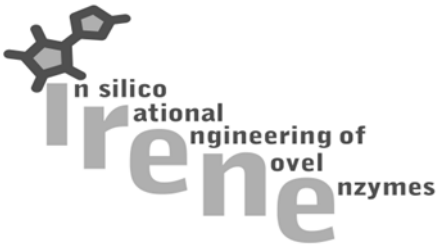
PROJECT FINAL REPORT



Grant Agreement number: *227279*

Project acronym: *IRENE*

Project title: *In silico Rational Engineering of Novel Enzymes*

Funding Scheme: *Collaborative project - Collaborative project for specific cooperation actions dedicated to international cooperation partner countries (SICA)*

Period covered: **from 1/04/2009** **to 31/03/2012**

Name of the scientific representative of the project's co-ordinator, Title and Organisation:

Prof.ssa Lucia GARDOSSI, Università degli Studi di Trieste - Italy

Tel: *+39 0405583103*

Fax: *+39 04052572*

E-mail: *gardossi@units.it*

Project website address: ***http://www.irene-fp7.eu/***

4.1 Final publishable summary report

4.1.1 Executive summary

The IRENE project is a SICA project for specific cooperation between EU and Russia and is financially supported by EC and Russian Federal Agency for Science and Innovation (FASI). IRENE project has developed computational methods and strategies which were applied to rationally design and produce biocatalysts endowed with new promiscuous properties functional to industrial applications.

IRENE aim: expanding the use of sustainable bio-catalysts in industry
Enzymes are increasingly used to perform a range of chemical reactions. These catalysts from nature are sustainable, selective and efficient, and offer a variety of benefits such as environmentally friendly manufacturing processes, reduced use of solvents, lower energy requirement, high atom efficiency and reduced cost. However, natural biocatalysts are often not optimally suited for industrial applications. To boost the use of enzymes in industrial processes, it is important to expand the range of reactions catalyzed by enzymes and to improve their properties for industrial applications.

IRENE's targets: efficient promiscuous enzymes for industry
The project led to computational methods to re-design rationally and produce promiscuous enzymes, namely able to catalyze new reactions or accept different substrates

What are the solutions provided by IRENE?

1. Software:
   - MUTATE Software program for fast and massive modeling and screening of mutant libraries
   - PROPKA 3.1, to predict of pH-dependent properties of proteins and protein-ligand complexes.
   - EFMO GAMESS module, to be used to estimate barrier heights
   - FragIt, (graphical) user interface that allows for easy setup of EFMO calculations. .
   - Scripts for fast estimation of barrier heights, in the rational design of enzyme catalysts.
2. Computational methods:
   - simulation, visualization and conceptualization for description of enzyme promiscuity
   - "ZEBRA" Bioinformatic analysis of subfamily specific positions and correlations between specific domains
   - Hybrid methods combining statistics and molecular simulation (GRID-PCA; 3D-QSAR)) for tri-dimensional description and prediction of enzyme properties
   - In-silico methodologies for integrating software and modeling tasks for automatic *in silico* mutants design and screening
3. New engineered enzymes endowed with industrially relevant properties:
   - Amidase activity engineered into a robust lipase scaffold (**reaction promiscuity**)
   - Lipase able to catalyze polycondensation of lactide (**substrate promiscuity**)
   - Penicillin amidase with improved synthesis/hydrolysis ratio in beta-lactam antibiotic synthesis (**substrate promiscuity**)
   - Glycoside hydrolases able to synthesize specific oligosaccharides of biological relevance and to enlarge their applicability in food chemistry (**substrate promiscuity**)
   - Enteropeptidase with enhanced selectivity and minimal non-specific cleavage (**substrate promiscuity**)
   - Penicillin amidase with improved enantiospecificity towards specific chiral amino compounds **(re-designed enantioselectivity)**

What will be the impact of IRENE project at economic and social level?
At present European companies supply about 70% of the world enzymes. To maintain and strengthen the European leadership in the biocatalysis sector, different routes must be pursued to make enzymes readily available for practical applications, thus making biocatalysis competitive as compared to conventional organic chemistry. Moreover the output of IRENE project will contribute to European policies for the construction of a "Knowledge based" society.

4.1.2: Project context and objectives

The IRENE project: http://www.irene-fp7.eu

IRENE consortium gathers a multidisciplinary group of European, Russian and Uzbek scientists with complementary expertise, covering all the major aspects of this front line research in a novel field of biocatalysis.

