

## Comparative binding energy analysis of haloalkane dehalogenase substrates: Modelling of enzyme-substrate complexes by molecular docking and quantum mechanical calculations

Jan Kmuníček<sup>1</sup>, Michal Boháč<sup>1</sup>, Santos Luengo<sup>2</sup>, Federico Gago<sup>2</sup>, Rebecca C. Wade<sup>3,4</sup> & Jiří Damborský<sup>1,\*</sup>

<sup>1</sup>National Centre for Biomolecular Research, Masaryk University, Kotlarska 2, 611 37 Brno, Czech Republic; <sup>2</sup>Department of Pharmacology, University of Alcalá, E-28871 Alcalá de Henares, Madrid, Spain; <sup>3</sup>European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnengasse 33, D-69118 Heidelberg, Germany; <sup>4</sup>European Molecular Biology Laboratory, Meyerhofstr. 1, D-69012 Heidelberg, Germany

Received 14 October 2002; Accepted 22 April 2003

### Summary

We evaluate the applicability of automated molecular docking techniques and quantum mechanical calculations to the construction of a set of structures of enzyme-substrate complexes for use in Comparative binding energy (COMBINE) analysis to obtain 3D structure-activity relationships. The data set studied consists of the complexes of eighteen substrates docked within the active site of haloalkane dehalogenase (DhIA) from *Xanthobacter autotrophicus* GJ10. The results of the COMBINE analysis are compared with previously reported data obtained for the same dataset from modelled complexes that were based on an experimentally determined structure of the DhIA-dichloroethane complex. The quality of fit and the internal predictive power of the two COMBINE models are comparable, but better external predictions are obtained with the new approach. Both models show a similar composition of the principal components. Small differences in the relative contributions that are assigned to important residues for explaining binding affinity differences can be directly linked to structural differences in the modelled enzyme-substrate complexes: (i) rotation of all substrates in the active site about their longitudinal axis, (ii) repositioning of the ring of epihalohydrines and the halogen substituents of 1,2-dihalopropanes, and (iii) altered conformation of the long-chain molecules (halobutanes and halohexanes). For external validation, both a novel substrate not included in the training series and two different mutant proteins were used. The results obtained can be useful in the future to guide the rational engineering of substrate specificity in DhIA and other related enzymes.

**Abbreviations:** COMBINE – Comparative binding energy; DhIA – haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10;  $E_a$  – activation energy;  $\Delta H$  – change in reaction enthalpy;  $K_m$  – apparent dissociation constant; PLS – Partial Least-Squares Projection to Latent Structures;  $Q^2$  – cross-validated correlation coefficient;  $R^2$  – correlation coefficient; RMSEP – root mean square error of predictions; SDEP – standard deviation of error of predictions.

### Introduction

Comparative binding energy (COMBINE) analysis is a computational technique for deriving detailed quantitative structure-activity relationships (QSAR) from a

set of three-dimensional structures of protein-ligand complexes [1]. COMBINE analysis systematically investigates the relationships between experimental binding affinities and the interaction energies for a set of enzyme-ligand complexes. COMBINE analysis was originally developed for drug design applications and nowadays it is well established as a

\*To whom correspondence should be addressed. E-mail: jiri@chemi.muni.cz

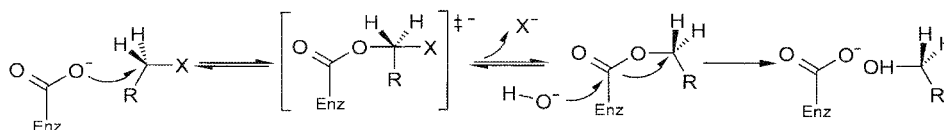


Figure 1. Scheme of the reaction mechanism of hydrolytic dehalogenation catalyzed by the haloalkane dehalogenases. Enz - enzyme.

standard 3D-QSAR method [1–12]. The application of this methodology in the protein-engineering field was first described in a study of the substrate specificity of haloalkane dehalogenase Dh1A (Figure 1) from *Xanthobacter autotrophicus* GJ10 [8]. As the crystal structure of the Dh1A-dichloroethane complex was known [13], the Michaelis complexes for the other substrates studied could be constructed by structural alignment to this structure [8]. However, a broader application of COMBINE analysis can be hindered by the fact that there is a large number of enzymes for which the experimental structure of an enzyme-substrate complex is unknown and positioning of the substrate in the active site may not be straightforward. This problem can become especially important for broad-specificity enzymes with a large active site in which a large variety of ligands can be accommodated in different binding modes. This is the case, for example, for the haloalkane dehalogenases DhaA from *Rhodococcus* sp. [14] and LinB from *Sphingomonas paucimobilis* UT26 [15], both of which have active sites between 2 and 2.5 times larger than that in Dh1A [16].

The problem outlined above indicates the necessity for a method capable of finding favorable orientations for a set of substrates inside the active site of an enzyme that will lead to the construction of robust QSAR models by COMBINE analysis. Automated molecular docking methods may provide a solution to this problem but validation is an important issue. The purpose of this study is to evaluate the applicability of a well-established docking program for the preparation of structures of Dh1A-substrate complexes and complementary quantum mechanical calculations for the selection of binding modes that are suitable for a COMBINE-type analysis.

## Methods

### Experimental data

Apparent dissociation constants ( $K_m$ ) were taken as a measure of binding affinities for a set of eighteen Dh1A substrates. The  $K_m$  values determined by

Table 1. Apparent dissociation constants (in mM) of haloalkane dehalogenase Dh1A<sup>a</sup>

No	Compound	$K_m$
1	1-chlorobutane	2,2
2	1-chlorohexane	1,4
3	1-bromobutane	0,06
4	1-bromohexane	0,3
5	1,1-dichloromethane	100
6	1,2-dichloroethane	0,53
7	1,1-dibromomethane	2,4
8	1,2-dibromoethane	0,01
9	1,2-dichloropropane	13
10	1,2-dibromopropane	1,3
11	2-chloroethanol	400
12	2-bromoethanol	11
13	epichlorohydrine	48
14	epibromohydrine	2,2
15	2-chloroacetonitrile	6,3
16	2-bromoacetonitrile	0,49
17	2-chloroacetamide	100
18	2-bromoacetamide	20

<sup>a</sup>Schanstra, J.P., Kingma, J. and Janssen, D.B., *J. Biol. Chem.*, 271 (1996) 14747.

Schanstra *et al.* [17] were logarithmically transformed (Table 1). 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.

