Masaryk University Brno, Faculty of Science

Department of Physical Chemistry
National Center for Biomolecular Research

RATIONAL RE-DESIGN OF HALOALKANE DEHALOGENASES GUIDED BY COMPARATIVE BINDING ENERGY ANALYSIS

Ph.D. Thesis

Brno 2003 Jan Kmuníček

Supervisor: **Prof. RNDr. Jaroslav Koča, DrSc.**
National Center for Biomolecular Research
Faculty of Science, Masaryk University Brno

Supervisor-specialist: **Prof. Dr. Jiří Damborský**
National Center for Biomolecular Research
Faculty of Science, Masaryk University Brno

SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT
FACULTY OF SCIENCE
MASARYK UNIVERSITY BRNO
CZECH REPUBLIC
2003

To my parents

Who does not know,
and does not know that does not know,
is fool.
Avoid him.

Who does not know,
and knows that does not know,
is not mature.
Enlighten him.

Who knows,
and does not know that knows,
is sleeper.
Wake him up.

Who knows,
and knows that knows,
is a wise man.
Follow him.

ARABIC PROVERB

ACKNOWLEDGEMENTS

My dear parents, I would like to deeply thank you in forefront, because you have never stopped to believe I could be a decent human being. Thank you for all!

Marcelka, thank you – my lovely sweetheart – that you were by my side in good and bad time even though you did not have to. Love seems to be really fantastic. Thanks for your tolerance. This all is your credit!

Hobbit, I thank you for being as you are. You are an inspiration for me.

Professor Jaroslav Koča, thank you very much that you decided to give me a chance and accepted me in your lab. You opened for me the door leading to the exciting discovery in the world of molecular modelling and computational chemistry. I am your debtor.

Thank you, **Jiří**, that you allowed me to gain invaluable life experience and not only in scientific sphere. Honestly thank you for your great leadership! This work would have never been done without your supervision, help and patience.

All great and friendly **NCBR members**, thank you for making my stay in the center really enjoyable and fruitful by forming the greatest working environment I could ever wish. Up to now the time we shared together was simply the best period of my life!

Dr. Rebecca Wade and **Dr. Federico Gago** are gratefully acknowledged for providing my research stays in European Molecular Biology Laboratory, Heidelberg and University of Alcalá, Alcalá, respectively. Thank you very much for your kindness and support! It was a great honor for me to work under supervision of such prominent scientists!

My dear **German and Spanish friends**: Roger Abseher, Francesca De Rienzo, Lutz Ehrlich, Timm Essigke, Stephania Ferrari, Masakazu Kobayashi, Jens Linge, Susanna Luedemann, Ting Wang, Peter Winn, Alexandro Fausto, Hector Guzman, Santos Manuel Luengo Regalado, Esther Marco Tejon, Jesus

Mendieta Gomez, Fatima Rodriguez Barrios and Sonsoles Martin Santamaria, thank you for the possibility to spend wonderful time with you and to learn a lot of new life experience from each of you. I really enjoyed the time we have spent together very much! Thank you for all your kindness. Hopefully sometimes somewhere I will have chance to pay it back.

Radek, thanks for your loyal friendship and all our sessions.

Steve and **Maverick**, thanks for those zeros and ones that make my free time really chilling-out.

Petr, thanks for invaluable advices concerning computational chemistry, C and C++ programming and especially for developing molecular modelling package Nemesis that has a bright future for sure!

Michal, thanks for "memetic" discussions.

Credits concerning friendly and effective help with LaTeXing go to following LaTeXperts: **Sherry**, **Shnek**, **Cerno** and **Petr Padrta**.

Thanks to **everybody** who somehow wilfully or even unwittingly contributed to the formation of this thesis and especially to persons who does not know about his/her contribution.

Thanks go to **Supercomputing Center Brno** for providing their computational facilities and technical support.

Lastly I would like to thank my **Alma Mater** for the education I got.

Thanks go to **Šot** for inventing GOD, typographic corrections and philosophical discussions.

Zamis is extremely appreciated for never-ending support in everything I decided to perform.

ABBREVIATIONS

AA	amino acid
Ad_N	nucleophilic addition
CC	coupled clusters
CI	configuration interaction
COMBINE	comparative binding energy
CoMFA	comparative molecular field analysis
DCE	1,2-dichloroethane
DFT	density functional theory
DhaA	haloalkane dehalogenase from *Rhodococcus rhodochrous* NCIMB 13064
DhlA	haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10
DhlB	L-2-haloacid dehalogenase
E	elimination
EVB	empirical valence bond
FEP	free energy perturbation
FF	force field

FH_4	tetrahydrofolate
GC–MS	gas chromatography–mass spectrometry
GSH	glutathione
HF	Hartree–Fock method
K_m	apparent dissociation constant
LinA	γ-hexachlorocyclohexane dehydrochlorinase
LinB	haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 1,3,4,6-tetrachloro-1,4-cyclohexadiene halidohydrolase
LinC	2,5-dichloro-2,5-cyclohexadiene-1,4-diol dehydrogenase
LinD	2,5-dichlorohydroquinone reductive dehalogenase
MD	molecular dynamics
MM	molecular mechanics
MO–LCAO	molecular orbital-linear combination of atomic orbitals
MPn	Møller–Plesset perturbation method of order n
Mw	relative molecular weight
NADH, NAD^+	nicotinamide adenine dinucleotide
PES	potential energy (hyper)surface
PLS	partial least-squares projection to latent structures

Q^2	cross-validated correlation coefficient
PQQ, $PQQH_2$	ubiquinone
QM	quantum mechanics
QM/MM	quantum mechanics/molecular mechanics
QSAR	quantitative structure–activity relationship
R^2	correlation coefficient
SCF	self-consistent field
S_N2	bimolecular nucleophilic substitution
TI	thermodynamic integration

Contents

1 Motivation **19**

2 Summary **21**

3 Shrnutí **23**

I Theory **25**

4 Computational chemistry **27**
4.1 Basic terms 27
4.2 Theoretical methods 29
4.2.1 Quantum mechanics 29
4.2.2 Molecular mechanics 33
4.2.3 Simulation techniques 34
4.3 COMBINE analysis 35

5 Microbial dehalogenation **41**
5.1 Halogenated aliphatic compounds 41
5.2 Dehalogenation mechanisms 43
5.3 Haloalkane dehalogenases 45
5.3.1 DhlA enzyme 52
5.3.2 LinB enzyme 56
5.4 Protein engineering 60

II Results **63**

6 Synopsis of results **65**

7 Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10 **69**

8 Comparative Binding Energy Analysis of Haloalkane Dehalogenase Substrates: Modelling of Enzyme–Substrate Complexes by Molecular Docking and Quantum Mechanical Calculations **105**

9 Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26 **127**

10 Rational Re-design of Haloalkane Dehalogenases Guided by Comparative Binding Energy Analysis **147**

11 Exploring the Structure and Activity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26: Evidence for Product and Water Mediated Inhibition **163**

III Appendixes **205**

12 Curriculum vitae **207**

13 List of publications **211**

14 List of presentations **213**

Chapter 1

Motivation

This thesis summarizes research work focused on study of catalytic properties of enzymes participating on degradation of environmental pollutants.

The study of microbial hydrolytic dehalogenases has been largely motivated by their potential use in biotechnological treatment of waste materials, bioremediation of contaminated soil and water, and industrial biocatalysis. These enzymes play a crucial role in the biodegradation of aliphatic halogenated compounds. They can selectively catalyze the release of halide ions from their substrates with water as the sole co-substrate.

Unfortunately, the wild type enzymes often do not acquire sufficiently high activity or specificity for degradation of environmental pollutants in industrial scale. As practical biocatalysts, these enzymes suffer from low activity and too narrow substrate specificity. Therefore numerous projects were initiated to understand the molecular basis of substrate specificity and to engineer their catalytic properties. 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) analysis has been shown to be a useful technique for deriving quantitative structure-activity relationships (QSARs) from a set of three-dimensional structures of enzyme–substrate complexes. This project is aimed to apply COMBINE analysis approach to study substrate binding to haloalkane dehalogenases for the improvement of their binding abilities.

Chapter 2

Summary

Due to environmental problems arising from the production and use of halogenated hydrocarbons, the study of microbial enzymes that can catabolize degradation of these compounds is of major interest. Haloalkane dehalogenases (EC 3.8.1.5) make up one such important class of enzymes because of their ability to attack halogenated aliphatic hydrocarbons, which are produced in several industrial processes. Haloalkane dehalogenases remove halides from organic compounds via a hydrolytic mechanism that results in the production of the corresponding alcohol, halide ion, and proton. 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. The adjustment of enzyme substrate specificity toward desired substrates requires identification of the amino acid residues that are involved in substrate binding. Such information can be deduced from three-dimensional structures of enzyme–substrate complexes using quantitative structure-activity relationships (QSARs) by means of Comparative Binding Energy (COMBINE) analysis. Obtained results demonstrate that COMBINE analysis is promising tool that can be used in construction of predictive and robust QSARs using atomic detail models of the enzyme–substrate complexes. The models provide insights into the important interactions for determining the relative binding affinities of the molecules, and suggestions for modification of these interactions to improve binding affinity. The computational requirements are sufficiently modest that the technique can be used to study tens to hundreds of enzyme–substrate complexes.

Chapter 3

Shrnutí

Důsledkem environmentálních problémů způsobených značnou produkcí a širokým použitím halogenovaných uhlovodíků je současný zájem o studium mikrobiálních enzymů, které katalyzují degradaci těchto látek. Haloalkan dehalogenázy (EC 3.8.1.5) tvoří jednu z takových důležitých tříd enzymů díky jejich schopnosti atakovat halogenované alifatické uhlovodíky, které jsou produkovány různými průmyslovými procesy. Haloalkan dehalogenázy jsou zodpovědné za odštěpení halogenidového aniontu hydrolytickým mechanismem, jehož produkty jsou odpovídající alkohol, halogenidový ion a proton. Enzymy divokého typu bohužel často nevykazují dostatečně vysokou aktivitu nebo specifitu pro degradaci environmentálních polutantů. Proteinové inženýrství je technologie, která může zlepšit katalytické vlastnosti příslušných enzymů. Pro úpravu enzymu za účelem zlepšení substrátové specifity je nutné identifikovat aminokyseliny podílející se na vazbě substrátu. Takovéto aminokyseliny lze poté následně modifikovat místně-cílenou mutagenezí. Užitečnou technikou umožňující odvození kvantitativních vztahů mezi strukturou a aktivitou použitím sady 3D struktur komplexů enzym–substrát je srovnávací analýza vazebných energií (angl. Comparative Binding Energy, COMBINE). Získané výsledky demonstrují, že COMBINE analýza je nástroj, který lze využít ke konstrukci prediktivních a robustních QSAR modelů s použitím struktur komplexů enzymu se substráty popsaných na atomové úrovni. Tyto modely nám umožňují pochopit důležité interakce určující relativní vazebné afinity molekul a navrhnout modifikace těchto interakcí za účelem zlepšení vazebné afinity. Výpočetní nároky COMBINE analýzy jsou dostatečně přiměřené k tomu, aby bylo možné tuto techniku aplikovat ke studiu desítek až stovek komplexů.

Part I

Theory

Chapter 4

Computational chemistry

This chapter explains basic terms related to application of different computational chemistry methods and describes fundamentals of specific computational approaches.

4.1 Basic terms

The term "computational chemistry" can be defined as a tool for solving chemical problems using computer. An attempt is made to model all aspects of real chemistry by computer as accurately as it is possible. It should not be mixed with the term "theoretical chemistry" that is more oriented for searching of theories that enable explanation or prediction of experimental data while computational chemistry [1] uses these theories in specific programs to model real molecules or molecular systems. It is apparent that both disciplines are not only complementary but they both need each other. Computational chemistry can be therefore defined as the set of implementations of specific theories. More particular is the term "molecular modelling".

Molecular modelling is a general term which covers a wide range of molecular graphics and computational chemistry techniques used to build, display, manipulate, simulate and analyze molecular structures, and to calculate properties of these structures [2]. Molecular modelling is used in a number of different areas, and therefore the term does not have a rigid definition [3]. To a chemical physicist, molecular modelling might imply performing a high level quantum mechanical calculation using a supercomputer on a structure with 4 or

5 atoms; to an organic chemist, molecular modelling might mean displaying and modifying a candidate drug molecule on a desktop computer; to a biochemist, molecular modelling might represent simulation of an enzymatic reaction. The criterion for a successful modelling experiment should not be how accurately the calculations are performed, but whether they are useful in rationalizing the behavior of the molecule, or in enhancing the creativity of the chemist in the design of novel compounds.

The first step in scientific method is to form a hypothesis. It can be just an educated guess or logical conclusion from known facts. If the hypothesis is found to be consistent with known facts, it is called a theory and usually published. A theory is one or more rules that are postulated to govern the behavior of physical systems. Scientific theories are quantitative in nature that allows them to be tested by experiment. The basic characteristics of most theories are that they explain observed phenomena, predict the results of future experiments and can be expressed in a mathematical form. There are two main constructs used in the scientific approach. The first commonly used construct are models. A model is a simple way of describing and predicting scientific results which is known to be an incorrect or incomplete description. Despite this fact models are extremely useful because they allow us to to predict and understand phenomena without performing the complex mathematical manipulations dictated by a rigorous theory. The valuableness of a model is not hidden in its intrinsic beauty but in its ability to solve practical problems. The second useful construct is an approximation. A lot of models typically involve deliberate introduction of simplifying approximations. Even though a theory may give a rigorous mathematical description of chemical phenomena, the mathematical difficulties can prevent feasibility of solving the specific problem exactly. One can leave out a particular part of the problem or apply specific approximation methods as simplified functions or fitting parameters [4].

It is important to realize that theories, models and approximations are suitable tools allowing us to achieve research aims. But one has always weight how the obtained results can be trusted with given degree of accuracy. The real strength of computational chemistry is the ability to generate data from which a human may gain insight and thereby rationalize the behavior of different class of molecules [5].

4.2 Theoretical methods

The cornerstone of all computational chemistry methods is the relationship between geometry of studied molecular system and its energy, reflecting all other static or dynamic properties. The dependence of potential energy of a molecule on the positions of atoms forming the molecule can be described by hypersurface in the N-dimensional space.

$$E_{POT} = f(x_1, x_2, \ldots, x_N) \tag{4.1}$$

The study of the potential energy (hyper)surface (PES) gives the information about molecular geometry and energy that can be used for the static (e.g., activation barriers of elementary reaction processes) and the dynamic (e.g., simulation of a system in time) reactivity studies. There are two main approaches within computational chemistry devoted to the energy determination and quantification: quantum mechanics (QM) and molecular mechanics (MM).

4.2.1 Quantum mechanics

In the quantum mechanical approach the molecule can be seen as system of atom nuclei and electrons that is completely described by Schrödinger equation. Schrödinger equation for one particle system can be written as follows:

$$(-\frac{\hbar}{2m}\nabla^2 + \mathcal{V})\Psi(\mathbf{r}, t) = E\Psi(\mathbf{r}, t) \tag{4.2}$$

E is total energy of a particle of mass m, moving through space under the influence of an external field $\mathcal{V}$. $\hbar$ is the Planck constant divided by 2π. Position vector $\mathbf{r}$ is given by $\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. Wavefunction Ψ describing the state of the system can be used for evaluating other properties. Symbol ∇^2 is an abbreviation for

$$\nabla^2 = \triangle = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{4.3}$$

Left-hand side of the Schrödinger equation is usually abbreviated to $\mathcal{H}\Psi$, where $\mathcal{H}$ is the Hamiltonian operator of energy describing relationships in the system of electrons and nuclei. Under the conditions that Hamiltonian operator $\mathcal{H}$ is not a function of time, it is possible to separate coordinates and time and the equation 4.2 can be reduced to

$$\mathcal{H}(\mathbf{r})\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \tag{4.4}$$

The energy can be determined by calculating integral

$$E = \frac{\int \Psi^* \mathcal{H} \Psi d\tau}{\int \Psi^* \Psi d\tau} \tag{4.5}$$

The wavefunction may be a complex number, so the complex conjugate notation is used in the previous expression. The square of the wavefunction at the point r is a probability of finding a particle at this point.[1] The probability of finding the particle over the whole space must be equal to one, therefore wavefunction must fulfill following condition:[2]

$$\int \Psi^* \Psi d\tau = 1 \tag{4.6}$$

Although the Schrödinger equation is a simple equation, it can be solved analytically for very simple problems only.[3] To be able to perform non-trivial numerical solution and obtain results in reasonable time several approximations had been introduced. The most used is Born–Oppenheimer approximation that separates electron movement from nuclei movement using the fact the electrons are much lighter than nuclei.[4] This means that electrons can respond to any change of nuclear positions almost instantaneously. The nuclei are considered to be fixed so that electronic wavefunction depends only on their positions and not on their momenta. Application of Born–Oppenheimer approximation allowed to define potential energy surface that can be used for solving a lot of chemical problem. Other type of approximations can be related to the shape of the wavefunction. One of the most widespread approximation of such type represented molecular orbital as linear combination of atomic orbitals (MO–LCAO). According to other used aproximations the methods based on quantum mechanics can be divided into four major groups:

Empirical methods (HMO, EHT and others) reduce the calculations only to necessary parts. They use very raw approximations that cause large mistakes so these methods are used rarely.

Semiempirical methods (CNDO, INDO, MINDO, NDDO, MNDO, PPP, AM1, PM3 and others) neglect some calculations and replace them by experimental data. These data form so-called parametrization of the method

[1] for the electronic wavefunction the probability is the electron density
[2] the wavefunction is said to be normalized
[3] particle in a box, harmonic oscillator, hydrogen atom and a few more
[4] the proton is 1836 times heavier than the electron

and can be obtained for example from spectral characteristic of the atoms. Only valence electrons are taken into account in the semiempirical methods so these methods can reach relatively high reduction of computational requirements comparing to classical *ab initio* methods.

Ab initio methods solve Schrödinger equation using minimum number of approximations (Born–Oppenheimer and MO–LCAO) and no experimental or empirical parameters are used. The main problem in *ab initio* approach is solving the electrostatic interaction between two electrons. Hartree–Fock (HF) method was introduced to answer that problem. The electrons system can be seen as an ensemble of electrons in which every single electron moves in average electrostatic field of other electrons (so-called core). The solution leads then to the system of HF equations solved by iterative process using self-consistent field method (SCF). The quality of result of HF method depends on the quality of wavefunction expansion in the basis set. The calculated energy will converge to so-called HF limit (the best energy reachable using HF method) with increasing expansion of the basis set. It is important to realize that even if we can afford to have long expansions of molecular orbitals and even we can reach the HF limit there are still defects in the wavefunctions which arise from approximations in the actual HF equations. The first error originates from the fact that the Schrödinger equation is not relativistically correct. Fast-moving inner electrons may move with speeds which are not negligible with the velocity of light and relativistic effects thus contribute; mass is not constant. Since most chemical and biological transformations of molecules do not involve core electrons this error causes no serious difficulty. The second error is more serious and is called the correlation energy error. The correlation energy is defined as difference between energy corresponding to exact calculation including correlation effects and HF energy. The post-HF methods had been developed for inclusion of correlation energy into calculation or for calculation of the correlation energy itself. To the three most popular post-HF methods belong configuration interaction method (CI), coupled clusters method (CC) and Møller–Plesset perturbation method (MP).

Density functional theory (DFT) methods are alternative to the *ab initio* methods [6]. They are based on Hohenbeg–Kohn theorems. According to these theorems is wavefunction Ψ of non-degenerated ground state of many-electron system unambiguous functional of one-electron density ρ.

The equation for determination of E and Ψ is given by variational condition

$$\frac{\delta E}{\delta \Psi} = 0 \tag{4.7}$$

that can be rewritten using mentioned theorem as

$$\frac{\delta E}{\delta \rho} = 0 \tag{4.8}$$

The large importance of the Hohenberg–Kohn theorems stems from the fact that they shows that one need not calculate the full N-body wavefunction but only the total electron density in order to obtain all ground-state properties [7]. Concerning the fact that one-electron density is much simplier object (it is a function of three variables) one can expect lower computational demands. However, it is important to note that Hohenberg–Kohn theorems are "only" existence theorems: they prove the existence of the functionals but do not describe any way of obtaining them. The problem is hidden in the form of energy functional that is not explicitly known and therefore approximate forms have to be introduced. Total electronic energy can be then expressed as

$$E(\rho) = E_{KE}(\rho) + E_C(\rho) + E_H(\rho) + E_{XC}(\rho) \tag{4.9}$$

where $E_{KE}(\rho)$ is kinetic energy, $E_C(\rho)$ is an electron-nuclear interaction term, $E_H(\rho)$ is energy of electron-electron interaction and $E_{XC}(\rho)$ contains correlation and exchange contributions. $E_{XC}(\rho)$ can be divided into two separate terms referring to the exchange and correlation parts, but actually corresponding to same-spin and mixed-spin interactions, respectively.

$$E_{XC}(\rho) = E_X(\rho) + E_C(\rho) \tag{4.10}$$

Both components can be expressed by two distinct types; local functionals that depend on only the electron density ρ while gradient-corrected functionals depending on both the ρ and its gradient $\nabla\rho$.

4.2.2 Molecular mechanics

MM is based on classical physics model where the atoms are represented as mass points that mutually interact. The MM energy is written as a sum of such terms, each describing the energy required for distorting a molecule in a specific manner [5].

$$E_{MM} = E_{STR} + E_{BEND} + E_{TORS} + E_{VDW} + E_{ELE} + E_{CROSS} \tag{4.11}$$

E_{STR} is the energy function for stretching a bond between two atoms, E_{BEND} represents the energy required for bending an angle, E_{TORS} is the torsional energy for rotation around a bond, E_{VDW} and E_{ELE} describe the non-bonded atom-atom interactions and finally E_{CROSS} describes coupling between the first three terms. Any deviation of the model from the "ideal" molecular geometry that has zero energy will result in an increase in energy. Therefore this approach can give only relative energies between two different molecular structures. Stable molecules correspond to minima on the potential energy surface and therefore they can be located by minimizing E_{MM} as a function of the nuclear coordinates.

A set of equations defining how the potential energy of a system varies with the locations of its component atoms. The implementation of these equations is called a force field (FF). The comprehensive list of currently available force fields can be found in the book by Young [8]. Every force field contains two major types of parameters used for description of the behavior and properties of the system in the specific force field. These are:

- Series of atom types defining the characteristic of an element within a specific chemical context. Atom types prescribe different characteristics and behavior for an element depending upon its environment. It means that the atom type depends on hybridization, charge and the types of the other atoms to which it is bonded.
- One or more parameter sets which fit the equations and atom types to experimental data. Parameter sets define force constants that are values used in the equations to relate atomic characteristic to energy components and structural data such bond lengths and angles.

The total energy is given by the potential energy of the atomic nuclei[5] V only. Therefore MM calculations require knowledge of the constitution of the molecule and individual atom types. The consequence of this approach is the fact that MM is not able to describe any processes and properties concerning electrons[6] and therefore is not applicable for study of chemical reactivity. Nevertheless, there exist some exceptions allowing simulation of breaking bonds as for example force field ReaxFF [9] that is currently applicable only for small hydrocarbons molecules.

The main advantage of MM approach is its applicability for solving many computational chemistry problems that are too large for treating by QM methods that deal with electrons in the system. Even if some of the electrons are ignored, there still remain a large number of particles that have to be considered as is the case of biomolecules. Such calculations can not be performed using the QM approach due to extreme time requirements.

4.2.3 Simulation techniques

Simulation techniques used in computational chemistry can be further divided into three categories:

Molecular dynamics (MD) is based on solving Newton's laws of motion for individual atoms of simulated system. It can provide data obtained from the computer simulation of intra- and intermolecular forces between the particles in the system. To achieve such purpose it is necessary to know the force affecting the individual atoms. The evaluation of the forces acting upon atoms is done in MD using empirical force field. The configurations of the system are generated by integration of the Newton's laws of motion

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)}{m_i} \tag{4.12}$$

where

$$\mathbf{F}_i(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) = -\nabla_i V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \tag{4.13}$$

[5] as a function of nuclear positions
[6] bond forming and breaking, etc.

and

$$\nabla_i = \frac{\partial}{\partial x_i} + \frac{\partial}{\partial y_i} + \frac{\partial}{\partial z_i} \tag{4.14}$$

These equations describe the motion of a particle of mass m_i with position r_i. The force F_i acting on the particle (i.e., direction and magnitude of a force) is given by a potential energy gradient in 4.13. The particle is forced to move toward decreasing potential energy. The result is a trajectory showing the behavior[7] of the system in time.

***Ab initio* dynamics** has the same basics as molecular dynamics but the forces influencing atoms are calculated on quantum-chemical level.

Quantum dynamics is based on solving time-dependent Schrödinger equation. Due to extreme complexity of this equation is the application of quantum mechanics restricted only to very small molecular systems.

Molecular dynamics simulations are important for understanding the physical basis of the structure and function of biological macromolecules [10]. They can provide an ultimate details concerning individual particle motions as a function of time. Probably the most challenging task in the present biomolecular simulations is to address problem of enzyme catalysis and to calculate free energies. In the field of computer simulations of the enzyme catalysis there are several computational approaches for study of the enzyme catalysis on an atomistic level [11, 12]. Currently, two main strategies capable known to provide reliable insight into enzyme catalysis are the empirical valence bond (EVB) and hybrid molecular orbital quantum mechanics/molecular mechanics (QM/MM) methods. Two most rigorous methods currently available for calculating free energies are the free energy perturbation (FEP) and the thermodynamic integration (TI) [13].

4.3 COMBINE analysis

COMBINE Analysis Outline
The aim of COMBINE analysis [14] is to make full, simultaneous and systematic use of all available information describing/influencing binding process of

[7] variance of velocities and positions of particles

ligands into macromolecular targets. The analysis tries to correlate general biological property of the system with information hidden in the 3D structure of macromolecular target. Such biological property can be for example the binding affinity of small ligand to a macromolecule. The ligand can be either substrate molecule or inhibitor. The abbreviation "COMBINE" is composed from COMparative BINding Energy but this abbreviation forms also an acronym that refers to two aspects of the technique: (i) macromolecule–ligand structural data are combined with experimental binding data and (ii) empirical molecular mechanics energy calculations are combined with PLS chemometric analysis. This methodology provides qualitative and quantitative prediction of binding affinities without a necessity to perform time-consuming and expensive experiments and gives us atomic-detail insight into binding process of specific ligand.

COMBINE Analysis Overview
The COMBINE method can be applied to any dataset for which two necessary pieces of information are available: experimental binding or activity measurements for a series of ligands that bind to a macromolecular receptor and an experimentally determined three-dimensional structure of the target macromolecular receptor complexed to a representative ligand. The method relies on the assumptions that all ligands analyzed bind to the receptor to the same binding site and that the binding mode can be deduced by comparative modelling techniques (Figure 4.1). In the first step of COMBINE analysis a set of macromolecule–ligand complexes is prepared from single macromolecule and a set of single ligand molecules. For all macromolecule–ligand complexes total binding energy is then calculated. COMBINE works with total binding energy calculated for the set of macromolecule–ligand complexes using a molecular mechanic force field. The total binding energy, ΔU, may be assumed to be given by the sum of several terms: (i) the sum of intermolecular interaction energies (Δu_i) between the ligand and each macromolecule residue, E^{LM}_{INTER}, (ii) the change in intramolecular energy of the ligand upon binding to the macromolecule, ΔE^L, (iii) the change in intramolecular energy of the macromolecule upon ligand binding, ΔE^M, (iv) the desolvation energy of a ligand, E^L_{DESOLV}, and (v) the desolvation energy of a macromolecule, E^M_{DESOLV}.

$$\Delta U = E^{LM}_{INTER} + \Delta E^L + \Delta E^M + E^L_{DESOLV} + E^M_{DESOLV} \tag{4.15}$$

The second and third terms, describing changes in intramolecular energies upon binding, can be neglected in most cases because many ligands are rather

small molecules and there is no evidence for large differences in the structure of the macromolecule when different ligands are bound. Intermolecular energy contributions are then decomposed into van der Waals and electrostatic interactions. The following step is the decomposition of the macromolecule–ligand 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. Addition of other potential descriptors that could influence ligand binding process is also possible and methodologically easily viable.

This matrix is then projected on to a small number of orthogonal "latent variables" using partial least-squares projection to latent structures (PLS) analysis [15–17], and the original energy terms are given weights, w_i, according to their importance in the model, in the form of PLS pseudo-coefficients. The higher coefficients are more significant for explaining the variance in the experimental data. PLS method is applied for selection of energetic terms correlated with binding affinities. Quality of COMBINE model is described by two parameters—correlation coefficient (R^2) and cross-validated correlation coefficient (Q^2). R^2 determines the quality of a fit and its value quantifies percents of explained variability of experimental data by the latent variables while Q^2 characterizes predictive ability of a model [18].

COMBINE Analysis Applications
COMBINE analysis was originally developed for drug design purposes and the first application was described by Ortiz *et al.* [19, 20] for study of 26 inhibitors of the human synovial fluid phospholipase A2. Later the analysis was used by Perez *et al.* [21] for study of HIV-1 protease inhibitors and further extended by Pastor *et al.* [22]. The methodological cornerstone describing COMBINE methodology in detail was published by Wade *et al.* [14]. Later other COMBINE studies have been conducted with different molecular systems as transcription factors of the nuclear receptor family binding to DNA [23], heterocyclic amines and human cytochrome P450 1A2 [24] and human neutrophil elastase with the N3-substituted trifluoromethylketone-based pyridone inhibitors [25]. Recently Wang and Wade constructed two other COMBINE models for two subtypes and one mutant of neuraminidase from influenza virus complexed with 43 inhibitors [26] and for periplasmic oligopeptide binding component–peptide complexes [27].

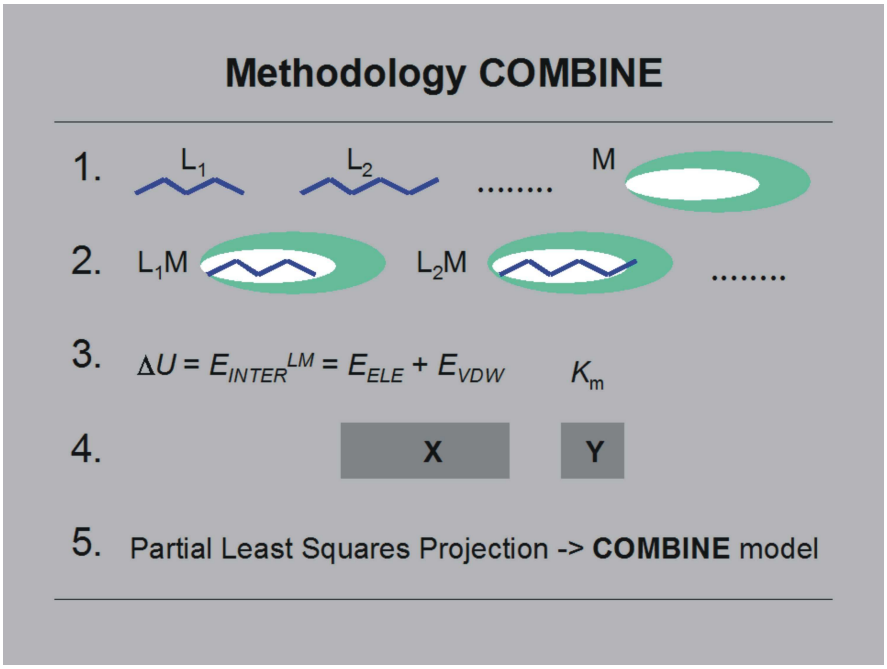

Figure 4.1: Simplified scheme of COMBINE analysis. L_1, L_2... – ligands; M – macromolecule; L_1M, L_2M... – macromolecule–ligand complexes; ΔU – total binding energy; E_{INTER}^{LM} – intermolecular interaction energies between the ligand and each macromolecule residue; E_{ELE} – electrostatic interaction energy; E_{VDW} – van der Waals interaction energy; K_m – logarithmic values of experimental binding affinity (apparent dissociation constant).

All COMBINE applications mentioned above concerned drug design. Our contribution to the field, outlined in this thesis, was the first application of COMBINE for study of enzyme catalysis and impact of mutations to catalytic efficiency. We applied COMBINE analysis for investigation of structure–function relationships of haloalkane dehalogenase DhlA and a set of halogenated substrates [28]. This study continued by application of COMBINE analysis for different enzyme species from the same protein family [29] and was finally extended by incorporation of methodological improvements [30].

Chapter 5

Microbial dehalogenation

5.1 Halogenated aliphatic compounds

Environmental Concerns
Halogenated aliphatic hydrocarbons are among the most frequently occurring environmental pollutants [31, 32]. Although large portion of these compounds is produced naturally by existing biotic and abiotic processes, extensive industry use of halogenated compounds caused presence of large amounts of these xenobiotics in the environment. Large quantities of these compounds are widely used for industrial purposes as solvents, extractive agents in the organic synthesis, side products from technological syntheses, parts of capacitors, dielectric impregnants, fire-resistant additives, inflammable hydraulic fluids, additives for lubricants and greasing oils, pesticides, fire retardants, heat transfer fluids and cleaning agents. The industrial production reaches several hundreds up to thousands of tons per year. This class of compounds represents one of the top group of compounds on the U.S. Environmental Protection Agency's list of Priority Pollutants [33, 34].

Sources of Halogenated Compounds
Halogenated aliphatic compounds occur in all compartments of biosphere, in underground water and surface water, atmosphere, soil, sediments and they are transported into food chains as well. They are detectable in city air. Halogenated hydrocarbons are produced in large quantities by natural and anthropogenic way.

Abiotic natural sources are especially forest and bush fires and volcanic activity. Biogenically are these compounds produced by sea organisms, higher plants, ferns, insects, bacteria, fungi and mammals [35]. Living organisms produce minimally 2 000 different halogenated derivatives [36]. Sea organisms commonly synthesize brominated compounds while terrestrial organisms produce rather chlorinated compounds. The main sources of volatile organic bromides in the atmosphere are seaweeds. Iodinated compounds do not occur so often and fluorinated compounds are very rare.

The natural sources of halogenated compounds were underestimated for a long time although their production is in specific cases even substantially higher than anthropogenic production. Natural production of chloromethane is such an example. It is estimated as 28 millions of tons per year while anthropogenic production is only 0.6 millions of tons per year [32]. The production of chlorinated phenols by lignin-degrading fungi is also larger than their anthropogenic production [37]. Biogenic sediments from various geological epochs contain quantities of organohalogens. The content of halogenated organic compounds in sediments is either primary or the compounds are synthesized *de novo* by biotic or abiotic way. Organisms exploit halogenated compounds as frightening agents or irritants. Halogenated alkanes, alcohols, ketones, carboxylic acids, aldehydes, epoxides, alkenes and amides of carboxylic acids can be isolated from seaweeds.

Properties of Halogenated Compounds
Halogenated organic compounds are ubiquitous in the environment. They exhibit antibacterial, antifungal, herbicidal, insecticidal and moluscicidal effects but on the other hand they are environmentally dangerous and are hazardous to humans due their toxic, genotoxic, teratogenic and irritating effects. Short-term exposition by halogenated aliphatic compounds causes skin irritation, depression of central nervous system, damages of liver and kidneys. For those reasons, ability to persist in the environment for a long time and their high mobility, big attention is currently devoted to this group of compounds. Halogenated hydrocarbons are generally more persistent compared to the equivalent non-halogenated hydrocarbons. The persistence in the environment is caused by the presence of stable carbon-halogen bond. The biological recalcitrance of halogenated compounds is related to the number, type and position of the halogen substituents. As a general rule, the carbon-halogen bond is regarded as increasingly recalcitrant with increased electronegativity of the substituent. Also

halogenated substances with one or few substituents are expected to be more easily degradable than the corresponding polyhalogenated compounds.

There are several possible reasons for the persistence of a pollutant in the environment. These include: (i) the absence of microorganisms with the capability of degrading a compound, (ii) the presence of unfavorable environmental conditions for biodegradation (permeability of the subsurface for air and water, temperature, pH, salinity and water content, presence of inhibitory chemicals, etc.), (iii) unfavorable substrate concentration (the pollutant may be present in too high concentration, leading to toxicity, or in too low concentration, failing to induce the degradative enzymes), and (iv) lack of bioavailability of the pollutant because of incorporation into humic substances or strong adsorption to soil particles.

5.2 Dehalogenation mechanisms

Microorganisms evolutionary developed broad range of enzymatic pathways for transformation of halogenated substrates [38–40]. There are several diverse groups of enzymes utilizing halogenated compounds as a growth substrate whose function is a carbon-halogen bond cleavage [41–44]. That cleavage can serve only as first but critical biotransformational step or performs complete biodegradation. Enzymatically catalyzed dehalogenation reactions can be divided according to reaction mechanism into following categories: oxidation, substitution, reduction, dehydrodehalogenation, hydratation and methyl transfer (Figure 5.1).