Beneficiary Number *	Beneficiary name	Beneficiary short name	Country
1 (coordinator)	UNIVERSITA DEGLI STUDI DI TRIESTE	UNITS	IT
2	KUNGLIGA TEKNISKA HOEGSKOLAN	KTH	SE
3	Københavns Universitet	UCPH	DK
4	TECHNISCHE UNIVERSITEIT DELFT	TUD	NL
5	NOVOZYMES A/S	NZ	DK
6	The National University of Uzbekistan named after Mirzo Ulugbek	NUU	UZ
7	Belozersky Institute of Physicochemical Biology	MSU	RU
8	B.P.Konstantinov PETERSBURG NUCLEAR PHYSICS INSTITUTE RUSSIAN ACADEMY OF SCIENCES	PNPI	RU
9	SHEMYAKIN AND OVCHINNIKOV INSTITUTE OF BIOORGANIC CHEMISTRY - RUSSIAN ACADEMY OF SCIENCE	IBCH	RU
10	Molecular Technologies	MLT	RU
11	Bio/ Techologies, Innovations, Researches LTD	BIOTIR	RU

The problem addressed by IRENE project

Industrial manufacturing of chemical intermediates, polymer building blocks, agrochemicals, and pharmaceuticals is on the brink of a paradigm shift, both in terms of use of resources and the nature of process technologies. First, there will be a shift from the use of petrochemical resources to bio-based platform chemicals, which are produced by fermentation from biomass components such starch or cellulose (second generation processes), or by bioprocessing of pre-treated biomass. Because of synergy in processes leading to biofuels, platform chemicals, and precursors for fine chemicals (like bioactive compounds), this **switch to sustainable resources** is expected to influence the whole chemistry production chain (market pull). Second, the **ongoing revolution in life sciences** will have a huge expanding effect on the possibilities to develop sustainable, environment-friendly, energy saving, clean bioprocesses (technology push):
Biocatalysts represent one of the tools that Industrial Biotechnologies provide to the new sustainable chemistry. Besides being highly efficient, selective and sustainable catalysts, enzymes enable to utilize **renewable resources** while improving the **environmental sustainability** of productive processes as well as the safety aspects associated to workers' health in industry.
Traditionally, in the past, new enzymes for desired reactions were obtained by tedious and time consuming screening of microbial cultures, often following enrichment and isolation of new cultures. Due to the genomics revolution, **massive sequencing combined with appropriate use of databases and efficient predictive bioinformatics tools have the potential to replace the current laborious screening** approaches. The technological advances in the field offer an array of tools, which nowadays still have to express their full applicative potential. As a matter of fact,

time-consuming, expensive, investment-intensive screening in the laboratory is expected to be replaced by *in silico* screening using computer programs, ranking, design and automated DNA synthesis, thus allowing a much shorter time from process idea to feasibility judgment with considerable savings on research costs and thereby contributing to a competitive European manufacturing sector.
To fully exploit the enormous developments in life sciences, **technologies and information must be used according to more effective and integrated strategies,** so that designing, developing and applying new and better enzymes for industrial processes become a faster and more effective practice. The achievement of this goal is of crucial importance for the technological and economic competitiveness of industrial biotechnological processes. Furthermore, better or new enzymes are needed, in some cases catalyzing still unknown reactions.
The major **bottlenecks** for replacement of conventional synthetic routes with biocatalysts in industry and to the full exploitation of enzymes' catalytic potential can be identified in the following points:

- requirement of multidisciplinary expertise for the implementation of the whole process, which cannot be afforded by most part of small and medium size enterprises
- longer time for new process development
- the still highly empirical nature of catalyst selection.

The concept: how IRENE has approached the problem

IRENE has addressed the previously mentioned **bottlenecks** by the convergence of different expertise for developing computational methods and strategies for rationally design and produce biocatalysts for industrial applications.
Due to the strong interaction between theoretical groups and experimentalists all computational tools used in this project were validated by experiments. Failures and successes were used for methods' evaluation, correction, tuning, comparison and combination, in an iterative process that led to the development of new methods and strategies, but also to the definition of practical guidelines, for specific enzyme design issues.

The targets of IRENE project: engineered promiscuous enzymes

IRENE project has developed computational methods that were used for re-designing rationally promiscuous enzymes, namely able to catalyze new reactions or accept different substrates
The experimental and computational work was focused on different specific enzymes, which are expected to have strong impact on the chemical, pharmaceutical and food industry, as well as in the applications to bio-transformations and biomass conversions.
From the conceptual point of view, IRENE project had three major **design subjects**:

- the introduction of new activities in specific enzyme scaffolds (**reaction promiscuity**),
- the improvement of catalytic activity towards specific targets (**substrate promiscuity**)
- the redesign of **enantioselectivity**.

The work on each different design subject involved more than one enzyme and proceeded in parallel.