### Modelling of enzyme-substrate complexes by automated molecular docking

The structures of the enzyme-substrate complexes were modelled using the program AUTODOCK 3.0 [18] which consists of three main separate modules (AUTOTORS, AUTOGRID, AUTODOCK) and a set of additional scripts. Initially, program AUTOTORS [18] was applied to the eighteen substrates to specify their rotatable bonds. The crystal structure of the Dh1A-dichloroethane complex (PDB-ID 2DHC) was obtained from the Brookhaven Protein Database [19]. Polar hydrogen atoms were added using the program WHATIF 5.0 [20]. His289 was singly protonated on

Table 2. Activation energy barriers and reaction enthalpies (in kcal/mol) for the substrates with multiple orientations

No	Compound	Orientation	$E_a$	$\Delta H$	$R^2$ <sup>d</sup>	$Q^2$ <sup>d</sup>	SDEP
1	1-chlorobutane	1_1 <sup>a</sup>	21,5	-10,8	0,97	0,95	0,7059
		1_2	22,9	-19,1	0,99	0,94	0,4575
		1_3	22,6	-16,2	0,98	0,93	0,6758
2	1-chlorohexane	2_1	23,1	-10,5	0,42	0,25	1,2917
		2_2 <sup>a</sup>	23,4	-6,0	0,98	0,96	1,0403
		2_3	25,5	-11,6	0,41	0,24	1,2946
3	1-bromobutane	3_1 <sup>a</sup>	23,0	-7,0	0,98	0,95	0,6882
		3_2	25,2	-8,7	0,99	0,96	0,4875
		3_3	26,1	-8,2	0,98	0,96	0,4887
4	1-bromohexane	4_1	28,6	-8,3	0,44	0,31	1,2710
		4_2 <sup>a</sup>	26,3	-7,3	0,98	0,94	1,1776
10	1,2-dibromopropane	10_1 <sup>b</sup>	31,2	0,8	0,99	0,94	0,6684
		10_2 <sup>ac</sup>	24,9	-9,1	0,99	0,66	0,7457

<sup>a</sup>orientation selected for COMBINE analysis.

<sup>b</sup>Dehalogenation from  $C_\beta$ .

<sup>c</sup>Dehalogenation from  $C_\alpha$ . <sup>d</sup>Parameter after adding this orientation to the model based on single-orientation substrates.

$N_8$  in accordance with its catalytic function. Non-polar hydrogen atoms were added using AMBER 5.0 [21]. The script *q.kollua* was used for addition of partial charges on all atoms of the enzyme and the script *addsol* was used to assign solvation parameters to the carbon atoms in the protein structure [18]. Grid maps were calculated for the atom types present in the substrates using the AUTOGRID [18] program with a grid of  $81 \times 81 \times 81$  points and a grid spacing of 0.25 Å. For docking, a Lamarckian Genetic Algorithm [18] was employed with a population of 50 individuals, a maximum number of  $1.5 \times 10^6$  energy evaluations, a maximum number of generations of 27 000, an elitism value of 1, a mutation rate of 0.02, and a cross-over rate of 0.80. The local search was based on a pseudo-Solis and Wets algorithm [22] with a maximum of 300 iterations per local search. Fifty docking runs were performed for each enzyme-substrate complex. Calculated substrate orientations from each run were clustered with the clustering tolerance for the root-mean-square positional deviation set to 0.5 Å. The geometry of the selected enzyme-substrate complexes was optimized using the molecular mechanics program AMBER and the Cornell *et al.* force field [23]. 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  $0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ . A non-bonded

cutoff of 10 Å and a distance-dependent dielectric constant ( $\epsilon = 4r_{ij}$ ) were used.

#### *Evaluation of multiple substrate orientations by quantum mechanics*

A simplified representation of the enzyme active site was used consisting of selected amino acid residues and the bound substrate molecule as obtained from the molecular docking protocol described above. The size of the system was selected in such a way that the substrate molecule was completely surrounded by the amino acid residues. The final cavity consisted of twenty amino acid residues because preliminary tests showed that a smaller cavity did not properly describe the reaction's energy profile. The following residues were included in the calculations: Glu56, Asp124, Trp125, Phe128, Leu129, Phe164, Val165, Phe172, Trp175, Phe190, Trp194, Phe222, Pro223, Met225, Val226, Asp260, Leu262, Leu263, His289 and Phe290. The total charge of the system was set to  $-3e$ . The semi-empirical quantum chemical program MOPAC 2000 [24], with subroutine DRIVER 1.0 [25] interfaced to the program TRITON 2.0 [26], was used for mapping the reaction pathway of the first dehalogenation step, as described previously [27]. Briefly, the  $S_N2$  reaction was modelled by a step-wise (0.05 Å) reduction of the distance between the nucleophilic oxygen from Asp124 and the attacked carbon on each substrate. After each step, the structure

Table 3. External predictions of apparent dissociation constants (in mM) for haloalkane dehalogenase mutants

No <sup>a</sup>	Substrate	wt	Phe172Trp		
		experiment <sup>b</sup>	experiment <sup>b</sup>	prediction <sup>d</sup>	prediction <sup>e</sup>
6'	1,2-dichloroethane	0,53	5,13	7,94	5,13
8'	1,2-dibromoethane	0,01	0,03	0,19	0,06
19'	1-bromo-2-chloroethane	0,07	0,10	1,20	0,55
		wt	Trp175Tyr		
		experiment <sup>c</sup>	experiment <sup>c</sup>	prediction <sup>d</sup>	prediction <sup>e</sup>
6''	1,2-dichloroethane	0,53	2,85	0,74	5,50
8''	1,2-dibromoethane	0,01	0,06	0,03	0,10

<sup>a</sup>A single prime corresponds to Phe172Trp, and a double prime corresponds to Trp175Tyr.

<sup>b</sup>Schanstra, J.P., et al., *Biochemistry*, 35 (1996) 13186.

<sup>c</sup>Krooshof, G.H., et al., *Biochemistry*, 37 (1998) 15013.

<sup>d</sup>Predicted using model<sub>p</sub>.

<sup>e</sup>Predicted using model<sub>N</sub>.

was fully optimized except for the driven coordinate and the heavy atoms of the protein backbone. The following keywords were used to control the MOPAC calculation: AM1, DEBUG, BFGS, GEO-OK, MMOK, MOZYME, NODIIS, NOINTER, NOXYZ and CHARGE.

#### *Construction of the COMBINE models by partial least-squares projection to latent structures (PLS)*

In COMBINE analysis, the computed molecular mechanics interaction energy for each energy minimized complex is decomposed into terms according to physical property and location, in this case, on a per residue basis. As the total interaction energy will not in general correlate with binding affinity, PLS analysis is used to derive a QSAR model in which log  $K_m$  values are correlated with the sum of selected weighted energy terms. The PLS analysis thus permits identification and ranking of interactions important for the differences in dissociation constant between the complexes, and provides a model that can be used to predict dissociation constants for new enzyme-substrate complexes. The program Q2 4.5.11 (Multivariate Infometric Analysis, Italy) was used for data pre-treatment and building of PLS models. The matrix of  $X$  variables consisted of 620 columns (van der Waals and electrostatic energy contributions for 310 amino acid residues) and 15 rows (enzyme-substrate complexes). The dependent variable  $y$  was represented by 15 logarithmically transformed values of the apparent dissociation constants  $K_m$ . All data were block-unscaled and variables with low magnitude en-

ergies (sum of squares lower than  $10^{-7}$  kcal/mol) were eliminated before the analysis. The program SIMCAP 8.0 (Umetri, Sweden) was used for permutation validation and external prediction of binding affinities for mutant enzymes. The quality of the models was described by the correlation coefficient ( $R^2$ ) and by the cross-validated correlation coefficient ( $Q^2$ ).  $R^2$  is a descriptor of the quality of fit and takes values up to a maximum of 1, which corresponds to a perfect fit. A value higher than 0.5 is generally considered as statistically significant.  $Q^2$ , which provides an estimate of the predictive power of a model, was computed using a Leave-One-Out cross-validation procedure. A value higher than 0.4 is generally considered as statistically significant.