Oxidation. Oxidation is mostly used for degradation of halogenated aromatic compounds. Oxygenases non-specifically oxidase halogenated aromatics and produce nonstable intermediates that spontaneously eliminate halogen. Oxygenolytic dehalogenation is often random reaction of enzymes with broad substrate specificity. The products of those reactions are not consequently metabolized by dehalogenating bacteria.

Hydrolytic, glutathione-dependent and intramolecular substitution. Halogenated aliphatic compounds undergo substitution in water in the absence of inorganic or biochemical catalysts. In general, these reactions proceed slowly with half-lives of days to centuries. The hydrolytic dehalogenases catalyze a nucleophilic displacement reaction with water as the sole co-substrate. Hydrolytic dehalogenation is often considered to be

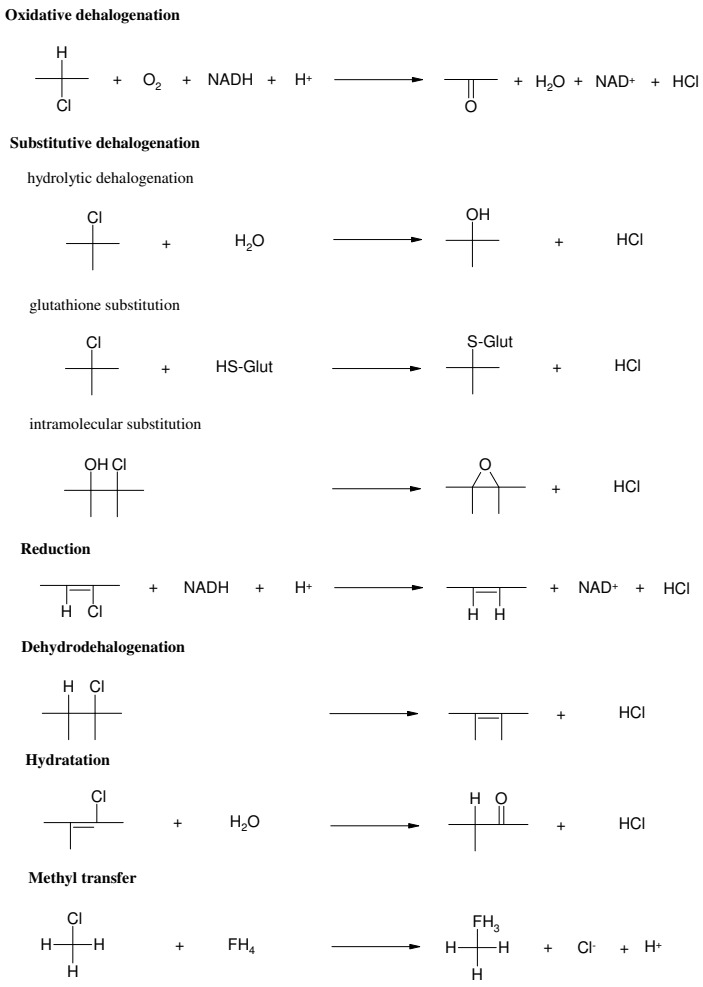

Figure 5.1: Enzyme-catalyzed dehalogenation reactions in microbial cultures that grow on halogenated organic compounds. Reproduced from the publication of Janssen *et al.* [38].

the first step in an aerobic degradation of halogenated aliphatic hydrocarbons and it allows entering of halogenated hydrocarbons into the central metabolism. Therefore some bacteria use halogenated hydrocarbons as a source of carbon and energy. Glutathione transferases play a key role in the detoxification of xenobiotic compounds and reactive products of oxidative transformation reactions in higher organisms. Their role in degradation of halogenated compounds is the nucleophilic displacement of a halogen substituent. Haloalcohol lyases, also called halohydrin hydrogen halide lyases, catalyze the intramolecular nucleophilic displacement of *vic*-haloalcohols to their corresponding epoxides.

Reduction. Reductive mechanism is more typical for anaerobic environment. It is often the only way how degrade some polychlorinated industrial substances. The reduction can be either performed by hydrogenolysis (replacement of a halogen atom by a hydrogen atom) or by dihaloelimination (elimination of two halogen atoms with simultaneous formation of a double bond).

Dehydrodehalogenation. Dehydrodehalogenases eliminate HCl from haloorganic substrate leading to the formation of a double bond. This reaction consists of removal of a halogen from one carbon atom and concomitant or subsequent removal of a hydrogen atom from an adjacent carbon atom.

Hydratation. A hydratation type of dehalogenation reaction has been proposed for the bacterial conversion of a few compounds that carry a halogen substituent on an unsaturated carbon atom. These compounds include derivatives of acrylic acids and aromatic compounds.

Methyl transfer. Chloromethane and dichloromethane are known to support growth of strictly anaerobic bacteria. The methylotrophic strain *Acetobacterium dehalogenans* uses chloromethane as sole energy source, producing acetate. A chloromethane dehalogenase, which is inducible by chloromethane, transfers the methyl group of its substrate onto tetrahydrofolate, producing methyltrihydrofolate and chloride.

5.3 Haloalkane dehalogenases

Probably the most widespread mechanism of microbial dehalogenation of halogenated aliphatic hydrocarbons is hydrolytic dehalogenation. This reaction is catalyzed by enzymes called haloalkane dehalogenases (EC 3.8.1.5). Haloalkane

dehalogenases [45, 46] cleave carbon-halogen bond in the presence of one water molecule as the only co-factor. The products of the dehalogenation reaction are corresponding alcohol, halide ion and proton. Substrate specificity of the haloalkane dehalogenases is very broad [47]. These enzymes catalyze transformation of halogenated alkanes, alkenes, cycloalkanes, ethers, alcohols, hydrines and carboxylic acids. Some of these compounds (e.g., 1,2-dichloroethane, 1,2-dichloropropane or 1,2,3-trichloropropane) are important environmental pollutants and haloalkane dehalogenases should serve as biocatalysts for their removal. Currently these enzymes are employed in bioreactors inoculated by dehalogenases-producing bacteria for decontamination of ground water [48], in chemical syntheses during removal of unwanted intermediates [41] and in biosensors for detection of halogenated compounds in the free environment. Broader application of haloalkane dehalogenases in practice is hindered by their low activity toward target substrates. The primary aim of structure-function relationships of these enzymes is proposal and *in vitro* construction of mutant enzymes with higher dehalogenation activity and broader substrate specificity. These modified enzymes would markedly help to increase efficiency of dehalogenation processes without environment exposition of newly formatted intermediates.

Haloalkane dehalogenases belong to α/β-hydrolase fold proteins [49, 50]. The core of each enzyme is similar and it consists of two different domains: α/β-fold domain (main domain) which is conserved in all α/β hydrolases and so-called cap-domain. The main domain is composed of an β-sheet constituting of eight β-strands surrounded by six α-helices. The cap-domain is composed of an additional five α-helices connected by loops. The active site is located between these two domains of the enzyme in an internal, predominantly hydrophobic cavity and can be reached from the solvent through a tunnel. At least three different groups of haloalkane dehalogenases can be distinguished according to different substrate specificity [51]. Each of these categories has its own representative with known three-dimensional structure: haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10 (DhlA) [52–55], haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 (LinB) [56–58] and haloalkane dehalogenase from *Rhodococcus rhodochrous* NCIMB 13064 (DhaA) [59]. The ratio of volumes of the active sites for three substrate specificity representatives DhlA : DhaA : LinB was determined as 1 : 2 : 2.5 [51].

Reaction Mechanism
The carbon-halogen bond in a substrate molecule is cleaved and corresponding alcohol is released during hydrolytic dehalogenation. The complete reaction

mechanism (Figure 5.2) involves three subsequent steps—nucleophilic substitution (S_N2), nucleophilic addition (Ad_N) and elimination (E). In the first step, nucleophilic oxygen atom of aspartate residue attacks carbon atom bound to halogen atom in a substrate molecule. Aspartate serves as a nucleophile, attacking the substrate molecule, and is located in a very sharp turn called the nucleophile elbow. Such attack causes formation of alkylenzyme ester—in which aspartate residue is esterified to alkyl group—and halide anion. In the second step, substrate–enzyme ester is attacked by nucleophilic water molecule. In the last step, tetrahedral intermediate decomposes and alkohol plus halide anion is released. Proton originating from the reaction protonates the histidine residue.

Three structural features of haloalkane dehalogenases were recognized as essential for their catalytic performance: (i) a catalytic triad, (ii) an oxyanion hole, and (iii) halide-stabilizing residues [51, 52]. The catalytic triad is formed from three amino acid residues that are directly involved in the reaction mechanism. This triad is highly conserved in all members of α/β-hydrolase fold superfamily. The catalytic triad is composed of a nucleophile, a base and an acid. The nucleophile initiates the dehalogenation reaction by nucleophilic attack on carbon atom bound to halogen in substrate. The base activates water molecule which subsequently hydrolyzes previously formed alkyl–enzyme intermediate. Histidine serves as base, abstracting the hydrogen from the hydrolytic water molecule. The catalytic acid (Asp in DhlA; Glu in LinB and DhaA) is hydrogen bonded to histidine, and increases the basicity of nitrogen in the imidazole ring. The acid helps a base to stabilize the charge that develops on base ring and to keep the base in proper orientation.

Another catalytically important structural feature that is conserved in all α/β-hydrolases [49] is a small gap called oxyanion hole [60] near the amidic nitrogen of amino acid following the nucleophile. This oxyanion hole is formed by two backbone nitrogen atoms. Amidic NH groups from tryptophan next to the nucleophile and from another residue located after strand $\beta3$ (Glu56 in DhlA, Asn38 in LinB, and Asn41 in DhaA) are hydrogen bonded to $O_{\delta2}$ of the nucleophile, leading to an increased nucleophilicity of $O_{\delta1}$ and stabilization of the negatively charged transition states. The oxyanion hole is believed to be a major contributing factor in lowering the free energy of the activated complex and stabilization of the tetrahedral alkyl–enzyme intermediate. The overview of amino acid residues essential for the catalytic function of DhlA and LinB enzymes is show in Table 5.1.

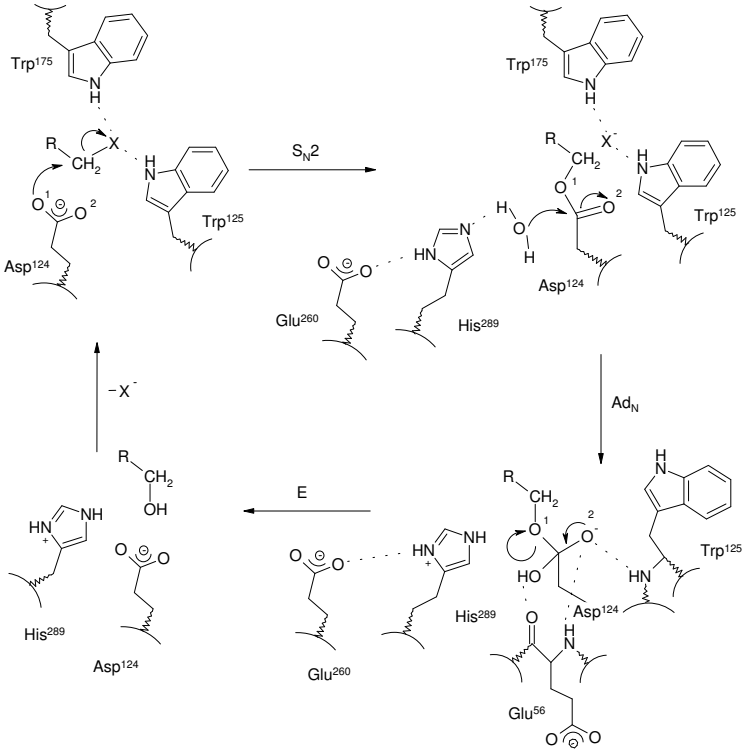

Figure 5.2: Scheme of haloalkane dehalogenase reaction mechanism. DhlA enzyme is used as illustrative example.

Table 5.1: Amino acid residues essential for catalysis.

enzyme	DhlA	LinB	DhaA
catalytic triad	Asp124	Asp108	Asp106
	Asp260	Glu132	Glu130
	His289	His272	His272
oxyanion hole	Glu56	Asn38	Asn41
	Trp125	Trp109	Trp107
primary halide stabilization	Trp125	Trp109	Trp107
	Trp175	Asn38	Asn41

Unlike residues of the catalytic triad and oxyanion hole, the halide-stabilizing residues are not present on equivalent positions in different haloalkane dehalogenases [51]. Trp125 and Trp175 were identified as essential for stabilization of halide ion in DhlA. The former one is also conserved in LinB and DhaA while the latter one is not conserved due to different architecture of the cap domain. The overall reaction rate of wild type enzymes is very slow so they can convert only several substrate molecules per second. The reaction mechanism has been studied in detail by protein crystallography [52–55, 61, 62], kinetic measurements [63–69], thermodynamic analysis [70], site-directed mutagenesis [71–78], structure–function relationships [47, 79–84] and by molecular modelling [51, 85–99].

The differences in the primary and tertiary structure of the enzymes—especially in the active site—play crucial role in their substrate specificity and activity toward their substrates. The understanding of relationships between the structure and the haloalkane dehalogenases substrate specificity has fundamental importance for improvement of the environmental catalysis by means of protein engineering. The substrate specificity can be a function of several properties [51] and following three were recognized as important also for haloalkane dehalogenases: (i) size and geometry of the active site, (ii) size and geometry of the entrance tunnel to the active site and (iii) efficiency of stabilization of the transition state and the halide anion by active site amino acids.

Bacterial Strains Containing Haloalkane Dehalogenases
Haloalkane dehalogenases are already known for more than 20 years. The first dehalogenating enzyme has been isolated from *Xanthobacter autotrophi-*

cus GJ10, bacterium capable to use 1,2-dichloroethane as a sole carbon and energy source. Up to now there are about fifteen different bacterial strains showing dehalogenase activity that had been discovered and characterized. The currently known bacterial strains producing haloalkane dehalogenases are listed in the Table 5.2.

Table 5.2: List of bacterial strains producing haloalkane dehalogenases.

Organism	Strain	Gene	Protein	AA	Mw
Xanthobacter autotrophicus	GJ10	*dhlA*	DhlA	310	35 144
Xanthobacter autotrophicus	GJ70	*dhlA*	DhlA	310	35 144
Ancylobacter aquaticus	AD20	*dhlA*	DhlA	310	35 144
Ancylobacter aquaticus	AD25	*dhlA*	DhlA	310	35 144
Sphingomonas paucimobilis	UT26	*linB*	LinB	296	33 107
Rhodococcus erythropolis	Y2	*dhaA*	DhaA	293	33 246
Rhodococcus rhodochrous	NCIMB 13064	*dhaA*	DhaA	293	33 246
Rhodococcus (formerly *Corynebacterium*) sp.	m15-3	*dhaA*	DhaA	293	33 246
Rhodococcus (formerly *Arthrobacter*) sp.	HA1	*dhaA*	DhaA	293	33 246
Rhodococcus (formely *Acinetobacter*) sp.	GJ70	*dhaA*	DhaA	293	33 246
Pseudomonas pavonaceae	170	*dhaA*	DhaA	293	33 246
Mycobacterium sp.	GP1	$dhaA_f$	DhaA_f	307	34 725
Rhodococcus sp.	TB2	*dhaA*	DhaA	293	33 246
Mycobacterium avium	N85	*dhmA*	DhmA	301	33 841
Mycobacterium avium	104	*106*	DhmA	301	33 841
Mycobacterium tuberculosis	H37Rv	*rv2296*	Rv2296	300	33 358
Mycobacterium tuberculosis	CDC1551	*mt2353*	Rv2296	300	33 358
Mycobacterium tuberculosis	H37Rv	*rv2579*	Rv2579	300	33 728
Mycobacterium tuberculosis	CDC1551	*mt2656*	Rv2579	300	33 728
Mycobacterium bovis	MU11	*iso-Rv2579*	Iso-Rv2579	300	33 714

5.3.1 DhlA enzyme

The haloalkane dehalogenase DhlA is a globular protein that was isolated from the nitrogen-fixing soil bacterium *Xanthobacter autotrophicus* GJ10 [52]. DhlA participates in biodegradation pathway (Figure 5.3) for utilization of 1,2-dichloroethane and 2-chloroethylvinylether as the only carbon source [38]. The structure of a free enzyme was solved for the first time by X-ray crystallography at 1.9 Å resolution [100].

DhlA is composed of 310 residues and its molecular mass is about 36 kDa (Figure 5.4). It hydrolyzes terminally chlorinated alkanes with a chain length up to four carbons and brominated alkanes with a chain length up to ten carbon atoms to primary alcohols. The optimal activity for the catalysis occurs at the pH of 8.2 and the temperature 37 °C. The catalytic triad of DhlA (Figure 5.5) is composed of Asp124–His289–Glu260 [52]. The nucleophile Asp124 is located between β-strand 5 and α-helix 3 and the base His289 between β-strand 8 and α-helix 10. The catalytic acid is positioned after β-strand 7. Bound substrates, transition states, and product structures are primarily stabilized by hydrogen bonds from the Trp125 and Trp175. Only one tunnel is connecting buried active site of DhlA with the surface of the enzyme [94].

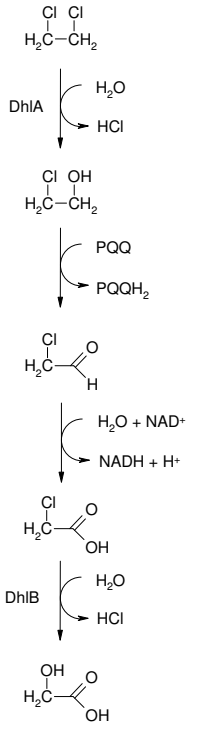

Figure 5.3: Degradation pathway of 1,2-dichloroethane catalyzed by the enzymes of bacterium *Xanthobacter autotrophicus* GJ10.

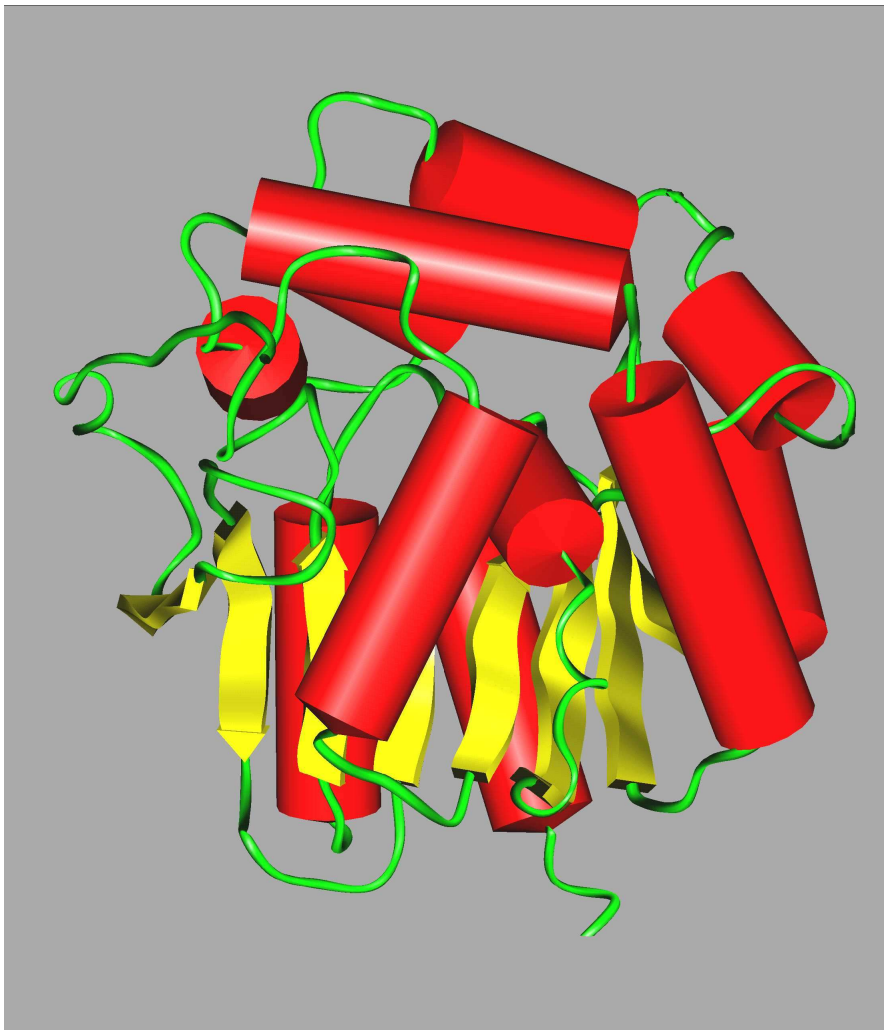

Figure 5.4: View of secondary structure elements making up the three-dimensional structure of DhlA enzyme. α-Helices are in red, β-strands in yellow, and random coils in green.

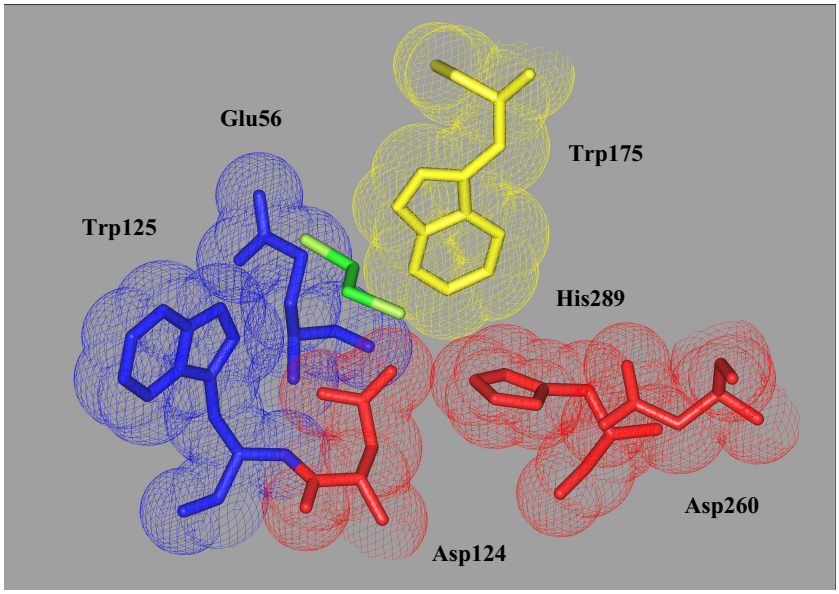

Figure 5.5: Detailed view of amino acid residues making up the DhlA active site (shown as sticks and van der Waals surface). Catalytic triad residues are in red, oxyanion hole residues in blue and primary halide-stabilizing residue that is not part of the oxyanion hole is represented in yellow. Crystallographic position (PDB-ID 2DHC) of 1,2-dichloroethane colored by atom-type is shown in the center of the active site.

5.3.2 LinB enzyme

Haloalkane dehalogenase LinB originates from *Sphingomonas paucimobilis* UT26 that was found in soil exposed for 10 years to lindane (γ-hexachlorocyclohexane) [101]. Lindane had been abundantly used as insecticide but its production and application was forbidden due to its toxicity and persistence. LinB is the second enzyme of biochemical degradation of lindane which is used by bacterium as a carbon and energy source (Figure 5.6).

LinB belongs to the same protein family as DhlA (Figure 5.7). These two proteins differ both by their structures and their catalytic properties. LinB is composed of 296 residues and its molecular mass is about 34 kDa (Figure 5.7). The catalytic triad in LinB (Figure 5.8) is composed of Asp108–His272–Glu132 [102]. The nucleophile Asp108 is located at the turn between β-strand 5 and α-helix 3. The base His272 is localized at the turn connecting β-strand 8 and α-helix 10. The catalytic acid Glu132 is positioned after β-strand 6 [103]. Bound substrates, transition states, and product structures are primarily stabilized by hydrogen bonds from the Trp109–Asn38 pair. The active site of LinB is 2.5 times larger than the active site of DhlA and is less buried inside the protein core [51]. There are at least three tunnels leading from the protein surface into the active site of LinB. LinB shows broader substrate specificity than DhlA, i.e., it is more active toward larger and β-substituted haloalkanes, and therefore it should be more suitable for the design of efficient catalysts for the target compounds carrying a halogen in the β-position.

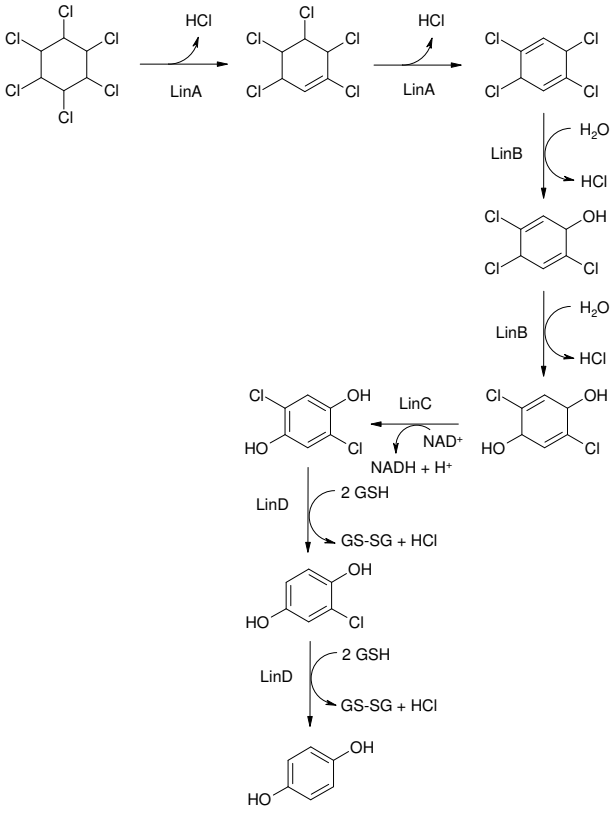

Figure 5.6: Degradation pathway of lindane catalyzed by the enzymes of bacterium *Sphingomonas paucimobilis* UT26.

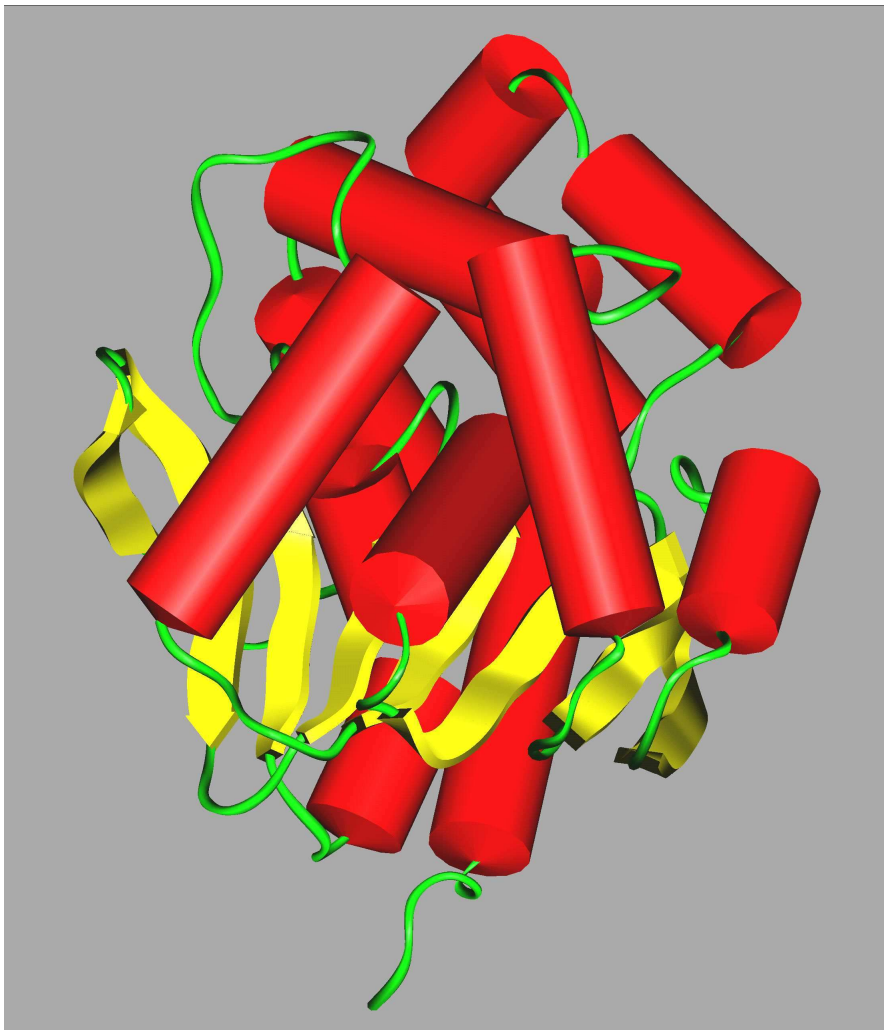

Figure 5.7: View of secondary structure elements making up the three-dimensional structure of LinB enzyme. α-Helices are in red, β-strands in yellow and random coil in green.

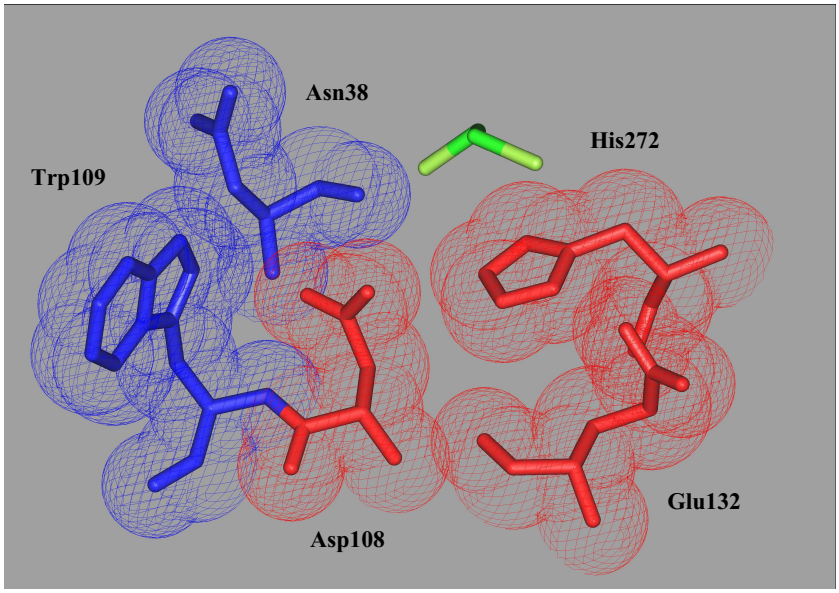

Figure 5.8: Detailed view of amino acid residues making up the LinB active site (shown as sticks and van der Waals surface). Catalytic triad residues are in red, oxyanion hole/primary halide-stabilizing residues in blue. Crystallographic position (PDB-ID 1G5F) of 1,2-dichloroethane colored by atom-type is shown in the center of the active site. The substrate molecule is bound in the non-reactive position.

5.4 Protein engineering

Protein engineering is focused on modification of proteins resulting in improved catalytic efficiency, increased enzymatic activity, thermostability, improved possibility of immobilization, changed pH or temperature optimum or targeted drug improvement with lowered side effects. Generally there are two complementary strategies how to achieve the aims of biomolecular engineering [105], i.e., how to obtain enzymes with intentionally modified and/or improved desired physical and catalytic properties. These are rational re-design and directed evolution.

Rational Re-design
Rational re-design [106] is very information-intensive as it usually requires both the availability of the structure of the enzyme and knowledge about the relationships between sequence, structure and mechanism/function. On the other hand, rapid progress in solving protein structures by X-ray crystallography and NMR spectroscopy and enormously increasing number of sequences stored in public databases have significantly eased access to essential data. In rational re-design [107] precise changes in amino acid sequence are preconceive based on a detailed knowledge of protein structure, function and mechanism and are then introduced using site-directed mutagenesis. The enzymes with desired properties are generated by individual amino acid substitution or secondary structure engineering in rational re-design approach. Depending on the purpose of the mutagenesis, amino acid substitutions are often selected by sequence comparison with homologous sequences [106]. However, the results have to be carefully interpreted because minor sequence changes by a single point-mutation may cause significant structural disturbances. Thus, comparison of the three-dimensional structures of mutant and wild-type enzymes are necessary to ensure that a single mutation is really site-directed.

Directed Evolution
Directed evolution [108–110] has emerged as a popular alternative of rational re-design in the field of protein engineering. Directed evolution techniques allow to genuinely mimic molecular evolution *in vitro*. Mimicking natural evolution, an initial parent gene is chosen and a diverse library of offspring genes is created through mutagenesis and recombination. A screen or selection is applied to the library and the mutants that exhibit the greatest improvement in the desired properties are chosen to become the parents to the next generation. This iterative search has generated large improvements in properties such as stability [111], substrate specificity and activity [109, 112, 113], enantioselecti-

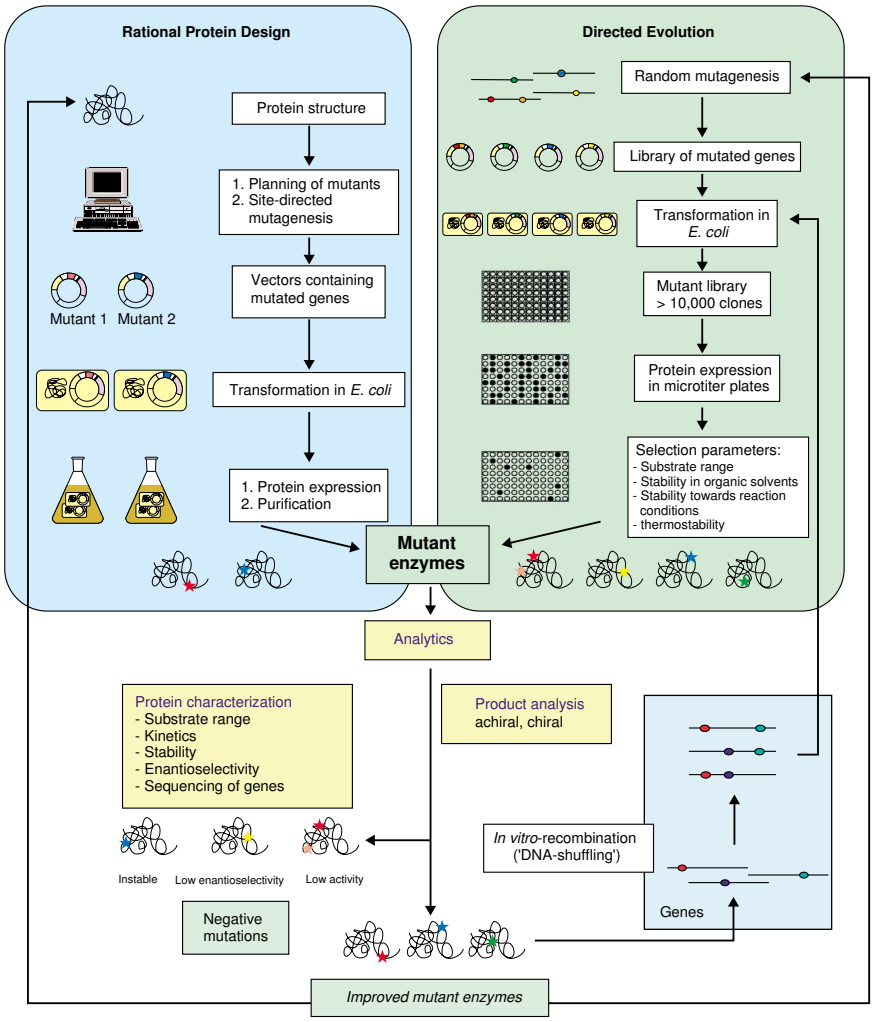

Figure 5.9: Comparison of rational protein design and directed evolution. Reprinted from the article of Bornscheuer *et al.* published in Current Opinion in Chemical Biology [106].

vity [106] and thermostability [114–116]. Directed evolution often discovers these improvements by making a few amino acid substitutions that collectively have an important functional effect. For this reason, the method generally requires a starting protein with some activity toward the desired reaction.

The scheme comparing both approaches is shown in Figure 5.2. During rational protein design, mutants are planned on the basis of their protein structure and then prepared by site-directed mutagenesis. After transformation in the host organism, the variant is expressed, purified and analyzed for desired properties. Directed evolution starts with the preparation of mutant gene libraries by random mutagenesis, which are then expressed in the host organism. Protein libraries are usually screened in microtiter plates using a range of selection parameters. Protein characterization and product analysis sort out desired and negative mutations. *In vitro* recombination by DNA shuffling, for example, can be used for further improvements. Both protein engineering approaches can be repeated or combined until biocatalysts with desired properties are generated.

