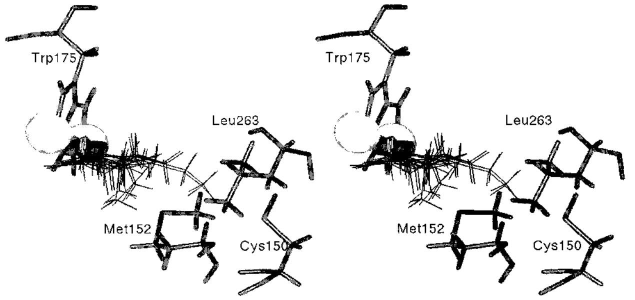

FIGURE 8: Stereoview of the substrate molecules docked in the enzyme active site. Unfavorable van der Waals interactions between long-chain substrates, i.e., butanes and hexanes, and Trp175 are represented by van der Waals surfaces of interacting atoms. Favorable interactions of the same long-chain molecules with Leu263, Met152, and Cys150 positioned on the opposite side of the active site can be seen.

substrates. This residue is not in direct contact with the substrate molecules, but makes important interactions with Trp125 and Phe172. The importance of these interactions has been experimentally demonstrated by Schanstra and co-workers (*29*). Many interactions significant for the third component also participate in the fourth component, for example, Val226, Phe172, Phe222, Glu56, Leu263, or Trp175. Phe172 contributes to stabilization of the transition state and the product (*32*, *62*). Schanstra and co-workers (*20*) performed mutational analysis at position 172 and constructed 16 different point mutants, some of which had modified activity and substrate range. Quantitative structure−function relationship (QSFR) analysis conducted with the same set of point mutants identified physicochemical properties critical for position 172: aromaticity, main-chain flexibility, refractivity, and bulkiness (*63*). Like Trp175 and Leu263, Phe172 also makes unfavorable van der Waals interactions with the long-chain substrates 1-bromohexane and 1-chlorohexane. Phe172 is among the most important residues of DhlA as it displayed both significant van der Waals and electrostatic interactions.

Two of the interactions with negative coefficients, Asp124^{ele} and Trp125^{vdw}, are very important for the first and second component, respectively. Asp124^{ele} together with the substrate electrostatic desolvation term dominates the first component of the model employing both AMBER-calculated electrostatic interaction energies and the change in the electrostatic desolvation energy of desolvation of the substrate and the enzyme upon complex formation (e.g., model 22). The variables 124^{ele} and $\Delta G_{desolv}{}^{S}$ are negatively correlated in this model and explain mainly the variability in the dissociation constants for the substrates 2-bromoacetamide, 2-chloroethanol, and 2-bromoethanol. Desolvation of these polar molecules is energetically demanding, resulting in poor binding affinity. The negative coefficient of Trp125 can be attributed to a different behavior of the energy changes associated with this residue in comparison with the rest of the variables. This can be observed in the partial weights and loading plots rather than in a different slope in the correlation with the external vector. The difference may have a structural origin, since Trp125 is located in a loop buried in the protein core, while most of the rest of the important interactions are associated with α-helices. Trp125 appears to be important for explaining differences between chlorinated and brominated derivatives. The essential role of Trp125 for binding of the halogen substituent, stabilization of the transition state, and halide ion release upon reaction has been postulated from crystallographic and fluorescence quenching studies (*64*), from site-directed mutagenesis experiments (*28*), and from molecular modeling (*32*, *36*). The electronegative aromatic indole nitrogens of the tryptophans provide polarization of the N−H bond, resulting in a slightly positive hydrogen that can interact with the halogen.

In summary, the most influential active site residues can be divided into two classes, with respect to their interaction with the substrates. The first class is formed by residues separating chlorinated derivates from brominated derivates. These residues include Trp125, Trp175, and Pro223, and form the halogen binding site in the protein, which is more selective for brominated derivatives. Mutations affecting these residues should be used to modulate the halogen specificity of the enzyme. The second set of residues discriminates substrates by their interactions with the substrate alkyl side chain. It includes Phe164, Phe172, Phe222, and Leu263, and there is a contribution from Asp124 as well. Mutations affecting these residues can be used to tune the activity of the enzyme for different side-chain specificities.

All amino acid residues discussed so far belong to the so-called first shell of residues, i.e., residues lining the active site of DhlA. Their significance is not unexpected since these residues make direct contacts with the substrate molecules and their possible role could be inferred from the X-ray structure. The identification of second-shell residues might be more useful for protein design purposes. In DhlA, such residues were also identified with the COMBINE models and include Phe222, Leu179, Lys176, Lys224, Val165, and Pro57 (listed in order of their significance; see Table 4). These residues represent suitable targets for future site-directed mutagenesis experiments.