## Results

### *Modelling of the enzyme-substrate complexes*

Individual substrate molecules were docked automatically into the active site of Dh1A (Figure 2). The automated docking procedure provided mechanistically viable orientations for fifteen out of eighteen substrate molecules. No biologically relevant substrate orientations inside the active site were found for 2-bromoacetonitrile and the two dihaloacetamides. The substrate molecules with satisfactory orientations can be further divided into two groups. The first group contains the substrates for which only a single orientation suitable for nucleophilic attack was found: 2-chloroacetonitrile, 1,2-dichloropropane,

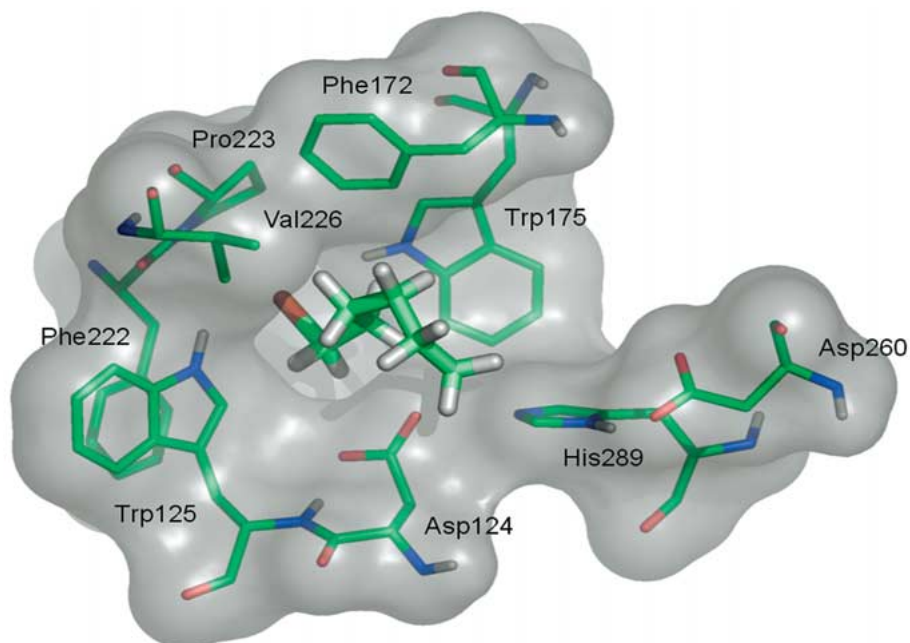


Figure 2. Active site cavity of DhIA with docked substrate 1-bromohexane. The electrophilic carbon of 1-bromohexane is oriented towards the nucleophilic oxygen of Asp124 in a position suitable for  $S_N2$  attack. Asp124 together with Asp260 and His289 constitutes the catalytic triad (13). The scissile bond is pointing towards the halide-binding pocket made of primary halide-stabilizing residues Trp125 and Trp175, and secondary halide-stabilizing residues Phe172, Pro223 and Val226 (35).

dihalomethanes, dihaloethanes, haloethanols and epihalohydrines. The second group is composed of substrates for which several satisfactory orientations were found: 1,2-dibromopropane, halobutanes and halohexanes. The two binding modes that were obtained for 1,2-dibromopropane represented orientations suitable for dehalogenation from either the  $\alpha$  or the  $\beta$  carbon atom.

#### Selection of correct binding modes

The orientations of 1,2-dibromopropane, the halobutanes and the halohexanes were further studied by quantum-mechanical calculations to select one unique enzyme-substrate complex per molecule. Semi-empirical modelling of the  $S_N2$  dehalogenation reaction was conducted for each of these enzyme-substrate complexes (Table 2). The change in heat of formation of the system during the reaction was taken as a criterion to select the binding mode with the most appropriate geometry for nucleophilic substitution, i.e. the initial reaction step. The activation energy ( $E_a$ ) of the reaction was approximated by the difference in heat of formation between the Michaelis complex and the transition state structures. The enthalpy change ( $\Delta H$ ) was expressed as the difference in heat of form-

ation between the Michaelis complex and the enzyme-product structures. The activation energy barriers and the reaction enthalpies are listed in Table 2. Stabilization of the released halide anion by two tryptophan residues located in the enzyme active site (Trp125 and Trp175) is critical for the first reaction step of DhIA, as demonstrated in previous experimental [28–31] and theoretical [27, 32–35] studies. The magnitude of this stabilization, as quantitated by  $\Delta H$ , was therefore the first criterion for evaluation of different substrate orientations. If tryptophan stabilization was missing, then  $\Delta H$  displayed more positive values. Stabilization was fulfilled by all binding modes except the mode that assumed the nucleophilic attack on the  $\beta$  carbon atom of 1,2-dibromopropane (orientation 10\_1). For this orientation, a positive enthalpy change of the reaction (0.8 kcal/mol) was found. Orientation 10\_2, which assumes dehalogenation from the  $\alpha$  carbon atom of 1,2-dibromopropane, was therefore selected for the COMBINE analysis. The orientations with the lowest activation barrier (1\_1, 3\_1 and 4\_2) were selected for inclusion in the COMBINE model for substrates 1-chlorobutane, 1-bromobutane and 1-bromohexane, respectively. Two out of three binding modes for 1-chlorohexane (2\_1 and 2\_2) showed similar activ-

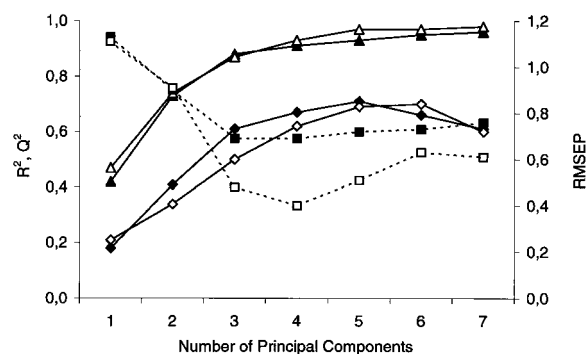


Figure 3. Comparison of correlation coefficient ( $R^2$ , triangles), cross-validated correlation coefficient ( $Q^2$ , diamonds) and root-mean-square error of prediction (RMSEP, squares) for model<sub>P</sub> (filled objects) and model<sub>N</sub> (empty objects).

ation barriers (23.1 and 23.4 kcal/mol) and therefore each mode was independently explored in alternative datasets that contained the remaining substrates with unambiguous orientation. This dual modelling was used as an additional selection criterion (Table 2). The COMBINE model constructed for the substrates with single orientation (dihalomethanes, dihaloethanes, haloethanols and epihalohydrines) and the 2\_1 binding mode showed significantly worse statistics ( $R^2 = 0.42$  and  $Q^2 = 0.25$ ) compared to the model with the 2\_2 orientation ( $R^2 = 0.98$  and  $Q^2 = 0.96$ ).