Choosing the most effective approach for a particular enzyme-engineering task depends on the level that the mechanistic base of the desired property is understood. With the rapidly increasing number of 3D protein structures available in databases and the development of powerful protein modelling tools, rational re-design will become more efficient and broadly applicable. Meanwhile, emergence of novel high-throughput screening processes and strategies for increasing sequence diversity will extend the application of directed evolution to many more industrial enzymes and increase the feasibility for creating new functions.

Part II

Results

Chapter 6

Synopsis of results

General Considerations
The study of the catalytic properties of microbial dehalogenases may help us to understand and assess the potential for degradation of xenobiotics using microorganisms. Such insight might provide biotechnological solutions to deal with environmental problems. In order to design effective bioremediation systems, all known parameters influencing biodegradation processes must be taken into consideration and novel systems capable to treat large quantities of waste materials properly and efficiently must be proposed.

A lot of computational effort had been recently invested into understanding of haloalkane dehalogenase reaction mechanism and more complete insight to the factors affecting their catalytic efficiency. The enzymes with hydrolytic type of dehalogenation reaction mechanism are currently targets of intensive research due to their potential application for bioremediation techniques that propose inexpensive and fast way how to remove halogenated pollutants from contaminated areas.

Improvement and modification of catalytic properties of dehalogenases can be proposed by means of computational chemistry and molecular modelling in combination with experimental studies. Moreover, the computational site-directed mutagenesis method for *in silico* construction of the protein mutants and estimation of their catalytic efficiency can save experimental time, material and financial resources.

COMBINE Applicability Testing
In this study [**Chapter 7**] we conducted COMBINE analysis for 18 substrates of the haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10 (DhlA). The (intermolecular) enzyme–substrate interaction energy was decomposed into residue-wise van der Waals and electrostatic contributions and complemented by surface area-dependent and electrostatic desolvation terms. COMBINE 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 and can be advantageously used to propose amino acid residues contributing most significantly to the substrate specificity of DhlA. Such residues were identified and they represent suitable targets for modification by site-directed mutagenesis. COMBINE model predicted correctly the change in apparent dissociation constants upon single-point mutation of DhlA for six enzyme–substrate complexes.

COMBINE Methodological Improvements
In following study [**Chapter 8**] we tried to solve the problematic aspect hindering broader application of COMBINE analysis. We evaluated applicability of automated molecular docking and quantum mechanical calculations to the construction of a set of structures of enzyme–substrate complexes for use in COMBINE analysis. The results of the COMBINE analysis were compared with previously reported data obtained for the same dataset with complexes based on an experimental structure of the DhlA-dichloroethane complex. It was revealed that 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. Moreover, both COMBINE models displayed a similar composition of the principal components.

Broadening of COMBINE Applicability
This study [**Chapter 9**] deals with substrate specificity of haloalkane dehalogenase isolated from *Sphingomonas paucimobilis* UT26 (LinB). COMBINE analysis for LinB enzyme and a set of twenty five subtrates was performed. The three principal components model explained 95% (91% cross-validated) of experimental data variability. The analysis was capable to explain differences in binding affinity arising from the length of the carbon chain. The study revealed that the substrate length optimum for LinB is six carbon atoms. Modelling of LinB is slightly more difficult due to its larger active site and less specific binding of substrates. Unfortunately, the COMBINE model did not explain sufficiently

affinity differences for analogous substrates differentiating by bearing halogen atom. The explanation of these differences will possibly require additional studies as measurement of inhibition kinetic constants, more precise description of desolvation effects and/or application of different docking algorithm and force field.

COMBINE Analysis in Protein Design
This study [**Chapter 10**] summarize the results of the project focused on the study of engineering of haloalkane dehalogenases. Haloalkane dehalogenases (E.C. 3.8.1.5) are microbial enzymes that utilize water as the only co-substrate to transform halogenated alkanes into halide ions and alcohols. These enzymes have a potential application in detoxification of subsurface pollutants, recovery of industrial side products and biosensors. Modification of their substrate specificity and activity is required for optimization of catalytic properties for biotechnological applications. The major objective of the project was to understand the structural determinants of substrate specificity of these enzymes. Two COMBINE models for haloalkane dehalogenases DhlA and LinB are compared and their correspondences and differences are evaluated. This study classed COMBINE analysis into the framework of protein engineering.

Explanation of Non-productive Binding
In next study [**Chapter 11**] we kinetically and structurally examined the haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 (LinB) in complex with 1,2-dichloroethane, 1,2-dichloropropane and butane-1-ol, that is the reaction product of the turnover of 1-chlorobutane. It was revealed by inhibition analysis that both 1,2-dichloroethane and 1,2-dichloropropane act as simple competitive inhibitors of the substrate 1-chlorobutane and that 1,2-dichloroethane binds to LinB with lower affinity than 1,2-dichloropropane. Subsequently we verified by docking calculations conducted with the enzyme in the absence of active site water molecules and halide ions that studied substrates could bind productively. After inclusion of water molecules and halide ions in the calculations, the compounds were docked in the manner similar to that observed in the crystal structure.

Chapter 7

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

Jan Kmuníček, Santos Luengo, Federico Gago, Angel R. Ortiz, Rebecca C. Wade, and Jiří Damborský

Biochemistry **40**: 8905–8917 (2001)

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

Jan Kmuníček,[1] Santos Luengo,[2] Federico Gago,[2] Angel R. Ortiz,[3] Rebecca C. Wade,[4] and Jiří Damborský[1,5]

National Center for Biomolecular Research, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic; Department of Pharmacology, University of Alcalá, E-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, USA; and European Molecular Biology Laboratory, Meyerhofstrasse 1, D-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, 2-chloroacetonitrile, epichlorohydrine, 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 modelled 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

[1]Masaryk University.
[2]University of Alcalá.
[3]Mount Sinai School of Medicine.
[4]European Molecular Biology Laboratory. Current address: European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany.
[5]To whom correspondence should be addressed. Telephone: +420-5-41129377. Fax: +420-5-41129506. E-mail: jiri@chemi.muni.cz.

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.

INTRODUCTION

Haloalkane dehalogenases are microbial enzymes that catalyze dehalogenation reactions [31, 32, 38], which are important for the degradation of environmental pollutants [41–43]. 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

[1]Abbreviations: BUW, block unscaled weights; COMBINE, comparative binding energy; DhlA, haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10; E^{ES}_{ELE}, 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^{S}_{DESOLV}, desolvation energy of a substrate; E^{E}_{DESOLV}, desolvation energy of an enzyme; FFD, fractional factorial design; ΔG_{ELE}, overall electrostatic free energy change upon binding; G^{ES}_{ELE}, electrostatic energy for all atoms in the enzyme–substrate complex; G^{S}_{ELE}, electrostatic energy for substrate atoms; G^{E}_{ELE}, electrostatic energy for enzyme atoms; ΔG^{S}_{DESOLV}, change in desolvation energy of the substrate upon binding; ΔG^{E}_{DESOLV}, 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;

been shown to be a useful technique for deriving quantitative structure–activity relationships from a set of three-dimensional structures of enzyme–ligand complexes [14, 19–23]. Here, we use COMBINE analysis to derive a predictive model for substrate binding specificity in which 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 [117]. 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 [49, 50] 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 co-factor 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 [52, 53, 55], kinetic measurements [63–66, 70], site-directed mutagenesis [69, 71–76, 103] and by molecular modelling [51, 85–88, 91, 92].

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

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.

Table 7.1: Steady-State Dissociation Constants (in mM) of Haloalkane Dehalogenase[a].

	compound	log K_m		compound	log K_m
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 reference [67]

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.* [67] were logarithmically transformed (Table 7.1). 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 [67].

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^{ES}_{INTER},

(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^S_{DESOLV}, and (v) the desolvation energy of the enzyme, E^E_{DESOLV}.

$$\Delta U = E^{ES}_{INTER} + \Delta E^S + \Delta E^E + E^S_{DESOLV} + E^E_{DESOLV} \tag{7.1}$$

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 [15, 16], 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):

$$\Delta G = \sum w_i \Delta u_i + C \tag{7.2}$$

Parametrization of Halogenated Substrates
The all-atom AMBER molecular mechanics force field [118] was used throughout, and consistent parameters for the haloalkanes were derived to describe the bonded and non-bonded interactions. For each molecule, molecular electrostatic potentials (MEPs) were calculated from the corresponding *ab initio* wave functions (RHF MP2//6-31G*) using Gaussian94 [119] following full energy minimization. Partial atomic charges were then derived by fitting each MEP

to a monopole–monopole expression using the RESP methodology [120, 121]. One conformation of each molecule (*trans*) was employed in the fit, except for 1,2-dichloroethane 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 7.1). 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 [122] was employed. Nonbonded parameters for halogen atoms were developed and tested following a previously reported procedure [123] 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 [124]. 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 [125], which allowed an integration time step of 2 fs to be used. A non-bonded cutoff of 10 Å was employed, and the lists of non-bonded pairs were updated every 25 steps. Density values were provided directly by the SANDER module. ΔH_{VAP} values were calculated according to the equation

$$\Delta H_{VAP} = RT - E_{INTER} \tag{7.3}$$

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 modelled with AMBER 5.0 [126] using the structure [52] of DhlA complexed with the substrate 1,2-dichloroethane (DCE) (PDB entry 2DHC) as a template. The WHATIF 5.0 program [127] was used for adding the polar hydrogen atoms. His289 was singly protonated in the δ-position in accordance with its catalytic function. Non-polar 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 non-bonded cutoff of 10.0 Å and a distance-dependent dielectric constant ($\varepsilon = 4r_{ij}$) 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 ($E_{DESOLV-SUR}$) 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 [128]. This program is an implementation of the method of Lee and Richards. Appropriate hydrophobicity coefficients were taken from the literature [129, 130] 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 [131, 132]: one for all the atoms in the complex (G^{ES}_{ELE}), one for the substrate atoms alone (G^{S}_{ELE}) and a third one for the enzyme atoms alone (G^{E}_{ELE}). 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:

$$\Delta G_{ELE} = G^{ES}_{ELE} - \left(G^{S}_{ELE} + G^{E}_{ELE}\right) \tag{7.4}$$

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 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 [131–133]: (i) the enzyme–substrate interaction energy in the presence of the surrounding solvent (E^{ES}_{ELE}), (ii) the change in desolvation energy of the substrate upon binding (ΔG^{S}_{DESOLV}), and (iii) the change in desolvation energy of the enzyme upon binding (ΔG^{E}_{DESOLV}):

$$\Delta G_{ELE} = E^{ES}_{ELE} + \left(\Delta G^{S}_{DESOLV} + \Delta G^{E}_{DESOLV}\right) \tag{7.5}$$

This decomposition is exact, contains all cross terms, and can be profitably used in COMBINE analysis. The first term in equation 7.5, 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

$$E^{ES}_{ELE} = \sum_{i=1}^{E} q_i \phi_i \tag{7.6}$$

where q represents the atomic point charges of the ith atom of the enzyme and ϕ_i is the electrostatic potential at the ith atom of the enzyme created by the S atoms of the substrate. Equation 7.6 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:

$$E^{ES}_{ELE} = \sum_{n=1}^{N} \sum_{k=1}^{K} q_{nk} \phi_{nk} = \sum_{n=1}^{N} e_n \tag{7.7}$$

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:

$$\Delta G_{ELE} = \sum_{n=1}^{N} e_n + \Delta G^{S}_{DESOLV} + \Delta G^{E}_{DESOLV} \tag{7.8}$$

The electrostatic potentials used in equations 7.6 and 7.7 can be calculated either as:

$$\phi_i = 332 \sum_{j=1}^{L} \frac{q_j}{\varepsilon r_{ij}} \tag{7.9}$$

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:

$$\nabla[\varepsilon(\overrightarrow{r})\nabla\phi(\overrightarrow{r})] = -4\pi\rho(\overrightarrow{r}) + \kappa^{-2}\phi(\overrightarrow{r}) \tag{7.10}$$

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 $\varepsilon(\overrightarrow{r})$ and $\phi(\overrightarrow{r})$ are the dielectric constant and the electrostatic potential, respectively, as a function of position. The solvent-corrected potential calculated with equation 7.10 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_{ELE}^{ES} described in equation 7.7. 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_{ELE}^{ES} values computed with equations 7.4 and 7.8.

The latter two terms of equations 7.5 and 7.8, 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 equation 7.8, 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 [20].

The Poisson–Boltzmann equation was solved using a finite difference method, as implemented in the DelPhi [133] 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 as low-dielectric medium ($\varepsilon = 4$) whereas the surrounding solvent was treated as a high-dielectric medium ($\varepsilon = 80$) 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 [131, 132]. Solvent-accessible surfaces [134], calculated with a spherical probe with a 1.4 Å radius [135], 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 [131, 132].

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) [136]. The program SIMCA-P 8.0 (Umetri, Sweden) was used for permutation validation [137]. 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_{INT} and SDEP_{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 the equations 7.11 and 7.12, respectively.

$$R^2 = 1 - \frac{\sum_i (y_{icalc} - y_{iobs})^2}{\sum_i (y_{iobs} - y_{imean})^2} \tag{7.11}$$

$$SDEC = \left[\sum_i \frac{(y_{icalc} - y_{iobs})^2}{N} \right]^{1/2} \tag{7.12}$$

R^2 takes values up to a maximum value 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 equation 7.13.

$$Q^2 = 1 - \frac{\sum_i (y_{ipred} - y_{iobs})^2}{\sum_i (y_{iobs} - y_{imean})^2} \tag{7.13}$$

A value higher than 0.4 is generally considered statistically significant. The parameters SDEP_{INT} and SDEP_{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 Halogenated Alkanes
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 7.1) 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 modelled, and each complex was energy-minimized. The positions of the substrates inside the active site after energy minimization are shown in Figure 7.2. 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 7.2.

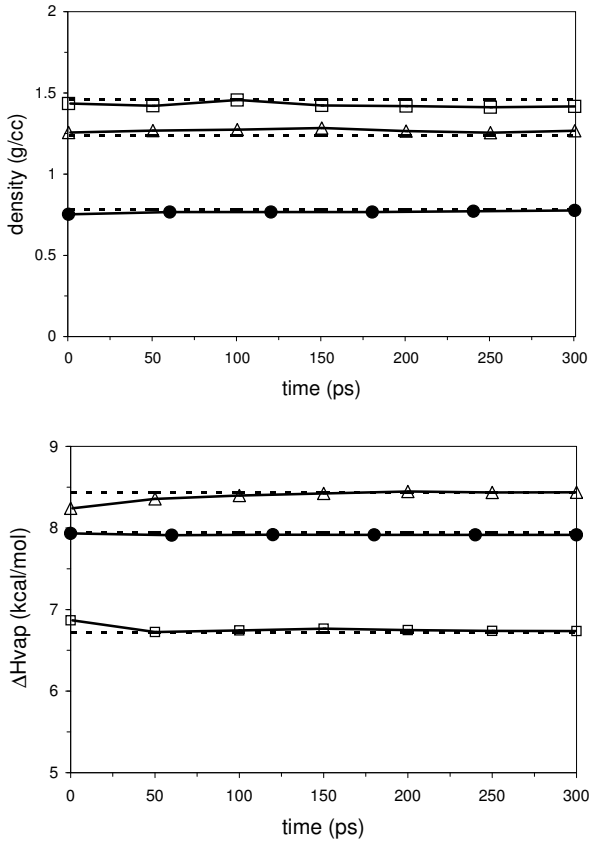

Figure 7.1: (A) Time evolution of the calculated density (grams per cubic centimeter) during the molecular dynamics simulations of the solvent boxes: bromoethane (□, exptl, 1.460), 1,2-dichloroethane (△, exptl, 1.235), and acetonitrile (•, exptl, 0.786). (B) Time evolution of the calculated enthalpy 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.

Table 7.2: Summary of the COMBINE Models.

model[a]	BUW[b]	UV[c]	FFD[d]	E^{e}_{VDW}	E^{f}_{ELE}	$E^{g}_{DESOLV-SUR}$	$E^{ES\ h}_{ELE}$	$\Delta G^{S\ i}_{DESOLV}$	$\Delta\ G^{E\ j}_{DESOLV}$	obj[k]	var[l]	A^{m}	R^2	SDEC	Q^2	$SDEP_{int}$
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

Table 7.2: Summary of the COMBINE Models—Continued.

model[a]	BUW[b]	UV[c]	FFD[d]	E^{e}_{VDW}	E^{f}_{ELE}	$E^{g}_{DESOLV-SUR}$	$E^{ES\ h}_{ELE}$	$\Delta G^{S\ i}_{DESOLV}$	$\Delta\ G^{E\ j}_{DESOLV}$	obj[k]	var[l]	A^{m}	R^2	SDEC	Q^2	$SDEP_{int}$
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

Table 7.2: Summary of the COMBINE Models—Continued.

model[a]	BUW[b]	UV[c]	FFD[d]	E^{e}_{VDW}	E^{f}_{ELE}	$E^{g}_{DESOLV-SUR}$	$E^{ES\ h}_{ELE}$	$\Delta G^{S\ i}_{DESOLV}$	$\Delta\ G^{E\ j}_{DESOLV}$	obj.[k]	var[l]	A^{m}	R^2	SDEC	Q^2	SDEP_{int}
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

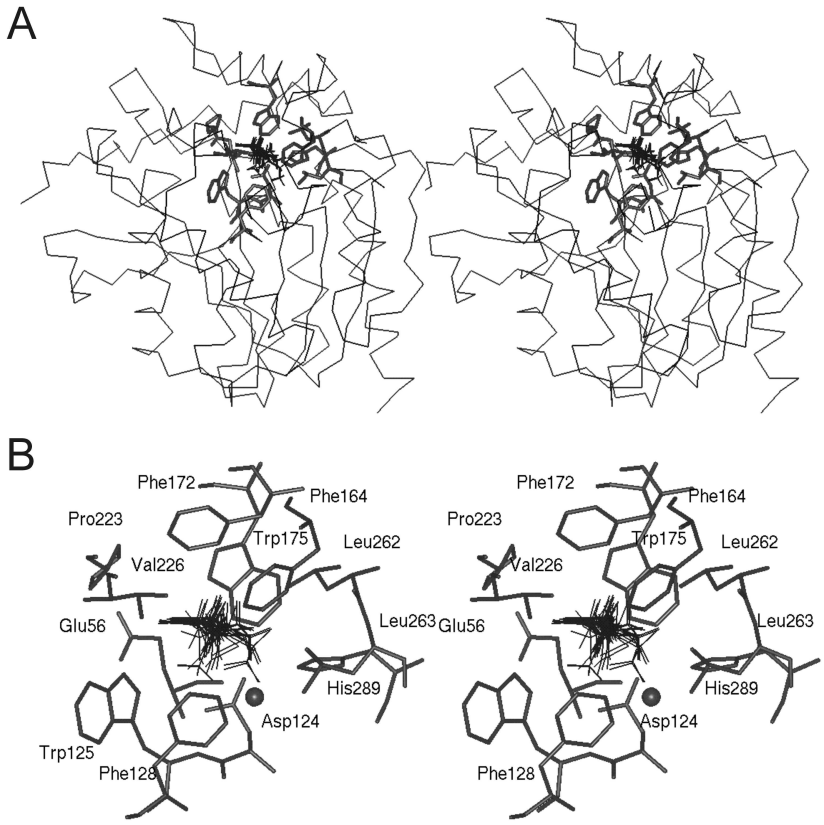

Figure 7.2: Stereoview of the DhlA haloalkane dehalogenase in complex with ligands. The 18 ligands are superimposed in the positions obtained from energy refinement. The protein backbone is represented by a C_α trace (A). The side chains of first-shell residues lining the active site that are identified as significant contributors to substrate specificity in COMBINE analysis are shown in stick representation, and in a magnified view in (B) where they are labelled according to the DhlA sequence. The active site water molecule that is included in the COMBINE analysis is shown as a sphere.

Effect of Scaling on the Predictive Ability
The unscaled models have statistical criteria similar to the BUW-scaled models (Table 7.2). 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 7.3). 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 7.4. 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.

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 7.2), 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 greater robustness of the models derived without variable selection (Table 7.3).

Effect of Surface Desolvation Energy on the Predictive Ability
Inclusion of the surface desolvation energy term ($E_{DESOLV-SUR}$) 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 7.2). The variable $E_{DESOLV-SUR}$ showed low weighted regression coefficients and was eliminated from the models that employed variable selection. Addition of $E_{DESOLV-SUR}$ 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 mo-

Table 7.3: External Validation and Permutation Test of COMBINE Models.

model[a]	obj[b]	var[c]	A^d	R^2	SDEC	Q^2	SDEP_{int}	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 a 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

Table 7.4: 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
172^{vdw}	**124**^{vdw}	**124**^{vdw}	**125**^{vdw}	**124**^{vdw}	**124**^{vdw}	**164**^{vdw}	**125**^{vdw}	**125**^{vdw}
125^{vdw}	**125**^{vdw}	**125**^{vdw}	**172**^{vdw}	**125**^{vdw}	**125**^{vdw}	**172**^{vdw}	**164**^{vdw}	120^{vdw}
164^{vdw}	**172**^{vdw}	**164**^{vdw}	**164**^{vdw}	**172**^{vdw}	**164**^{vdw}	**125**^{vdw}	120^{vdw}	126^{vdw}
124^{vdw}	**164**^{vdw}	**172**^{vdw}	**124**^{vdw}	**164**^{vdw}	**172**^{vdw}	171^{vdw}	145^{vdw}	**164**^{vdw}
222^{vdw}	**175**^{vdw}	222^{vdw}	222^{vdw}	**175**^{vdw}	222^{vdw}	97^{ele}	126^{vdw}	97^{ele}
175^{vdw}	**289**^{vdw}	**289**^{vdw}	**175**^{vdw}	**289**^{vdw}	**289**^{vdw}	261^{vdw}	97^{ele}	**172**^{vdw}
128^{vdw}	222^{vdw}	**175**^{vdw}	**128**^{vdw}	222^{vdw}	**175**^{vdw}	196^{vdw}	**172**^{vdw}	171^{vdw}
289^{vdw}	**223**^{vdw}	**262**^{vdw}	**289**^{vdw}	**223**^{vdw}	176^{ele}	**128**^{vdw}	171^{vdw}	196^{vdw}
223^{vdw}	**125**^{ele}	176^{ele}	**223**^{vdw}	125^{ele}	293^{ele}	173^{vdw}	127^{vdw}	121^{ele}
226^{vdw}	260^{ele}	293^{ele}	**226**^{vdw}	260^{ele}	64^{ele}	**175**^{vdw}	261^{vdw}	59^{ele}
263^{vdw}	176^{ele}	64^{ele}	**263**^{vdw}	176^{ele}	168^{vdw}	120^{vdw}	130^{vdw}	261^{vdw}
262^{vdw}	**226**^{ele}	168^{vdw}	**262**^{vdw}	226^{ele}	167^{vdw}	145^{vdw}	196^{vdw}	189^{vdw}
124^{ele}	**172**^{ele}	167^{vdw}	**124**^{ele}	**262**^{vdw}	271^{vdw}	189^{vdw}	**128**^{vdw}	**128**^{vdw}
179^{vdw}	**175**^{ele}	271^{vdw}	179^{vdw}	**172**^{ele}	230^{ele}	168^{vdw}	121^{ele}	273^{ele}
125^{ele}	**262**^{vdw}	230^{ele}	**125**^{ele}	**175**^{ele}	291^{ele}	104^{vdw}	**175**^{vdw}	104^{vdw}
176^{vdw}	179^{vdw}	291^{ele}	176^{vdw}	179^{vdw}	116^{ele}	184^{vdw}	104^{vdw}	203^{vdw}
172^{ele}	**263**^{ele}	116^{ele}	**172**^{ele}	176^{vdw}	72^{ele}	203^{vdw}	184^{vdw}	184^{vdw}
226^{ele}	311^{ele}	72^{ele}	**226**^{ele}	**263**^{ele}	244^{ele}	170^{vdw}	203^{vdw}	163^{ele}
260^{ele}	**164**^{ele}	244^{ele}	**56**^{vdw}	**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}

Table 7.4: Most Important X Variables[a] in COMBINE Models—Continued.

	model 1	model 2	model 3	model 4	model 5	model 6	model 7	model 8	model 9
	56^{vdw}	**128**^{vdw}	134^{vdw}	224^{vdw}	**128**^{vdw}	274^{vdw}	167^{vdw}	122^{vdw}	57^{ele}
	224^{vdw}	**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

dels 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 equation 7.7, 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, ΔG^E_{DESOLV} and ΔG^S_{DESOLV}, were incorporated in the analysis as additional terms. Addition of ΔG^E_{DESOLV} and ΔG^S_{DESOLV} 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 ΔG^E_{DESOLV} and ΔG^S_{DESOLV} 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 7.4). 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 SDEP_{INT} of 0.59, and shows good external validation (Tables 7.4 and 7.5). Validation 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 7.3, while the loading plots are shown in Figure 7.4. 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 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 7.5. The goodness of fit and external predictive ability of the model is presented in Figure 7.6. A stereoview of the energy contributions selected by COMBINE analysis with the assigned type of energy contributions is shown in Figure 7.7. 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, Cys150) on the opposite side of the active site is shown in Figure 7.8.

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 [63, 64]. Four substrate molecules with available experimental binding constants were modelled in the active sites of the mutant proteins. The experimental binding constants were calculated only for models with the best 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 7.5). 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 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 [19]. In this study, the applicability of COMBINE analysis for protein engineering purposes has been investigated. COMBINE

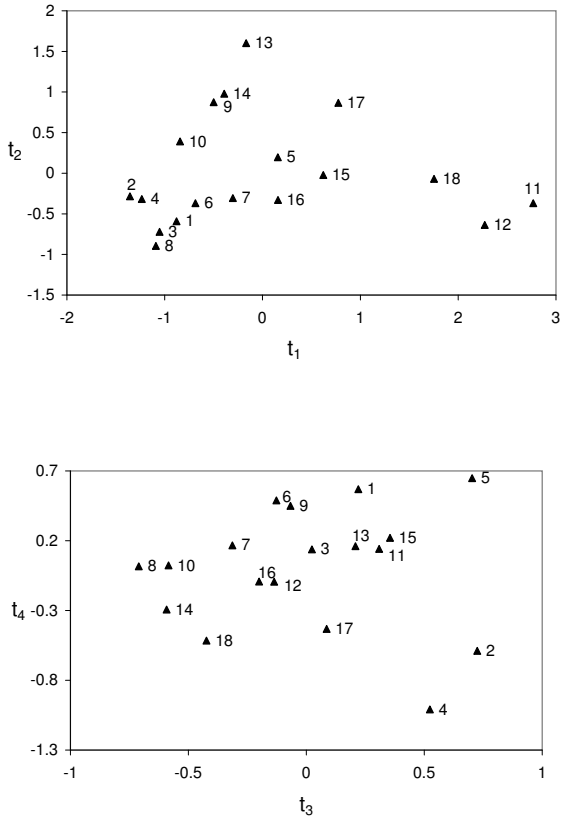

Figure 7.3: Scores plots for latent variables t_1 *versus* t_2 (A) and t_3 *versus* t_4 (B) for model 4. The objects (compounds) are numbered according to Table 7.1.

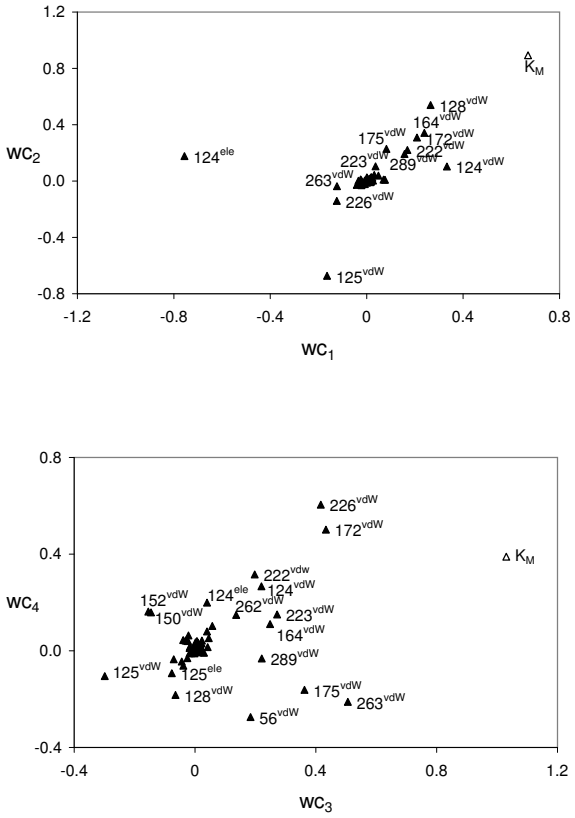

Figure 7.4: Loading plots wc_1 *versus* wc_2 (A) and wc_3 *versus* wc_4 (B) for model 4. Selected variables (energy contributions) are numbered according to the DhlA sequence.

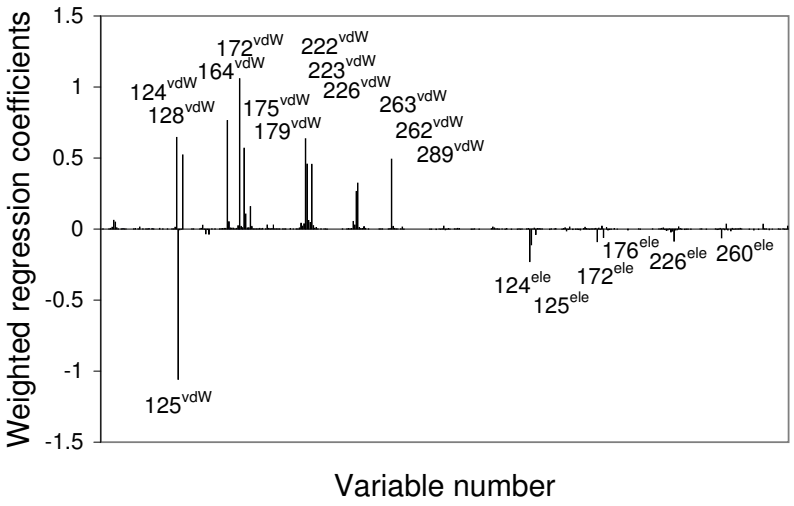

Figure 7.5: Plot of weighted regression coefficients for model 4. Selected variables (energy contributions) are numbered according to the DhlA sequence.

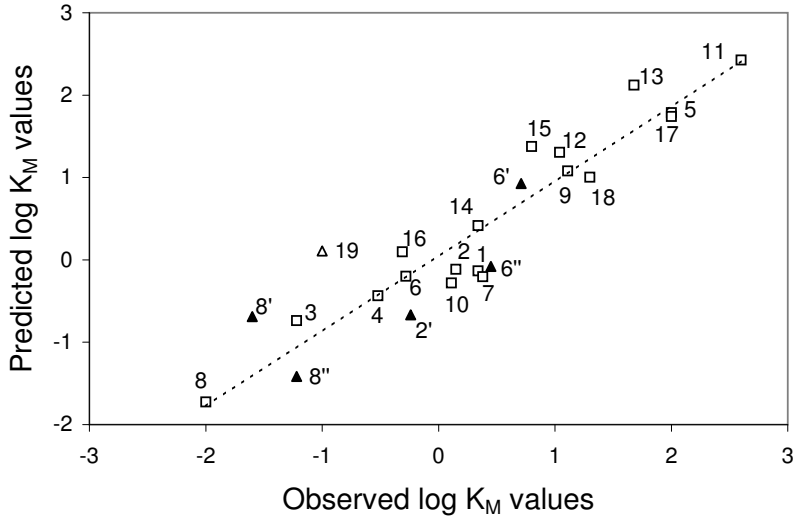

Figure 7.6: Plot of observed *versus* predicted K_m values for model 4. The objects (compounds) are numbered according to Tables 7.1 and 7.5. Compounds from the working (training) set are represented by as white squares, and compounds from the prediction set are represented as triangles (the white triangle 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.

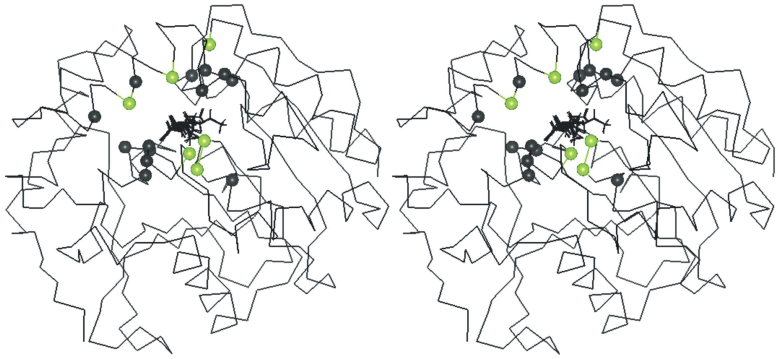

Figure 7.7: Stereoview of DhlA with assigned energy contributions. The protein backbone is represented by the C_α trace. The C_α atoms of the residues showing the most important energy contributions in model 4 are shown as balls (see Table 7.4). Dark-colored residues have only van der Waals energy contributions, while light-colored residues have both van der Waals and electrostatic energy contributions. The substrate molecules are in the positions obtained from energy minimization.

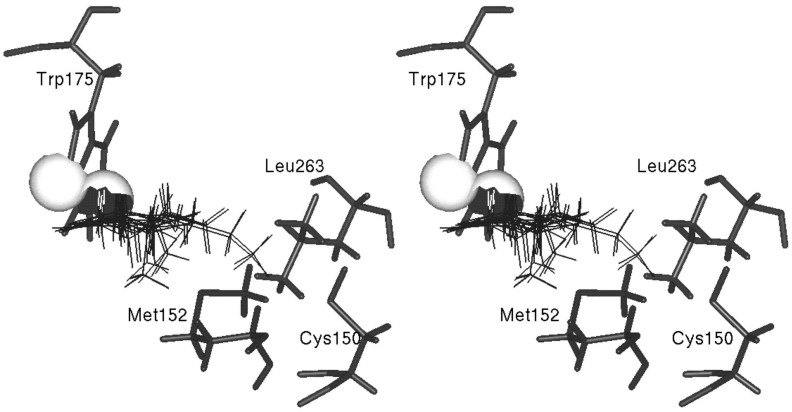

Figure 7.8: Stereoview of the substrate molecules docked in the enzyme active site. Unfavorable van der Waals interactions between long-chain substrates, i.e., butanes and hexanes, and Trp175 are represented by van der Waals surfaces of interacting atoms. Favorable interactions of the same long-chain molecules with Leu263, Met152 and Cys150 positioned on the opposite side of the active site can be seen.

Table 7.5: External Predictions of Apparent Dissociation Constants for Haloalkane Dehalogenase Mutants Using Models 4 and 40.

		wt	Phe172Trp		
no.[a]	substrate	expt[b]	expt[b]	pred[d]	pred[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-chlorohexane	0.07	0.10	1.28	1.38
		wt	Trp175Tyr		
no.[a]	substrate	expt[c]	expt[c]	pred[d]	pred[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 reference [64]; [c] From reference [63]; [d] Predicted using model 4; [e] Predicted using model 40

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 $\text{SDEP}_{INT} \leq 0.59$), 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 modelling 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 equa-

tion improved the quality of the models and resulted in a COMBINE model that achieved very good predictive ability ($Q^2 = 0.74$, $\text{SDEP}_{INT} = 0.59$, $\text{SDEP}_{EXT} = 0.67$). 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 ($\text{SDEP}_{INT} = 0.54$) and slightly worse predictions for objects not included in model development ($\text{SDEP}_{EXT} = 0.82$).

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 7.3), loading plots (Figure 7.4), weighted regression coefficients plot (Figure 7.5), and the structures of enzyme–substrate complexes (Figure 7.7) 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 to 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 by COMBINE analysis [14, 19–23] 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.