4.1.3: Main S&T results/foreground

The IRENE project produced **conceptual advances in the understanding** of enzyme function and properties, which will be of aid in the full exploitation of catalytic potential of enzymes.
An **advanced platform of software and computational methods** was produced, dedicated to end-users who need to design and screen effective bio-catalysts for industrial applications.
The integration of experimental and computation studies allowed drawing **general guidelines,** which represent a state-of- the- art reference for effective *in silico* enzyme design and screening.
As a result, **new enzymes** were developed to address specific industrial needs.
The platform of **computational tools** developed by IRENE project is directed to scientists and industrial operators for being applied to:

The **new engineered enzymes** developed by IRENE project have been tailored to meet the following industrially relevant properties:

- Hydrolases enzymes endowed with amidase activity engineered into a robust lipase scaffold (**reaction promiscuity**)
- Lipase able to catalyze polycondensation of lactide (**substrate promiscuity**)
- Penicillin amidase with improved synthesis/hydrolysis ratio in beta-lactam antibiotic synthesis (**substrate promiscuity**)
- Glycoside hydrolases able to synthesize specific oligosaccharides of biological relevance and to enlarge their applicability in food chemistry (**substrate promiscuity**)
- Enteropeptidase with enhanced selectivity and minimal non-specific cleavage (**substrate promiscuity**)
- Penicillin amidase with improved enantiospecificity towards specific chiral amino compounds **(re-designed enantioselectivity)**

The **combined experimental and computational approach** was essential not only to create new biocatalysts but also to increase the understanding of the molecular basis of enzyme action. This will form the basis for further developments in the field of promiscuity and rational enzyme design.

a) Computational methods developed for rational enzyme engineering and function-activity correlations

The work done demonstrated how different computational methods can be applied or even integrated for achieving “in silico screening” and quantitative scoring of calculated acitivities/selectivities.
These methods have been validated and refined through the production of two generations of mutants that were experimentally characterized. Results, failures and successes induced the IRENE consortium to discuss and deliver guidelines for the rational design of enzymes. They are the output of the research task aiming at introducing amidase activity into lipase scaffolds, namely engineering “reaction promiscuity”. This task has been the most challenging within the IRENE project, since new catalytic machinery had to be design within an enzyme active site, for the stabilization of a different transition state. The task was coordinated by Partner NZ, who integrated the efforts and the different methodological approaches of a large extent of IRENE’s Partners. The following scheme summarizes the conceptual results.

GUIDELINES FOR THE RATIONAL DESIGN OF ENZYMES

The general strategy relies on three conceptual and methodological steps, necessary to model the activity of the enzyme, which can be summarized as follows:

1. Defining the transition state (QM) or a model of the transition state (e.g. TI)
2. Fitting the transition state modeled into the enzyme structure
3. Then optimizing enzyme-substrate interactions for near-attack complex

From the methodological point of view, the modeling of the activity of the enzyme/mutant is the result of three basic simulation methods:

Method	**Procedure**	**output**
Docking	+/- guidance	variants with lowest energy conformers
QM	with guidance	variants with lowest energy barrier
MD	no guidance- except guided starting point	variants with best orientation and interactions of crucial atoms

On that basis, activity of the enzyme/mutant can be defined as the function of:

Activity = f(docking) + f(QM) + f(MD)

Where:

f(docking) = f[(best fit of the substrate into the enzyme active site; the substrate is free to move generating different conformers for the interaction energy evaluation (ES). The best generated poses are evaluated by a subsequent Molecular Dynamic (MD) simulation of the ES complex for analysing its stability and the correct orientation for the near-attack complex)]

f(QM) = f[(energetic barrier given the path from docked form of near attack complex to transition state (TS). The energetic profile of the path is computed by QM calculation)]

f(MD)= f[(The TS or its model (TI) is subjected to MD simulation in order to evaluate the orientation of key residues and atoms and their possible contacts; the correct stabilisation of the TS is estimated. The substrate specificity should be accounted by MD simulation: how to keep the substrate in place to have a chance of orienting the key atoms correctly)]

The computational methods were developed in parallel by Partners, often addressing the same engineering target. A platform of alternative and/or complementary methods is now available and they can be summarized as follows:

1 Molecular simulation, visualization and conceptualization
2 Conceptualization based on extraction of molecular descriptors
3 Energy based design and screening
4 Bioinformatics analysis to identify relevant motifs and correlations between specific domains
5 Multivariate statistical analysis and 3D QSAR

6 Automatic mutants generation based on “multiobjective optimization softwares” able to integrate all methods listed above to construct a high-throughput scheme (under development).

More details on the specific computational methodologies and their potential are described in the folloeing section.

a.1 PM6-based methodology for *in silico* designing and screening
(developed by UCPH and validated on the basis of mutants produced by NZ)

- Work out the reaction mechanism for the wild-type with PM6 and confirm at a higher level of theory with the developed EFMO method. This could be done in 2-3 weeks with 200-300 cores (the bulk of the time is the EFMO calculations).
- Using PM6 perform all possible single mutations within a certain distance to the active site. Each mutant can be evaluated in 24 hrs using ca 5 cores. 19 mutants at 50 different positions were tested using 200 cores, this would take ca 25 days. Then take the top (lowest barrier) 25 single mutants and test all possible pairs. This will take about a week on 200 cores. Then take the top 5 single and double mutants and test them with EFMO.
- Time scale: about 1 week per mutant using 200 cores.