#### Construction of the COMBINE model

An initial COMBINE model was derived for the fifteen molecules for which a suitable enzyme-substrate complex could be obtained by the automated molecular docking procedure. In cases that resulted in more than one binding mode, the mode selected based on the quantum-mechanical results was chosen. It was revealed that simultaneous inclusion of dihalopropanes and haloethanes into the training set led to unstable models. Each of these pairs have different locations in the active site. The dihalopropanes bind near the entrance tunnel whereas the haloethanes penetrate deeper into the binding site of the protein and are oriented towards the nucleophile. Fixation of the haloethane molecules in their docked position during molecular mechanics energy refinement of the enzyme-substrate complexes solved this problem and enabled inclusion of both halopropanes and haloethanes in the final model. This fixation stopped translation of haloethanes in the active site and retained the favorable position obtained from the automated docking procedure.

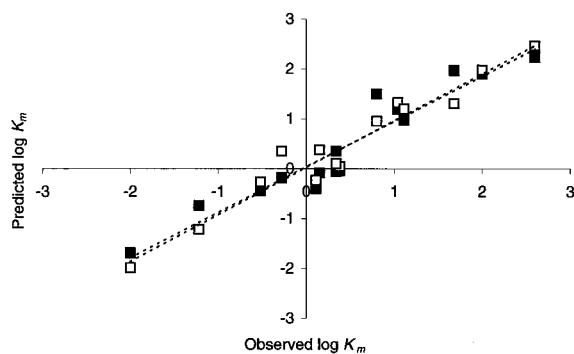


Figure 4. Comparison of observed versus predicted  $K_m$  values for model<sub>P</sub> (filled squares) and model<sub>N</sub> (empty squares).

## Discussion

In order to validate the applicability of automated molecular docking techniques to the preparation of enzyme-substrate complexes and the use of quantum-mechanical calculations for selection of the most suitable binding modes, the newly constructed model (model<sub>N</sub>) was compared to our previously reported COMBINE model (model<sub>P</sub>) [8] that was based on a manual superposition of the substrates onto the experimentally determined structure of 1,2-dichloroethane in its complex with Dh1A. To allow direct comparison of the models, model<sub>P</sub> was re-calculated with the same fifteen substrates that are present in model<sub>N</sub>. The models were compared in terms of their statistics, important variables and structural features of the enzyme-substrate complexes.

#### Statistical comparison of COMBINE models

Selected statistical criteria, i.e., correlation coefficient,  $R^2$ ; cross-validated correlation coefficient,  $Q^2$ ; and root-mean-square error of prediction, RMSEP are compared in Figure 3. The quality of fit for the four-component models is shown in Figure 4. Both approaches provided models with very similar statistics, except for a significantly improved predictive power for model<sub>N</sub>, as indicated by the lower RMSEP values (in models with more than two principal components). The better predictive ability of model<sub>N</sub> was also demonstrated for an external test set involving two different mutant proteins and a set of five halogenated substrates that included the novel substrate **19'** (1-bromo-2-chloroethane) which was not present in the original training set (Table 3).

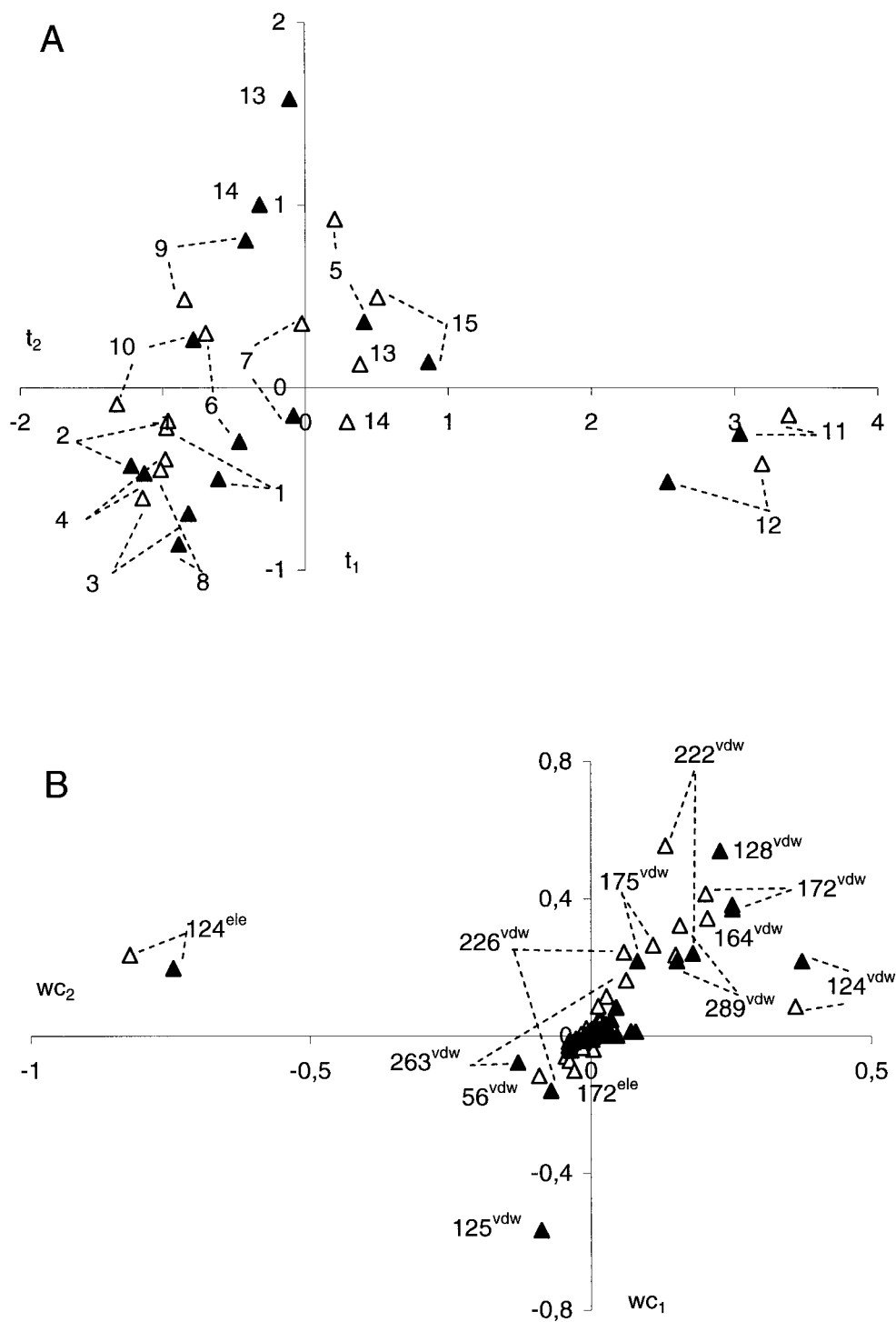


Figure 5. Comparison of score (A) and loading (B) plots for the first and second principal components of model<sub>p</sub> (filled triangles) and model<sub>N</sub> (empty triangles). The compounds are numbered according to Table 1 and energy contributions are numbered according to DhIA sequence.