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 7.4 and Figure 7.7). 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 [52, 72]. This attack leads to formation of a covalent alkyl–enzyme ester and a halide ion. Asp124 is positioned on a nucleophile elbow [49] 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 [51]. 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 [75]. 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 $HN_{\varepsilon 1}$. Mutagenesis of Trp175 results in proteins with low activity [63, 75]; therefore improved binding for these substrates can only be achieved by mutations in neighboring residues (helix 5). Nine out of 12 *in vivo* mutants of DhlA with improved activity toward 1-chlorohexane [74] carried modifications in helix 5 or its close surroundings. Priest and co-workers 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 7.4 and Figure 7.7). The importance of Val226 could not be directly attributed to the binding of specific 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 [76]. 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 [85, 89]. Schanstra and co-workers [64] performed mutational analysis in position 172 and constructed 16 different point mutants, some of which had modified activity and substrate range. Quantitative structure–function relationships (QSFR) analysis conducted with the same set of point mutants identified physico-chemical properties critical for position 172: aromaticity, main-chain flexibility, refractivity, and bulkiness [79]. 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 ΔG^{S}_{DESOLV} are negatively correlated in this model and explain mainly the variability in the dissociation constants for the sub-

strates 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 [61], from site-directed mutagenesis experiments [75], and from molecular modelling [85, 92]. 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.

CONCLUSIONS

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 7.4). These residues represent suitable targets for future site-directed mutagenesis experiments.

ACKNOWLEDGEMENTS

This work was supported by NATO Linkage Grant MTECH.LG.974701.

SUPPORTING INFORMATION AVAILABLE

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

Chapter 8

Comparative Binding Energy Analysis of Haloalkane Dehalogenase Substrates: Modelling of Enzyme–Substrate Complexes by Molecular Docking and Quantum Mechanical Calculations

Jan Kmuníček, Michal Boháč, Santos Luengo, Federico Gago, Rebecca C. Wade, and Jiří Damborský

Journal of Computer-Aided Molecular Design, **17**: 299–311 (2003)

Comparative Binding Energy Analysis of Haloalkane Dehalogenase Substrates: Modelling of Enzyme–Substrate Complexes by Molecular Docking and Quantum Mechanical Calculations

Jan Kmuníček,[1] Michal Boháč,[1] Santos Luengo,[2] Federico Gago,[2] Rebecca C. Wade,[3] and Jiří Damborský[1,4]

National Center for Biomolecular Research, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic; Department of Pharmacology, University of Alcalá, E-28871 Alcalá de Henares, Madrid, Spain; and European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany

ABSTRACT

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 (DhlA) 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 DhlA-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

[1]Masaryk University.
[2]University of Alcalá.
[3]European Media Laboratory.
[4]To whom correspondence should be addressed. Telephone: +420-5-41129377. Fax: +420-5-41129506. E-mail: jiri@chemi.muni.cz.

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 DhlA and other related enzymes.

INTRODUCTION

Comparative binding energy (COMBINE)[1] analysis is a computational technique for deriving detailed quantitative structure-activity relationships (QSAR) from a set of three-dimensional structures of protein-ligand complexes [19]. 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 standard 3D-QSAR method [19–29, 150]. The application of this methodology in the protein-engineering field was first described in a study of the substrate specificity of haloalkane dehalogenase DhlA (Figure 8.1) from *Xanthobacter autotrophicus* GJ10 [28]. As the crystal structure of the DhlA-dichloroethane complex was known [52], the Michaelis complexes for the other substrates studied could be constructed by structural alignment to this structure [28]. 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. [59] and LinB from *Sphingomonas paucomobilis* UT26 [56], both of which have active sites between 2 and 2.5 times larger than that in DhlA [51].

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

[1] Abbreviations: COMBINE, comparative binding energy; DhlA, haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10; E_a, activation energy; ΔH, 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

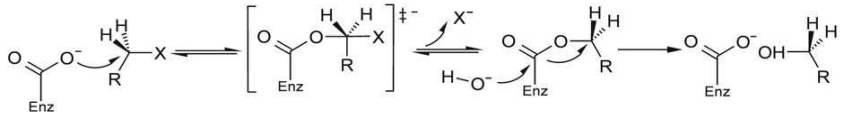

Figure 8.1: Scheme of the reaction mechanism of hydrolytic dehalogenation catalyzed by the haloalkane dehalogenases. Enz – enzyme.

evaluate the applicability of a well-established docking program for the preparation of structures of DhlA-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 DhlA substrates. The K_m values determined by Schanstra *et al.* [67] were logarithmically transformed (Table 8.1). 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 [138] which consists of three main separate modules (AUTOTORS, AUTOGRID, AUTODOCK) and a set of additional scripts. Initially, program AUTOTORS [138] was applied to the eighteen substrates to specify their rotatable bonds. The crystal structure of the DhlA-dichloroethane complex (PDB-ID 2DHC) was obtained from the Brookhaven Protein Database [139]. Polar hydrogen atoms were added using the program WHATIF 5.0 [127]. His289 was singly protonated on N_δ in accordance with its catalytic function. Non-polar hydrogen atoms were added using AMBER 5.0 [126]. 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 [138]. Grid maps were calculated for the atom types present in

Table 8.1: Apparent Dissociation Constants (in mM) of Haloalkane Dehalogenase DhlA[a].

no.	compound	K_m	no.	compound	K_m
1	1-chlorobutane	2.20	10	1,2-dibromopropane	1.30
2	1-chlorohexane	1.40	11	2-chloroethanol	400.00
3	1-bromobutane	0.06	12	2-bromoethanol	11.00
4	1-bromohexane	0.30	13	epichlorohydrine	48.00
5	1,1-dichloromethane	100.00	14	epibromohydrine	2.20
6	1,2-dichloroethane	0.53	15	2-chloroacetonitrile	6.30
7	1,1-dibromomethane	2.40	16	2-bromoacetonitrile	0.49
8	1,2-dibromoethane	0.01	17	2-chloroacetamide	100.00
9	1,2-dichloropropane	13.00	18	2-bromoacetamide	20.00

[a] From reference [67]

the substrates using the AUTOGRID [138] program with a grid of 81 × 81 × 81 points and a grid spacing of 0.25 Å. For docking, a Lamarckian Genetic Algorithm [138] was employed with a population of 50 individuals, a maximum number of 1.5×10^6 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 [140] 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 [118]. 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 non-bonded cutoff of 10 Å and a distance-dependent dielectric constant ($\varepsilon = 4r_{ij}$) 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 se-

lected 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 [141], with subroutine DRIVER 1.0 [142] interfaced to the program TRITON 2.0 [143], was used for mapping the reaction pathway of the first dehalogenation step, as described previously [85]. Briefly, the S_N2 reaction was modelled by a stepwise (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 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 energies (sum of squares lower than 10^{-7} kcal/mol) were eliminated before the analysis. The program SIMCA-P 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 correla-

tion 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 DhlA (Figure 8.2). 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, 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 α or the β 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 8.2). 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 (ΔH) was expressed as the difference in heat of formation between the Michaelis complex and the enzyme-product structures. The activation energy barriers and the reaction enthalpies are listed in Table 8.2. 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

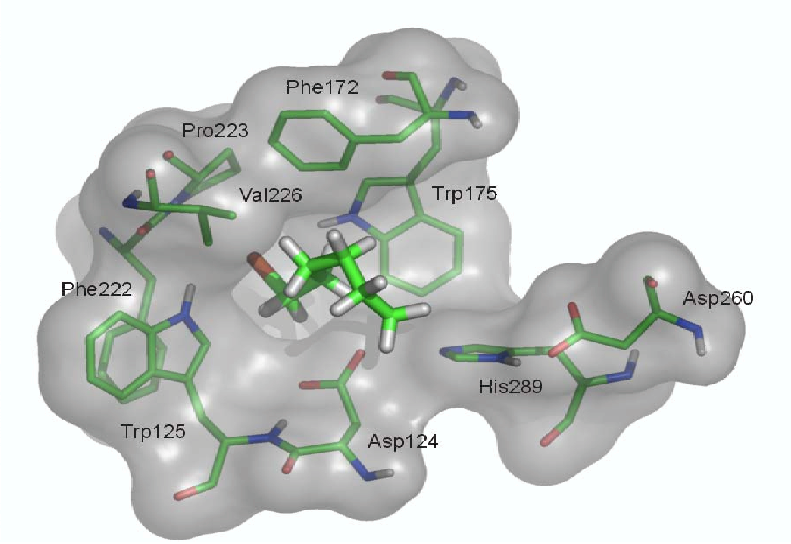

Figure 8.2: Active site cavity of DhlA with docked substrate 1-bromohexane. The electrophilic carbon of 1-bromohexane is oriented toward the nucleophilic oxygen of Asp124 in a position suitable for S_N2 attack. Asp124 together with Asp260 and His289 constitutes the catalytic triad [52]. The scissile bond is pointing toward the halide-binding pocket made of primary halide-stabilizing residues Trp125 and Trp175, and secondary halide-stabilizing residues Phe172, Pro223 and Val226 [90].

of DhlA, as demonstrated in previous experimental [61, 63, 65, 75] and theoretical [90, 91, 95, 144] studies. The magnitude of this stabilization, as quantitated by ΔH, was therefore the first criterion for evaluation of different substrate orientations. If tryptophan stabilization was missing, then ΔH displayed more positive values. Stabilization was fulfilled by all binding modes except the mode that assumed the nucleophilic attack on the β 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 α 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 activation 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 8.2). 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 halohexanes 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 halohexanes penetrate deeper into the binding site of the protein and are oriented toward the nucleophile. Fixation of the halohexane 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 halohexanes in the final model. This fixation stopped translation of halohexanes in the active site and retained the favorable position obtained from the automated docking procedure.

Table 8.2: Activation Energy Barriers and Reaction Enthalpies (in kcal/mol) for the Substrates with Multiple Orientations.

no.	compound	orient	E_a	ΔH	$R^{2\ d}$	$Q^{2\ d}$	SDEP
1	1-chlorobutane	1_1[a]	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[a]	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[a]	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[a]	26.3	−7.3	0.98	0.94	1.1776
10	1,2-dibromopropane	10_1[b]	31.2	0.8	0.99	0.94	0.6684
		10_2[ac]	24.9	−9.1	0.99	0.66	0.7457

[a] Orientation selected for COMBINE analysis; [b] Dehalogenation from C_β; [c] Dehalogenation from C_α; [d] Parameter after adding this orientation to the model based on single-orientation substrates

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 ($model_N$) was compared to our previously reported COMBINE model ($model_P$) [28] that was based on a manual superposition of the substrates onto the experimentally determined structure of 1,2-dichloroethane in its complex with DhlA. To allow direct comparison of the models, $model_P$ was re-calculated with the same fifteen substrates that are present in $model_N$. 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 8.3. The quality of fit for the four-component models is shown in Figure 8.4. Both approaches provided models with very similar statistics, except for a significantly improved predictive power for $model_N$, as indicated by the lower RMSEP values (in models with more than two principal components). The better predictive ability of $model_N$ 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 8.3).

Chemometric Comparison of the COMBINE Models
Score plots of the first two components (Figure 8.5A) 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 $model_N$. Haloethanols (compounds 11 and 12) are the two most polar substrates and their binding affinities are described by the first principal component (Figure 8.5A), which is mainly determined by the electrostatic interaction energy with Asp124 (Figure 8.5B). There is a hydrogen bond present between the hydroxyl groups of the haloethanols and the nucleophilic oxygen of Asp124. Asp124 also shows significant favorable 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 8.6A) differ significantly between the two models with regard to the positions of the halohexanes 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 8.6B) 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 favorable van der Waals interactions with the halogen atoms of every substrate molecule in model$_N$. Residues Leu262 and Leu263, on the other hand, make favorable van der Waals interactions only with selected substrates, that is, the halobutanes (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 halohexanes and halobutanes, thus resolving the steric hindrance of these molecules with Leu263 in model$_P$. Two van der Waals energy contributions that are significant in model$_P$ (Cys150 and Met152) do not have an analogous counterpart in model$_N$ (Figure 8.6B). This is due to the direct contact of the carbon tail of halohexanes with Cys150 and Met152 in model$_P$ that is missing in model$_N$. These missing van der Waals interactions are better seen on the graph of weighted regression coefficients (Figure 8.7). 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$_N$. 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$_N$. The described electrostatic interactions with epihalohydrines have a long-range character, as seen from the distribution of interacting residues within the structure of DhlA (Figure 8.8).

Structural Comparison of the COMBINE Models

The structures of DhlA–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 8.9). The substrates in the complexes obtained by means of the automated molecular docking program are generally adjusted to the size and

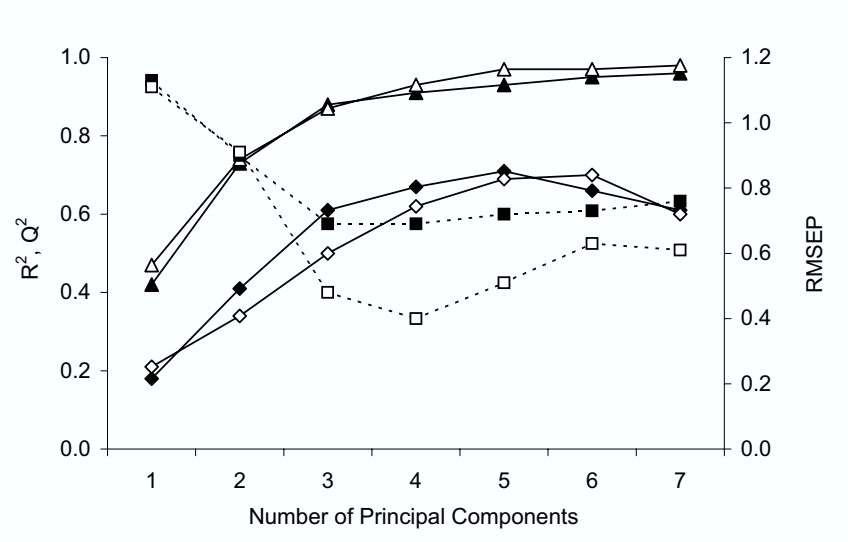

Figure 8.3: Comparison of correlation coefficient (R^2, triangles), cross-validated correlation coefficient (Q^2, diamonds) and root-mean-square error of prediction (RMSEP, squares) for model$_P$ (filled objects) and model$_N$ (empty objects).

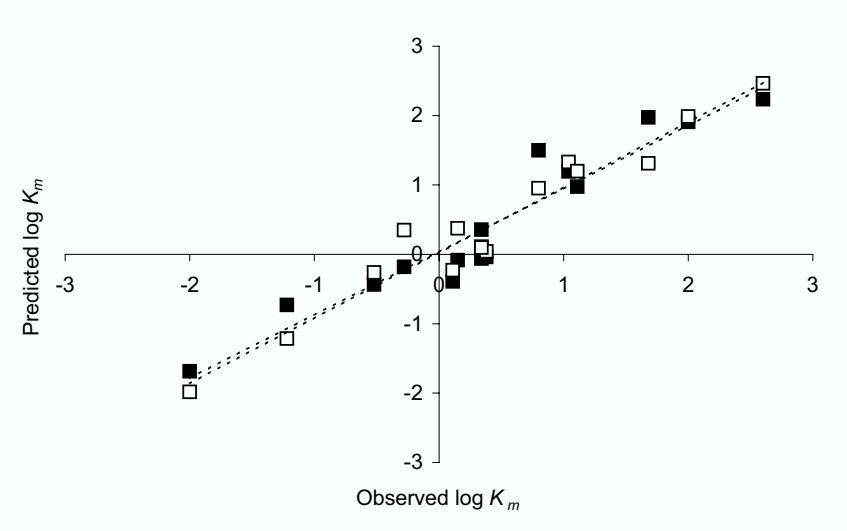

Figure 8.4: Comparison of observed *versus* predicted K_m values for model$_P$ (filled squares) and model$_N$ (empty squares).

Table 8.3: External Predictions of Apparent Dissociation Constants (in mM) for Haloalkane Dehalogenase Mutants.

		wt	Phe172Trp		
no.[a]	substrate	expt[b]	expt[b]	pred[d]	pred[e]
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-chlorohexane	0.07	0.10	1.20	0.55
		wt	Trp175Tyr		
no.[a]	substrate	expt[c]	expt[c]	pred[d]	pred[e]
6"	1,2-dichloroethane	0.53	2.85	0.74	5.50
8"	1,2-dibromoethane	0.01	0.06	0.03	0.10

[a] A single prime corresponds to Phe172Trp, and a double prime corresponds to Trp175Tyr; [b] From reference [64]; [c] From reference [63]; [d] Predicted using model$_P$; [e] Predicted using model$_N$

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 8.9). This rotation places the substrate molecules in both a position and a conformation (e.g., *gauche* for 1,2-dichlororethane) perfectly suited for S_N2 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_N2 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 8.9).

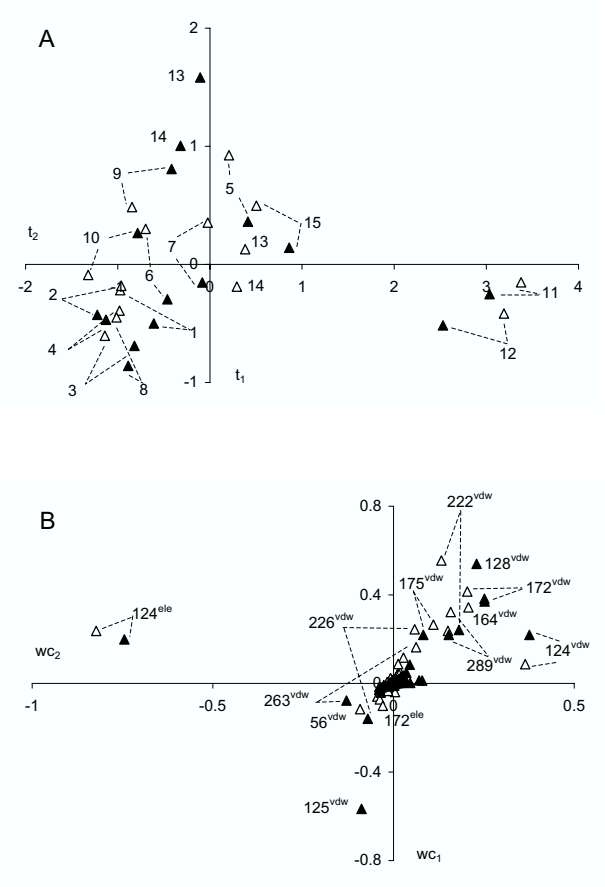

Figure 8.5: Comparison of score (A) and loading (B) plots for the first and second principal components of model$_P$ (filled triangles) and model$_N$ (empty triangles). The compounds are numbered according to Table 8.1 and energy contributions are numbered according to DhlA sequence.

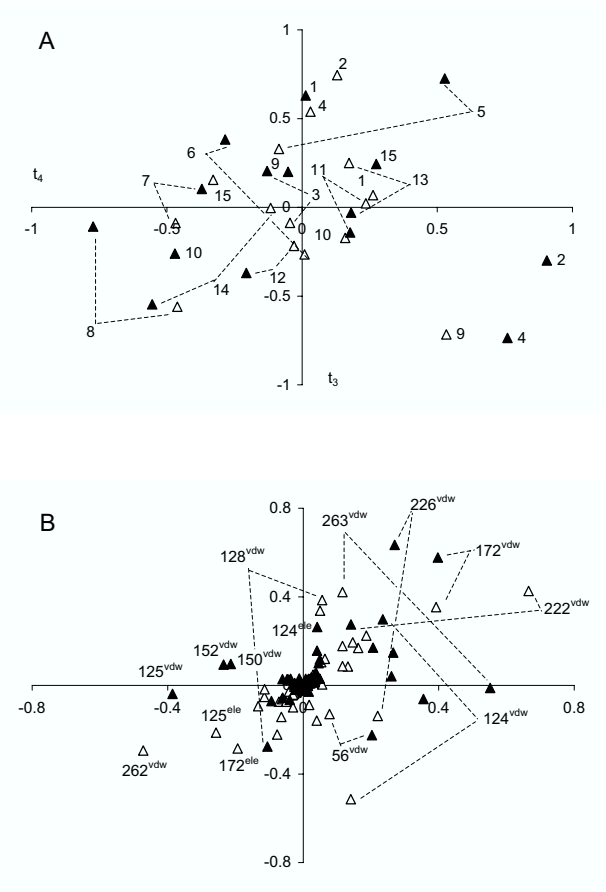

Figure 8.6: Comparison of score (A) and loading (B) plots for the third and fourth principal components of model$_P$ (filled triangles) and model$_N$ (empty triangles). The compounds are numbered according to Table 8.1 and energy contributions are numbered according to DhlA sequence.

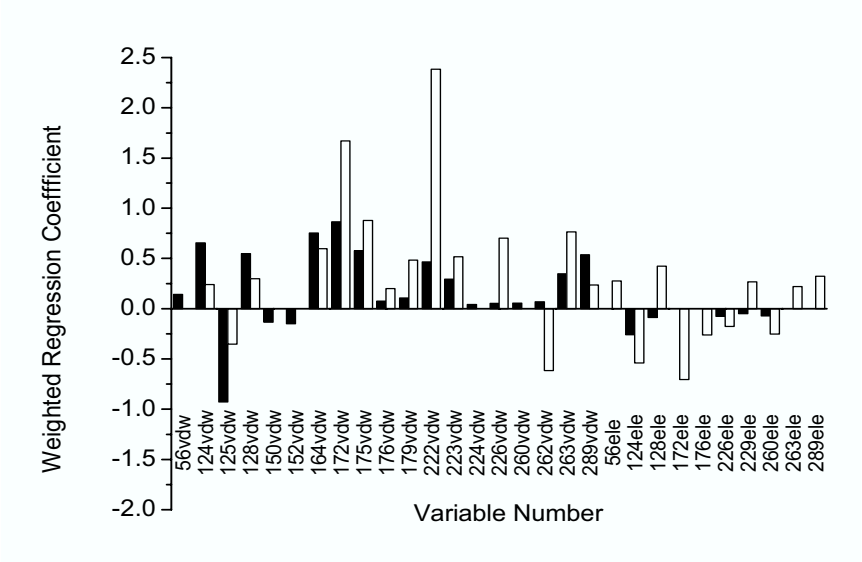

Figure 8.7: Comparison of the weighted regression coefficients for model$_P$ (filled rectangles) and model$_N$ (empty rectangles). Energy contributions are numbered according to the DhlA sequence.

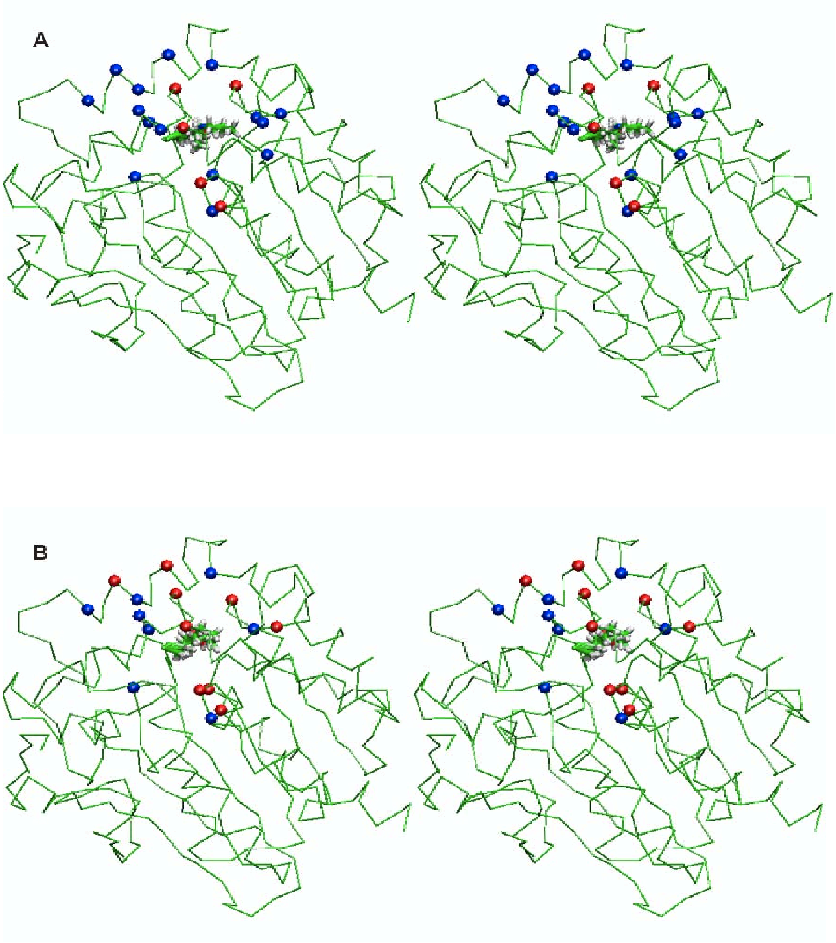

Figure 8.8: Comparison of the energy contributions in model$_P$ (A) and model$_N$ (B). The C_α atoms of the residues showing the most important van der Waals energy contributions (blue) and the residues showing both van der Waals and electrostatic energy contributions (red) in each model are shown as balls.

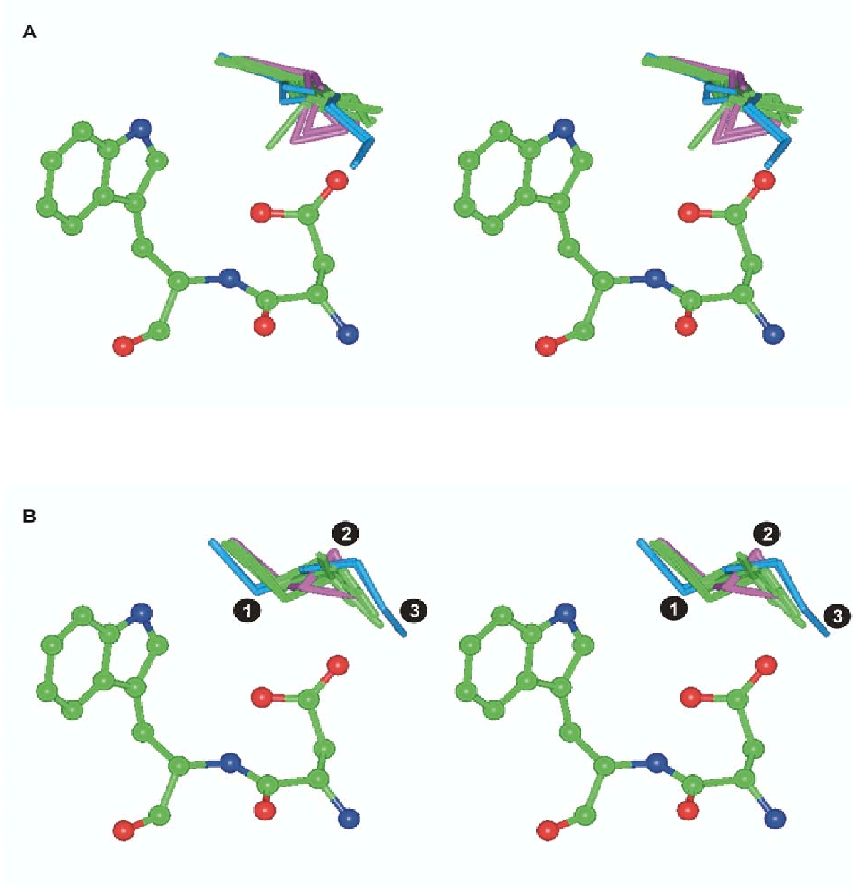

Figure 8.9: Comparison of the orientations and conformations of substrate molecules in the active site of DhlA used for construction of model$_P$ (A) and model$_N$ (B). Two amino acid residues of DhlA (the nucleophile Asp124 and the halide-stabilizing Trp125) are shown in ball-and-stick representation. 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.

CONCLUSIONS

An automated molecular docking procedure provided structures for a set of DhlA–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 orientations 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 DhlA and related dehalogenases. For this purpose, the active site residues of DhlA 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 AVAILABLE

The datasets used for COMBINE analysis and the reaction paths calculated for multiple binding modes. This material is available via Internet at following URL: http://ncbr.chemi.muni.cz/∼jiri.

Chapter 9

Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26

Jan Kmuníček, Tomáš Jedlička, Santos Luengo, Federico Gago, Rebecca C. Wade, and Jiří Damborský

in preparation

Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26

Jan Kmuníček,[1] Tomáš Jedlička,[1] Santos Luengo,[2] Federico Gago,[2] Rebecca C. Wade,[3] and Jiří Damborský[1,4]

National Center for Biomolecular Research, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic; Department of Pharmacology, University of Alcalá, E-28871 Alcalá de Henares, Madrid, Spain; and European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany

ABSTRACT

Haloalkane dehalogenases are microbial enzymes cleaving a carbon-halogen bond in halogenated compounds. Haloalkane dehalogenase LinB isolated from *Sphingomonas paucimobilis* UT26 is a broad-specificity enzyme. Comparative binding energy (COMBINE) analysis for LinB enzyme and a set of twenty five substrates was performed. Two outliers with kinetically different mechanism of dehalogenation [*bis*(2-chloroethyl)ether and chlorocyclohexane] were identified and excluded from the analysis. The two-components COMBINE model explained 92% of experimental data variability (89% cross-validated) and primarily described differences in binding affinities caused by variability in chain length. Binding affinity is increasing with increasing substrate size up to chain length of six carbon atoms and then is decreasing again. The amino acid residues important for explanation of variability in binding affinities of studied substrate were identified: (i) the first shell residues—Asn38, Asp108, Trp109, Glu132, Ile134, Phe143, Phe151, Phe169, Val173, Trp207, Pro208, Ile211, Leu248, and His272, (ii) the tunnel residues—Pro144, Asp147, Leu177, and Ala247 and (iii) the second shell residues—Pro39 and Phe273. The tunnel and the second shell residues are the most attractive targets for substitutions since their replacement will not lead to loss of functionality by disruption of the active site architecture.

[1]Masaryk University.
[2]University of Alcalá.
[3]European Media Laboratory.
[4]To whom correspondence should be addressed. Telephone: +420-5-41129377. Fax: +420-5-41129506. E-mail: jiri@chemi.muni.cz.

INTRODUCTION

Haloalkane dehalogenases are microbial enzymes cleaving a carbon-halogen bond in halogenated alkanes, cycloalkanes, alkenes, selected ethers, and alcohols [47]. Haloalkane dehalogenases act by hydrolytic mechanism using a water molecule as the only co-substrate. These enzymes can be used for protection of the environment, e.g., bioremediation of contaminated areas [48], removal of intermediates of chemical syntheses [41] and in biosensors. Haloalkane dehalogenases show low activity and restricted substrate range with industrially important substrates and therefore represent suitable targets for protein engineering studies.

Computer-assisted protein design followed by site-directed mutagenesis experiment has been proven as powerful method for enzyme modification [105–107, 110, 145–147]. Site-directed mutagenesis allows replacement of selected amino acids residues in the protein structure in a predefined way. The residues appropriate for mutagenesis must be identified in advance using some computational approach. Here we use Comparative binding energy (COMBINE) analysis [19] to derive a predictive model for substrate specificity in which important interactions for binding are designated. COMBINE analysis has been previously used in drug design field [14, 19–27], while its applicability for protein design purposes has been tested only recently. COMBINE model has been constructed for the haloalkane dehalogenase DhlA from *Xanthobacter autotrophicus* GJ10 [28]. This model explained 91% (73% cross-validated) of the quantitative variance in the apparent dissociation constants of eighteen substrates and identified the residues contributing most significantly to the substrate specificity of DhlA: Asp124, Trp125, Phe164, Phe172, Trp175, Phe222, Pro223, and Leu263. This model was later further improved by modified methodology for the construction of enzyme–substrate complexes using automated molecular docking techniques and quantum mechanical calculations to the modelling of enzyme–substrate complexes [30].

Haloalkane dehalogenases belong to α/β-hydrolase fold proteins [49, 50]. The core of each enzyme in a family is similar and consists of two different domains: α/β-fold (main) domain and so-called cap domain. The main domain is composed of an β-sheet constituting of eight β-strands surrounded by six α-helices. The cap domain is composed of an additional five α-helices connected by loops. The active site is located between these two domains in an internal, predominantly hydrophobic cavity and can be reached from the solvent through a

tunnel. At least three different groups of haloalkane dehalogenases can be distinguished according to different substrate specificity [51]. Each of these categories has its own representative with known three-dimensional structure: haloalkane dehalogenase DhlA [52–55], haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26 [56–58], and haloalkane dehalogenase DhaA from *Rhodococcus rhodochrous* NCIMB 13064 [59]. The ratio of volumes of the active sites for three substrate specificity representatives DhlA : DhaA : LinB was determined as 1 : 2 : 2.5 [51]. The differences in substrate specificity for the three classes are mainly due to the differences in composition and geometry of the active site, the halide-stabilizing residues, and the entrance tunnel connecting the active site with the protein surface. It is therefore interesting to compare systematically and quantitatively structural determinants of the substrate specificity in different haloalkane dehalogenases. In this paper we apply COMBINE analysis to study of substrate specificity of the haloalkane dehalogenase LinB and compare it with the COMBINE model derived previously for DhlA.

METHODS

Experimental Data
Apparent dissociation constants (K_m) were taken as a measure of binding affinities for a set of 25 substrates (Table 9.1). This set was selected from the 196 compounds using the statistical experimental design (Damborsky *et al.*, in preparation). The K_m values were measured using steady-state kinetics with purified LinB enzyme. The values were logarithmically transformed prior to statistical analysis.

Preparation of Enzyme Structure
The crystal structure of LinB enzyme (PDB-ID 1D07) was obtained from the Brookhaven Protein Database. Polar hydrogen atoms were added using the program WHATIF 5.0 [127]. His272 was singly protonated on N_δ in accordance with its catalytic function. Non-polar hydrogen atoms were added using AMBER 5.0 [126]. 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 [138].

Preparation of Substrate Structures
All substrate molecules were built using program INSIGHTII 95 (Accelrys, USA). The partial atomic charges were calculated by AM1 semi-empirical quantum chemical method using program MOPAC implemented in INSIGHTII. Pro-

gram AUTOTORS of the program package AUTODOCK 3.0 [138] was applied to assign rotatable bonds in the twenty five substrate molecules.

Construction of Enzyme–Substrate Complex Structures
The enzyme–substrate complex structures were prepared using an automated docking procedure of the AUTODOCK 3.0 [138]. Grid maps were calculated for the atom types present in the substrates using the AUTOGRID program of AUTODOCK with a grid of 81 × 81 × 81 points and a grid spacing of 0.25 Å. A Lamarckian Genetic Algorithm [138] was employed for docking with population of 50 individuals, a maximum number of 1.5×10^6 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 [140] 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.

Refinement of Complex Structures and Calculation of Desolvation Energies
The geometry of the selected enzyme–substrate complexes was optimized using AMBER 5.0 and the Cornell *et al.* force field [118]. 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} $Å^{-1}$. Non-bonded cutoff was set to 10 Å and a distance-dependent dielectric constant to $\varepsilon = 4r_{ij}$. The enzyme–substrate interaction energy in the presence of the surrounding solvent together with the change in desolvation energies of the substrate and the enzyme upon binding were estimated. The approach for calculation of the electrostatic contributions to the free energies of binding and the changes of enzyme and substrate desolvation energies requires solving of linear form of the Poisson–Boltzmann equation and is described in more details in the publication of Kmunicek *et al.* [28].

Construction of the COMBINE Models
COMBINE analysis was applied to construct a model correlating log K_m values with the interaction energies calculated for enzyme–substrates complexes on a per residue basis. The statistical method Partial least-squares projection to latent structures (PLS) [15, 16] was used for identification and ranking of interactions important for the differences in apparent dissociation constant among enzyme–substrate complexes. The X matrix consisted of either 594 columns (van der Waals and electrostatic energy contributions for 296 amino acid residues

plus two energy contributions from a catalytic water molecule) or 1 753 columns (the matrix above plus energy contributions from 1 159 crystalographically resolved water molecules) and 25 rows (enzyme–substrate complexes). The dependent variable y was represented by 25 logarithmically transformed values of the apparent dissociation constants K_m. Variables with low magnitude energies (sum of squares lower than 10^{-7} kcal/mol) were eliminated before the analysis. All PLS models were constructed using statistical program SIMCA 8.0 (Umetri, Sweden). 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 provides an estimate of the predictive power of a model, when a value higher than 0.4 is generally considered as statistically significant.