In order to implement the methodology, the following software/tools were developed/optimized by UCPH Partner:
a.1.1 EFMO GAMESS module, http://www.msg.chem.iastate.edu/GAMESS/
The EFMO method is a QM/MM method that can be used to estimate barrier heights (and hence activity) of enzymatic reactions quite accurately. The calculations can be set up with little human intervention using the FragIt GUI (see below) and is therefore suitable for use in industry. The EFMO method is implemented as a module in the GAMESS-US program, which is distributed free of charge to the scientific community (including industry). Due to its ease of use and lack of empirical parameters we expect the EFMO/FragIt method to become the QM/MM method of choice for industry.
The EFMO method is a polarizable force-field where the parameters are derived on-the-fly from quantum mechanics. For chemically relevant parts of a system, quantum mechanics is used and less important parts are treated classically. The goal of EFMO is to be a next-generation polarizable force-field. Input files can be easily generated using the FragIt tool.
a.1.2 FragIt, www.fragit.org
FragIt is a (graphical) user interface that allows for easy setup of EFMO calculations. FragIt is released under and open source license and is also available as a web service. Due to its ease of use and lack of empirical parameters we expect the EFMO/FragIt method to become the QM/MM method of choice for industry.
The aim of the FragIt tool is to generate input files for the FMO and EFMO method (see below) in GAMESS. The fragmentation is done by the use of SMILES and SMARTS to search for substructures in proteins. The software itself will be released as open source as soon as the paper is submitted. The source code can be obtained through links on the web page when that happens.
a.1.3 PROPKA 3.1, http://propka.org
PROPKA can predict of pH-dependent properties of proteins and protein-ligand complexes. PROPKA can be used by anyone to rapidly predict pH-dependent properties such as acidity of functional groups, and protein charge and stability. PROPKA is released on an open source license and is freely available to anyone. The code has been downloaded more than 300 times and the web interface receives about 2000 page views per month.
a.1.4 Scripts for fast estimation of barrier heights, https://github.com/mzhKU/Enzyme-Screening
UCPH has developed a fast high throughput method for estimating barrier heights. The method relies on the Mopac2009 and PyMol software packages, which are interface with a series of scripts. These scripts are available under and open source license. Our method is currently the only way to rapidly estimate barrier heights so we expect it to be used by anyone interested in the rational design of enzyme catalysts.

All software developed at Copenhagen University under the IRENE project and used in the studies has been deposited at the below URL: https://github.com/mzhKU/Enzyme-Screening
The scripts are written either for the Bash environment or for the Python programming language. One script is written for the PyMOL program. A very brief description of each script is given in the "README" file which is found at the above adress. In addition, detailed usage and instructions for all scripts are provided in the file "tutorial_screening.pdf", which is available at the same location. All software is completely open source and free to use

a.2 Bioinformatic analysis approach for hot spot identification
(developed by MSU, making use of software by MLT and validated on the basis of mutants produced by MSU, NZ, TUD, PNPI, IBCH and characterized by BIOTIR)
Enzymes within single family share a common function but differ in more specific properties and can be divided into subfamilies with different specificity, enantioselectivity, stability, etc. Consequently it would be important to identify positions that demonstrate specific variations inside enzyme family – in other words, are conserved only within subfamilies, but different between families. For those cases we suggest using a term “subfamily-specific position(s)” or SSP(s) to outline that distribution of amino acid types in those columns is specific to functional subfamilies. A novel scoring function is suggested to consider both physicochemical and alphabetical conservation of functional subfamilies. Structural information is used to reward scores of SSPs for neighboring other subfamily-specific and conserved positions. Bernoulli statistics is used to rank the results. Algorithm does not require pre-defined subfamilies as classification is proposed automatically by graph-based clustering. Large-scale evaluation of the method showed high accuracy of ranking significant SSPs. Subfamily specific positions are evolutionary flexible and variable in nature to improve or change enzymatic functions and seem to be a good explanation for protein evolution. Switch of amino acid states in specific positions between mechanistically different subfamilies may be used to produce a functionally promiscuous intermediate while introduction of new amino acid types may invoke a novel catalytic effect. Thus, new bioinformatic analysis methodology has been suggested to identify hotspot for rational design together with the exact mutation to introduce new function. Subfamily-specific positions can be used to limit the number of changes during directed evolution and random mutagenesis experiments but also to search for new catalytic properties.
Web-based Java application Zebra, documentation and benchmark datasets are freely available for non-commercial use at http://biokinet.belozersky.msu.ru/zebra.
Details on time scale and performance of the method are available from the Table here below.