SUPPORTING INFORMATION AVAILABLE

Parameters derived for halogenated compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Janssen, D. B., Pries, F., and Van der Ploeg, J. R. (1994) *Annu. Rev. Microbiol. 48*, 163−191.
2. Fetzner, S., and Lingens, F. (1994) *Microbiol. Rev. 58*, 641−685.
3. Fetzner, S. (1998) *Appl. Microbiol. Biotechnol. 50*, 633−657.
4. Copley, S. D. (1998) *Curr. Opin. Chem. Biol. 2*, 613−617.
5. Swanson, P. E. (1999) *Curr. Opin. Biotechnol. 10*, 365−369.
6. Vlieg, J. E. T. H., Poelarends, G. J., Mars, A. E., and Janssen, D. B. (2000) *Curr. Opin. Microbiol. 3*, 257−262.
7. Ortiz, A. R., Pisabarro, M. T., Gago, F., and Wade, R. C. (1995) *J. Med. Chem. 38*, 2681−2691.
8. Ortiz, A. R., Pastor, M., Palomer, A., Cruciani, G., Gago, F., and Wade, R. C. (1997) *J. Med. Chem. 40*, 1136−1148.
9. Pastor, M., Perez, C., and Gago, F. (1997) *J. Mol. Graphics Modell. 15*, 364−371.
10. Perez, C., Pastor, M., Ortiz, A. R., and Gago, F. (1998) *J. Med. Chem. 41*, 836−852.
11. Wade, R. C., Ortiz, A. R., and Gago, F. (1998) in *3D QSAR in Drug Design* (Kubinyi, H., Folkers, G., and Martin, Y. C., Eds.) pp 19−34, Kluwer Academic Publishers, Dordrecht, The Netherlands.
12. Tomic, S., Nilsson, L., and Wade, C. R. (2000) *J. Med. Chem. 43*, 1780−1792.
13. Keuning, S., Janssen, D. B., and Witholt, B. (1985) *J. Bacteriol. 163*, 635−639.
14. Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G., and Goldman, A. (1992) *Protein Eng. 5*, 197−211.
15. Nardini, M., and Dijsktra, B. W. (1999) *Curr. Opin. Struct. Biol. 9*, 732−737.
16. Verschueren, K. H. G., Seljee, F., Rozeboom, H. J., Kalk, K. H., and Dijkstra, B. W. (1993) *Nature 363*, 693−698.
17. Verschueren, K. H. G., Franken, S. M., Rozeboom, H. J., Kalk, K. H., and Dijkstra, B. W. (1993) *J. Mol. Biol. 232*, 856−872.
18. Ridder, I. S., Rozeboom, H. J., and Dijkstra, B. W. (1999) *Biol. Crystallogr. 55*, 1273−1290.
19. Krooshof, G. H., Ridder, I. S., Tepper, A. W. J. W., Vos, G. J., Rozeboom, H. J., Kalk, K. H., Dijsktra, B. W., and Janssen, D. B. (1998) *Biochemistry 37*, 15013−15023.
20. Schanstra, J. P., Ridder, I. S., Heimeriks, G. J., Rink, R., Poelarends, G. J., Kalk, K. H., Dijkstra, B. W., and Janssen, D. B. (1996) *Biochemistry 35*, 13186−13195.
21. Schindler, J. F., Naranjo, P. A., Honaberger, D. A., Chang, C. H., Brainard, J. R., Vanderberg, L. A., and Unkefer, C. J. (1999) *Biochemistry 38*, 5772−5778.
22. Schanstra, J. P., and Janssen, D. B. (1996) *Biochemistry 35*, 5624−5632.
23. Krooshof, G. H., Floris, R., Tepper, A., and Janssen, D. B. (1999) *Protein Sci. 8*, 355−360.
24. Pries, F., Kingma, J., Pentega, M., VanPouderoyen, G., Jeronimus-Stratingh, C. M., Bruins, A. P., and Janssen, D. B. (1994) *Biochemistry 33*, 1242−1247.
25. Pries, F., Kingma, J., and Janssen, D. B. (1995) *FEBS Lett. 358*, 171−174.
26. Pries, F., Kingma, J., Krooshof, G. H., Jeronimus-Stratingh, C. M., Bruins, A. P., and Janssen, D. B. (1995) *J. Biol. Chem. 270*, 10405−10411.
27. Hynkova, K., Nagata, Y., Takagi, M., and Damborsky, J. (1999) *FEBS Lett. 446*, 177−181.
28. Kennes, C., Pries, F., Krooshof, G. H., Bokma, E., Kingma, J., and Janssen, D. B. (1995) *Eur. J. Biochem. 228*, 403−407.
29. Schanstra, J. P., Ridder, A., Kingma, J., and Janssen, D. B. (1997) *Protein Eng. 10*, 53−61.
30. Pries, F., VandenWijngaard, A. J., Bos, R., Pentenga, M., and Janssen, D. B. (1994) *J. Biol. Chem. 269*, 17490−17494.
31. Krooshof, G. H., Kwant, E. M., Damborsky, J., Koca, J., and Janssen, D. B. (1997) *Biochemistry 36*, 9571−9580.
32. Damborsky, J., Kuty, M., Nemec, M., and Koca, J. (1997) *J. Chem. Inf. Comput. Sci. 37*, 562−568.
33. Lightstone, F. C., Zheng, Y.-J., Maulitz, A. H., and Bruice, T. C. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 8417−8420.
34. Kuty, M., Damborsky, J., Prokop, M., and Koca, J. (1998) *J. Chem. Inf. Comput. Sci. 38*, 736−741.
35. Damborsky, J. (1998) *Pure Appl. Chem. 70*, 1375−1383.
36. Lightstone, F. C., Zheng, Y. J., and Bruice, T. C. (1998) *J. Am. Chem. Soc. 120*, 5611−5621.
37. Damborsky, J., and Koca, J. (1999) *Protein Eng. 12*, 989−998.
38. Damborsky, J., Kuty, M., Nemec, M., and Koca, J. (1997) in *Quantitative Structure-Activity Relationships in Environmental Sciences* (Chen, F., and Schüürmann, G., Eds.) Vol. VII, pp 5−20, SETAC Press, Pensacola, FL.
39. Schanstra, J. P., Kingma, J., and Janssen, D. B. (1996) *J. Biol. Chem. 271*, 14747−14753.
40. Wold, S., Johansson, E., and Cocchi, M. (1993) in *3D QSAR in Drug Design: Theory, Methods and Application* (Kubinyi, H., Ed.) pp 523−550, ESCOM, Leiden, The Netherlands.
41. Geladi, P., and Kowalski, B. R. (1986) *Anal. Chim. Acta 185*, 1−17.
42. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
43. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M. W., Johnson, B. G., Robb, M. A., Cheeseman, J. R., Keith, T., Petersson, G. A., Montgomery, J. A., Raghavachari, K., Al-Laham, M. A., Zakrzewski, V. G., Ortiz, J. V., Foresman, J. B., Cioslowski, J., Stefanov, B. B., Nanayakkara, A., Challacombe, M., Peng, C. Y., Ayala, P. Y., Chen, W., Wong, M. W., Andrés, J. L., Replogle, E. S., Gomperts, R., Martin, R. L., Fox, D. J., Binkley, J. S., Defrees, D. J., Baker, J., Stewart, J. P., Head-Gordon, M., González, C., and Pople, J. A. (1995) *Gaussian94*, Gaussian, Inc., Pittsburgh, PA.
44. Bayly, C. I., Cieplak, P., Cornell, W. D., and Kollman, P. A. (1993) *J. Phys. Chem. 97*, 10269−10280.
45. Cornell, W. D., Cieplak, P., Bayly, C. I., and Kollman, P. A. (1993) *J. Am. Chem. Soc. 115*, 9620−9631.
46. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., Ferguson, D. M., Seibel, G. L., Singh, C., Weiner, P. K., and Kollman, P. A. (1995) *AMBER*, University of California, San Francisco.
47. Fox, T., and Kollman, P. A. (1998) *J. Phys. Chem. B 102*, 8070−8079.
48. Lide, D. R. (1997) *CRC Handbook of Chemistry and Physics*, CRC Press, Boca Raton, FL.
49. Ryckaert, J. P., Ciccoti, G., and Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327−341.
50. Case, D. A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E., III, Ross, W. S., Simmerling, C. L., Darden, T. A., Merz, K. M., Stanton, R. V., Cheng, A. L., Vincent, J. J., Crowley, M., Ferguson, D. M., Radmer, R. J., Seibel, G. L., Singh, U. C., Weiner, P. K., and Kollman, P. A. (1997) *AMBER*, version 5.0, University of California, San Francisco.