RESULTS

Modelling of Enzyme–Substrate Complexes

The substrate set consisted of monohalogenated alkanes up to the chain length of eight carbon atoms, dihalogenated propanes, ethane and pentane, monohalogenated cyclohexanes, monohalogenated ether and dihalogenated propene. The automated docking procedure provided positionaly suitable orientations for nineteen substrates: 1-chlorobutane (2), 1-chlorohexane (3), 1-chloroheptane (4), 1-chlorooctane (5), 1-bromopropane (6), 1-bromobutane (7), 1-bromohexane (8), 1,3-dichloropropane (11), 1,5-dichloropentane (12), 1,2-dibromoethane (13), 1,3-dibromopropane (14), 1-bromo-3-chloropropane (15), 1,2-dibromopropane (16), 1-bromo-2-methylpropane (18), *bis*(2-chloroethyl)ether (19), 4-bromobutyronitrile (22), 3-chloro-2-methylpropene (23), 3-chloro-2-(chloromethyl)-1-propene (24), and 2,3-dichloropropene (25). Extended docking (128 runs) had to be used for obtaining the reactive conformation of 1-iodohexane (10), chlorocyclohexane (20), and bromocyclohexane (21). No suitable orientations even in extended dockings were found for three substrates: 1-chloropropane (1), 1-iodopropane (9), and 2-bromo-1-chloropropane (17). The orientations for 1-chloropropane and 1-iodopropane were prepared from selected orientation of the 1-bromopropane (7) by manual exchange of halogen atoms and minimization; orientation of 2-bromo-1-chloropropane was prepared in the same manner from selected orientation of 1,2-dibromopropane (16). The substrate orientations obtained from automated docking spatially formed one cluster occupying position properly suited for the dehalogenation reaction. During minimization some substrates drifted away from the original positions and formed a separate

cluster that was less suited for the nucleophilic attack. An alternative approach for the minimization of complexes was therefore applied which kept the substrate molecules in their docked orientations and adjusted by minimization only protein structures (see next paragraph).

Construction and Optimization of COMBINE Models
Initial COMBINE models were built for the set of 25 substrates. The data X were composed of the interaction energies of substrates with the amino acid residues of LinB plus solvent molecules. The logarithmic values of apparent dissociation constant formed data y. The effect of several factors on the statistical quality of models was tested: (I) effect of data pretreatment, (II) effect of fixation, (III) effect of a solvent, and (IV) effect of objects selection.

I. Several different types of data pretreatment were applied on X matrix: (i) no data pre-treatment, (ii) centering only, (iii) scaling to unit variance only, (iv) scaling to unit variance plus centering, (v) pareto scaling only, and (vi) pareto scaling plus centering. Statistical parameters of COMBINE models based on the data with different pretreatment are summarized in Table 9.2. All models employing centering (models M2, M4, M6, M8, M10, and M12) performed poorly. Out of all non-centered models, pareto-scaled models (models M5, M6, M11, and M12) showed slightly better internal predictive power than the models without scaling and with the scaling to unit variance. However, overall statistical quality of all non-centered models was comparable. It was therefore decided to use the data without any pretreatment for construction of the final model.

II. Comparison of models without (M1–M6) and with (M7–M12) fixation of the substrate molecules during minimization of enzyme–substrate complexes revealed that the models without fixation possessed somewhat better statistical parameters. Chemometric analysis of fixed and unfixed models together with careful inspection of the structures of enzyme–substrate complexes lead us propose combined model based on unfixed structures of complexes for all but two longest substrates in the set: 1-chloroheptane (4) and 1-chlorooctane (5). These molecules make close contacts with the tunnel residues resulting in their drift from the reactive position during minimization.

III. The influence of solvent on the COMBINE model was initially tested implicitly by inclusion of 1 159 crystallographic water molecules into the data matrix, i.e., the water molecules acted as the objects in the PLS model, and

later explicitly by adding desolvation energies. Water molecules significantly increased complexity of the PLS model but did not improve its statistical quality (data not shown). Therefore, all water molecules were excluded from further analysis besides catalytic water bound near the catalytic triad. Addition of desolvation energies of both a substrate and an enzyme to the X matrix also did not improve the models and were not incorporated in the final model.

IV. Two outlying objects *bis*(2-chloroethyl)ether and chlorocyclohexane were systematically identified in the PLS models. These two molecules have the lowest k_{cat}/K_m with LinB out of all 25 substrates analyzed in this study and probably posses kinetically different mechanism of dehalogenation (see discussion). Furthermore, the same two compounds were identified as outliers also in the classical quantitative structure-activity relationships analysis relating K_m values to the physico-chemical properties of substrate molecules (Damborsky *et al.*, in preparation).

Chemometric Analysis of Final COMBINE Model
The optimized parameters were used for the construction of the final model: X matrix was uncentered and unscaled; y vector was logarithmically transformed; enzyme–substrate complexes were optimized without fixation for all substrates but 1-chloroheptane and 1-chlorooctane; solvent molecules were not modelled either implicitly or explicitly; two outliers *bis*(2-chloroethyl)ether and chlorocyclohexane were excluded from the analysis. The final model was constructed for the matrix of 23 objects and 595 variables and was composed of two principal components with the correlation coefficient $R^2 = 0.92$ and cross-validated correlation coefficient $Q^2 = 0.89$. The relationship between the experimental and predicted values of the binding affinities is shown in Figure 9.1. Significant interactions for substrate specificity of LinB were identified by calculating weighted regression coefficients (Figure 9.2). Twenty x variables (interaction energies) have been assigned as the most important contributions based on the coefficient values: fifteen of them correspond to van der Waals terms while five correspond to electrostatic terms. These coefficients also provide information about the direction of the effect: fourteen interaction energies show positive contribution while negative interaction. The scores plot (Figure 9.3) displays the distribution of objects (substrates) according to first and the second component of the model. The first principal component separates compounds horizontally into three groups: (i) 1-chlorohexane (3), 1-bromohexane (8), 1-iodohexane (10), and 1,5-dichloropentane (12); (ii) 1-chloropropane (1), 1-bromopropane (6), 1-iodopropane (9), and 2,3-dichloropropene (25); and (iii) substrates not

Table 9.1: Steady-state kinetic parameters of the haloalkane dehalogenase LinB.

no.	compound	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1}.mM^{-1}$)
1	1-chloropropane	1.10	1.12	1.02
2	1-chlorobutane	0.13	1.02	7.94
3	1-chlorohexane	0.01	1.03	206.51
4	1-chloroheptane	0.02	3.35	222.73
5	1-chlorooctane	0.02	2.75	133.17
6	1-bromopropane	0.23	5.52	23.88
7	1-bromobutane	0.04	3.41	78.95
8	1-bromohexane	0.01	1.05	105.05
9	1-iodopropane	0.11	3.94	35.79
10	1-iodohexane	0.01	2.33	232.73
11	1,3-dichloropropane	0.16	0.95	5.94
12	1,5-dichloropentane	0.02	2.45	130.29
13	1,2-dibromoethane	1.90	6.12	3.22
14	1,3-dibromopropane	0.04	6.60	165.00
15	1-bromo-3-chloropropane	0.21	6.89	32.79
16	1,2-dibromopropane	0.14	0.84	6.02
17	2-bromo-1-chloropropane	0.55	1.37	2.49
18	1-bromo-2-methylpropane	0.05	1.60	31.98
19	*bis*(2-chloroethyl)ether	0.87	0.36	0.42
20	chlorocyclohexane	0.25	0.14	0.57
21	bromocyclohexane	0.02	1.34	63.29
22	4-bromobutyronitrile	0.21	3.70	17.90
23	3-chloro-2-methylpropene	0.34	3.08	9.06
24	3-chloro-2-(chloromethyl)-1-propene	0.08	8.17	103.46
25	2,3-dichloropropene	0.54	1.63	3.01

Table 9.2: COMBINE Models.

model	scaling	centering	fixation	objects	variables	A	R^2	Q^2
M1	–	–	–	25	595	2	0.85	0.79
M2	–	+	–	25	595	1	0.49	0.34
M3	unit variance	–	–	25	595	2	0.86	0.82
M4	unit variance	+	–	25	595	1	0.63	0.54
M5	pareto	–	–	25	595	2	0.86	0.82
M6	pareto	+	–	25	595	1	0.56	0.42
M7	–	–	+	25	595	2	0.86	0.81
M8	-	+	+	25	595	1	0.54	0.42
M9	unit variance	–	+	25	595	2	0.85	0.81
M10	unit variance	+	+	25	595	1	0.59	0.49
M11	pareto	–	+	25	595	2	0.87	0.82
M12	pareto	+	+	25	595	1	0.57	0.46

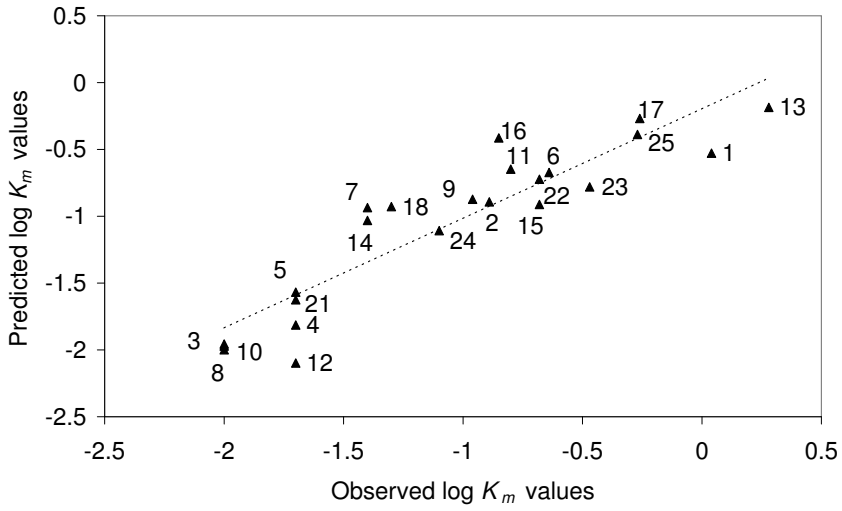

Figure 9.1: Plot of observed *versus* predicted K_m values.

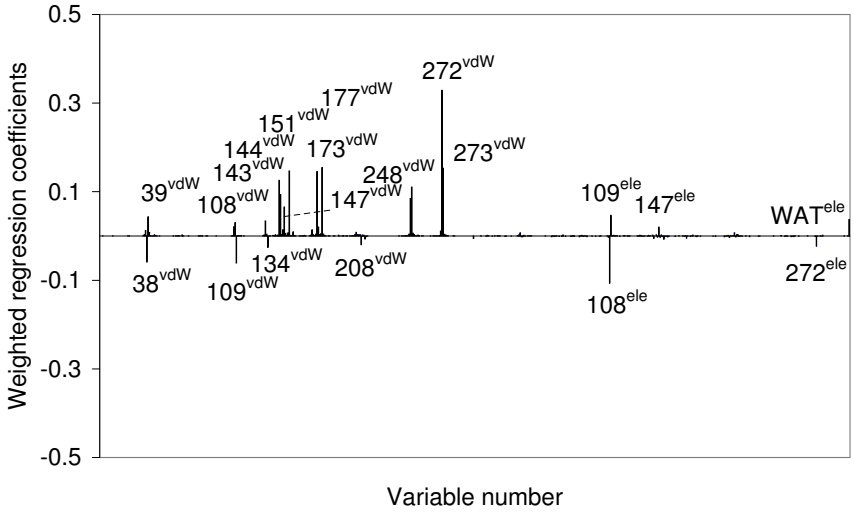

Figure 9.2: Plot of weighted regression coefficients for final COMBINE model. Selected variables (energy contributions) are numbered according to the LinB sequence.

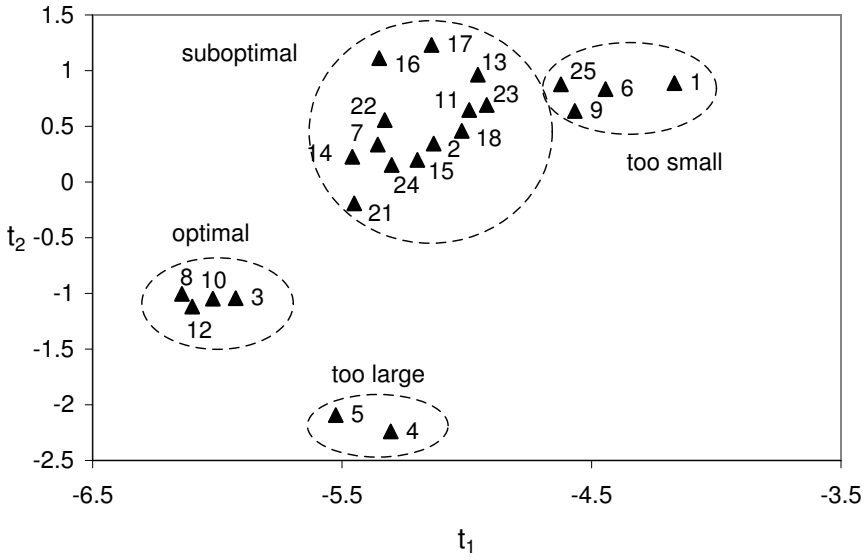

Figure 9.3: Scores plot for latent variables t_1 *versus* t_2. The objects (compounds) are numbered according to Table 9.1.

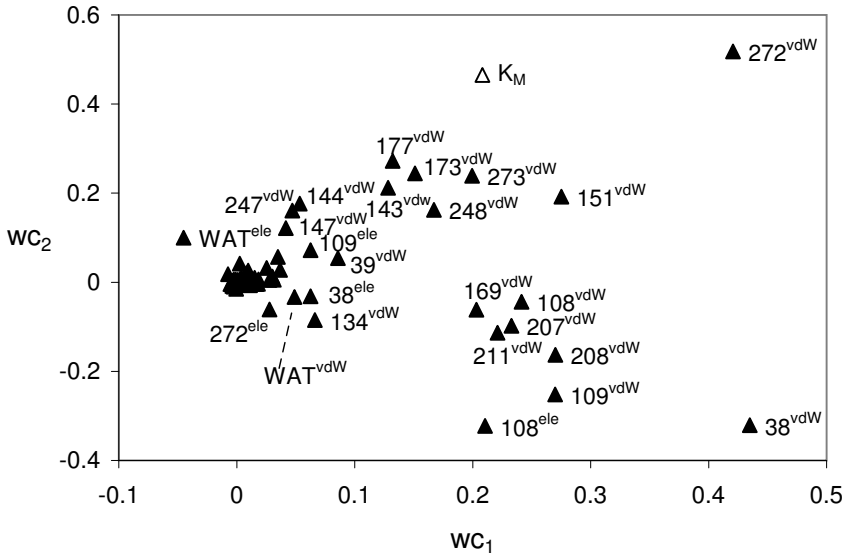

Figure 9.4: Loading plot wc_1 *versus* wc_2. Selected variables (energy contributions) are numbered according to the LinB sequence.

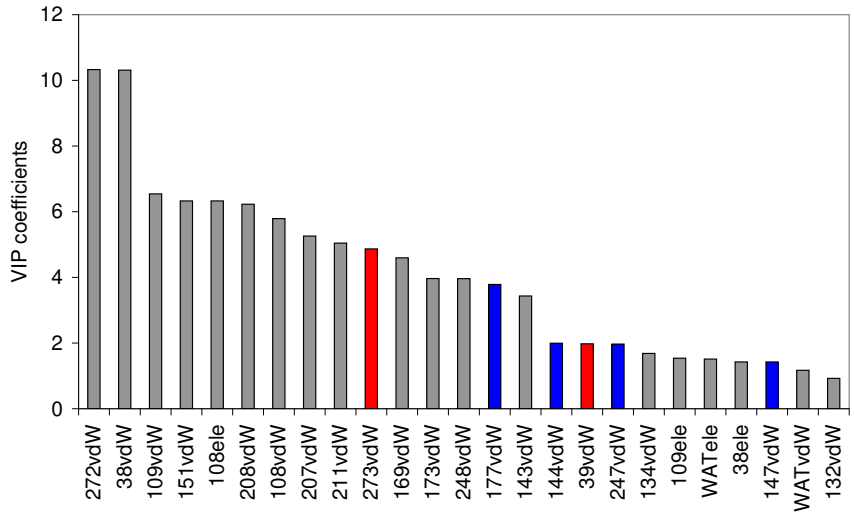

Figure 9.5: Plot of variable importance in the projection. Selected variables (energy contributions) are numbered according to the LinB sequence. The interaction energy contributions shown as red columns (Pro39 and Phe273) correspond to the second shell residues (Pro39, Phe273), interaction energy contributions shown as blue columns (Pro144, Asp147, Leu177 and Ala247) correspond to the tunnel residues and all other interaction energy contributions correspond to the first shell residues.

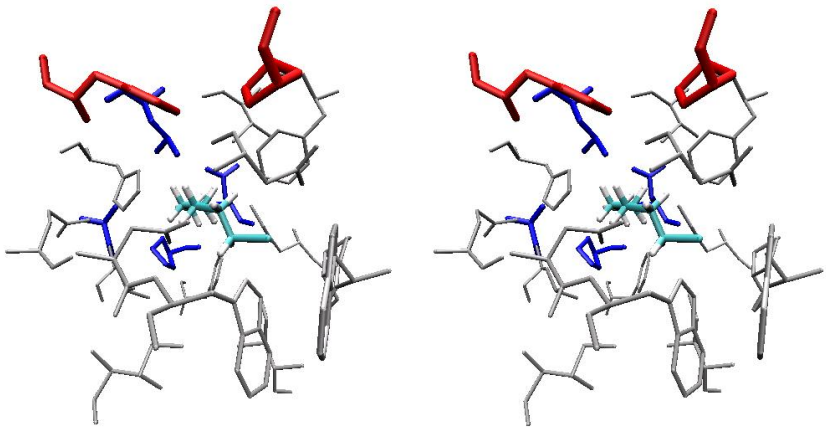

Figure 9.6: Stereoview of LinB active site with amino acid residues selected by COMBINE analysis. The residues are colored according to the type of interaction energy contributions.

included in any of two former groups. The second component separated objects vertically also into three groups: (i) 1-chloroheptane (4) and 1-chlorooctane (5), (ii) 1-chloropropane (1), 1-bromopropane (6), 1-iodopropane (9), and 2,3-dichloropropene (25), and (iii) substrates left over. The loading plot (Figure 9.4) shows the distribution of variables according to their extent of contribution to the individual principal components. The most significant contribution to the first principal component is provided mainly by van der Waals interaction energies 38^{vdw}, 39^{vdw}, 108^{vdw}, 109^{vdw}, 143^{vdw}, 151^{vdw}, 169^{vdw}, 173^{vdw}, 177^{vdw}, 207^{vdw}, 208^{vdw}, 211^{vdw}, 248^{vdw}, 272^{vdw}, 273^{vdw}, and four electrostatic interaction energies 108^{ele}, 109^{ele}, 272^{ele}, and WAT^{ele}. The most important contribution to the second principal component is provided by van der Waals interaction energies 38^{vdw}, 109^{vdw}, 143^{vdw}, 173^{vdw}, 177^{vdw}, 208^{vdw}, 272^{vdw}, 273^{vdw}, and three electrostatic interaction energies 108^{ele}, 147^{ele}, and 272^{ele}.

DISCUSSION

Two outliers were detected during the initial modelling of twenty five substrates of the haloalkane dehalogenase LinB. *Bis*(2-chloroethyl)ether (19) and chlorocyclohexane (20) have the lowest k_{cat}/K_m and were detected as the outliers also in classic quantitative structure-activity relationships (QSAR) analysis (Damborsky *et al.*, in preparation). These evidences suggest that two discussed substrates may by converted by kinetically different dehalogenation mechanism. Kinetic comparison for two analogous substrates, chlorocyclohexane (20) and bromocyclohexane (21), revealed that the hydrolysis of alkyl-enzyme intermediate formed by dehalogenation of chlorocyclohexane is surprisingly 33-times slower compared to bromocyclohexane [148]. The final COMBINE model constructed for twenty three substrates of the haloalkane dehalogenase LinB explained 92% of variability in experimental data (89% cross-validated) and primarily described differences in binding affinities caused by variability in chain length. This property dominated both principal components.

The first principal component explained 78% of variability in experimental data and divided substrates into three groups (Figure 9.3; horizontal direction) according complementarity with the active site. Binding affinity is increasing with increasing substrate size up to chain length of six carbon atoms and then is decreasing again. The optimal length corresponds to group of long-chain substrates 1-chlorohexane (3), 1-bromohexane (8), 1-iodohexane (10), and 1,5-dichloropentane (12); most unfavorable is the group of mono substituted propanes 1-chloropropane (1), 1-bromopropane (6), 1-iodopropane (9),

and 2,3-dichloropropene (25); and suboptimal length corresponds to the rest of the substrates. The first component is to large extend dominated by the energy contributions from Asn38 and His272 (Figure 9.4; horizontal direction). These two amino acid residues are located on the opposite sides of the active site (Figure 9.6). Asn38 forms the bottom of the active site and together with Trp109 stabilizes a halogen [56, 95], while His272 is the base of the catalytic triad [103] and forms the important point of contact for the substrates entering the active site pocket *via* the entrance tunnel. His272 showed extremely strong van der Waals interaction with all studied substrates. The importance of this residue for binding of small ligands near the opening of the entrance tunnel was revealed already by the crystallographic studies [57, 58]. The first component also contributed to the distinction of hexanes from excessively long 1-chloroheptane (4) and 1-chlorooctane (5). This separation would not be possible without fixation of these molecules in the reactive position. Using such methodology we could capture non-linear relationships between the chain-length of the substrate molecules and their binding affinities. Non-linearity was experienced also during the classic QSAR analysis of the same data set (Damborsky *et al.*, in preparation). The second component explained 14% in data variability and also separated substrate molecules primarily according to their size (Figure 9.3; vertical direction). A number of amino acid residues were responsible for the separation of these groups by the second principal component: His272, Leu177, Val173, and Phe273. These residues are located on the side of the active site opposed to the halide-binding pocket (in direction of the tunnel) and make more favorable van der Waals interactions with long substrates then with short ones. Opposite trend holds for the electrostatic interaction energies of Asn38 and Asp108 plus van der Waals interaction energies of Trp109, Pro208, and Ile211. It is appropriate to assume that placing long substrates to the active site will bring their chain close to the wall of the tunnel thus maximizing van der Waals interactions. On the opposite side of the active site, the short substrates could freely occupy the best positions near the amino acid residues located at the bottom of the active site. Positional differences were observed for halogen-stabilizing residue Asn38 in different enzyme–substrate complexes. Asn38 displays high flexibility of both its side-chain and main-chain allowing good accommodation of the active site to different substrates.

Variable importance in projection (VIP) parameter quantifies the overall importance of each variable in PLS model and as such is well suited for identification of amino acid residues representing the best candidates for site-directed mutagenesis. Figure 9.5 shows twenty four energy contributions with the high-

est VIP. Three types of contributing residues (Figure 9.6) can be distinguished: the first shell residues (Asn38, Asp108, Trp109, Glu132, Ile134, Phe143, Phe151, Phe169, Val173, Trp207, Pro208, Ile211, Leu248, and His272), the tunnel residues (Pro144, Asp147, Leu177, and Ala247) and the second shell residues (Pro39 and Phe273). The tunnel and the second shell residues are the most attractive targets for substitutions since their replacement will not lead to loss of functionality by disruption of the active site architecture. The relevance of such a proposal has already been proven by mutants constructed previously using directed evolution and side-directed mutagenesis techniques. The equivalents of the tunnel residue L177 and the second-shell residue Phe273 in DhaA (Cys176 and Tyr273, respectively) were identified as hot spots for specificity of this enzyme in the directed evolution towards dehalogenation of 1,2,3-trichloropropane [149]. The importance of L177 for specificity has been proven by independent directed evolution [116], cumulative mutagenesis [158], and saturated mutagenesis [159] experiments. The new candidate for future mutagenesis studies identified by the COMBINE analysis is the second shell residue Pro39. This highly rigid residue is adjacent to Asn38, which is functionally one of the most important residues of LinB. Asn38 is involved in: (i) halogen binding, (ii) transition state and product stabilization, and (iii) co-ordination of catalytic water molecule. Modified flexibility of Asn38 achieved by mutation of Pro39 may produce novel catalytic properties. Construction of Pro39A mutant is currently in progress in our laboratory.

CONCLUSIONS

The COMBINE models constructed for DhlA [28] and LinB (this study) were compared. Only limited number of protein residues (6–8%) contributed significantly to the explanation of variability in K_m and van der Waals interaction energies dominated over electrostatic interaction energies in both models. Significant contributions provided by specific amino acid residues correspond well with the composition of the enzyme active site. Different halogen-stabilizing residues (Trp125 and Trp175 in DhlA; Asn38 and Trp109 in LinB) are known to be employed in substrate binding in different dehalogenases [95] and these were correctly identified by the models. Interesting is the difference in contributions provided by the catalytic base located on equivalent position of both proteins in the opening of entrance tunnel. His289 in DhlA is significantly less

important than His272 in LinB which can be attributed to the different orientation of the active site pocket. The pocket of DhlA is approximately orthogonal to the entrance tunnel while in LinB is in line.

ACKNOWLEDGEMENTS

This work was supported by grants from the Czech Ministry of Education J07/98:143100005 and ME551 (JD).

Chapter 10

Rational Re-design of Haloalkane Dehalogenases Guided by Comparative Binding Energy Analysis

Jiří Damborský, **Jan Kmuníček**, Tomáš Jedlička, Santos Luengo, Federico Gago, Angel R. Ortiz, and Rebecca C. Wade

In Svendsen, A. (Ed.), *Enzyme Functionality: Design, Engineering and Screening*, Marcel Dekker, New York, pp. 79–96 (2003)

Rational Re-design of Haloalkane Dehalogenases Guided by Comparative Binding Energy Analysis

Jiří Damborský,[1,2] Jan Kmuníček,[1] Tomáš Jedlička,[1] Santos Luengo,[3] Federico Gago,[3] Angel R. Ortiz,[4] and Rebecca C. Wade[5]

National Center for Biomolecular Research, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic; Department of Pharmacology, University of Alcalá, E-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, USA; and European Media Laboratory, Villa Bosch, Schloss-Wolfsbrunnenweg 33, D-69118 Heidelberg, Germany

Comparative Binding Energy Analysis

Comparative binding energy (COMBINE) analysis is a computational method for deducing quantitative structure–activity relationships using structural data from macromolecule–ligand complexes [19]. It can be applied to the formation of macromolecule-small molecule complexes and macromolecule–macromolecule complexes; in this article, these complexes will be referred to generically as macromolecule–ligand complexes. The "COMBINE" acronym refers to two aspects of the technique [150]: (i) macromolecule–ligand structural data are combined with experimental binding data and (ii) empirical molecular mechanics energy calculations are combined with partial least-squares projection to latent structures (PLS) chemometric analysis. COMBINE analysis systematically explores the relationships between experimental binding affinities for a set of ligands and selected interaction energies with the macromolecule. COMBINE analysis is formally similar to CoMFA (Comparative Molecular Field Analysis) [151] in as much as both methods deal with data matrices containing a large number of energy descriptors that are subjected to chemometric analysis. On the other hand, the energy descriptors differ: in CoMFA they are interaction fields calculated for the ligand alone, whereas in COMBINE analysis they represent residue-based ligand–receptor interactions. Compared to classical molecular mechanics calculations of binding energies, the advantages

[1]Masaryk University.

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

[3]University of Alcalá.

[4]Mount Sinai School of Medicine.

[5]European Media Laboratory.

of subjecting macromolecule–ligand interaction energies to statistical analysis are that the noise due to inaccuracies in the potential energy functions and molecular models can be reduced and that mechanistically important interactions can be identified. Compared to classical quantitative structure–activity relationships (QSAR) analysis, COMBINE is expected to be more predictive as it incorporates more physically relevant information about the energetics of ligand–receptor interactions [19].

To estimate the total binding energy for each macromolecule–ligand complex, ΔU, a molecular mechanics force field is used to calculate the following terms: (i) the sum, E^{LM}_{INTER}, of intermolecular interaction energies (Δu_i) between the ligand and each macromolecule residue, each of which consists of van der Waals and electrostatic contributions; (ii) the change in intramolecular energy of the ligand upon binding to the macromolecule, ΔE^L; and (iii) the change in intramolecular energy of the macromolecule upon ligand binding, ΔE^M. In addition, a measure of the cost in electrostatic free energy of desolvating the apposing surfaces of both interacting partners upon complex formation [21] is estimated using a continuum electrostatics method that provides two extra terms: (iv) the desolvation energy of the ligand, E^L_{DESOLV}, and (v) the desolvation energy of the macromolecule, E^M_{DESOLV}.

$$\Delta U = E^{LM}_{INTER} + \Delta E^L + \Delta E^M + E^L_{DESOLV} + E^M_{DESOLV} \tag{10.1}$$

The COMBINE analysis methodology is schematized in Figure 9.1. The energy descriptors obtained from the set of experimentally determined or modelled macromolecule–ligand complexes are used to construct a matrix in which the rows represent the different ligands and the columns contain the two blocks of residue-based molecular mechanics energy information (van der Waals and electrostatic) plus the additional desolvation energy terms and a last column containing the experimental binding affinities/activities. This matrix is then projected to a small number of latent variables using the PLS method [15], and the original energy terms are given weights, w_i, according to their importance in the model.

Application of COMBINE Analysis in Drug Design
COMBINE analysis was initially used for the study of protein–inhibitor complexes. Ortiz *et al.* [19] applied COMBINE analysis to a series of 26 inhibitors of the human synovial fluid phospholipase A_2. The COMBINE model explained

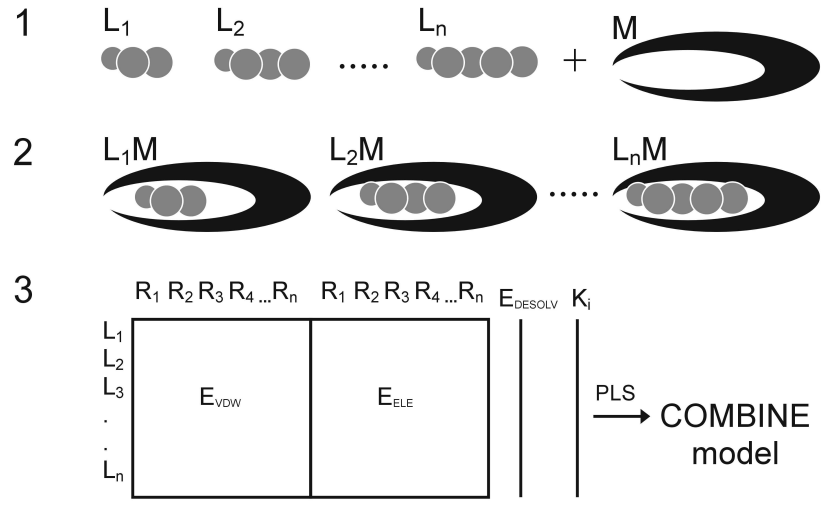

Figure 10.1: Scheme of the COMBINE analysis methodology. $L_1...L_n$ – ligands; M – macromolecule; $L_1M...L_nM$ – macromolecule–ligand complexes; $R_1...R_n$ – residue of a macromolecule; E_{VDW} – van der Waals interaction energy; E_{ELE} – electrostatic interaction energy; E_{DESOLV} – desolvation energy of the ligand and the macromolecule; K_i – log (experimental binding affinity); PLS – partial least-squares projection to latent structures analysis.

92% (82% cross-validated) of the quantitative variability of binding constants and provided insight into the mechanism of phospholipase inhibition. Only 2% of the energy terms were required for explaining the differences in activity. The model indicated that the calcium ion present in the enzyme active site is important for inhibitory activity as is the steric accommodation of the inhibitors in the binding site of the enzyme. Perez *et al.* [21] conducted COMBINE analysis with a set of 33 HIV-1 protease inhibitors and externally validated their models using an additional 16 inhibitors. Incorporation of electrostatic desolvation effects in the model resulted in significant improvement of its predictive ability. The model constructed for a merged set of 49 inhibitors explained 91% (81% cross-validated) of the quantitative variability of the experimental data. This study was further extended by Pastor *et al.* [22] who incorporated the two possible binding modes of the HIV-1 inhibitors into the COMBINE model. This was achieved by manipulation of the data matrix used to describe the interaction energies and provided a model with improved external predictive ability and simplified interpretability. Tomic *et al.* [23] developed a COMBINE model for the binding specificity of transcription factors of the nuclear receptor family to DNA. They analyzed experimental data for the interaction of 20 mutant glucocorticoid receptor DNA-binding domains with 16 different response elements in a total of 320 complexes. The analysis revealed that specificity of binding of the transcription factor to DNA is largely determined by the energy cost of DNA desolvation and is tuned by intermolecular electrostatic interactions and conformational changes. Lozano *et al.* [24] applied in parallel COMBINE and GRID/GOLPE analyzes to a series of 12 heterocyclic amines and human cytochrome P450 1A2. The resultant COMBINE model explained 90% (74% cross-validated) of the quantitative variability of the activity data and corresponded well with the GRID/GOLPE model explaining 96% (79% cross-validated) of the quantitative variability of the activity data. The study showed that the combined use of two 3D-QSAR approaches for model construction acts as a mutual validation procedure and allows a more reliable and detailed interpretation of the results. Cuevas *et al.* [25] studied 40 complexes of human neutrophil elastase with the N3-substituted trifluoromethylketone-based pyridone inhibitors. The authors carried out Poisson–Boltzmann computations and derived two additional descriptors representing the electrostatic energy contributions to the partial desolvation of both the receptor and the ligands, and solvent-screened electrostatic interactions. Incorporation of these descriptors into the model improved its statistical parameters. Most recently, Wang and Wade [26] constructed a COMBINE model for two subtypes and one mutant of neuraminidase from influenza virus complexed with 43 inhibitors. The model

highlighted 12 protein residues and 1 bound water molecule as particularly important for inhibitory activity and indicated the potential for using COMBINE analysis to investigate species specificity and resistant mutants.

Application of COMBINE Analysis in Protein Engineering
A primary goal of protein engineering is to alter the physico-chemical and functional properties of proteins by modification of their structures. Protein structures can be engineered either by directed evolutionary approaches [113, 114], which do not require any *a priori* knowledge of protein–function relationships, or by rational design which is based on the knowledge of these relationships. Protein structures and structure–function relationships are often so complex that it is difficult to study them without the use of computer graphics and computer modelling. COMBINE analysis quantitatively explores residue-based protein–ligand interactions and provides quantitative information about the importance of every residue in a macromolecule for the binding of different substrates. Mutagenesis of the residues with the highest importance in a COMBINE model should lead to the most significant changes in substrate specificity. The molecular models of mutant structures can be constructed *in silico* and the effects of substitution on substrate binding can be predicted prior to experiment using the COMBINE model. The application of COMBINE analysis to the study of structure–function relationships and engineering of haloalkane dehalogenase DhlA has been recently reported by Kmunicek *et al.* [28] and is further extended in this contribution.

Protein Engineering of Haloalkane Dehalogenases
Haloalkane dehalogenases are microbial enzymes that catalyze the cleavage of a carbon–halogen bond by a hydrolytic mechanism (Figure 9.2). Haloalkane dehalogenases require a water molecule as the only co-factor for the reaction that is considered to be a critical step for the biological degradation of various haloalkanes [38]. Haloalkanes are widely used as solvents, degreasing agents, intermediates in chemical synthesis, and pesticides. Therefore haloalkane dehalogenases could find application in bioremediation technologies and chemical syntheses [39, 41, 48]. Different haloalkane dehalogenases have been isolated from various bacteria [68, 101, 117, 152–156], but none of them shows sufficient activity toward some of the technologically interesting compounds, such as 1,2-dichloropropane, 2-chloropropane, 2-chlorobutane, and 1,2,3-trichloropropane, although these substances have the potential to be good substrates for haloalkane dehalogenases from the reaction mechanism standpoint.

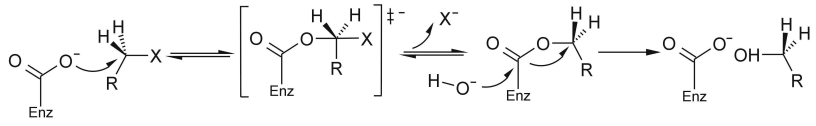

Figure 10.2: Reaction scheme of hydrolytic dehalogenation catalyzed by haloalkane dehalogenases. Enz – enzyme.

Site-directed mutagenesis experiments were initiated to study structure–function relationships and re-design of haloalkane dehalogenases [63–65, 69, 71–76, 78, 102, 103, 157–159]. These studies identified some functional residues, such as the catalytic triad or pairs of transition-state and product stabilizing residues, but to our knowledge none of them provided enzymes with significantly improved activities toward target substances. Structural studies have been conducted to determine the 3D structures of the wild type [52–57, 59, 61, 62, 100] and mutant proteins [63, 64, 160]. The haloalkane dehalogenases are composed of two domains. The core of the main domain consists of an eight-stranded β-pleated sheet with seven parallel strands and one antiparallel strand (Figure 9.3). This β-sheet is surrounded by α-helices. The cap domain is lying on top of the main domain and consists of five α-helices. A buried, mainly hydrophobic cavity is located between these two domains. Three-dimensional structures provide not only a good starting point for the rational design of site-directed mutations and for the interpretation of results from mutagenesis experiments, but also essential data for computer-modelling studies. Molecular docking [57, 88], quantitative structure-function relationships [79], quantum-mechanical calculations [85, 86, 89, 91, 161, 162], and molecular dynamics simulations [92–94] have brought insights into the binding of substrates to the enzyme active site, the mechanism of the dehalogenation reaction, and the conformational behavior of several dehalogenase enzymes at atomic resolution. Although the haloalkane dehalogenases are currently being intensively studied and engineered, an effective catalyst for some target compounds has not been obtained yet. Another approach to improve catalytic performance is to modify the reaction conditions. Gray *et al.* reported construction of a thermostable haloalkane dehalogenase DhaA [116] suitable for dehalogenation at elevated temperatures.