Method	**Availability**	**Time scale estimation**	**Open source**
Bioinformatic analysis of subfamily specific positions	http://biokinet.belozersky.msu.ru/zebra Free for non-commercial use	1-60 minutes for 100-10000 sequences	No
NAMD molecular dynamics simulation	http://www.ks.uiuc.edu/Research/namd/ Free	200 minutes / ns (~80 000 atoms)	Yes
Autodock molecular docking	http://autodock.scripps.edu/ Free	20-50 minutes for 100 dockings of a substrate with 2-10 degrees of freedom	Yes
Mutate	http://www.moltech.ru/ Commercial	300 mutants per minute	No

a.2.1 *software for highly accurate QM/MM calculations*
MSU partner has also developed a software for highly accurate QM/MM calculations: flexible effective fragment potential QM/MM method. The initial point was the QM/MM computer program is based on the combination of the PC GAMESS quantum chemistry package (the INTEL-specific

version of the GAMESS(US) code) and the molecular modeling program TINKER. An essential feature of this QM/MM approach is the effective fragment potential (EFP) representation of the molecular mechanical part. The peptide chains of the protein matrices assigned to the MM subsystem are partitioned into small effective fragments whose contributions to the QM Hamiltonian are taken into account. In turn, interaction between effective fragments is modeled with the conventional force field parameters. Implementation of such procedure to the QM/MM version based on the flexible EFP allows one to perform accurate and fast calculations of energy profiles of the enzymatic reactions. Secondly, libraries of EFP parameters consistent with the density functional theory approximations in the QM subsystems should be created.

a.3 *Optimization of mutant libraries by automatic integration of modelling and statistical analysis (Developed by UNITS, exploiting structural models by UCPH and KTH, validated on the basis of mutants produced by NZ)*

This task had the objective of exploring novel strategies for integrating concepts, software, modelling and statistics methods for faster design and screening of enzymes. Recent advances in computational sciences have led to novel sophisticated and refined methods that are able to describe the biocatalyst machinery in detail. Solutions of research problems in molecular modeling of enzymes can be found within different time frames and accuracy levels, which will depend on the computational techniques used. No single computational method can be considered as a comprehensive, quantitative tool for enzyme engineering but rather each of the four main families of computational methods, QM (Quantum Mechanics), MM (Molecular Mechanics), QSAR (Quantitative Structure Activity Relationships) and Bioinformatics should be exploited for their inherent potential.

The approach uses in parallel two methodologies, which combine modeling methods (producing chemical/structural descriptors) and statistics methods. The resulting tools we used to "monitor" how mutations affect the enzyme properties at two different levels:

1) Residues close to "catalytic machinery" : by correlating structure and activity by 3DQSAR statistical model
2) Active site "environment" for correct placing of substrate and TS stabilization: by analyzing statistically the structural/chemical differences of the enzyme active sites by Principal Component Analysis -PCA
3) Then the two methodologies are integrated in an "authomatic work-flow" for in silico generation of virtual mutants using the 3DQSAR model and the PCA as scoring functions for selecting best virtual mutants

<u>1) Optimizing mutagenesis strategy (second generation of mutants) at residues close to catalytic machinery</u>

a) Construct a 3DQSAR model on the basis of experimental data coming from the first generation of mutants, thus correlating mutant structure (chemical descriptors) and their activity
b) Learrn from 3DQSAR model how mutations affected the "catalytic efficiency"
c) Refine mutagenesis strategies on the basis of statistic analysis

<u>2) Optimizing active site "environment"</u>

a) Describe the chemical nature of enzymes endowed with catalytic properties/specificity you want to confer to your mutants and the properties of enzyme you use as a scaffold for mutagenesis by studying chemical interactions and then extract molecular descriptors (i.e. chemical descriptors=molecular interaction fields)
b) Analyze statistically (PCA) the descriptors, thus clustering the different enzyme classes on the basis of their chemical properties (ability to establish interactions which will be able to accept, orient, hold the substrate with the optimum geometry and thus stabilize the rate determining step)
c) Project in the PCA space the mutants obtained or *in silico* predicted in step 1 (3DQSAR model) and check whether "active site environment" responds to properties of the family of enzymes considered as target

d) Refine active-site properties on the basis of projections on PCA clusters by automatic *in silico* evolution of virtual enzymes (step 3).

3)Integration of steps 1 and 2 in an automatic work flow for *in silico* evolution of libraries of virtual mutants.