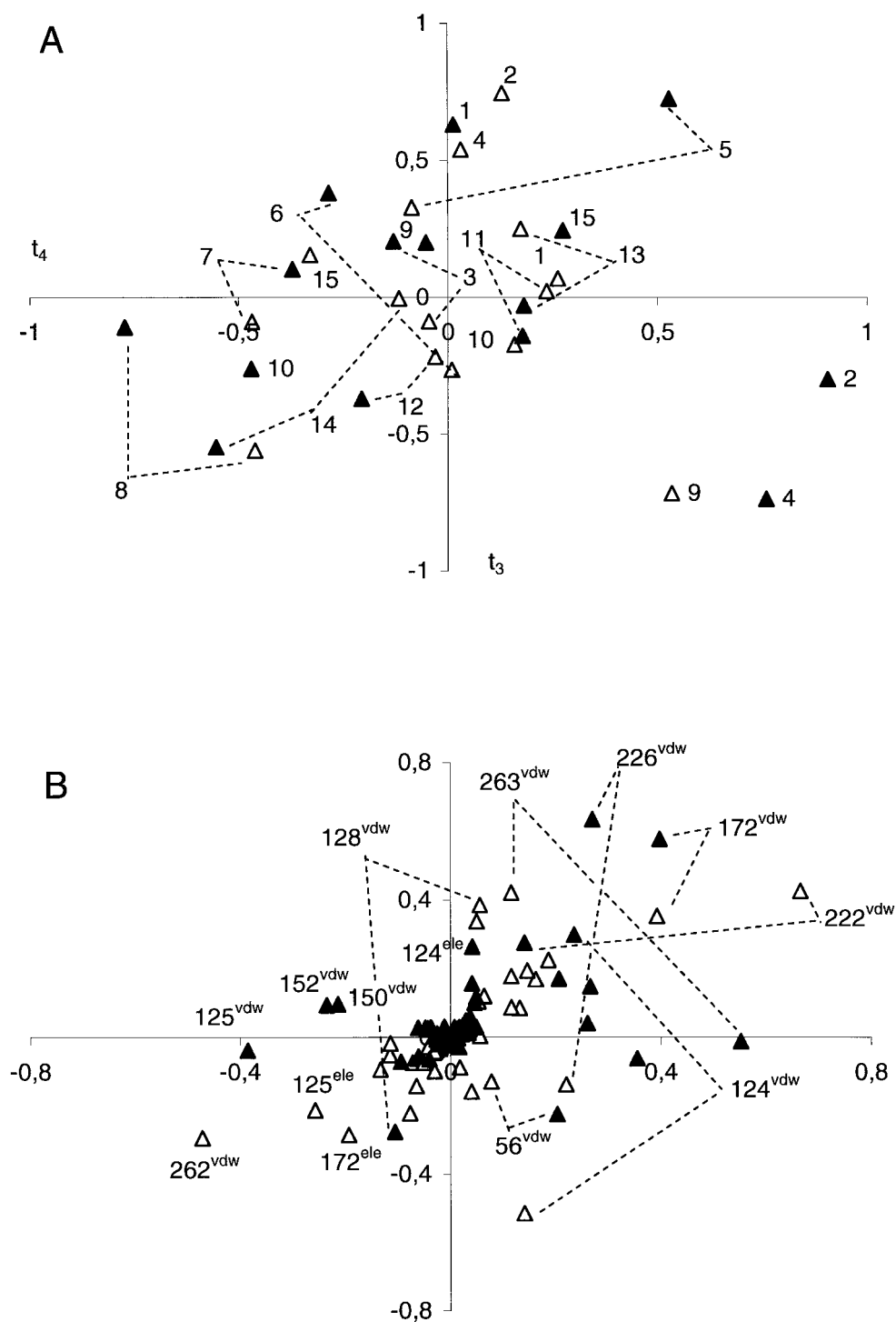


Figure 6. Comparison of score (A) and loading (B) plots for the third and fourth principal components of  $model_P$  (filled triangles) and  $model_N$  (empty triangles). The compounds are numbered according to Table 1 and energy contributions are numbered according to DhIA sequence.