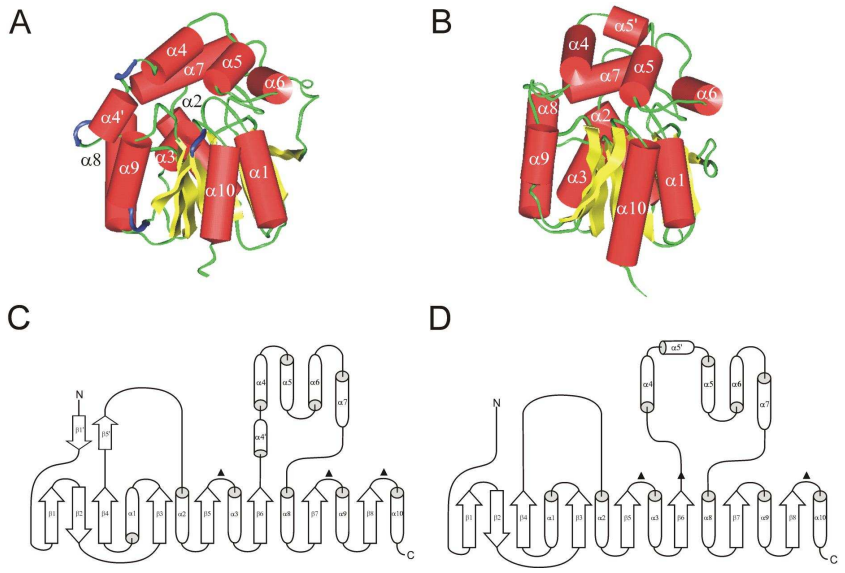

Figure 10.3: Three-dimensional model of the haloalkane dehalogenase DhlA (A) and LinB (B), and topological arrangement of secondary elements in DhlA (C) and LinB (D). The structures were determined by protein crystallography [54, 56]. Numbering of the secondary elements respects the evolutionary changes in the cap domains [94]. The triangles indicate position of the catalytic triad residues.

COMBINE Model for the Haloalkane Dehalogenase DhlA

COMBINE analysis was conducted to identify the protein residues responsible for the differences in binding affinities of 18 chlorinated and brominated aliphatic substrates of haloalkane dehalogenase DhlA from *Xanthobacter autotrophicus* GJ10 [28]. Experimental data for the following compounds were extracted from the literature [67]: 1-chlorobutane, 1-chlorohexane, 1-bromobutane, 1-bromohexane, 1,1-dichloromethane, 1,2-dichloroethane, 1,1-dibromomethane, 2-chloroethanol, 1,2-dibromoethane, 1,2-dichloropropane, 1,2-dibromopropane; 2-bromoethanol, 2-chloroacetonitrile, 2-chloroacetamide, epichlorohydrine, epibromohydrine, 2-bromoacetonitrile, and 2-bromoacetamide. The values of apparent dissociation constants (K_m) varied by three orders of magnitude. The substrate molecules were positioned in the active site of DhlA in such a way that their C–X bonds aligned with the corresponding bond as found in the experimental structure of 1,2-dichloroethane in the Michaelis–Menten complex with DhlA [52]. Manually prepared enzyme–substrate complexes were energy minimized and van der Waals and electrostatic interaction energies between the protein and the substrates were calculated and decomposed on a per residue basis using the program AMBER 5.0 [126]. The data matrix composed of these intermolecular interaction energies, together with the desolvation energies calculated using an electrostatic continuum method, was correlated with K_m values using the PLS method. A four-component model explained 91% (73% cross-validated) of the quantitative variance in K_m (Figure 9.4A). The first dimension mainly projected out electrostatic term of Asp124, which contributes substantially to the energy variance but has a poor contribution to the K_m correlation. Asp124 is the nucleophile that initiates the dehalogenation reaction and is in very close contact with the electrophilic carbon of each substrate (Figure 9.4B). Analysis of the second, third, and fourth principal components showed that only a few energy variables, involving only a few protein residues, are important for explaining the differences in binding among substrates (1% of the enzyme's amino acids explained 91% of variance in K_m). These residues can be divided into two classes, with respect to their interaction with the substrates. The first class is formed by residues separating chlorinated from brominated derivatives: Trp125, Trp175, and Pro223. These residues form the halogen binding site in the enzyme. Mutations affecting these residues should be primarily used to modulate the halogen specificity of the enzyme. Phe222 also contributes to the separation of chlorinated derivates from brominated derivates, together with Leu179 (Figure 9.5). The second set of residues discriminates substrates by their interactions with the substrate alkyl chains. These are mainly Phe172, Phe222, and Phe164, with a contribution from Asp124 as well. Mutations affecting these

residues can be used to tune activity of the enzyme for different chain specificity. α-Helix 4 has the largest concentration of residues involved in explaining the K_m differences: Phe172, Trp175, Lys176, and Leu179 (see Figure 9.3 for position of α-helix 4). This finding is in a good agreement with experimental observations by Priest *et al.* [74], who isolated 12 *in vivo* mutants of DhlA with improved activity toward 1-chlorohexane and 9 of them carried modifications in α-helix 4 or its close surroundings. Priest *et al.* suggested that α-helix 4 is critical for the specificity of DhlA. The applicability of the COMBINE models to predictions was validated using two mutants of DhlA for which the crystal structures had been determined [63, 69]. Four substrate molecules with available experimental binding constants were modelled in the active sites of the mutant proteins and their K_m values were predicted using the COMBINE model. The trends in changes of binding affinity due to mutation were predicted correctly without exception [28].

The main disadvantage of the methodology described above is the need for at least one experimental structure of an enzyme–substrate complex and the assumption that all substrates bind to the active site in the same mode. There are probably many enzymes for which the structural information on the enzyme–substrate complex is missing or which bind their substrates in different orientations, e.g., broad-specificity enzymes. An additional study was therefore conducted with DhlA in which all substrate molecules were automatically and independently positioned inside the active site using a computational method. The remaining part of the COMBINE analysis procedure was the same as described above. The automated molecular docking program AUTODOCK 3.0 [138] was used for positioning 18 halogenated substrates into the active site of DhlA. The docking calculations provided suitable orientations for 15 out of these 18 substrate molecules, as no suitable orientations were found for dihaloacetamides and 2-bromoacetonitrile, i.e., they could not be docked with an orientation close to that necessary for catalysis. Multiple orientations were found for several substrates: 1,2-dibromopropane, halobutanes, and halohexanes. The orientations for the subsequent COMBINE analysis were selected using quantum mechanical calculations, which discriminated between binding modes on the basis of their suitability for the ensuing S_N2 dehalogenation reaction. Dehalogenation reactions were simulated inside a reduced model of the active site of DhlA composed of 20 amino acids (Figure 9.6) using the semi-empirical quantum mechanics program MOPAC 6.0 [141] interfaced by TRITON 2.0 [143]. The selected orientations were further optimized by energy minimization and were found to be in very good agreement with the expected reaction mechanism of

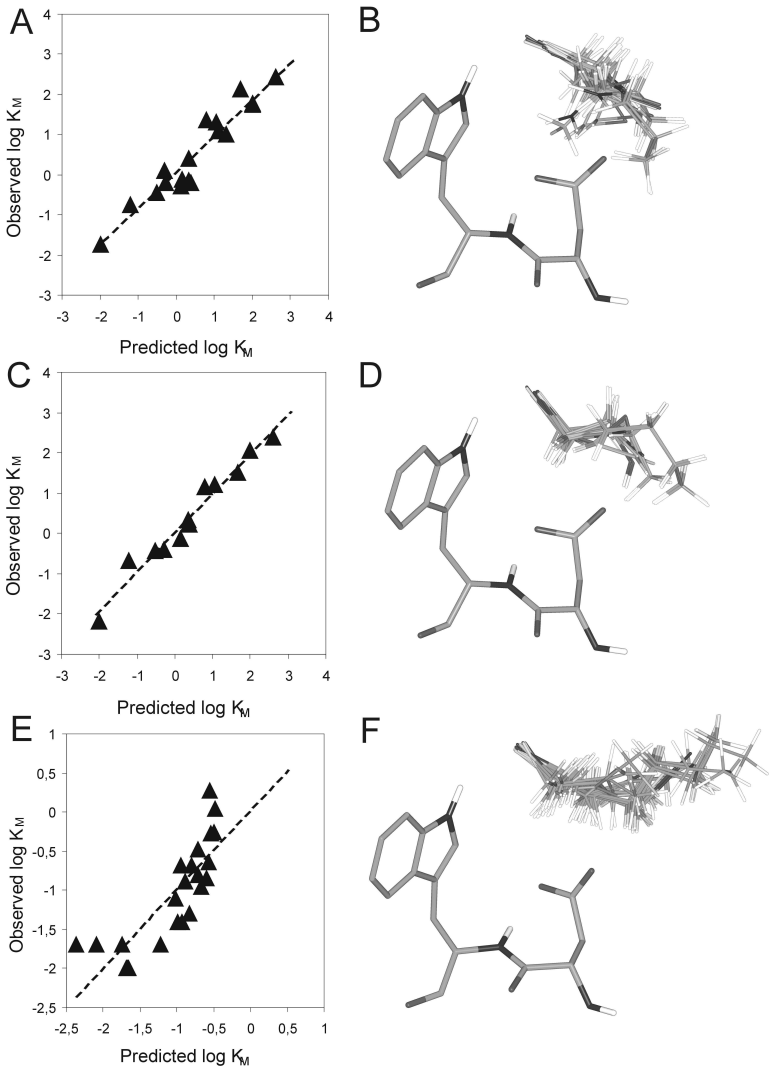

Figure 10.4: Plots of observed *versus* predicted K_m values and the models of Michaelis complexes for structure-based model of DhlA (A, B), docking-based model of DhlA (C, D), and docking-based model of LinB (E, F). The nucleophile (Asp) and the halide-stabilizing residue (Trp) are shown in stick representation.

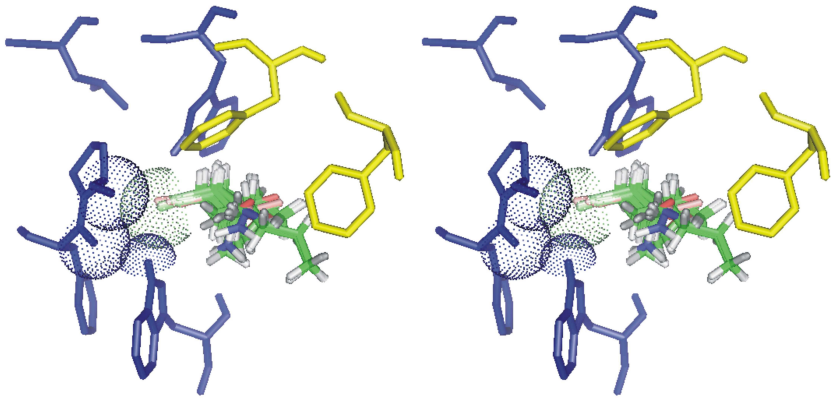

Figure 10.5: Stereoview of the active site of haloalkane dehalogenase DhlA with bound ligands. Residues separating chlorinated from brominated derivatives are shown as blue sticks: Trp125, Trp175, Leu179, Phe222, and Pro223. Residues separating substrates according to the size and shape of their carbon chains are shown as yellow sticks: Phe164 and Phe172. The van der Waals surface of the protein atoms in direct contact with the halogen atom is represented by dots.

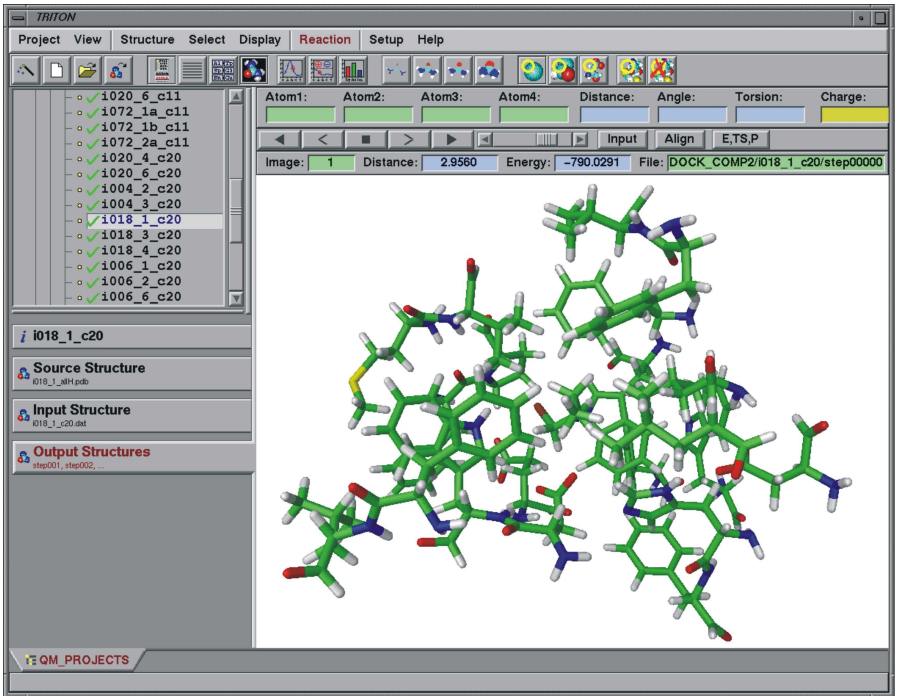

Figure 10.6: Three-dimensional model of the active site of haloalkane dehalogenase DhlA used for quantum mechanical calculations as displayed in the main window of the TRITON program. This program is used for preparation of the input data for calculation of reaction coordinates, for monitoring of the progress of calculation, and for analysis of output data. The software is freely available at http://ncbr.chemi.muni.cz/triton/triton.html.

DhlA (Figure 9.4D). Furthermore, bound substrates resembled the reactive conformation of 1,2-dichloroethane described by Lau *et al.* [93]. The goodness of the fit to the experimental data in the COMBINE model constructed from these selected docked orientations was comparable to that previously obtained with the structure-based model (compare Figures 4A and C). The model explained 96% (67% cross-validated) of the quantitative variance in K_m, and two outliers (dihalopropanes) had to be removed from the model. The composition of the latent variables extracted and the importance of amino acid residues for explaining K_m values, however, were similar in both models, leading to a similar biochemical interpretation. Interestingly, the automatic docking-based model employed more electrostatic contributions than the structure-based alignment model. The largest difference in van der Waals interactions was noted for the residues in direct contact with the halogenated hexanes (Phe222 and Leu263) due to the different orientations and conformations adapted by these long substrates in the small active site.

COMBINE Model for the Haloalkane Dehalogenase LinB
Haloalkane dehalogenase LinB from *Sphingomonas paucimobilis* UT26 belongs to the same protein family as DhlA. These two proteins differ both by their structures and their catalytic properties. The catalytic triad of LinB is composed of Asp108–His272–Glu132 [102], while the catalytic triad of DhlA consists of Asp124–His289–Asp260 [62]. The catalytic acid is positioned after β-strand 6 in LinB(Figure 9.3D) and after β-strand 7 in DhlA (Figure 9.3C). Bound substrates, transition states, and product structures are primarily stabilized by hydrogen bonds from the Trp109–Asn38 pair in LinB and the Trp125–Trp175 pair in DhlA. The active site of LinB is 2.5 times larger than the active site of DhlA and is less buried inside the protein core [51]. There are at least three tunnels leading to the active site of LinB, but only one tunnel in DhlA [94]. LinB shows broader substrate specificity than DhlA, i.e., it is more active toward larger and β-substituted haloalkanes, and therefore it should be more suitable for the design of efficient catalysts for the target compounds carrying a halogen in the β-position. Currently, there is no 3D structure of a Michaelis complex for LinB. Furthermore, it is not safe to assume that all substrates bind to the large active site in a similar way. An automated docking-based methodology was therefore used for the construction of a COMBINE model for LinB. Experimental data (K_m values) were determined for 25 substrates: 1-chloropropane, 1-chlorobutane, 1-chlorohexane, 1-chloroheptane, 1-chlorooctane, 1-bromopropane, 1-bromobutane, 1-bromohexane, 1-iodopropane, 1-iodohexane, 1,3-dichloropropane, 1,5-dichloropentane, 1,2-dibromoethane, 1,3-

dibromopropane, 1-bromo-3-chloropropane, 1,2-dibromopropane, chlorocyclohexane, 1-bromo-2-methylpropane, *bis*(2-chloroethyl)ether, bromocyclohexane, 2-bromo-1-chloropropane, 4-bromobutyronitrile, 3-chloro-2-methylpropene, 3-chloro-2-(chloromethyl)-1-propene, and 2,3-dichloropropene. A preliminary model, consisting of only one principal component, explained 91% (87% cross-validated) of the quantitative variability in K_m values (Figure 9.4E). Two outliers, *bis*(2-chloroethyl)ether and chlorocyclohexane, had to be removed from the model (Figure 9.4F). Extreme k_{cat}/K_m values were repeatedly measured with these substrates. The model explained variability in K_m values resulting from the different lengths of the substrate molecules but could not deal properly with the variability originating from the different halogens. In DhlA, halogen substituents are tightly bound between two opposing tryptophans and the COMBINE model constructed for this protein could distinguish between chlorinated and brominated substrates. More research is needed to refine the model for LinB, e.g., by investigating the contribution of electrostatic desolvation energies, the effects of the energy minimization on conformation of substrates in Michaelis complexes, or the effects of explicit inclusion of the water molecules in Michaelis complexes. In fact, active exchange of water molecules between the active site of LinB and the bulk solvent was observed in nanosecond-scale molecular dynamics simulation [94]. The lesson learned so far from comparison of DhlA and LinB COMBINE models is that exactly the same methodology to generate structures of the complexes cannot necessarily be applied, even to closely related proteins, but the modelling protocol must be adjusted with to respect to the proteins distinguishing structural and biochemical features.

CONCLUSIONS

COMBINE analysis quantitatively explores macromolecule–ligand interactions on a residue basis and provides quantitative information about the importance of every residue in a macromolecule for binding of different substrates. COMBINE analysis identified a number of specificity-determining amino acid residues in the haloalkane dehalogenase DhlA. Trp125, Trp175, Leu179, Phe222, and Pro223 are important in distinguishing chlorinated and brominated derivatives. Mutations affecting these residues should modulate the halogen specificity of the enzyme. A second set of residues (Phe164, Phe172, and Phe222) are found to discriminate substrates by their interactions with the carbon chain. The predictive ability of the COMBINE model derived for DhlA was confirmed with two site-directed point mutants and four novel substrates. Modelling the specificity of the haloalkane dehalogenase LinB using the same methodology is slightly

more difficult due to its larger active site and less specific binding of its substrates. Our current COMBINE model differentiates between molecules of different chain length but cannot properly distinguish substrates bearing a different halogen atom. To achieve this goal we are currently tailoring our modelling protocol for the LinB enzyme.

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).

Chapter 11

Exploring the Structure and Activity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26: Evidence for Product and Water Mediated Inhibition

Aaron J. Oakley, Zbyněk Prokop, Michal Boháč, **Jan Kmuníček**,
Tomáš Jedlička, Marta Monincová, Ivana Kutá-Smatanová,
Yuji Nagata, Jiří Damborský, and Matthew C. J. Wilce

Biochemistry **41**: 4847–4855 (2002)

Exploring the Structure and Activity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26: Evidence for Product and Water Mediated Inhibition[1]

Aaron J. Oakley,[2] Zbyněk Prokop,[3] Michal Boháč,[3] Jan Kmuníček,[3] Tomáš Jedlička,[3] Marta Monincová,[3] Ivana Kutá-Smatanová,[3,4] Yuji Nagata,[5] Jiří Damborský,[3,6] and Matthew C. J. Wilce[2,6]

Department of Pharmacology/Crystallography Center, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia; National Center for Biomolecular Research, Masaryk University, Kotlářská 2, CZ-611 37 Brno, Czech Republic; and Graduate School of Life Sciences, Tohoku University, Katahira, Sendai 980-8577, Japan

Received August 30, 2001; Revised Manuscript Received January 23, 2002

ABSTRACT

The hydrolysis of haloalkanes to their corresponding alcohols and inorganic halides is catalyzed by α/β-hydrolases called haloalkane dehalogenases. The study of haloalkane dehalogenases is vital for the development of these enzymes if they are to be utilized for bioremediation of organohalide-contaminated industrial waste. We report the kinetic and structural analysis of the haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26 (LinB) in complex with each of 1,2-dichloroethane and 1,2-dichloropropane and the reaction product of 1-chlorobutane turnover. Activity studies showed very weak but detectable activity of LinB with 1,2-dichloroethane [0.012 nmol s^{-1} (mg of enzyme)$^{-1}$] and

[1]Coordinate and X-ray amplitudes have been deposited at the Research Collaboratory for Structural Bioinformatics Protein Data Bank: PDB entry 1G5F for the complex with 1,2-dichloroethane, PDB entry 1G42 for the complex with 1,2-dichloropropane, and PDB entry 1G4H for the complex with butane-1-ol.

[2]University of Western Australia.

[3]Masaryk University.

[4]Current address: Photosynthesis Research Center, South Bohemian University, Branišovská 31, 370 05 České Budějovice, Czech Republic.

[5]Tohoku University.

[6]To whom correspondence should be addressed:
JD: Fax: +420-5-41129506. E-mail: jiri@chemi.muni.cz.
MCJW: Fax: +618-9346-3469. E-mail: mwilce@receptor.pharm.uwa.edu.au.

1,2-dichloropropane [0.027 nmol s^{-1} (mg of enzyme)$^{-1}$]. These activities are much weaker compared, for example, to the activity of LinB with 1-chlorobutane [68.2 nmol s^{-1} (mg of enzyme)$^{-1}$]. Inhibition analysis reveals that both 1,2-dichloroethane and 1,2-dichloropropane act as simple competitive inhibitors of the substrate 1-chlorobutane and that 1,2-dichloroethane binds to LinB with lower affinity than 1,2-dichloropropane. Docking calculations on the enzyme in the absence of active site water molecules and halide ions confirm that these compounds could bind productively. However, when these moieties were included in the calculations, they bound in the manner similar to that observed in the crystal structure. These data provide an explanation for the low activity of LinB with small, chlorinated alkanes and show the importance of active site water molecules and reaction products in molecular docking.

INTRODUCTION

Due to environmental problems arising from the production and use of halogenated hydrocarbons, the study of microbial enzymes that can catabolize these compounds is of major interest [42]. Haloalkane dehalogenases (EC 3.8.1.5) are one such important class of enzyme because of their ability to attack polychlorinated aliphatic hydrocarbons, which are produced in several industrial processes [31, 32]. Haloalkane dehalogenases remove halides from organic compounds via a hydrolytic mechanism that results in the production of the corresponding alcohol, inorganic halide, and hydrogen ion. One such haloalkane dehalogenase, LinB[1] is isolated from a γ-hexachlorocyclohexane degrading bacterial strain *Sphingomonas paucimobilis* UT26 [101]. It is the second enzyme in the biochemical pathway enabling the bacterium to utilize γ-hexachlorocyclohexane as its sole carbon and energy source [163].

LinB is a haloalkane dehalogenase of the α/β hydrolase family, with a main domain consisting of an eight-stranded β-sheet flanked with α-helices, and a cap domain which is comprised of α-helices and several loops [49, 56]. The active site is formed by the close interactions of the two domains (Figure 10.1). Catalysis by this family of dehalogenases is thought to take place in two steps [52]. Once substrate binds with the scissile halide in the binding site formed by Trp109,

[1]Abbreviations: DhlA, haloalkane dehalogenase from *Xanthobacter autotrophicus* GJ10; DhaA, haloalkane dehalogenase from *Rhodococcus erythropolis* strains Y2, m15-3, HA1, GJ70, NCIMB13064, and TB2; LinB, haloalkane dehalogenase from *Sphingomonas paucimobilis* UT26; RMSD, root-mean-square deviation; 1,2-DCP, 1,2-dichloropropane; 1,2-DCE, 1,2-dichloroethane; 1-CB, 1-chlorobutane.

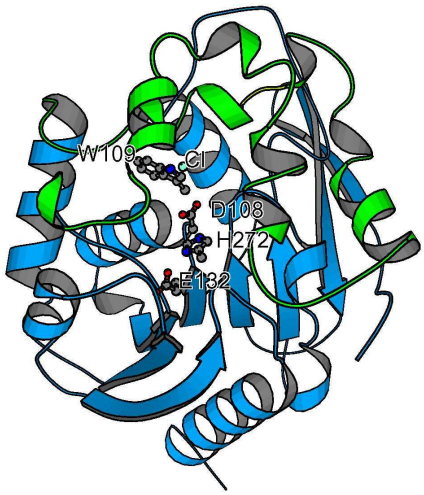

Figure 11.1: The LinB structure in Richardson representation. The core (main) domain is coloured blue. The cap domain is coloured green. The active site residues (Asp108, Trp109, Glu132, and His272) are shown in ball-and-stick representation. The halide binding site is indicated by the Cl^- ion.

Asn38, and Pro208, the carboxylate group of Asp108 makes a nucleophilic attack upon the scissile carbon. Inorganic halide is released with the concomitant formation of an ester intermediate formed by the alkane and Asp108. A nearby water molecule undergoes deprotonation by His272 and attacks the ester linkage, resulting in the formation of the alcohol product and free Asp108. Site-directed mutagenesis of LinB confirms that Asp108, His272, and Glu132 comprise the catalytic triad in that enzyme [103].

The broad substrate specificity of LinB could make it useful in bioremediation; however, LinB is inhibited by the key pollutants 1,2-dichloroethane (1,2-DCE) and 1,2-dichloropropane (1,2-DCP). These compounds are economically important; ∼7 230 000 tonnes of 1,2-DCE was produced in the United States in 1992 by various industrial processes [164]. 1,2-DCP is formed as a by-product in the synthesis of propylene oxide and epichlorohydrin [165], and is a proble-

matic pollutant because it is not catalyzed efficiently by any of the haloalkane dehalogenases. 1,2-DCP is recalcitrant under aerobic conditions [44, 165]. We have investigated the structural and kinetic aspects of LinB with 1,2-DCE and 1,2-DCP to determine the reasons for the low activity of LinB with these compounds. For comparison, we have also attempted to determine the structure of LinB in complex with the model substrate 1-chlorobutane (1-CB). This information is crucial in the effort to design an enzyme with improved catalytic properties for these compounds.

MATERIALS AND METHODS

Crystallisation
LinB was cloned, overexpressed, purified, and crystallized as described previously [166]. Briefly, crystals of LinB grew as plates of dimension 0.8 × 0.4 × 0.2 mm using the hanging drop vapor diffusion method. The mother liquor contained 18% PEG 6 000, 0.2 M calcium acetate and 0.1 M Tris (pH 8.9).

Soaking Experiments
Crystals of LinB were transferred using a cryoloop (Hampton Research) from the mother liquor to a 2 μL hanging drop of cryoprotectant [0.1 M tris-(hydroxymethyl)-aminomethane buffer (pH 8.9), 0.2 M calcium acetate, 18% polyethylene glycol 6 000, 20% (v/v) glycerol] on a siliconised cover slip. The compound of interest (10 μl) was placed into a well containing 1 mL of cryoprotectant solution. The halogenated compounds then entered the crystal via vapour diffusion from the reservoir. The cover slip was sealed on top using vaseline. The soak times and temperature were chosen according to the enzymatic activity of LinB toward the compound of interest. Information for each soak is shown in Table 10.1.

X-ray Data Collection and Processing
Protein crystals were mounted with cryoprotectant in a loop (Hampton Research) and kept frozen using an Oxford Cryostream Cooler (Oxford Cryosystems). Data were collected using a MAR345 area detector (Mar Research) using Cu Kα X-rays ($\lambda = 1.5418$ Å) from a Rigaku RU-200 rotating anode generator. X-rays were focused using nickel-coated gold mirrors (Mar Research). Consecutive batches of 0.25° oscillations were collected. Data were processed using the HKL package [167].

Table 10.1: Soaking and Data Collection Statistics for the LinB–Ligand Complex.

	1,2-DCE	1,2-DCP	1-CB
Data Collection			
soak time (min)	60	30	60
soak temperature (°C)	20	20	4
space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
cell dimensions			
a (Å)	50.9	50.8	50.8
b (Å)	71.9	72.0	72.0
c (Å)	73.3	73.3	73.3
maximum resolution (Å)	1.8	1.8	1.8
no. of crystals	1	1	1
total no. of observations	54 733	55 091	57 835
no. of unique reflections	23 846	24 918	24 212
completeness of data (%)[a]	93.2 (91.0)	97.0 (95.7)	94.3 (95.9)
data $> 3\sigma_1$ (%)	65.8 (12.5)	77.0 (41.3)	75.3 (50.5)
I/σ_1	10.8 (2.1)	12.8 (3.0)	19.6 (3.5)
multiplicity	2.3 (2.2)	2.2 (2.1)	2.4 (2.3)
R_{merge} (%)[b]	7.2 (25.9)	5.4 (24.1)	4.0 (34.5)
Refinement Statistics			
no. of non-hydrogen atoms	2 673	2 795	2 611
no. of protein	2 356	2 346	2 357
no. of ligand atoms	8	15	15
no. of solvent molecules	291	422	237
no. of ions	4	2	3
R_{factor} (%)[c]	17.5 (27.3)	16.4 (22.4)	18.5 (28.0)
R_{free} (%)[d]	20.4 (30.1)	19.8 (25.7)	20.5 (27.9)

Table 10.1: Soaking and Data Collection Statistics for the LinB–Ligand Complex—Continued.

Refinement Statistics			
RMSD from ideal geometry			
bonds (Å)	0.005	0.005	0.005
angles (deg)	1.3	1.3	1.3
dihedrals (deg)	22.5	22.4	22.5
impropers (deg)	0.85	0.85	0.84
bonded B's (Å^2)	1.457	1.15	0.50
residues in most favorable region of a Ramachandran plot	87.4	86.6	85.8

[a] Numbers in parentheses refer to the highest-resolution bin (1.9–1.8 Å);

[b] $R_{merge} = \sum_h \sum_i |I_{hi} - < I_h >| / \sum_h \sum_i I_h$;

[c] $R_{factor} = \sum_h ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$ where F_{obs} and F_{calc} are observed and calculated structure factors, respectively; [d] R_{free} was calculated based on 5% of the diffraction data not used in the refinement

Structure Solution and Refinement
Structures were determined using the wild type LinB structure (PDB entry 1CV2) with all nonprotein atoms removed. At all stages during refinement, R_{free} validation based on randomly selected reflections (5% of the total set) was used. The structures were refined as follows. Molecular replacement models were subjected to rigid body, positional, and B-factor refinement in CNS [168], followed by 20 cycles of automatic refinement by CCP4/ARP [169] that included the automatic building of water molecules. Water molecules that were built into the active site and halide binding site were omitted from the models, and further positional and B-factor refinement was conducted to reduce phase bias. Sigma-A weighted electron density maps were generated from these models and inspected using O [170].

Some large spherical peaks in the solvent were interpreted as Ca^{2+} ions (a component of the crystallization mixture), as they appeared to be coordinated by water molecules. These were built into the density and included in later rounds of refinement. Ligand models were built into electron density after extensive inspection of $2F_o - F_c$ and $F_o - F_c$ maps. The locations of chlorine atoms of the compounds were identified by their increase peak heights in the $F_o - F_c$ maps. Bond and angle parameters for the ligands were derived using XPLO2D [171] from models produced with WebLab Viewer Pro 3.5 (Molecular Simulations Inc.). For all ligands, there was more electron density than the binding of the ligand in one orientation could account for, and hence, multiple models for the ligand were inserted. Special care was taken in the modelling of these ligands because of overlap.

Structure Analysis
Intermolecular contacts between ligand and protein were assessed using CONTACT [172]. Model quality was assessed using CNS [168] and MOLEMAN [173]. The quality of the structures was analysed using PROCHECK [172] and MOLEMAN [173]. LSQMAN [174] was used to superimpose models for comparison.

Activity Measurements
Enzymatic activity of LinB was assessed by determination of the substrate and product concentrations using gas chromatography–mass spectrometry (GC–MS). The reaction was conducted in 1.5 mL screw-cup vessels. The reaction was initiated by adding 200 μL of enzyme solution (0.4 mM) into 0.8 mL of substrate solution [0.57 mM of 1,2-DCE and 0.51 mM of 1,2-DCP, respectively, dissolved in a glycine buffer (pH 8.6)]. The mixture was incubated at 37 ^{o}C and analyzed

every 20 min for 4 h for product formation. At each time point, 1 μL of the reaction mixture was analyzed by GC–MS (Trace MS 2000, Thermo Finnigan) equipped with an SSL Injector and DB5-MS 25 m $\times$ 0.25 mm $\times$ 0.25 μm capillary column (J&W Scientific). The reaction mixture without enzyme served as an abiotic control.

Inhibition Kinetics
Inhibition kinetics of LinB were studied by determination of the substrate and product concentrations using gas chromatography (Trace GC 2000, Thermo Finnigan) equipped with a flame ionization detector and the DB-FFAP 30 m $\times$ 0.25 mm $\times$ 0.25 μm capillary column (J&W Scientific). Dehalogenation reactions were performed in 25 mL Reacti-Flasks closed by Mininert valves in a shaking water-bath at 37 °C. The reaction mixture (pH 8.6) consisted of the enzyme preparation and varied concentrations (0.05–6.5 mM) of substrate 1-chlorobutane and varied concentrations (0–26 mM) of inhibitor 1,2-DCE or 1,2-DCP. The reaction was stopped by the addition of ethanol. The data, measured in triplicate, were fitted to different inhibition models. The steady-state inhibition constants were calculated using the computer program EZ-Fit version 1.1.[2]

Molecular Docking
Substrate and product molecules were docked in the active sites using the program AUTODOCK 3.0 [175]. Crystal water molecules retained in the enzyme active site in some of the dockings were treated as a part of the rigid enzyme. Substrates and products were treated as flexible molecules. The WHATIF 5.0 program [127] was used for adding the polar hydrogen atoms on protein structures. His289 was singly protonated on N_δ in accordance with its catalytic function. Charges were added on all enzyme atoms using the script *q.kollua*. Solvation parameters were added to the atoms in the protein structure by the script *addsol*. The grid maps were calculated using AUTOGRID 3.0 with 81 $\times$ 81 $\times$ 81 points and grid spacing 0.25 Å. The Lamarckian genetic algorithm [138] was used with a population of 50 individuals, a maximum number of 1.5×10^6 energy evaluations, a maximum number of generations of 27 000; an elitism value of 1, a mutation rate of 0.02, and cross over rate of 0.50. The local search was based on the pseudo-Solis and Wets algorithm [140] with a maximum of 300 iterations per local search. Final orientations from every docking were clustered with the clustering tolerance for the root-mean-square positional deviation (0.5 Å). Fifty docking runs were performed for each enzyme–substrate complex.

[2]By F. W. Perrella.

RESULTS

1,2-DCE Complex Crystal Structure

Despite this compound being an inhibitor (4), strong density interpretable as a Cl^- ion was observed in the halide binding site. The distances between the Cl^- ion and Trp109 $N_{\varepsilon 1}$, Asn38 $N_{\delta 2}$ and Pro208 N were 3.2 Å, 3.5 Å, and 3.5 Å, respectively. A large volume of electron density was seen in the active site of LinB, and was interpreted as 1,2-DCE bound in two orientations (Figure 10.2a). Both ligand models are oriented with one well-ordered chlorine atom binding in a pocket in the active site comprising residues Leu248, Ala247, Pro144, and Leu177. Each ligand was given an occupancy of 0.5. The other chlorine atoms are placed divergently between the two ligand models pointing toward the halide binding pocket. They appear less well ordered as their electron density is spread out more. A water molecule located close to these ligands was observed to be 2.6 Å away from Asp147 $O_{\delta 2}$ atom, which was shifted 1.8 Å (measured from atom $O_{\delta 2}$) parallel to the ligand compared with the apo-enzyme structure. No other shifts in the active-site residues were observed. The inhibitor molecules contact residues Leu177, Asp108, and His272. The small number of contacts of the ligand with the protein would explain the relative disorder of this molecule. This ligand displaces three water molecules (observed in the free enzyme structure) from the active site.