51. Vriend, G. (1990) *J. Mol. Graphics 8*, 52−56.
52. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
53. Eisenberg, D., Weiss, R. M., Terwilliger, T. C., and Wilcox, W. (1982) *Faraday Symp. Chem. Soc. 17*, 109−120.
54. Eisenberg, D., Wesson, M., and Yamashita, M. (1989) *Chem. Scr. 29A*, 217−221.
55. Gilson, M. K., Sharp, K. A., and Honig, B. H. (1987) *J. Comput. Chem. 9*, 327−335.
56. Gilson, M. K., and Honig, B. H. (1988) *Proteins: Struct., Funct., Genet. 4*, 7−18.
57. Nicholls, A., and Honig, B. H. (1991) *J. Comput. Chem. 12*, 435−445.
58. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
59. Connolly, M. L. (1983) *J. Appl. Crystallogr. 16*, 548−558.
60. Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R., and Clementi, S. (1993) *Quant. Struct.-Act. Relat. 12*, 9−20.
61. Eriksson, L., Johansson, E., and Wold, S. (1997) in *Quantitative Structure-Activity Relationships in Environmental Sciences* (Chen, F., and Schüürmann, G., Eds.) Vol. VII, pp 5−20, SETAC Press, Pensacola, FL.
62. Damborsky, J., Bohac, M., Prokop, M., Kuty, M., and Koca, J. (1998) *Protein Eng. 11*, 901−907.
63. Damborsky, J. (1997) *Quant. Struct.-Act. Relat. 16*, 126−135.
64. Verschueren, K. H. G., Kingma, J., Rozeboom, H. J., Kalk, K. H., Janssen, D. B., and Dijkstra, B. W. (1993) *Biochemistry 32*, 9031−9037.

BI010464P