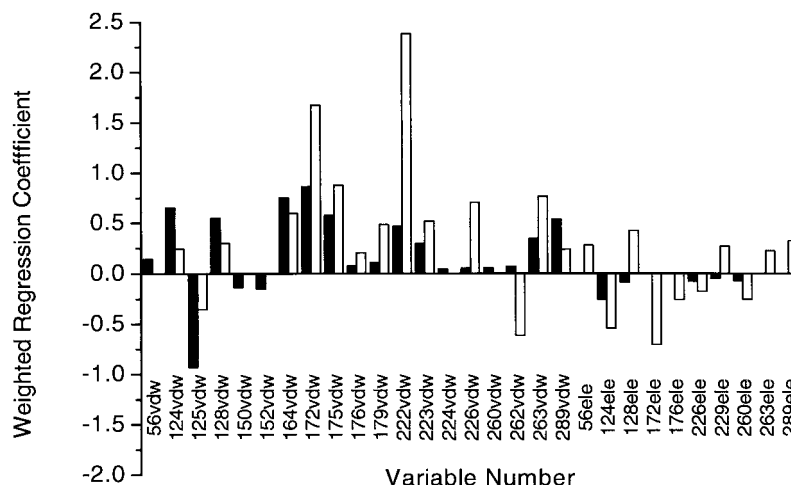
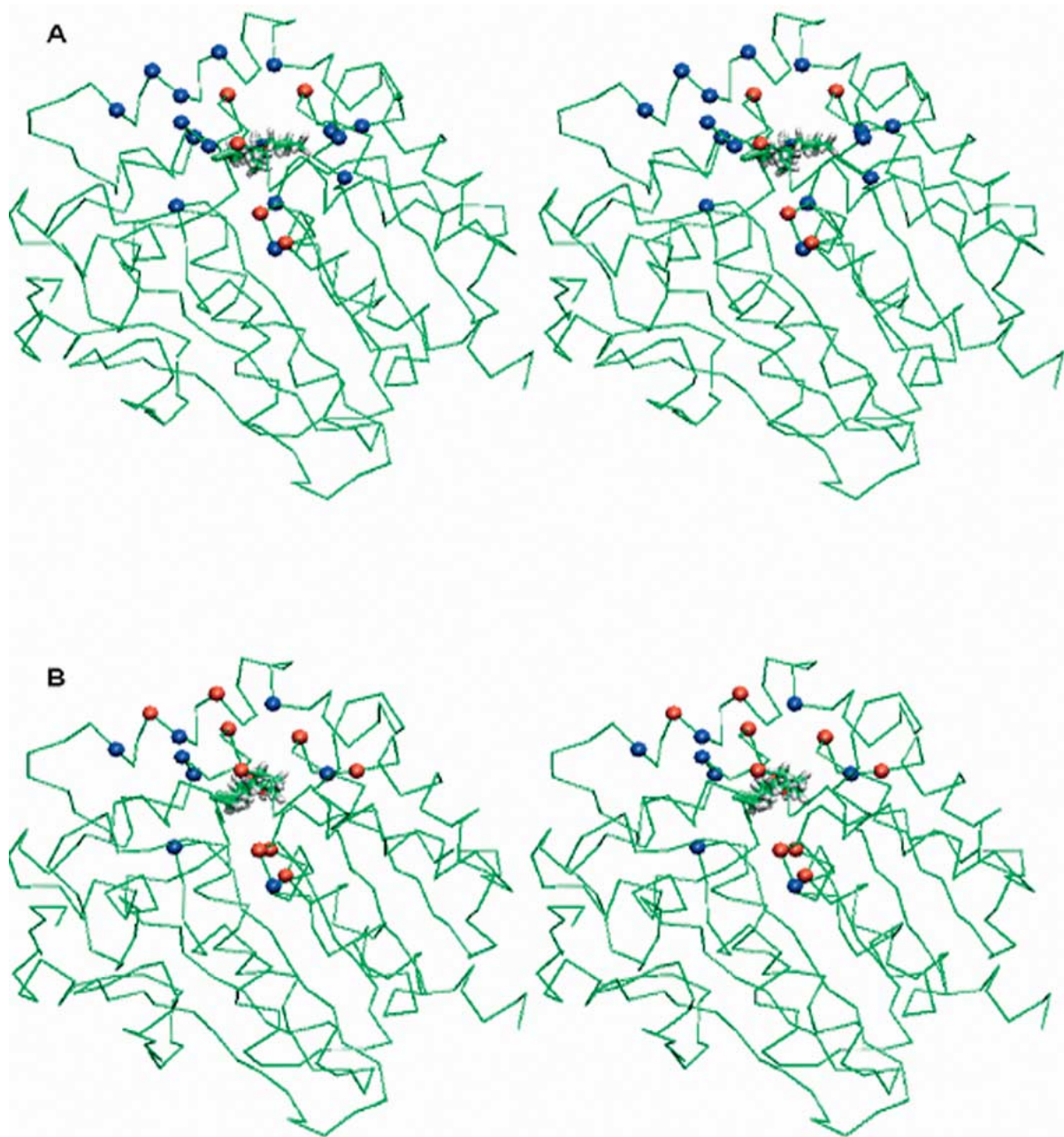


Figure 7. Comparison of the weighted regression coefficients for model<sub>p</sub> (filled rectangles) and model<sub>N</sub> (empty rectangles). Energy contributions are numbered according to the DhIA sequence.

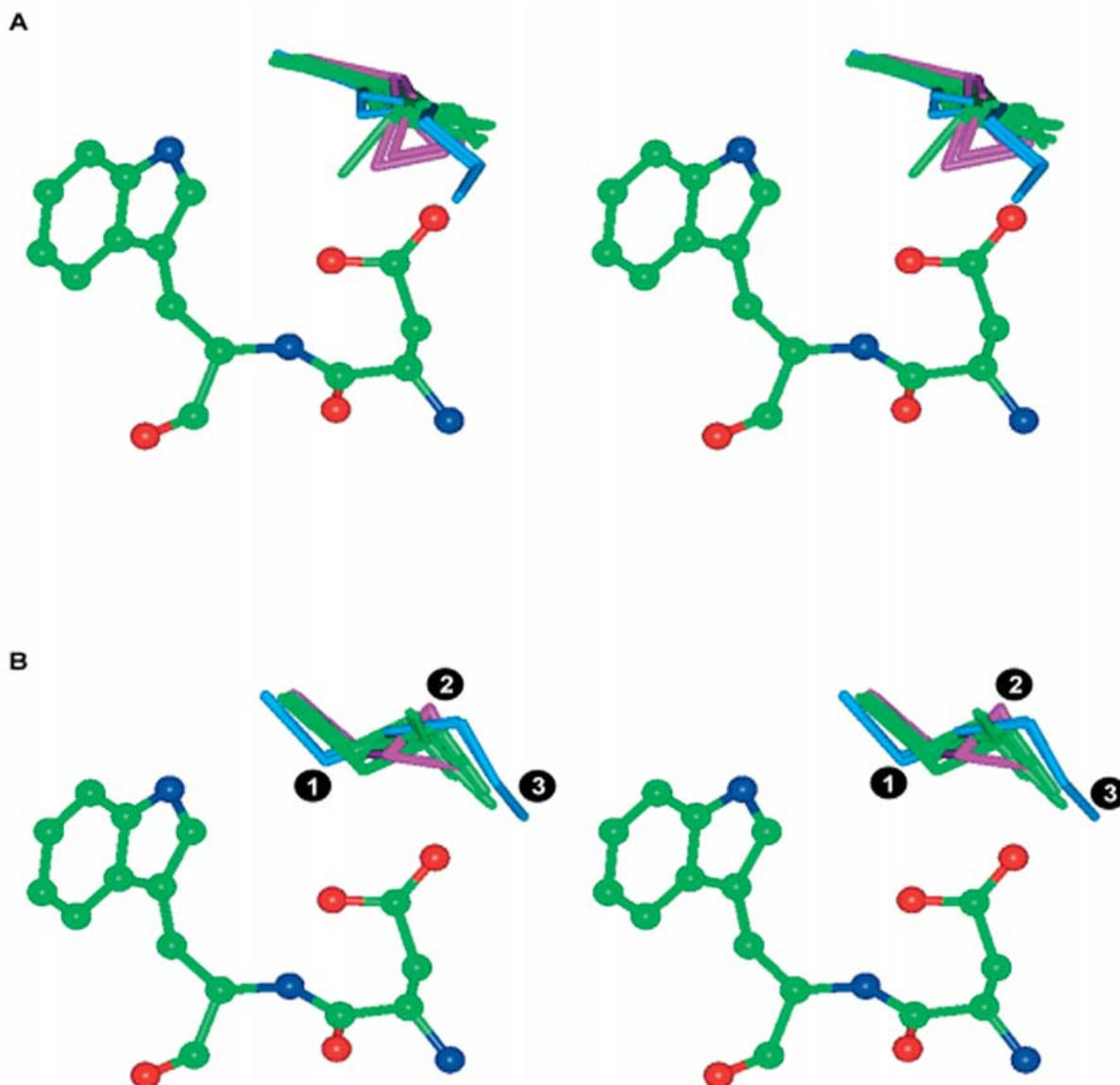
### Chemometric comparison of the COMBINE models