1,2-DCP Complex Crystal Structure

The electron density in the active site was interpreted as 1,2-DCP bound in at least two conformations (Figure 10.3a). There was no evidence that the compound had been hydrolyzed and no evidence for the covalent modification of catalytic residue Asp108. The halide binding site contained an area of spherical density that was interpreted as a chloride ion. Since a racemic mixture was used, the two possible enantiomers were built into the active site in overlapping conformations. The refined model yielded excellent quality electron density with no peaks in $F_o - F_c$ maps. Attempts to remove one or more of the models and subject the model to further refinement gave peaks ($> 3\sigma$) in resulting $F_o - F_c$ maps. Each ligand model was given an occupancy of 0.5. None of the ligand models are bound productively. They are located adjacent to the face of the catalytic His272 residue (Figure 10.3a). The closest ligand Cl atom is 5.2 Å away from the chloride binding site and 5.0 Å away from the Asp108 $O_{\delta 1}$ atom. A catalytic water molecule was observed 2.89 Å from Asp108 $O_{\delta 1}$. The models contact residues Val173, Leu177, Phe169, Asp108, and His272. Two differences

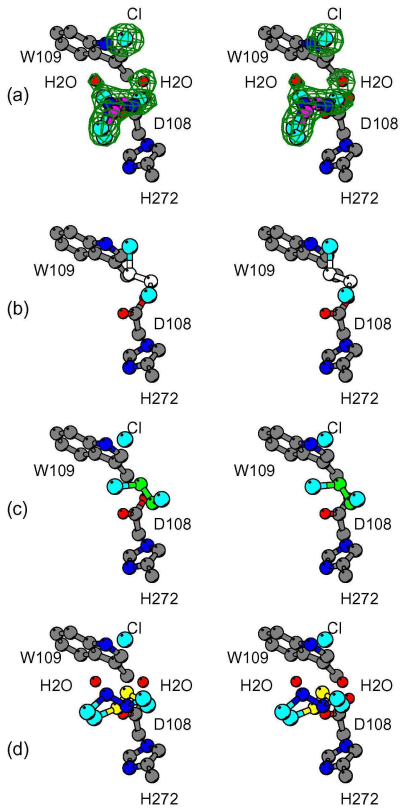

Figure 11.2: Stereodiagrams of experimentally determined and docked models of 1,2-DCE binding in the active-site of LinB. Important catalytic residues are indicated in ball-and-stick representation. (a) Crystallographically determined binding of 1,2-DCE with surrounding water and Cl^- ion. A $2F_o - F_c$ electron density map contoured at 1σ shown for all ligands represented in green. The two ligands are represented with carbon atoms colored blue and magenta and Cl atoms in cyan. Panels b–d represent the different modeling experiments with various combinations of Cl^- and water molecules included in the calculations. The models are represented with carbon atoms colored as follows: (b) white for model 1, (c) green for model 2, and (d) yellow for model 3 and blue for model 4. The details of the modelling experiments are in Table 10.3.

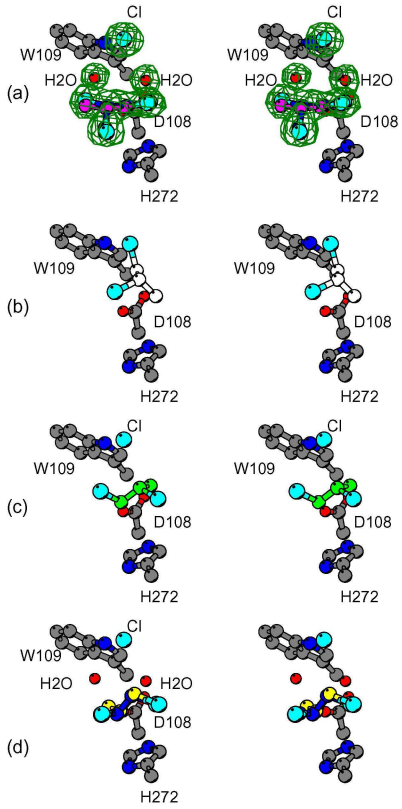

Figure 11.3: Stereodiagrams of experimentally determined and docked models of 1,2-DCP binding in the active site of LinB. Important catalytic residues are indicated in ball-and-stick representation. (a) Crystallographically determined binding of 1,2-DCP with surrounding water and Cl^- ion. A $2F_o - F_c$ electron density map contoured at 1σ shown for all ligands represented in green. The two ligands are represented with carbon atoms colored blue and magenta and Cl atoms in cyan. Panels b–d represent the different modelling experiments with various combinations of Cl^- and water molecules included in the calculations. The models are represented with carbon atoms colored as follows: (b) white for model 1, (c) green for model 2, and (d) yellow for model 3 and blue for model 4. The details of the modeling experiments are in Table 10.3.

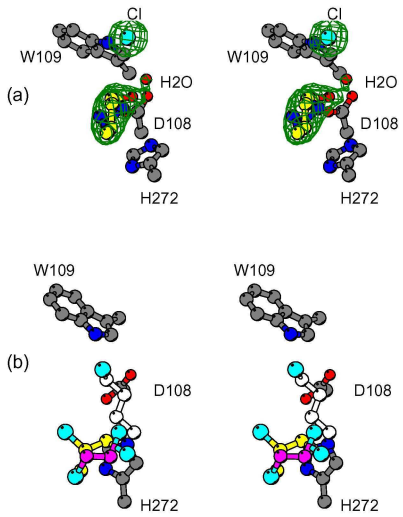

Figure 11.4: (a) Stereodiagram of crystallographically determined binding location of 1-CB turnover product 1-butanol. A $2F_o - F_c$ electron density map contoured at 1σ shown for all ligands represented in green. The two conformers are shown with carbon atoms colored blue and yellow. Important catalytic residues are indicated in ball-and-stick representation. (b) Comparison of crystallographically determined 1,2-DCE and 1,2-DCP structures *versus* 1-CB docked in the active site as it would be found prior to S_N2 attack. Chlorine atoms are represented in cyan, with carbon atoms colored according to molecule: yellow for 1,2-DCP, magenta for 1,2-DCE, and white for 1-CB.

were noted in the side chain configurations of the 1,2-DCP complex compared with the free enzyme structure. The side-chain of Asp147 is observed to lie in two conformations, one of which matches the free enzyme structure; the other has shifted parallel to the ligand such that it forms part of the tunnel to the active site (the shift in the atom Asp147 $O_{\delta 2}$ was observed to be 1.9 Å). Phe143 has shifted (the shift in C_ζ was 1.3 Å) such that it is closer to the ligand. The ligands displace three water molecules found in the active site in the free enzyme structure.

1-Butanol Complex Crystal Structure
The compound has previously been observed to act as a substrate of LinB with an activity of 68.2 nmol s^{-1} (mg of enzyme)$^{-1}$. Despite the compound being soaked at 4 °C, the density in the active site indicated that the compound had been hydrolyzed. A single chloride ion was placed between Trp109, Asn38, and Pro208 (distances between Cl^- and the Trp109 $N_{\varepsilon 1}$, Asn38 $N_{\delta 2}$, and Pro208 N atoms are 3.3 Å, 3.6 Å, and 3.4 Å, respectively). A 1-butanol molecule was modelled in the active site in two positions with each model given an occupancy of 0.5 (Figure 10.4a). The distances between 1-butanol oxygen and Asp108 $O_{\delta 1}$ oxygen are 3.5 Å and 3.8 Å for each conformation. A water molecule 2.7 Å from Asp108 $O_{\delta 1}$ appears to have replaced the substrate water molecule forming the OH group of 1-butanol in the reaction. The only observed shift in the active site was the deflection of the Asp147 side chain toward the 1-butanol molecule (the distance between the positions of $O_{\delta 2}$ of the wild type and the complex structure is 2.1 Å). The distance between Asp147 $O_{\delta 1}$ and atom C_4 of the reaction product is 3.2 Å. The 1-butanol models contact residues Asp108, Pro144, Asp147, Ala247, and His272. The 1-butanol molecule displaces two water molecules in the free enzyme active site. A 1-butanol molecule was also observed bound on the surface of the enzyme away from the active site, contacting residues Thr264, Phe289, and Arg292.

Structure Comparisons
The structures were compared with the free enzyme (PDB entry 1CV2) and the enzyme complexed with propane-1,3-diol (PDB entry 1D02). The 1,2-DCP, 1-butanol and 1,2-DCE complexes superimpose on the free enzyme with RMSDs of 0.50, 0.17, and 0.58 Å over all C_α atoms, respectively. The greatest divergences between the structures are at the N- and C-termini, which are relatively disordered in the structures. The active site residues superimpose closely in all cases. The greatest shifts are in the side chain position of Asp147 and Phe143 (described above). The ligands described here roughly superimpose, all being

Table 10.2: Kinetic Constants of the Haloalkane Dehalogenase LinB with the Substrate 1-Chlorobutane in the Presence of Inhibitors 1,2-Dichloroethane and 1,2-Dichloropropane.

	K_m (mM)	K_i (mM)	k_{cat} (s^{-1})
1,2-DCE	0.25 ± 0.057	2.31 ± 0.550	1.05 ± 0.053
1,2-DCP	0.26 ± 0.002	0.97 ± 0.085	0.91 ± 0.027

located adjacent to the His272 side chain; however, the propane-1,3-diol molecule described previously [56] is closer to the Asp108 carboxylate group. One of the OH groups of propane-1,3-diol is 3.1 Å away from Asp108 $O_{\delta 1}$, as opposed to 3.5 Å in the case of 1-butanol described above.

Dehalogenation of 1,2-DCE and 1,2-DCP by LinB
No detectable activity was observed in previous kinetic analyzes of enzymatic dehalogenation of 1,2-DCE and 1,2-DCP by LinB [101]. The presence of chloride ions in the active site of LinB complexed with 1,2-DCE and 1,2-DCP (see above) suggested that these molecules may be converted by LinB, and thus, improvements to the experimental design were made to lower the detection limit of the activity assay. This measurement showed very weak, but detectable activity of LinB with 1,2-DCE [0.012 nmol s^{-1} (mg of enzyme)$^{-1}$] and 1,2-DCP [0.027 nmol s^{-1} (mg of enzyme)$^{-1}$]. These activities are much weaker compared to the activity of LinB with 1-CB [68.169 nmol s^{-1} (mg of enzyme)$^{-1}$].

Inhibition of LinB by 1,2-DCE
The inhibition of 1-CB hydrolysis by 1,2-DCE was studied by steady-state kinetic analysis (Table 10.2 and Figure 10.5a). The identical slopes of the Hanes–Woolf plot suggest that the limiting initial velocity in the presence of 1,2-DCE was equal to the limiting initial velocity in its absence; however, the apparent K_m increased in the presence of 1,2-DCE. This suggests that, at any concentration of 1,2-DCE, a portion of the enzyme exists in the enzyme–inhibitor form which had no affinity for 1-CB. In addition, the initial velocity was driven to zero by increasing 1,2-DCE concentrations at a fixed substrate concentration (Figure 10.5c). The experimental data were analyzed by fitting the equations for competitive ($SE_{fit} = 0.0014$), uncompetitive ($SE_{fit} = 0.0025$), noncompetitive ($SE_{fit} = 0.0023$) and mixed noncompetitive ($SE_{fit} = 0.0039$) inhibition.

The 1,2-DCE inhibition data fitted best to the competitive model with a K_i of 2.31 ± 0.55 mM. LinB has 10 times lower affinity for 1,2-DCE than for 1-CB (K_m = 0.25 ± 0.057 mM; Table 10.2). The simple competitive inhibition of 1-CB hydrolysis by 1,2-DCE was deduced from these findings.

Inhibition of LinB by 1,2-DCP
An inhibition of 1-CB hydrolysis by 1,2-DCP was probed by steady-state kinetic analysis (Table 10.2). The Hanes–Woolf plot shows the effect of 1,2-DCP on the apparent K_m while limiting initial velocity was equal in both the presence and absence of the inhibitor (Figure 10.5b). In addition, the initial velocity was driven to zero by increasing 1,2-DCP concentrations at a fixed substrate concentration (Figure 10.5c). The apparent K_m increased in the presence of 1,2-DCP, suggesting that at any concentration of 1,2-DCP, a portion of the enzyme existed in enzyme–inhibitor form with no affinity for 1-CB. The data were analyzed by fitting the equations for competitive (SE_{fit} = 0.0007), uncompetitive (SE_{fit} = 0.0015), noncompetitive (SE_{fit} = 0.0016), and mixed noncompetitive (SE_{fit} = 0.0012) inhibition. The 1,2-DCP inhibition data fitted best to the competitive model with a K_i of 0.97 ± 0.085 mM (Table 10.2). The simple competitive inhibition of 1-CB hydrolysis by 1,2-DCP was deduced from these findings.

Docking Calculations
Initially, the ligands were docked in the free protein molecule without any cofactors (water molecules and ions). These calculations did not reproduce crystallographically determined positions of the ligands in the active site. Ligand molecules were docked in positions suitable for a S_N2 dehalogenation reaction with a halogen atom in a position equivalent to the Cl^- ion and adjacent carbon atom oriented toward the nucleophile Asp108 (Figures 10.2 and 10.3, model 1). Detailed exploration of the enzyme–substrate and enzyme–product complexes revealed the possible importance of the active-site waters and halide ion for binding of small ligands in the active site. We decided to study the effects of the presence and absence of halogen anion in the enzyme active site and the presence and absence of active site water molecules on ligand orientations. The result of the docking calculation employing the specific combination of influential factors was quantitatively expressed as the RMSD of the docked orientation and the crystallographically determined orientation (Table 10.3).

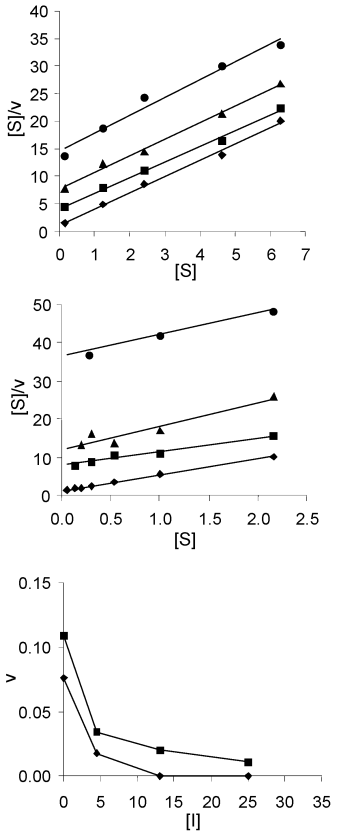

Figure 11.5: 1,2-DCE and 1,2-DCP inhibition kinetics. (a) Hanes–Woolf plot ([S] *versus* [S]/v) demonstrating the effect of 1,2-DCE on the hydrolysis of 1-CB by LinB. The velocities (v) were determined in the presence of 0 (⧫), 5 (■), 14 (▲) and 26 mM 1,2-DCE (●). (b) Hanes–Woolf ([S] *versus* [S]/v) plot demonstrating the effect of 1,2-DCP on 1-CB hydrolysis. Rates were obtained in the presence of 0 (⧫), 4 (■), 12 (▲) and 22 mM 1,2-DCP (●). (c) Plot of velocity of 1-CB hydrolysis (v) *versus* inhibitor concentration ([I]) at substrate concentration 0.15 mM for 1,2-DCE inhibition (■) and 0.14 mM for 1,2-DCP inhibition (⧫).

DISCUSSION

X-ray analyses of LinB complexed with 1,2-DCE and 1,2-DCP show that these compounds bind non-productively in the active site of the enzyme, while for 1-CB, the reaction products are observed. These observations are in agreement with previous kinetic analyzes [101]. The data from highly sensitive GC–MS analysis, however, indicate very weak activity of LinB with 1,2-DCE and 1,2-DCP, the values of which are 5 500 and 2 500 times slower than 1-CB dechlorination activity, respectively. The presence of Cl^- in the halide binding site in the crystallographically determined 1,2-DCP and 1,2-DCE complexes may derive from the weak activity of LinB toward these substrates or may be a residue from protein purification buffers (which contained Cl^- ions).

With the exception of residue Asp147, there is little movement of active site residues when comparing the complexes of LinB with the structure of the free enzyme. Slight movements were detected in Asp147; however, Asp147 is not close to the catalytic triad, and it would appear unlikely that this residue assists in the formation of a productive enzyme–substrate complex.

Inhibition kinetics confirm that both 1,2-DCE and 1,2-DCP compete for the active site with 1-CB. The kinetic data for both 1,2-DCE and 1,2-DCP inhibition of 1-CB turnover can be explained by a scheme in which either inhibitor or substrate binds to the active site. The substrate and inhibitors do not bind strictly into the same binding position (Figure 10.4b); however, the binding positions for inhibitors and substrate overlap.

The docking calculations help explain the non-productive binding of DCE and DCP in the active site of LinB. Docking should reveal the most energetically favorable binding sites within the active site. In the absence of a water molecule and halide in the docking calculations, the molecules bind productively in terms of both its position and conformation (Figures 10.2b and 10.3b). No alternative binding modes were found in these docking calculations. Yet as water and the halide ion are added to the calculation, the ligands adopt orientations increasingly similar to the crystallographically determined positions (Table 10.3 and Figures 10.2c and 10.3c). This result indicates that it is the relative inability of 1,2-DCE and 1,2-DCP to displace active site waters and halide ions compared to the ability of 1-CB that determines their low rate of turnover. Two water molecules in particular appear to be important. These can be seen in Figure 10.2a and 10.3a, and are located adjacent to the halide ions observed

in these structures. These water molecules appear to form a barrier between 1,2-DCE, 1,2-DCP, and the halide binding site. The water molecules are 3.3 and 3.0 Å from the halide ion and 2.9 and 3.0 Å from the carboxylate oxygen atoms of Asp108, respectively. These stabilizing interactions must be overcome if the ligand is to displace water molecules and bind productively. Interestingly, only one of these water molecules is seen in the complex with 1-butanol (Figure 10.4a). This water molecule is also observed in the apoenzyme structure [56].

Notably, small chlorinated substrates (1,1-dichloromethane, 1,2-DCE, and 1-chloropropane) exhibited outlying behavior in the quantitative model of structure–biodegradability relationships constructed for 27 substrates of the haloalkane dehalogenase DhaA [176]. These substrates exhibited unexpectedly low activity, which could not be directly related to any intrinsic property of these molecules, such as their reactivity, structural, or physicochemical properties. It was proposed that these molecules cannot adopt the proper orientation in the enzyme active site [176]. The data presented here agree with this proposal.

The production of enzymes with high activity toward 1,2-DCE and 1,2-DCP is desirable. Mutations in the cap domain of LinB could be introduced to make the non-productive binding site observed for 1,2-DCE and 1,2-DCP less energetically favorable. Specifically, the mutation of Phe143, Phe151, Val173, Leu177, and Phe169 (all in the cap domain) to bulkier amino acids could have this effect without disrupting the catalytic triad or occluding the active site. These five amino-acids form part of the binding pockets for 1,2-DCE and 1,2-DCP and line the active site cavity.

Comparisons with Other Dehalogenases
The catalytic activity of LinB contrasts with that of DhlA. While 1,2-DCE is an extremely poor substrate for LinB, it is a much better substrate for DhlA. Direct comparisons between LinB and DhlA [52] are difficult because the cap domains of the two proteins are very different [51]. One consequence of this difference is that the active site of DhlA dehalogenase (112 $Å^3$) is much smaller than that of LinB (276 $Å^3$). The small size of the DhlA active site appears only to allow productive binding of 1,2-DCE. LinB, with its relatively large active site, does not restrict 1,2-DCE to bind productively, and appears to provide pockets for the non-productive binding of 1,2-DCE and other small halocarbons.

The active site of LinB is similar to the *Rhodococcus* dehalogenase DhaA [59], which has an active site cavity only slightly smaller (246 $Å^3$) than that of LinB.

Table 10.3: Root-Mean-Square Deviations (Å) between Crystal and Docked Ligand Orientations[a].

	complexes		substrates				product	
	halogen anion	active site waters	1,2-DCE		1,2-DCP		1-butanol	
			molecule I	molecule J	molecule I	molecule J	molecule I	molecule J
model 1	–	–	4.4	4.7	4.5	4.1	NA[b]	NA[b]
model 2	+	–	3.6	3.7	3.6	3.8	4.2	4.0
model 3	+	+	2.7	2.4	1.4	1.8	2.5	2.3
model 4	–	+	2.9	2.5	1.2	1.7	NA[b]	NA[b]

[a] Molecule I and molecule J refer to the two ligand molecules in the crystal structures; [b] Not applicable

Importantly, the residues of the catalytic triad are the same in the two, and many of the hydrophobic residues lining the cavity are conserved [51]. Activity of DhaA to 1,2-DCE has not been detected [65], and therefore, a mechanism of water-mediated inhibition observed in LinB could act in DhaA. Further experiments are required to confirm this proposal. Two substitutions that occur in the active site of DhaA are Trp152 and Cys187 (corresponding to Phe143 and Leu177, respectively, in LinB). The existence of these substitutions support the proposal that mutations of residues in the cap domain of LinB can be created that retain catalytic activity.

CONCLUSIONS

This study shows that the industrially important molecules 1,2-DCE and 1,2-DCP can bind non-productively in the active site of LinB, and that the low activity of LinB for these substrates appears to result from their inability to displace water and halide molecules from the active site. The water- and product-mediated effects observed in this study could be important for other dehalogenases such as DhaA, which has an active site very similar to that of LinB. The activity of LinB toward small halocarbons such as 1,2-DCE could be improved through the introduction of mutations in the cap domain of the protein that make non-productive binding less favorable.

ACKNOWLEDGEMENTS

Financial support from the Australian Research Council (Grant A00001093) and Czech Ministry of Education (Grant LN00A016) is acknowledged. AJO is the recipient of an Australian Research Council Postdoctoral Research Fellowship. MCJW is supported by the Raine Foundation. MB acknowledges the MCB 2000 Studentships from University of Western Australia. Dr. Igor Kučera (Biochemistry Department, Masaryk University) is gratefully acknowledged for valuable advice on the presentation of enzyme kinetics data and Jana Sýkorová (National Center for Biomolecular Research, Masaryk University) for technical assistance with protein purification.

Bibliography

[1] G. H. Grant and W. G. Richards. *Computational Chemistry*. Oxford University Press, New York, 1995.

[2] A. R. Leach. *Molecular Modelling: Principles and Applications*. Addison Wesley Longman Limited, England, 1996.

[3] M. Remko. *Molekulové modelovanie: Princípy a aplikácie*. Slovak Academic Press, Bratislava, 2000.

[4] Ch. J. Cramer. *Essentials of Computational Chemistry: Theories and Models*. Wiley-Interscience, England, 2002.

[5] F. Jensen. *Introduction to Computational Chemistry*. Wiley-Interscience, England, 1999.

[6] W. Koch and M. C. Holthausen. *A Chemist's Guide to Density Functional Theory*. Wiley-Interscience, Germany, 2001.

[7] M. Springborg. Density-Functional Formalism—Foundations, Limitations and Applications. In M. Springborg, editor, *Density-Functional Methods in Chemistry and Material Sciences*, pages 1–18. Wiley-Interscience, England, 1997.

[8] D. C. Young. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*. Wiley-Interscience, England, 2001.

[9] A. C. T. Van Duin, S. Dasgupta, F. Lorant, and W. A. Goddard. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A*, 105:9396–9409, 2001.

[10] M. Karplus and J. A. McCammon. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.*, 9:646–652, 2002.

[11] A. Warshel. Computer Simulations of Enzyme Catalysis: Methods, Progress, and Insights. *Annu. Rev. Biophys. Biomol. Struct.*, 32:425–443, 2003.

[12] A. Warshel. Electrostatic Origin of the Catalytic Power of Enzymes and the Role of Preorganized Active Sites. *J. Biol. Chem.*, 273:27035–27038, 1998.

[13] W. Wang, O. Donini, C. M. Reyes, and P. A. Kollman. Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions. *Annu. Rev. Biophys. Biomol. Struct.*, 30: 211–243, 2001.

[14] R. C. Wade, A. R. Ortiz, and F. Gago. Comparative Binding Energy Analysis. In H. Kubinyi, G. Folkers, and Y. C. Martin, editors, *3D QSAR in Drug Design*, pages 23–28. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

[15] S. Wold, E. Johansson, and M. Cocchi. PLS—Partial Leasts-Squares Projections to Latent Structures. In H. Kubinyi, editor, *3D QSAR in Drug Design: Theory, Methods and Application*, pages 523–550. ESCOM, Leiden, The Netherlands, 1993.

[16] P. Geladi and B. R. Kowalski. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta*, 185:1–17, 1986.

[17] G. Cruciani, S. Clementi, and M. Baroni. Variable Selection in PLS Analysis. In H. Kubinyi, editor, *3D QSAR in Drug Design: Theory, Methods and Application*, pages 551–564. ESCOM, Leiden, The Netherlands, 1993.

[18] S. Wold. Validation of QSAR's. *Quant. Struct.-Act. Relat.*, 10:191–193, 1991.

[19] A. R. Ortiz, M. T. Pisabarro, F. Gago, and R. C. Wade. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.*, 38:2681–2691, 1995.

[20] A. R. Ortiz, M. Pastor, A. Palomer, G. Cruciani, F. Gago, and R. C. Wade. Reliability of Comparative Molecular Field Analysis Models: Effects of Data Scaling and Variable Selection using a Set of Human Synovial Fluid Phospholipase A2 Inhibitors. *J. Med. Chem.*, 40:1136–1148, 1997.

[21] C. Perez, M. Pastor, A. R. Ortiz, and F. Gago. Comparative Binding Energy Analysis of HIV-1 Protease Inhibitors: Incorporation of Solvent Effects and Validation as a Powerful Tool in Receptor-Based Drug Design. *J. Med. Chem.*, 41:836–852, 1998.

[22] M. Pastor, C. Perez, and F. Gago. Simulation of Alternative Binding Modes in a Structure-Based QSAR Study of HIV-1 Protease Inhibitors. *J. Mol. Graphics Model.*, 15:364–371, 1997.

[23] S. Tomic, L. Nilsson, and R. C. Wade. Nuclear Receptor-DNA Binding Specificity: A COMBINE and Free-Wilson QSAR Analysis. *J. Med. Chem.*, 43:1780–1792, 2000.

[24] J. J. Lozano, M. Pastor, G. Cruciani, K. Gaedt, N. B. Centeno, F. Gago, and F. J. Sanz. 3D-QSAR Methods on the Basis of Ligand-Receptor Complexes. Application of COMBINE and GRID/GOLPE Methodologies to a Series of CYP1A2 Ligands. *J. Comput.-Aid. Mol. Design*, 14:341–353, 2000.

[25] C. Cuevas, M. Pastor, C. Perez, and F. Gago. Comparative Binding Energy (COMBINE) Analysis of Human Neutrophil Elastase Inhibition by Pyridone-Containing Trifluormethylketones. *Combin. Chem. High Through. Screen.*, 4:627–642, 2001.

[26] T. Wang and R. C. Wade. Comparative Binding Energy (COMBINE) Analysis of Influenza Neuraminidase-Inhibitor Complexes. *J. Med. Chem.*, 44:961–971, 2001.

[27] T. Wang and R. C. Wade. Comparative Binding Energy (COMBINE) Analysis of OppA-Peptide Complexes to Relate Structure to Binding Thermodynamics. *J. Med. Chem.*, 45:4828–4837, 2002.

[28] J. Kmuníček, S. Luengo, F. Gago, A. R. Ortiz, R. C. Wade, and J. Damborský. Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10. *Biochemistry*, 40:8905–8917, 2001.

[29] J. Damborský, J. Kmuníček, T. Jedlička, S. Luengo, F. Gago, A. R. Ortiz, and R. C. Wade. Rational Re-Design of Haloalkane Dehalogenases Guided by Comparative Binding Energy Analysis. In A. Svendsen, editor, *Enzyme Functionality: Design, Engineering and Screening*, pages 79–96. Marcel Dekker, New York, 2003.

[30] J. Kmuníček, S. Luengo, F. Gago, A. R. Ortiz, R. C. Wade, and J. Damborský. Comparative Binding Energy Analysis of Haloalkane Dehalogenase Substrates: Modelling of Enzyme-Substrate Complexes by Molecular Docking and Quantum Mechanical Calculations. *J. Comput.-Aid. Mol. Design*, 17:299–311, 2003.

[31] S. Fetzner and F. Lingens. Bacterial Dehalogenases: Biochemistry, Genetics, and Biotechnological Applications. *Microbiol. Rev.*, 58:641–685, 1994.

[32] S. Fetzner. Bacterial Dehalogenation. *Appl. Microbiol. Biotechnol.*, 50: 633–657, 1998.

[33] D. J. Hardman. Biotransformation of Halogenated Compounds. *Crit. Rev. Biotechnol.*, 11:1–40, 1991.

[34] J. H. Slater, A. T. Bull, and D. J. Hardman. Microbial Dehalogenation. *Biodegradation*, 6:181–189, 1995.

[35] N. Winterton. Chlorine: The Only Green Element—towards a Wider Acceptance of Its Role in Natural Cycles. *Green Chem.*, 2:173–225, 2000.

[36] G. W. Gribble. The Natural Production of Chlorinated Compounds. *Environ. Sci. Technol.*, 28:311–319, 1994.

[37] J. A. Field, F. J. M. Verhagen, and E. de Jong. Natural Organohalogen Production by Basidiomycetes. *Trends Biotechnol.*, 13:451–456, 1995.

[38] D. B. Janssen, F. Pries, and J. R. Van der Ploeg. Genetics and Biochemistry of Dehalogenating Enzymes. *Annu. Rev. Microbiol.*, 48:163–191, 1994.

[39] D. B. Janssen and J. P. Schanstra. Engineering Proteins for Environmental Applications. *Curr. Opin. Biotech.*, 5:253–259, 1994.

[40] D. B. Janssen, J. E. Oppentocht, and G. J. Poelarends. Microbial Dehalogenation. *Curr. Opin. Biotechnol.*, 12:254–258, 2001.

[41] P. E. Swanson. Dehalogenases Applied to Industrial-Scale Biocatalysis. *Curr. Opin. Biotechnol.*, 10:365–369, 1999.

[42] S. D. Copley. Microbial Dehalogenases: Enzymes Recruited to Convert Xenobiotic Substrates. *Curr. Opin. Chem. Biol.*, 2:613–617, 1998.

[43] J. E. T. H. Vlieg, G. J. Poelarends, A. E. Mars, and D. B. Janssen. Detoxification of Reactive Intermediates during Microbial Mechanism of Halogenated Compounds. *Curr. Opin. Microbiol.*, 3:257–262, 2000.

[44] F. Pries, J. R. van der Ploeg, J. Dolfing, and D. B. Janssen. Degradation of Halogenated Aliphatic Compounds: The Role of Adaptation. *FEMS Microbiol. Rev.*, 15:279–295, 1994.

[45] T. Leisinger and R. Bader. Microbial Dehalogenation of Synthetic Organohalogen Compounds: Hydrolytic Dehalogenases. *Chimia*, 47:116–121, 1993.

[46] B. C. Dravis, K. E. Le Jeune, A. D. Hetro, and A. J. Russell. Enzymatic Dehalogenation of Gas Phase Substrates with Haloalkane Dehalogenase. *Biotechnol. Bioeng.*, 69:235–241, 2000.

[47] J. Damborský, E. Rorije, A. Jesenská, Y. Nagata, G. Klopman, and W. J. G. M. Peijnenburg. Structure-Specificity Relationships for Haloalkane Dehalogenases. *Environ. Toxicol. Chem.*, 20:2681–2689, 2001.

[48] G. Stucki and M. Thuer. Experiences of a Large-Scale Application of 1,2-Dichloroethane Degrading Microorganisms for Groundwater Treatment. *Environ. Sci. Technol.*, 29:2339–2345, 1995.

[49] D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, K. H. G. Verschueren, and A. Goldman. The α/β Hydrolase Fold. *Prot. Engng.*, 5:197–211, 1992.

[50] M. Nardini and B. W. Dijkstra. α/β Hydrolase Fold Enzymes: The Family Keeps Growing. *Curr. Opin. Struct. Biol.*, 9:732–737, 1999.

[51] J. Damborský and J. Koča. Analysis of the Reaction Mechanism and Substrate Specificity of Haloalkane Dehalogenases by Sequential and Structural Comparisons. *Prot. Engng.*, 12:989–998, 1999.

[52] K. H. G. Verschueren, F. Seljee, H. J. Rozeboom, K. H. Kalk, and B. W. Dijkstra. Crystallographic Analysis of the Catalytic Mechanism of Haloalkane Dehalogenase. *Nature*, 363:693–698, 1993.

[53] K. H. G. Verschueren, S. M. Franken, H. J. Rozeboom, K. H. Kalk, and B. W. Dijkstra. Refined X-ray Structures of Haloalkane Dehalogenase at pH 6.2 and pH 8.2 and Implications for the Reaction Mechanism. *J. Mol. Biol.*, 232:856–872, 1993.

[54] S. M. Franken, H. J. Rozeboom, K. H. Kalk, and B. W. Dijkstra. Crystal Structure of Haloalkane Dehalogenase: An Enzyme to Detoxify Halogenated Alkanes. *EMBO J.*, 10:1297–1302, 1991.

[55] I. S. Ridder, H. J. Rozeboom, and B. W. Dijkstra. Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10 Refined at 1.15 Å Resolution. *Biol. Crystallogr.*, 55:1273–1290, 1999.

[56] J. Marek, J. Vévodová, I. Kutá-Smatanová, Y. Nagata, L. A. Svensson, J. Newman, M. Takagi, and J. Damborský. Crystal Structure of the Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26. *Biochemistry*, 39:14082–14086, 2000.

[57] A. J. Oakley, Z. Prokop, M. Boháč, J. Kmuníček, T. Jedlička, M. Monincová, I. Kutá-Smatanová, Y. Nagata, J. Damborský, and M. C. J. Wilce. Exploring the Structure and Activity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26: Evidence for Product and Water Mediated Inhibition. *Biochemistry*, 41:4847–4855, 2002.

[58] V. A. Streltsov, Z. Prokop, J. Damborský, Y. Nagata, A. Oakley, and M. C. J. Wilce. Haloalkane Dehalogenase LinB from *Sphingomonas paucimobilis* UT26: X-ray Crystallographic Studies of Dehalogenation of Brominated Substrates. *Biochemistry*, 42:10104–10112, 2003.

[59] J. Newman, T. S. Peat, R. Richard, L. Kan, P. E. Swanson, J. A. Affholter, I. H. Holmes, J. F. Schindler, C. J. Unkefer, and T. C. Terwilliger. Haloalkane Dehalogenase: Structure of a *Rhodococcus* Enzyme. *Biochemistry*, 38:16105–16114, 1999.

[60] P. Bryan, M. W. Pantoliano, S. G. Quill, H.-Y. Hsiao, and T. Poulos. Site-Directed Mutagenesis and the Role of the Oxyanion Hole in Subtilisin. *Proc. Natl. Acad. Sci. U.S.A.*, 83:3743–3745, 1986.

[61] K. H. G. Verschueren, J. Kingma, H. J. Rozeboom, K. H. Kalk, D. B. Janssen, and B. W. Dijkstra. Crystallographic and Fluorescence Studies of the Interaction of Haloalkane Dehalogenase with Halide Ions. Studies with Halide Compounds Reveal a Halide Binding Site in the Active Site. *Biochemistry*, 32:9031–9037, 1993.

[62] K. H. G. Verschueren, S. M. Franken, H. J. Rozeboom, K. H. Kalk, and B. W. Dijkstra. Non-Covalent Binding of the Heavy Atom Compound $[Au(CN)_2]^-$ at the Halide Binding Site of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10. *FEBS Lett.*, 323:267–270, 1993.

[63] G. H. Krooshof, I. S. Ridder, A. W. J. W. Tepper, G. J. Vos, H. J. Rozeboom, K. H. Kalk, B. W. Dijkstra, and D. B. Janssen. Kinetic Analysis and X-ray Structure of Haloalkane Dehalogenase with a Modified Halide-Binding Site. *Biochemistry*, 37:15013–15023, 1998.

[64] J. P. Schanstra, I. S. Ridder, G. J. Heimeriks, R. Rink, G. J. Poelarends, K. H. Kalk, B. W. Dijkstra, and D. B. Janssen. Kinetic Characterization and X-ray Structure of a Mutant of Haloalkane Dehalogenase with Higher Catalytic Activity and Modified Substrate Range. *Biochemistry*, 35:13186–13195, 1996.