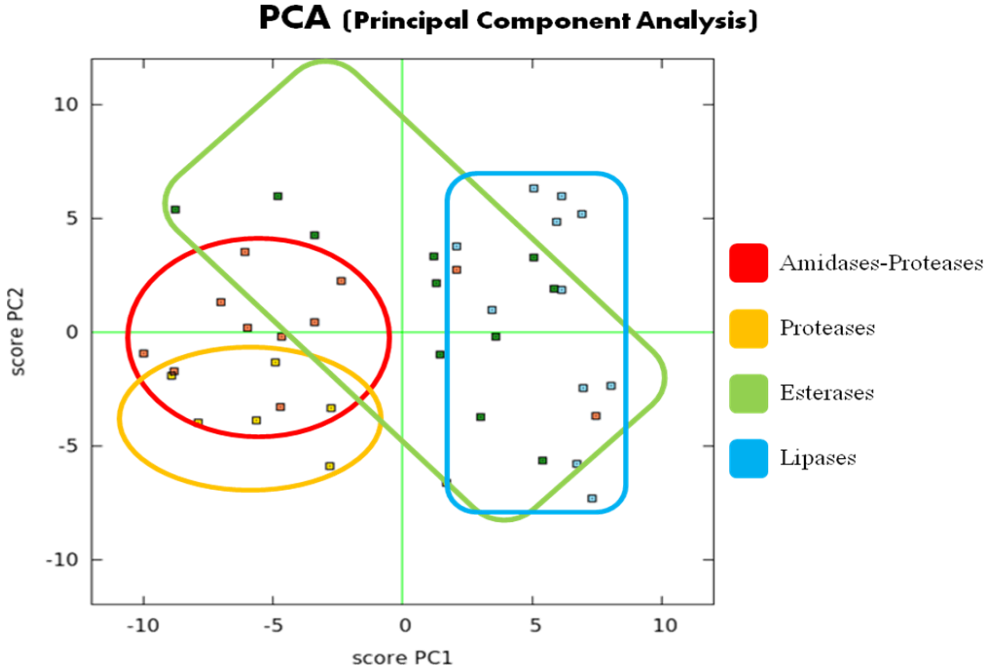
a) Construct the frame of the work-flow for integrating the software and computing actions (modelling + statistics)
b) automatic in silico generation of mutant structures
c) structure relaxation (MD)
d) automatic calculation of descriptors
e) computation of mutants within the 3DQSAR model
f) Projection of selected mutants inside the PCA space
g) Automatic screening of mutants by using 3DQSAR model and PCA clusters as scoring functions
h) Automatic mutant evolution until meeting scoring criteria

Computational resources required by each optimization step

Computational step	Time scale	Software	Open source ?
Generation of the mutant structure	1 hour on a 4 core processor	PyMOL or SCAP	yes
Mutant structure equilibration by Molecular Dynamic simulation	1 week on a 4 core processor	GROMACS	yes
Structure superimposition	1 hour on a 4 core processor	PyMOL	yes
Molecular descriptor calculation by GRID mapping	5 hour on a 4 core processor	GRID	GRID is a commercial software
Construct 3DQSAR model correlating mutant structure and their activity and variable analysis	10 days on 4 core processor	GOLPE	GOLPE is commercial software but similar open source alternatives is the R package (http://www.r-project.org/)
Molecular descriptor calculation by VolSurf	5 hour on a 4 core processor	VolSurf	VolSurf is a commercial software
Construction of the Principal Component Analysis model and variable analysis	1 week on 4 core processor	VolSurf	VolSurf is a commercial software
Automatization of the procedures and integration of the different softwares	10 days on any computer	modeFRONTIER	modeFRONTIER is a commercial software, but a valid open source alternative is represented by KNIME (http://www.knime.org/)
In-silico design and screening	From 6 to 8 mutants can be evaluated within a day	modeFRONTIER and all the integrated software	As above

In order to implement the strategy described above, Partner UNITS has developed the following computational methodologies:
a.3.1 *Screening of mutants on the basis of the chemical properties of active sites: Application of Principal Component Analysis (PCA) to the classification of active sites of different families of hydrolases* (UNITS)

A innovative computational method was developed aiming at describing and analyzing the chemical properties of active sites of enzymes. The method is based on the statistical analysis (PCA) of molecular descriptors. The clustering of different families of enzymes is based not on the primary sequence but on the properties of the "microenvironment" that surrounds the catalytic machinery. That means that the method is able to recognize and "fish out" enzymes that provide an environment able to stabilize effectively the transition states (TS) of similar reactions. Therefore the method can be also applied for obtaining quantitative information on how a certain mutation has affected the "microenvironment" of the active site. By using a data set of 39 different hydrolases, a map of the "minimal active site environment" for the stabilization of the TS of amide hydrolysis was done by using the "volsurf" descriptors (http://chemiome.chm.unipg.it/volsurf.html).
The aim was the identification of differences that might account for the fact that different classes of serin hydrolases catalyze different reaction, although they present the same catalytic triad. The active sites of enzymes sub-families were described in terms of spatial position in a multidimensional space that characterizes the environment able to stabilize intermediates and transition states. The PCA analysis clearly identified the classes of hydrolases and grouped them into clusters.