Score plots of the first two components (Figure 5A) show that the majority of substrates are positioned similarly in both sets of complexes. The only exceptions are both epihalohydrine molecules (compounds **13** and **14**) which have their epoxy rings re-orientated in model<sub>N</sub>. Haloethanols (compounds **11** and **12**) are the two most polar substrates and their binding affinities are described by the first principal component (Figure 5A), which is mainly determined by the electrostatic interaction energy with Asp124 (Figure 5B). There is a hydrogen bond present between the hydroxyl groups of the haloethanols and the nucleophilic oxygen of Asp124. Asp124 also shows significant favourable electrostatic interactions with other polar substrates: 2-epichlorohydrine, 2-epibromohydrine and 2-chloroacetonitrile (compounds **13**, **14** and **15**). The van der Waals energy contributions emanating from Phe222, Phe172 and Phe164 dominate the second principal component. The score plots of the third and fourth principal components (Figure 6A) differ significantly between the two models with regard to the positions of the haloethanols and dihalopropanes (compounds **2**, **4**, **9** and **10**). This is due to the different conformations that these molecules adopt in the enzyme active site. The main contributors to the third principal component (Figure 6B) are the van der Waals interactions involving residues Phe172, Phe222, Leu262, while the fourth principal component is primarily made up of van der Waals energy contributions from Asp124, Phe222 and Leu263. Phe172 and Phe222 make favourable van der Waals

interactions with the halogen atoms of every substrate molecule in model<sub>N</sub>. Residues Leu262 and Leu263, on the other hand, make favourable van der Waals interactions only with selected substrates, that is, the haloethanols (compounds **1** and **3**). The allowed flexibility of the substrate molecules during the docking calculation permits turning of the tail of the long carbon chains in haloethanols and haloethanols, thus resolving the steric hindrance of these molecules with Leu263 in model<sub>p</sub>. Two van der Waals energy contributions that are significant in model<sub>p</sub> (Cys150 and Met152) do not have an analogous counterpart in model<sub>N</sub> (Figure 6B). This is due to the direct contact of the carbon tail of haloethanols with Cys150 and Met152 in model<sub>p</sub> that is missing in model<sub>N</sub>. These missing van der Waals interactions are better seen on the graph of weighted regression coefficients (Figure 7). The same graph also reveals non-negligible differences in the importance of electrostatic interactions between the two models. The electrostatic contributions from residues Glu56, Phe128, Phe172, Phe176, Leu263, and His289 are significant only in model<sub>N</sub>. The bulk of these electrostatic contributions are due to interactions between active site residues and the epoxide ring of epihalohydrines. These are the only polar molecules that are exposed to interactions with residues occupying the space of the active site under the cap domain in model<sub>N</sub>. The described electrostatic interactions with epihalohydrines have a long-range character, as seen from the distribution of interacting residues within the structure of DhIA (Figure 8).



*Figure 8.* Comparison of the energy contributions in model<sub>P</sub> (A) and model<sub>N</sub> (B). The  $C_{\alpha}$  atoms of the residues showing the most important van der Waals energy contributions (blue) and the residues showing both van der Waals and electrostatic energy contributions (red) in each model are shown as balls.



*Figure 9.* Comparison of the orientations and conformations of substrate molecules in the active site of Dh1A used for construction of model<sub>P</sub> (A) and model<sub>N</sub> (B). Two amino acid residues of Dh1A (the nucleophile Asp124 and the halide-stabilizing Trp125) are shown in ball-and-stick representation. The numbers indicate three major differences between both sets of docked structures: (i) all substrates are rotated along their longitudinal axis, (ii) the rings of epihalohydrines (in purple) and the halogen substituents of 1,2-dihalopropanes are repositioned, and (iii) halohexanes (in blue) change the conformation of their carbon chain.

#### *Structural comparison of the COMBINE models*

The structures of Dh1A-substrate complexes obtained through automated molecular docking show three major differences when compared to the structures from the previous manual alignment: (i) all substrates without exception are rotated along their longitudinal axis, (ii) the rings of epihalohydrines and the halogen substituents of 1,2-dihalopropanes are repositioned, and (iii) long-chain molecules (halobutanes and halohexanes) change the conformation of their carbon chain (Figure 9). The substrates in the complexes obtained by means of the automated molecular docking program are generally adjusted to the size and shape of the active site and their orientations are in good agreement with the expected reactive conformation for the  $S_N2$  reaction. Rotation of the

tioned, and (iii) long-chain molecules (halobutanes and halohexanes) change the conformation of their carbon chain (Figure 9). The substrates in the complexes obtained by means of the automated molecular docking program are generally adjusted to the size and shape of the active site and their orientations are in good agreement with the expected reactive conformation for the  $S_N2$  reaction. Rotation of the

substrate molecules along their longitudinal axes positions their electrophilic carbon atoms closer (by about 1 Å) to the attacking oxygen of Asp124 while retaining the released halide ion in a position intermediate between two stabilizing amino acids, Trp125 and Trp175 (Figure 9). This rotation places the substrate molecules in both a position and a conformation (e.g., *gauche* for 1,2-dichloroethane) perfectly suited for S<sub>N</sub>2 displacement. In line with this observation, the geometry of the Michaelis complexes evolved smoothly to that of the enzyme-product complexes without any additional conformational change during the quantum mechanical calculations. The only exceptions were the 1,2-dihalopropanes, which changed their conformation from *trans* to *gauche* during the S<sub>N</sub>2 displacement reaction (*gauche* conformations were never obtained from the docking calculations). Epihalohydrines and 1,2-dihalopropanes were docked in orientations such that their favorable interactions with surrounding active site residues were maximized. Long-chain substrates adopted conformations that were adjusted to the relatively small size of the enzyme active site. The tail of the carbon chain distant from the cleaved carbon-halogen bond is rotated to prevent steric hindrance with the active site residues (Figure 9).

## Conclusions

An automated molecular docking procedure provided structures for a set of Dh1A-substrate complexes that was used to derive a robust COMBINE model. The procedure was not devoid of problems, however, as multiple binding modes were apparent for some substrates whereas biologically relevant binding modes were missing for others. Quantum-mechanical calculations were successfully used as an additional and complementary computational tool for selection of suitable binding modes, even though we note that this approach may not be equally applicable to enzyme inhibitors or other protein ligands. Fixation of the substrate molecule in the docked conformation during structural refinement of the enzyme-substrate complexes by energy minimization was found to be a useful alternative to full minimization in certain cases. Comparison of the present COMBINE model with a previous one based on an experimentally derived enzyme-substrate structure solved at 2.4 Å resolution and manual superposition of the substrates revealed that, despite significant differences in substrate ori-

entations and conformations, both models are similar in terms of overall fit and internal predictive power. It thus appears that positioning of the substrate molecules relative to each other is more crucial for construction of robust COMBINE models than absolute positioning and orientation of the substrates inside the active site. However, the new COMBINE model derived from the automatically docked structures, performed notably better in external prediction.