[65] J. F. Schindler, P. A. Naranjo, D. A. Honaberger, C. H. Chang, J. R. Brainard, L. A. Vanderberg, and C. J. Unkefer. Haloalkane Dehalogenases: Steady-State Kinetics and Halide Inhibition. *Biochemistry*, 38:5772–5778, 1999.

[66] J. P. Schanstra and D. B. Janssen. Kinetics of Halide Release of Haloalkane Dehalogenase: Evidence for a Slow Conformational Change. *Biochemistry*, 35:5624–5632, 1996.

[67] J. P. Schanstra, J. Kingma, and D. B. Janssen. Specificity and Kinetics of Haloalkane Dehalogenase. *J. Biol. Chem.*, 271:14747–14753, 1996.

[68] D. B. Janssen, J. Gerritse, J. Brackman, C. Kalk, D. Jager, and B. Witholt. Purification and Characterization of a Bacterial Dehalogenase with Activity toward Halogenated Alkanes, Alcohols and Ethers. *Eur. J. Biochem.*, 171:67–92, 1988.

[69] G. H. Krooshof, E. M. Kwant, J. Damborský, J. Koča, and D. B. Janssen. Repositioning the Catalytic Triad Acid of Haloalkane Dehalogenase: Effects on Activity and Kinetics. *Biochemistry*, 36:9571–9580, 1997.

[70] G. H. Krooshof, R. Floris, A. Tepper, and D. B. Janssen. Thermodynamic Analysis of Halide Binding to Haloalkane Dehalogenase Suggests the Occurrence of Large Conformational Changes. *Prot. Sci.*, 8:355–360, 1999.

[71] F. Pries, J. Kingma, M. Pentega, G. Van Pouderoyen, C. M. Jeronimus-Stratingh, A. P. Bruins, and D. B. Janssen. Site-Directed Mutagenesis and Oxygen Isotope Incorporation Studies of the Nucleophilic Aspartate of Haloalkane Dehalogenase. *Biochemistry*, 33:1242–1247, 1994.

[72] F. Pries, J. Kingma, and D. B. Janssen. Activation of an Asp-124 → Asn Mutant of Haloalkane Dehalogenase by Hydrolytic Deamidation of Asparagine. *FEBS Lett.*, 358:171–174, 1995.

[73] F. Pries, J. Kingma, G. H. Krooshof, C. M. Jeronimus-Stratingh, A. P. Bruins, and D. B. Janssen. Histidine 289 is Essential for Hydrolysis of the Alkyl-Enzyme Intermediate of Haloalkane Dehalogenase. *J. Biol. Chem.*, 270:10405–10411, 1995.

[74] F. Pries, A. J. van den Wijngaard, R. Bos, M. Pentenga, and D. B. Janssen. The Role of Spontaneous Cap Domain Mutations in Haloalkane Dehalogenase Specificity and Evolution. *J. Biol. Chem.*, 269:17490–17494, 1994.

[75] C. Kennes, F. Pries, G. H. Krooshof, E. Bokma, J. Kingma, and D. B. Janssen. Replacement of Tryptophan Residues in Haloalkane Dehalogenase Reduces Halide Binding and Catalytic Activity. *Eur. J. Biochem.*, 228:403–407, 1995.

[76] J. P. Schanstra, A. Ridder, J. Kingma, and D. B. Janssen. Influence of Mutations of Val226 on the Catalytic Rate of Haloalkane Dehalogenase. *Prot. Engng.*, 10:53–61, 1997.

[77] J. P. Schanstra, R. Rink, F. Pries, and D. B. Janssen. Construction of an Expression and Site-Directed Mutagenesis System of Haloalkane Dehalogenase in *Escherichia coli*. *Prot. Expr. Purif.*, 4:479–489, 1993.

[78] S. Marvanová, Y. Nagata, M. Wimmerová, J. Sýkorová, and K. Hynková. Biochemical Characterization of Broad-Specificity Enzymes using Multivariate Experimental Design and a Colorimetric Microplate Assay: Characterization of the Haloalkane Dehalogenase Mutants. *J. Microbiol. Methods*, 44:149–157, 2001.

[79] J. Damborský. Quantitative Structure-Function Relationships of the Single-Point Mutants of Haloalkane Dehalogenase: A Multivariate Approach. *Quant. Struct.-Act. Relat.*, 16:126–135, 1997.

[80] J. Damborský, A. Berglund, M. Kutý, A. Ansorgová, Y. Nagata, and M. Sjøstrøm. Mechanism-Based Quantitative Structure-Biodegradability Relationships for Hydrolytic Dehalogenation of Chloro- and Bromoalkenes. *Quant. Struct.-Act. Relat.*, 17:450–485, 1998.

[81] J. Damborský, K. Manová, A. Berglund, M. Sjøstrøm, M. Němec, and I. Holoubek. Biotransformation of Chloro- and Bromoalkenes by Crude Extracts of *Rhodococcus erythropolis* Y2. *Adv. Environ. Res.*, 1:50–57, 1997.

[82] J. Damborský, K. Manová, and M. Kutý. A Mechanistic Approach to Deriving Quantitative Structure-Biodegradability Relationships. A Case Study: Dehalogenation of Haloaliphatic Compounds. In W. J. G. M. Peijnenburg and J. Damborský, editors, *Biodegradability Prediction*, pages 75–92. Kluwer Academic Publishers, Netherlands, 1996.

[83] J. Damborský. Quantitative Structure-Function and Structure-Stability Relationships of Purposely Modified Proteins. *Prot. Engng.*, 11:21–30, 1998.

[84] J. Damborský, M. G. Nyandoroh, M. Němec, I. Holoubek, A. T. Bull, and D. J. Hardman. Some Biochemical Properties and Classification of a Range of Bacterial Haloalkane Dehalogenases. *Biotechnol. Appl. Biochem.*, 26:19–25, 1997.

[85] J. Damborský, M. Kutý, M. Němec, and J. Koča. A Molecular Modeling Study of the Catalytic Mechanism of Haloalkane Dehalogenase: 1. Quantum Chemical Study of the First Reaction Step. *J. Chem. Inf. Comput. Sci.*, 37:562–568, 1997.

[86] M. Kutý, J. Damborský, M. Prokop, and J. Koča. A Molecular Modeling Study of the Catalytic Mechanism of Haloalkane Dehalogenase: 2. Quantum Chemical Study of Complete Reaction Mechanism. *J. Chem. Inf. Comput. Sci.*, 38:736–741, 1998.

[87] J. Damborský. Computer Modelling of Microbial Hydrolytic Dehalogenation. *Pure. Appl. Chem.*, 70:1375–1383, 1998.

[88] J. Damborský, M. Kutý, M. Němec, and J. Koča. Molecular Modelling to Understand the Mechanisms of Microbial Degradation—Application to Hydrolytic Dehalogenation with Haloalkane Dehalogenases. In F. Chen and G. Schüürmann, editors, *Quantitative Structure-Activity Relationships in Environmental Sciences*, volume VII, pages 5–20. SETAC Press, Pensacola, 1997.

[89] J. Damborský, M. Boháč, M. Prokop, M. Kutý, and J. Koča. Computational Site-Directed Mutagenesis of Haloalkane Dehalogenase in Position 172. *Prot. Engng.*, 11:901–907, 1998.

[90] L. S. Devi-Kesavan and J. Gao. Combined QM/MM Study of Mechanism and Kinetic Isotope Effect of the Nucleophilic Substitution Reaction in Haloalkane Dehalogenase. *J. Am. Chem. Soc.*, 125:1532–1540, 2003.

[91] F. C. Lightstone, Y. J. Zheng, A. H. Maulitz, and T. C. Bruice. Non-Enzymatic and Enzymatic Hydrolysis of Alkyl Halides: A Haloalkane Dehalogenation Enzyme Evolved to Stabilize the Gas-Phase Transition State of an S_N2 Displacement Reaction. *Proc. Natl. Acad. Sci. U.S.A.*, 94:8417–8420, 1997.

[92] F. C. Lightstone, Y. J. Zheng, and T. C. Bruice. Molecular Dynamics Simulations of Ground and Transition States for the S_N2 Displacement of Cl^- from 1,2-Dichloroethane at the Active Site of *Xanthobacter autotrophicus* Haloalkane Dehalogenase. *J. Am. Chem. Soc.*, 120:5611–5621, 1998.

[93] E. Y. Lau, K. Kahn, P. Bash, and T. C. Bruice. The Importance of Reactant Positioning in Enzyme Catalysis: A Hybrid Quantum Mechanics/Molecular Mechanics Study of a Haloalkane Dehalogenase. *Proc. Natl. Acad. Sci. U.S.A.*, 97:9937–9942, 2000.

[94] M. Otyepka and J. Damborský. Functionally Relevant Motions of Haloalkane Dehalogenases Occur in the Specificity-Modulating Cap Domains. *Prot. Sci.*, 11:1206–1217, 2002.

[95] M. Boháč, Y. Nagata, Z. Prokop, M. Prokop, M. Monincová, M. Tsuda, J. Koča, and J. Damborský. Halide-Stabilizing Residues of Haloalkane Dehalogenases Studied by Quantum Mechanic Calculations and Site-Directed Mutagenesis. *Biochemistry*, 41:14272–14280, 2002.

[96] D. Robert, X. Gironés, and R. Carbó-Dorca. Quantification of the Influence of Single-Point Mutations on Haloalkane Dehalogenase Activity: A Molecular Quantum Similarity Study. *J. Chem. Inf. Comput. Sci.*, 40: 839–846, 2000.

[97] P. Paneth. Chlorine Kinetic Isotope Effects on Enzymatic Dehalogenations. *Acc. Chem. Res.*, 36:120–126, 2003.

[98] A. Soriano, E. Silla, and I. Tuñón. Internal Rotation of 1,2-Dichloroethane in Haloalkane Dehalogenase. A Test Case for Analyzing Electrostatic Effects in Enzymes. *J. Phys. Chem. B*, 107:6234–6238, 2003.

[99] K. Kahn and T. C. Bruice. Comparison of Reaction Energetics and Leaving Group Interactions during the Enzyme-Catalyzed and Uncatalyzed Displacement of Chloride from Haloalkanes. *J. Phys. Chem. B*, 107:6876–6885, 2003.

[100] H. J. Rozeboom, J. Kingma, D. B. Janssen, and B. W. Dijkstra. Crystallization of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10. *J. Mol. Biol.*, 200:611–612, 1988.

[101] Y. Nagata, K. Miyauchi, J. Damborský, K. Manová, A. Ansorgová, and M. Takagi. Purification and Characterization of Haloalkane Dehalogenase of a New Substrate Class from a γ-Hexachlorocyclohexane-Degrading Bacterium, *Sphingomonas paucimobilis* UT26. *Appl. Environ. Microbiol.*, 63: 3707–3710, 1997.

[102] Y. Nagata, K. Hynková, J. Damborský, and M. Takagi. Construction and Characterization of Histidine-Tagged Haloalkane Dehalogenase (LinB) of a New Substrate Class from a γ-Hexachlorocyclohexane-Degrading Bacterium, *Sphingomonas paucimobilis* UT26. *Prot. Expr. Purif.*, 17:299–304, 1999.

[103] K. Hynková, Y. Nagata, M. Takagi, and J. Damborský. Identification of the Catalytic Triad in the Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26. *FEBS Lett.*, 446:177–181, 1999.

[104] Y. Nagata, K. Miyauchi, and M. Takagi. Complete Analysis of Genes and Enzymes for γ-Hexachlorocyclohexane Degradation in *Sphingomonas paucimobilis* UT26. *J. Industr. Microbiol. Biotechnol.*, 23:380–390, 1999.

[105] D. D. Y. Ryu and D.-H. Nam. Recent Progress in Biomolecular Engineering. *Biotechnol. Prog.*, 16:2–16, 2000.

[106] U. T. Bornscheuer and M. Pohl. Improved Biocatalysts by Directed Evolution and Rational Protein Design. *Curr. Opin. Chem. Biol.*, 5:137–143, 2001.

[107] R. Chen. Enzyme Engineering: Rational Redesign versus Directed Evolution. *Trends Biotechnol.*, 19:13–14, 2001.

[108] D. N. Bolon, Ch. A. Voigt, and S. L. Mayo. *De Novo* Design of Biocatalysts. *Curr. Opin. Chem. Biol.*, 6:125–129, 2002.

[109] M. M. Altamirano, J. M. Blackburn, C. Aquayo, and A. R. Fersht. Directed Evolution of New Catalytic Activity using α/β-Barrel Scaffold. *Nature*, 403:617–622, 2000.

[110] H. Tao and V. W. Cornish. Milestones in Directed Enzyme Evolution. *Curr. Opin. Chem. Biol.*, 6:858–864, 2002.

[111] M. G. Pikkemaat, A. B. M. Linssen, H. J. C. Berendsen, and D. B. Janssen. Molecular Dynamics Simulations as a Tool for Improving Protein Stability. *Prot. Engng.*, 15:185–192, 2002.

[112] M. G. Pikkemaat and D. B. Janssen. Generating Segmental Mutations in Haloalkane Dehalogenase: A Novel Part in the Directed Evolution Toolbox. *Nucl. Acid Res.*, 30:35–40, 2002.

[113] O. Kuchner and F. H. Arnold. Directed Evolution of Enzyme Catalysts. *Trends Biotechnol.*, 15:523–530, 1997.

[114] F. H. Arnold. Design by Directed Evolution. *Acc. Chem. Res.*, 31:125–131, 1998.

[115] M. Lehmann and M. Wyss. Engineering Proteins for Thermostability: The Use of Sequence Alignments versus Rational Design and Directed Evolution. *Curr. Opin. Biotechnol.*, 12:371–375, 2001.

[116] K. A. Gray, T. H. Richardson, K. Kretz, J. M. Short, F. Bartnek, R. Knowles, L. Kan, P. E. Swanson, and D. E. Robertson. Rapid Evolution of Reversible Denaturation and Elevated Melting Temperature in a Microbial Haloalkane Dehalogenase. *Adv. Synth. Catal.*, 343:607–617, 2001.

[117] S. Keuning, D. B. Janssen, and B. Witholt. Purification and Characterization of Hydrolytic Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10. *J. Bacteriol.*, 163:635–639, 1985.

[118] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[119] M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andrés, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. González, and J. A. Pople. *GAUSSIAN 94*. Gaussian, Inc., Pittsburgh, PA, 1995.

[120] C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman. A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges—the RESP Model. *J. Phys. Chem.*, 97: 10269–10280, 1993.

[121] W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman. Application of RESP Charges to Calculate Conformational Energies, Hydrogen-Bond Energies, and Free-Energies of Solvation. *J. Am. Chem. Soc.*, 115:9620–9631, 1993.

[122] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, D. M. Ferguson, G. L. Seibel, C. Singh, P. K. Weiner, and P. A. Kollman. *AMBER 4.1*. University of California, San Francisco, 1995.

[123] T. Fox and P. A. Kollman. Application of the RESP Methodology in the Parametrization of Organic Solvents. *J. Phys. Chem. B*, 102:8070–8079, 1998.

[124] D. R. Lide. *CRC Handbook of Chemistry and Physics*. CRC Press, Boca Raton.

[125] J. P. Ryckaert, G. Ciccoti, and H. J. C. Berendsen. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.*, 23:327–341, 1977.

[126] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, D. M. Ferguson, R. J. Radmer, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. *AMBER 5.0*. University of California, San Francisco, 1997.

[127] G. Vriend. WHAT IF: A Molecular Modeling and Drug Design Program. *J. Mol. Graphics*, 8:52–56, 1990.

[128] B. Lee and F. M. Richards. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.*, 55:379–400, 1971.

[129] D. Eisenberg, R. M. Weiss, T. C. Terwilliger, and W. Wilcox. Hydrophobic Moments and Protein-Structure. *Faraday Symp. Chem. Soc.*, 17:109–120, 1982.

[130] D. Eisenberg, M. Wesson, and M. Yamashita. Interpretation of Protein Folding and Binding with Atomic Solvation Parameters. *Chem. Scr.*, 29A: 217–221, 1989.

[131] M. K. Gilson, K. A. Sharp, and B. H. Honig. Calculating the Electrostatic Potential of Molecules in Solution—Method and Error Assessment. *J. Comput. Chem.*, 9:327–335, 1988.

[132] M. K. Gilson and B. H. Honig. Calculation of the Total Electrostatic Energy of a Macromolecular System - Solvation Energies, Binding-Energies, and Conformational-Analysis. *Proteins: Struct. Funct. Genet.*, 4:7–18, 1988.

[133] A. Nicholls and B. H. Honig. A Rapid Finite-Difference Algorithm, Utilizing Successive Over-Relaxation to Solve the Poisson–Boltzmann Equation. *J. Comput. Chem.*, 12:435–445, 1991.

[134] F. M. Richards. Areas, Volumes, Packing and Protein Structure. *Annu. Rev. Biophys. Bioeng.*, 6:151–176, 1977.

[135] M. L. Connolly. Analytical Molecular Surface Calculation. *J. Appl. Crystallogr.*, 16:548–558, 1983.

[136] M. Baroni, G. Costantino, G. Cruciani, D. Riganelli, R. Valigi, and S. Clementi. Generating Optimal Linear PLS Estimations (GOLPE): An Advanced Chemometric Tool for Handling 3D-QSAR Problems. *Quant. Struct.-Act. Relat.*, 12:9–20, 1993.

[137] L. Eriksson, E. Johansson, and S. Wold. Quantitative Structure-Activity Relationship Model Validation. In F. Chen and G. Schüürmann, editors, *Quantitative Structure-Activity Relationships in Environmental Sciences*, volume VII, pages 381–397. SETAC Press, Pensacola, 1997.

[138] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated Docking using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.*, 19:1639–1662, 1998.

[139] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The Protein Data Bank. *Nucl. Acid. Res.*, 28:235–242, 2000.

[140] F. J. Solis and R. J. B. Wets. Minimization by Random Search Techniques. *Math. Oper. Res.*, 6:19–30, 1981.

[141] J. J. P. Stewart. MOPAC - A Semiempirical Molecular-Orbital Program. *J. Comput.-Aid. Mol. Design*, 4:1–45, 1990.

[142] M. Černohorský, M. Kutý, and J. Koča. A Multidimensional Driver for Quantum Chemistry Program MOPAC. *Comp. Chem.*, 21:35–45, 1996.

[143] J. Damborský, M. Prokop, and J. Koča. TRITON: Graphic Software for Rational Engineering of Enzymes. *Trends Biochem. Sci.*, 26:71–73, 2001.

[144] A. Shurki, M. Štrajbl, J. Villà, and A. Warshel. How Much Do Enzymes Really Gain by Restraining Their Reacting Fragments? *J. Am. Chem. Soc.*, 124:4097–4107, 2002.

[145] J. A. Brannigan and A. J. Wilkinson. Protein Engineering 20 Years On. *Nat. Rev. Mol. Cell Biol.*, 3:964–970, 2002.

[146] S. J. Anthony-Cahill and T. J. Magliery. Expanding the Natural Repertoire of Protein Structure and Function. *Curr. Pharm. Biotechnol.*, 3: 299–315, 2002.

[147] D. B. Janssen, J. E. Oppentocht, and G. Poelarends. Bacterial Growth on Halogenated Aliphatic Hydrocarbons: Genetics and Biochemistry. In M. M. Haggblom and I. D. Bossert, editors, *Dehalogenation: Microbial Processes and Environmental Applications*, pages 207–226. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2003.

[148] Z. Prokop, M. Monincová, R. Chaloupková, M. Klvaňa, Y. Nagata, D. B. Janssen, and J. Damborský. Catalytic Mechamism of the Haloalkane Dehalogenase LinB from *Sphingomonas paucimobilis* UT26. *J. Biol. Chem.*, 2003. In press.

[149] T. Bosma, M. G. Pikkemaat, J. Kingma, J. Dijk, and D. B. Janssen. Steady-State and Pre-Steady-State Kinetic Analysis of Halopropane Conversion by a *Rhodococcus* Haloalkane Dehalogenase. *Biochemistry*, 2003. In press.

[150] R. C. Wade. Derivation of QSARs using 3D Structural Models of Protein-Ligand Complexes by COMBINE Analysis. In H. D. Holtje and W. Sippl, editors, *Rational Approaches to Drug Design: 13th European Symposium on Quantitative Structure-Activity Relationships*, pages 23–28. Prous Science, Barcelona, 2001.

[151] R. D. Cramer, D. E. Patterson, and J. D. Bunce. Comparative Molecular Field Analysis (CoMFA). 1. Effect of Shape on Binding of Steroids to Carrier Proteins. *J. Am. Chem. Soc.*, 110:5959–5967, 1988.

[152] T. Yokota, T. Omori, and T. Kodama. Purification and Properties of Haloalkane Dehalogenase from *Corynebacterium* sp. strain m15-3. *J. Bacteriol.*, 169:4049–4054, 1987.

[153] R. Scholtz, T. Leisinger, F. Suter, and A. M. Cook. Characterization of 1-Chlorohexane Halidohydrolase, a Dehalogenase of Wide Substrate Range from an *Arthrobacter* sp. *J. Bacteriol.*, 169:5016–5021, 1987.

[154] P. J. Sallis, S. J. Armfield, A. T. Bull, and D. J. Hardman. Isolation and Characterization of a Haloalkane Halidohydrolase from *Rhodococcus erythropolis* Y2. *J. Gen. Microbiol.*, 136:115–120, 1990.

[155] G. J. Poelarends, M. Wilkens, M. J. Larkin, J. D. Van Elsas, and D. B. Janssen. Degradation of 1,3-Dichloropropene by *Pseudomonas cichorii* 170. *Appl. Environ. Microbiol.*, 64:2931–2936, 1998.

[156] A. Jesenská, M. Bartoš, V. Czerneková, I. Rychlík, I. Pavlík, and J. Damborský. Cloning and Expression of the Haloalkane Dehalogenase Gene dhmA from *Mycobacterium avium* N85 and Preliminary Characterization of DhmA. *Appl. Environ. Microbiol.*, 68:3724–3730, 2002.

[157] P. Holloway, K. L. Knoke, J. T. Trevors, and H. Lee. Alternation of the Substrate Range of Haloalkane Dehalogenase by Site-Directed Mutagenesis. *Biotechnol. Bioeng.*, 59:520–523, 1998.

[158] Y. Nagata, Z. Prokop, S. Marvanová, J. Sýkorová, M. Monincová, M. Tsuda, and J. Damborský. Reconstruction of Mycobacterial Dehalogenase Rv2579 by Cumulative Mutagenesis of Haloalkane Dehalogenase LinB. *Appl. Environ. Microbiol.*, 69:2349–2355, 2003.

[159] R. Chaloupková, J. Sýkorová, Z. Prokop, A. Jesenská, M. Monincová, M. Pavlová, Y. Nagata, M. Tsuda, and J. Damborský. Modification of Activity and Specificity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26 by Engineering of Its Entrance Tunnel. Submitted.

[160] M. G. Pikkemaat, I. S. Ridder, H. J. Rozeboom, K. H. Kalk, B. W. Dijkstra, and D. B. Janssen. Crystallographic and Kinetic Evidence of a Collision Complex Formed during Halide Import in Haloalkane Dehalogenase. *Biochemistry*, 38:12052–12061, 1999.

[161] A. H. Maulitz, F. C. Lightstone, Y. J. Zheng, and T. C. Bruice. Nonenzymatic and Enzymatic Hydrolysis of Alkyl Halides: A Theoretical Study of the S_N2 Reactions of Acetate and Hydroxide Ions with Alkyl Chlorides. *Proc. Natl. Acad. Sci. U.S.A.*, 94:6591–6595, 1997.

[162] M. Prokop, J. Damborský, and J. Koča. TRITON: *In Silico* Construction of Protein Mutants and Prediction of Their Activities. *Bioinformatics*, 16:845–846, 2000.

[163] Y. Nagata, T. Nariya, R. Ohtomo, M. Fukuda, K. Yano, and M. Takagi. Cloning and Sequencing of a Dehalogenase Gene Encoding an Enzyme with Hydrolase Activity Involved in the Degradation of γ-Hexachlorocyclohexane in *Pseudomonas paucimobilis*. *J. Bacteriol.*, 175:6403–6410, 1993.

[164] Anonymous. *Chem. Eng. News*, 71:38–46, 1993.

[165] F. von Wintzingerode, C. Schlotelburg, R. Hauck, W. Hegemann, and U. B. Gobel. Development of Primers for Amplifying Genes Encoding CprA- and PceA-Like Reductive Dehalogenases in Anaerobic Microbial Consortia, Dechlorinating Trichlorobenzene and 1,2-Dichloropropane. *FEMS Microbiol. Ecol.*, 35:189–196, 2001.

[166] I. Smatanová, Y. Nagata, A. L. Svensson, M. Takagi, and J. Marek. Crystallization and Preliminary X-ray Diffraction Analysis of Haloalkane Dehalogenase LinB from *Sphingomonas paucimobilis* UT26. *Acta Crystallogr. D*, 55:1231–1233, 1999.

[167] Z. Otwinowski and W. Minor. Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology: Macromolecular Crystallography, part A*, volume 276, pages 307–326. Academic Press, New York, 1997.

[168] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren. Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr. D*, 54:905–921, 1998.

[169] A. Perrakis, R. J. Morris, and V. S. Lamzin. Automated Protein Model Building Combined with Iterative Structure Refinement. *Nat. Struct. Biol.*, 6:458–463, 1999.

[170] T. A. Jones, J. Y. Zou, S. W. Cowan, and M. Kjeldgaard. Improved Methods for Building Protein Models in Electron Density Maps and the Location of Errors in These Models. *Acta Crystallogr. A*, 47:110–119, 1991.

[171] G. J. Kleywegt and T. A. Jones. Databases in Protein Crystallography. *Acta Crystallogr. D*, 54:1119–1131, 1998.

[172] Collaborative Computational Project (No. 4). *Acta Crystallogr. D*, 50: 760–763, 1994.

[173] G. J. Kleywegt. Validation of Protein Models from C-alpha Coordinates Alone. *J. Mol. Biol.*, 273:371–376, 1997.

[174] G. J. Kleywegt. Use of Non-Crystallographic Symmetry in Protein Structure Refinement. *Acta Crystallogr. D*, 52:842–857, 1996.

[175] D. S. Goodsell and A. J. Olson. Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins: Struct. Funct. Genet.*, 8: 195–202, 1990.

[176] J. Damborský, K. Manová, and M. Kutý. A Mechanistic Approach to Deriving Quantitative Structure-Biodegradability Relationships. A Case Study: Dehalogenation of Haloaliphatic Compounds. In *Biodegradability Prediction*, pages 75–92. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.

Part III

Appendixes

Chapter 12

Curriculum vitae

Person Identification

Name and surname	Jan Kmuníček
Title	MSc.
Date of birth	5 September 1975
Place of birth	Brno, Czech Republic
Nationality	Czech
Marital status	Single

Affiliation

Address	National Center for Biomolecular Research Masaryk University, Kotlářská 2 611 37 Brno, Czech Republic
Phone	+420-5-41129 588
Fax	+420-5-41129 506
E-mail	smeagol@chemi.muni.cz

Permanent Residence

Address	Táborské návrší 444 664 01 Bílovice nad Svitavou, Czech Republic

Education and Academic Qualifications

1997 Bc.	Department of Physical Chemistry Masaryk University, Brno, Czech Republic
1999 MSc.	Department of Physical Chemistry Masaryk University, Brno, Czech Republic

Molecular Modelling Skills

Molecular mechanics	AMBER
Quantum mechanics	MOPAC, GAUSSIAN semiempirical, ab initio, DFT methods
Molecular docking	AUTODOCK
Multivariate statistics	GOLPE, SIMCA
Molecular graphics	GAUSSVIEW, GOPENMOL, INSIGHT II MOIL-VIEW, MOLDEN, MOLEKEL MOLMOL, MOLSCRIPT, NACCESS PYMOL, RASMOL, SPARTAN, TRITON VMD, XLEAP, XMMOL

Foreign Sojourns

08–10/1999	EMBL, Heidelberg, Germany	(R. C. Wade group)
04–05/2000	EMBL, Heidelberg, Germany	(R. C. Wade group)
09–12/2002	University of Alcalá, Spain	(F. Gago group)

Attended Courses

11/2000	Humboldt School on Structural Biology, Berlin, Germany
08/2001	Summer School of Theoretical and Computational Chemistry, Prague, Czech Republic

Current Research Interests

Topics	computational chemistry, molecular modeling, 3D QSAR comparative binding energy (COMBINE) analysis computer assisted protein design, protein–ligand interactions

Appreciation by Scientific Community

1997	Award of Chemistry Section of the Faculty of Science
1998	Second prize in Student Scientific Contest '98
1999	Third prize in Student Scientific Contest '99
2000	Honorable Mention Award in Student Poster Ceremony at Fifth International Symposium and Exhibition on Environmental Contamination in Central and Eastern Europe
2003	Award of Rector of Masaryk University

Fellowships

NATO Science Programme	Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenases

Chapter 13

List of publications

Kmuníček, J., Luengo, S., Gago, F., Ortiz, A. R., Wade, R. C., Damborský, J., Comparative Binding Energy (COMBINE) Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10. *Biochemistry* **40**: 8905–8917 (2001).

Oakley A. J., Prokop Z., Boháč M., **Kmuníček J.**, Jedlička, T., Monincová M., Kutá-Smatanová I., Nagata Y., Damborský J., Wilce M. C. J., Exploring the Structure and Activity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26: Evidence for Product and Water Mediated Inhibition. *Biochemistry* **41**: 4847–4855 (2002).

Kmuníček, J., Boháč, M., Luengo, S., Gago, F., Wade R. C., Damborský J., Comparative Binding Energy Analysis of Haloalkane Dehalogenase Substrates: Modelling of Enzyme–Substrate Complexes by Molecular Docking and Quantum Mechanic Calculations. *Journal of Computer-Aided Molecular Design* **17**: 299–311 (2003).

Damborský, J., **Kmuníček, J.**, Jedlička, T., Luengo, S., Gago, F., Ortiz, A. R., Wade, R. C., 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, pages 79–96 (2003).

Kmuníček, J., Jedlička, T., Gago, F., Wade, R. C., Damborský, J., Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Sphingomonas paucimobilis* UT26. In preparation (2003).

Chapter 14

List of presentations

Kmuníček, J., Damborský, J., Wade, R. C., Luengo, S., Gago, F., Ortiz, A. R., Computer Design of Haloalkane Dehalogenase Using Comparative Binding Energy (COMBINE), *Fourth Workshop of Biochemists and Molecular Biologists*, February 9, 2000, Brno, Czech Republic (lecture).

Damborský, J., Jesenská, A., Hynková, K., **Kmuníček, J.**, Boháč, M., Kutá-Smatanová, I., Marek, J., Koča, J., Screening and Construction of Environmental Biocatalysts, *Biodegradation IV*, March 8–9, 2000, Seč, Czech Republic (lecture, presented by JD).

Damborský, J., Hynková, K., Smatanová, I., Marek, J., Wimmerová, M., Kutý, M., **Kmuníček, J.**, Boháč, M., Prokop, M., Koča, J., Nagata, Y., Takagi, M., Computer-Assisted Engineering of Haloalkane Dehalogenases, *The World Congress on Biotechnology: Biotechnology 2000*, September 3–8, 2000, Berlin, Germany (poster, presented by JD).

Damborský, J., **Kmuníček, J.**, Boháč, M., Prokop, M., Koča, J., Computer Design of Dehalogenating Enzymes, *Seventeenth Biochemical Congress*, September 7–10, 2000, Prague, Czech Republic (lecture, presented by JD).

Kmuníček, J., Damborský, J., Wade, R., Gago, F., Luengo, S., Ortiz, A. R., Design of Haloalkane Dehalogenase Using Comparative Binding Energy (COMBINE) Analysis, *Fifth International Symposium and Exhibition on Environmental Contamination in Central and Eastern Europe*, September 12–14, 2000, Prague, Czech Republic (poster).

Kmuníček, J., Luengo, S., Gago, F., Ortiz, A. R., Wade, R., Damborský, J., Application of Comparative Binding Energy (COMBINE) Analysis for the Design of Environmental Biocatalysts, *Ninth International Workshop on Quantitative Structure Activity Relationships in Environmental Sciences (QSAR 2000)*, September 16–20, 2000, Bourgas, Bulgaria (lecture).

Damborský, J., **Kmuníček, J.**, Boháč, M., Prokop, M., Koča, J., Computer Modelling in Environmental Biotechnology, *Environmental Biotechnology in 21st Century*, December 1–2, 2000, Tokyo, Japan (lecture, presented by JD).

Kmuníček, J., Luengo, S., Gago, F., Ortiz, A. R., Wade R., Damborský, J., Prediction of Binding Affinities for Mutant Proteins: Influence of Scaling and Variable Selection on Quality of COMBINE Models, *Fifth Workshop of Biochemists and Molecular Biologists*, February 14, 2001, Brno, Czech Republic (lecture).

Damborský, J., Hynková, K., Smatanová, I., Marek, J., Wimmerová, M., Kutý, M., **Kmuníček, J.**, Boháč, M., Prokop, M., Koča, J., Nagata, Y., Tagaki, M., Computer Design of Haloalkane Dehalogenases, *International School of Structural Biology and Magnetic Resonance*, June 5–15, 2001, Erice, Italy (poster, presented by JD).

Kmuníček, J., Boháč, M., Luengo, S., Gago, F., Ortiz, A. R., Wade, R. C., Damborský, J., Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10: Modelling of Enzyme–Substrate Complexes by Molecular Docking and Quantum-Mechanic Calculations, *Sixth Workshop of Biochemists and Molecular Biologists*, February 7, 2002, Brno, Czech Republic (lecture).

Damborský, J., Jesenská, A., Prokop, Z., **Kmuníček, J.**, Boháč, M., Prokop, M., Sýkorová, J., Chaloupková, R., Pavlová, M., Jedlička, T., Monincová, M., Strouhal, M., Budová, M., Koča, J., Otyepka, M., Nagata, Y., Wilce, M., Protein Engineering of Dehalogenating Enzymes, *First Meeting of Czech and Slovak Structural Biologists*, March 14–16, 2002, Nové Hrady, Czech Republic (lecture, presented by JD).

Kmuníček, J., Luengo, S., Gago, F., Ortiz, A. R., Wade, R., Damborský, J.: Comparative Binding Energy Analysis of the Substrate Specificity of Haloalkane Dehalogenase from *Xanthobacter autotrophicus* GJ10, *First Meeting of Czech and Slovak Structural Biologists*, March 14–16, 2002, Nové Hrady, Czech Republic (poster).

Kmuníček, J., Luengo, S., Gago, F., Ortiz, A., Wade, R. C., Damborský, J., Rational Re-Design of Enzymes Using Computational Tools, *SFC Eurochem 2002*, July 8–11, 2002, Toulouse, France (lecture).

Damborský, J., Jesenská, A., Prokop, Z., **Kmuníček, J.**, Boháč, M., Prokop, M., Sýkorová, J., Chaloupková, R., Pavlová, M., Jedlička, T., Monincová, M., Strouhal, M., Budová, M., Koča, J., Otyepka, M., Nagata Y., Wilce, M., Protein Engineering of Dehalogenating Enzymes, *Twelfth International Biodeterioration and Biodegradation Symposium*, July 14–18, 2002, Prague, Czech Republic (lecture, presented by JD).

Kmuníček, J., Boháč, M., Luengo, S., Gago, F., Wade, R. C., Damborský, J.: Modelling of Enzyme–Substrate Complexes for COMBINE Analysis of Haloalkane Dehalogenase by Molecular Docking and Quantum Mechanics, *Second Meeting of Czech and Slovak Structural Biologists*, March 13–15, 2003, Nové Hrady, Czech Republic (poster).

Damborský, J., Jesenská, A., Prokop, Z., **Kmuníček, J.**, Boháč, M., Chaloupková, R., Pavlová, M., Monincová, M., Strouhal, M., Jedlička, T., Těšínská, I., Klvaňa, M., Otyepka, M., Banáš, P., Nagata, Y., Oakley, A.: Computer-Assisted Engineering of Haloalkane Dehalogenases for Environmental Applications, *Sixth International Symposium on Biocatalysis and Biotransformations*, June 28–July 3, 2003, Olomouc, Czech Republic (poster, presented by JD).

Damborský, J., Jesenská, A., Prokop, Z., **Kmuníček, J.**, Boháč, M., Chaloupková, R., Pavlová, M., Monincová, M., Otyepka, M., Nagata, Y., Oakley, A.: Protein Engineering of Haloalkane Dehalogenases, *Advances in Organic, Bioorganic and Pharmaceutical Chemistry*, November 28–November 30, 2003, Liblice, Czech Republic (lecture, presented by JD).