The statistical analysis of variables pointed out the most relevant differences between the active sites of lipase and amidase/proteases family. The main difference resides in the ability to establish hydrogen bonds and being hydrated. Therefore the lipase active site provide an "environment" very hydrophobic and with a scarce tendency to bind water. This might also account for the fact that amidases/proteases are active only at high water activity values whereas lipases work efficiently even in nearly anhydrous environments.

a.3.2 Correlating mutants structure and their activity (3DQSAR) for In silico screening
Libraries of mutant and experimental data (hydrolytic activity) produced by Partner NZ were exploited by Partner UNITS for the construction of a mathematical model correlating the structure of the different mutants and their activities (3D-QSAR) and its validation. This was the first attempt to apply this methodology, largely employed in drug design, to design and screening of enzymes. The mathematical model confirmed the existence of a correlation between the introduced mutations and the observed variations of activity. The statistical analysis identified the variables (spatial positions corresponding to aminoacid residues) responsible of the observed effects.

a.3.3 *Solvent effect on lipase activity (UNITS in collaboration with NUU)*
Partner UNITS, in collaboration with NUU, has also addressed the problem of *in silico* prediction of the effect of reaction media on biocatalysts and biocatalyzed reactions. Effect of water on

conformation and catalytic activity of lipases was investigated *in silico*, confirming that diluted aqueous solutions are not the most suited media for handling lipases, enzymes evolved for acting in membranes or on hydrophobic phases. (*Adv. Synth. Catal*. 2011, *353*, 2466 – 2480).

a.3.4 Predicting solvent effect on thermodynamic of biocatalyzed reactions(UNITS in collaboration with MSU)

In collaboration with Partner MSU the "BESSICC" algorithm was developed (*Biotech. Bioeng., 2012, in press.* DOI: 10.1002/bit.24439), able to calculate the effect of medium composition on biocatalyzed reactions equilibrium. It is based on COSMO-RS calculation of activity coefficients of all the species in the reaction mixture and minimization of Gibbs free energy of the reaction.
Starting from one single experimental measurement of the equilibrium position for a given biocatalyzed reaction it can predict the yield of the reaction in any other solvent or solvent mixture.

a.3.5 Molecular descriptors for evaluation of entropic contribution of enantiodiscrimination (method developed by UNITS by analyzing mutants produced by KTH)

In order to generate the mathematical predictive model, a new molecular descriptor was developed able to account for entropy contribution. A 3DQSAR model was constructed able to correlate structure of substrates and enantioselectivity while taking into account entropy contribution

a.4 *Optimized methodology of high-performance in silico screening of mutant enzyme libraries by means of molecular docking for assessment of substrate specificity, stereoselectivity, synthetic efficiency*

(Developed by MLT, validated on the basis of mutants produced by MSU and characterized by BIOTIR)

The optimized methodology of high-performance *in silico* screening of mutant enzyme libraries includes the use of the following software tools, most of which have been developed within the IRENE project.

1. First, positions inside or nearby the enzyme active site are selected in order to screen for alternative amino acids conferring desired catalytic properties. Selection of such positions may be performed with the use of multiple sequence alignment analysis developed by MSU.
2. As a choice, the list of mutated residues can be selected from residues forming contacts with bound substrate.
3. Protein-ligand docking software (for example Lead software developed by MLT) can be used to model initial enzyme-substrate complex.
4. Particular amino acids to be placed at the point of mutagenesis can be selected from bioinformatics (MSU) or QSAR (UNITS) considerations.
5. As far as the list of potential amino acid substitutions at each position is formed, virtual screening of all possible mutants is performed with the use of Mutate software developed by MLT.

Mutate software has graphical user interface, which can be used by scientists (even without professional computational chemistry background) under Windows or Linux. The Mutate software exploits original graph-theoretical algorithm, called TSAR (thermodynamic sampling of amino acid residues), to sample the space of all potential mutant enzyme sequences (which grows exponentially to the number of chosen positions). TSAR models interaction of a bound substrate with mutated enzyme by using docking modules of the Lead software. Finally TSAR selects the best mutant sequences, which yield optimum interaction with the selected substrate. Depending on the particular task (improvement of substrate binding, enzyme activity, enzyme selectivity) different modeling approaches can be applied with Mutate. The overall rate of screening of mutant enzyme library is given on the Figure here below. It can be seen that while the number of mutant enzyme variants increases exponentially the CPU time needed for calculations grows much slower for the developed LEAD module for *in silico* enzyme engineering. On the contrary, other programs, such as well known Modeller software, need CPU time proportional to the number of mutant variants.