We expect that COMBINE-type models can aid in the rational engineering of substrate specificity of Dh1A and related dehalogenases. For this purpose, the active site residues of Dh1A can be conveniently divided into two groups. The first group contains the residues that are essential for catalysis: the catalytic triad (Asp124, Asp260 and His289), the oxyanion hole (Glu56 and Trp125) and the primary halide-stabilizing residues (Trp125 and Trp175). Modification of these residues is not recommended because it could lead to enzyme inactivation. The second group contains the residues involved in substrate binding: Phe222, Pro223, Val226, Leu262, and Leu263 (the first shell) and Lys176, Leu176 and Arg229 (the second shell). We would expect these second-shell residues to be especially suitable targets for site-directed mutagenesis studies.

## Acknowledgements

This work was supported by the NATO Linkage Grant MTECH.LG.974701 and grants from the Czech Ministry of Education J07/98:143100005 and ME551 (JD).

## Supporting information

The datasets used for COMBINE analysis and the reaction paths calculated for multiple binding modes. This material is available via Internet at <http://ncbr.chemi.muni.cz/~jjiri>.

## References

1. Ortiz, A.R., Pisabarro, M.T., Gago, F., and Wade, R.C., *J. Med. Chem.*, 38 (1995) 2681.
2. Ortiz, A.R., Pastor, M., Palomer, A., Cruciani, G., Gago, F., and Wade, R.C., *J. Med. Chem.*, 40 (1997) 1136.
3. Pastor, M., Perez, C., and Gago, F., *J. Mol. Graph. Model.*, 15 (1997) 364.

4. Perez, C., Pastor, M., Ortiz, A.R., and Gago, F., *J. Med. Chem.*, 41 (1998) 836.
5. Lozano, J.J., Pastor, M., Cruciani, G., Gaedt, K., Centeno, N.B., Gago, F., and Sanz, F., *J. Comput.-Aid. Mol. Design*, 14 (2000) 341.
6. Tomic, S., Nilsson, L., and Wade, C.R., *J. Med. Chem.*, 43 (2000) 1780.
7. Cuevas, C., Pastor, M., Perez, C., and Gago, F., *Comb. Chem. High Through. Screen.*, 4 (2001) 627.
8. Kmunicek, J., Luengo, S., Gago, F., Ortiz, A.R., Wade, R.C., and Damborsky, J., *Biochemistry*, 40 (2001) 8905.
9. Wade, R.C. 2001. Derivation of QSARs using 3D structural models of protein-ligand complexes by COMBINE analysis, *In* Holtje, H.-D. and Sippl, W. (eds.), *Rational approaches to drug design: 13th European symposium on Quantitative Structure-Activity Relationships*, Prous Science, Barcelona, p. 23.
10. Wang, T., and Wade, R.C., *J. Med. Chem.*, 44 (2001) 961.
11. Wang, T., and Wade, R.C., *J. Med. Chem.*, 45 (2002) 4828.
12. Damborsky, J., Kmunicek, J., Jedlicka, T., Luengo, S., Gago, F., Ortiz, A.R., and Wade, R.C. 2003. Rational re-design of haloalkane dehalogenases guided by comparative binding energy analysis, *In* Svendsen, A. (ed.), *Enzyme functionality: design, engineering and screening*, Marcel Dekker, New York, in press.
13. Verschuere, K.H.G., Seljee, F., Rozeboom, H.J., Kalk, K.H., and Dijkstra, B.W., *Nature*, 363 (1993) 693.
14. Newman, J., Peat, T.S., Richard, R., Kan, L., Swanson, P.E., Affholter, J.A., Holmes, I.H., Schindler, J.F., Unkefer, C.J., and Terwilliger, T.C., *Biochemistry*, 38 (1999) 16105.
15. Marek, J., Vevodova, J., Kuta-Smatanova, I., Nagata, Y., Svensson, L.A., Newman, J., Takagi, M., and Damborsky, J., *Biochemistry*, 39 (2000) 14082.
16. Damborsky, J., and Koca, J., *Prot. Engng.*, 12 (1999) 989.
17. Schanstra, J.P., Kingma, J., and Janssen, D.B., *J. Biol. Chem.*, 271 (1996) 14747.
18. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., and Olson, A.J., *J. Comput. Chem.*, 19 (1998) 1639.
19. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E., *Nucl. Acid Res.*, 28 (2000) 235.
20. Vriend, G., *J. Mol. Graphics*, 8 (1990) 52.
21. Case, D.A., Pearlman, D.A., Caldwell, J.W., Cheatham III, T.E., 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. AMBER 5.0, University of California, San Francisco (1997).
22. Solis, F.J., and Wets, R.J.B., *Math. Oper. Res.*, 6 (1981) 19.
23. 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., *J. Am. Chem. Soc.*, 117 (1995) 5179.
24. Stewart, J.J.P., *J. Comput.-Aid. Mol. Design*, 4 (1990) 1.
25. Cernohorsky, M., Kutý, M., and Koca, J., *Comp. Chem.*, 21 (1996) 35.
26. Damborsky, J., Prokop, M., and Koca, J., *Trends Biochem. Sci.*, 26 (2001) 71.
27. Damborsky, J., Kutý, M., Nemeč, M., and Koca, J., *J. Chem. Inf. Comp. Sci.*, 37 (1997) 562.
28. Verschuere, K.H.G., Kingma, J., Rozeboom, H.J., Kalk, K.H., Janssen, D.B., and Dijkstra, B.W., *Biochemistry*, 32 (1993) 9031.
29. Kennes, C., Pries, F., Krooshof, G.H., Bokma, E., Kingma, J., and Janssen, D.B., *Eur. J. Biochem.*, 228 (1995) 403.
30. Krooshof, G.H., Ridder, I.S., Tepper, A.W.J.W., Vos, G.J., Rozeboom, H.J., Kalk, K.H., Dijkstra, B.W., and Janssen, D.B., *Biochemistry*, 37 (1998) 15013.
31. Schindler, J.F., Naranjo, P.A., Honaberger, D.A., Chang, C.-H., Brainard, J.R., Vanderberg, L.A., and Unkefer, C.J., *Biochemistry*, 38 (1999) 5772.
32. Lightstone, F.C., Zheng, Y.-J., Maulitz, A.H., and Bruice, T.C., *Proc. Natl. Acad. Sci. USA*, 94 (1997) 8417.
33. Shurki, A., Strajbl, M., Villa, J., and Warshel, A., *J. Am. Chem. Soc.*, 124 (2002) 4097.
34. Bohac, M., Nagata, Y., Prokop, Z., Prokop, M., Monincova, M., Koca, J., Tsuda, M., and Damborsky, J., *Biochemistry*, 41 (2002) 14272.
35. Devi-Kesavan, L.S., and Gao, J., *J. Am. Chem. Soc.*, 125 (2003) 1532.