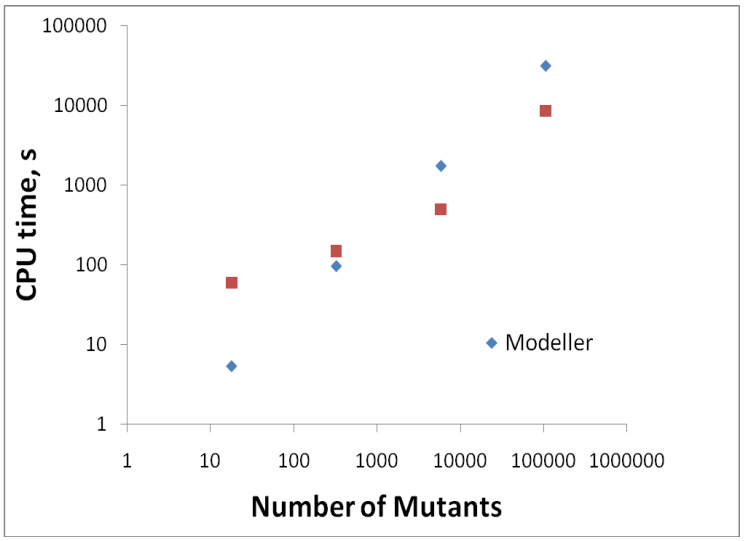


Figure 1. CPU time needed to screen the mutant library containing the given number of enzyme variants. As a comparison, software Modeller (blue rombs) requires time proportional to the number of mutant variants screened; the time required by LEAD module for in silico engineering (red squares) grows logarithmically to the number of mutants.

The MUTATE Software program has been developed by MLT and is commercially available. It works under Windows and Linux, to model structure and substrate specificity of mutant enzymes. Software is capable of screening large libraries of mutant enzymes at a speed of ~million mutant enzymes per processor per day
The main functionality of the Mutate software is the following:
-Modeling structure of side chains of a mutant protein (including reconstruction of all side chains) given a fixed protein backbone, or a structure of arbitrary library of mutant proteins;
-Modeling structure of mutant proteins (enzymes) with non-covalently or covalently bound ligands (substrates, intermediates, transition state analogues) and selection of mutants basing on the Lead's docking energy score;
The distinctive trait of the Mutate software is a combination of docking energy calculations, which allow fast molecular mechanics-based scoring and sampling of bound arbitrary ligands and protein side-chains, and innovative graph theoretic algorithm TSAR (Thermodynamic Sampling of Amino acid Residues) for selection of the optimal mutants and side chain configuration out of a library of mutants. TSAR algorithm allows dealing with astronomically large libraries of mutant libraries at a reasonable speed.

b) Engineered enzymes rationally designed and produced within the IRENE project.

As mentioned above, from the conceptual point of view, IRENE project had three major **design subjects**:

- the introduction of new activities in specific enzyme scaffolds (**reaction promiscuity**),
- the improvement of catalytic activity towards specific targets (**substrate promiscuity**)
- the redesign of **enantioselectivity**.
-

The work on each different design subject has involved more than one enzyme and proceeded in parallel.

The following list of **biocatalysts has been rationally designed and produced** inside the IRENE project because of their potential industrial impact:

1. amide forming enzymes with higher efficiency and different specificity as compared to the known proteases to be used in fine chemistry and in polymer chemistry (patented)
2. Lipases to be used in polycondensation of lactide, for the production of bio-based and compostable polymers (published)
3. amidases with increased synthetic efficiency and improved regioselectivity for more cost effective enzymatic synthesis of beta lactam antibiotics (patented)
4. glycoside hydrolases with enhanced synthetic efficiency for the production of glycoconjugates of relevance and to be used in the biotransformation of "reluctant" oligosaccharides in food industry (patented both enzyme and process)
5. enteropeptidases with higher activity and higher selectivity for biological applications (published)
6. Penicillin acylase mutants with improved enantioselectivity to be applied in specific enantio-resolutions and cascade reaction (patented)

The scientific activities of IRENE project have encountered technical and conceptual obstacles in the engineering of nitrilase activity into lipases and esterase activity into HNL enzymes. The research carried out so far has clearly indicated that the original objectives deserve to be addressed in a longer time-frame. Nevertheless, the successful work carried out in the parallel tasks has allowed developing and validating all the expected computational strategies necessary for investigating, explaining and predicting enzyme promiscuity. As a consequence, these tools will be ready for facing in the next future the toughest challenges.
The concepts, methodologies at the basis of the design of the new biocatalysts as well as their properties are described in the following paragraphs.

b.1 *Engineering amidase activity into lipase scaffold ((reaction promiscuity)*
The introduction of amidase functionalities into scaffolds of lipases well characterized and structurally known would allow the exploitation of their stability and catalytic efficiency, also in non aqueous media, Moreover, it was envisaged enzymes to develop enzymes able to accept a different range of acyl moieties as compared to peptidases, Last but not least, it is of fundamental interest to understand why enzymes having the same catalytic "machinery" (i.e. catalytic triad of Serin proteases) catalyze different reactions (i.e. lipase catalyze hydrolysis of esters but not of amides).
Reaction promiscuity is the most demanding form of promiscuity since it implies that the reaction mechanism has to be changed to stabilize a new transition state. This made this task the most challenging requiring a concerted and precisely tuned contribution of the largest number of Beneficiaries with their multiple expertises. A schematic representation of the integration of different Partners and methodologies is reported here below.

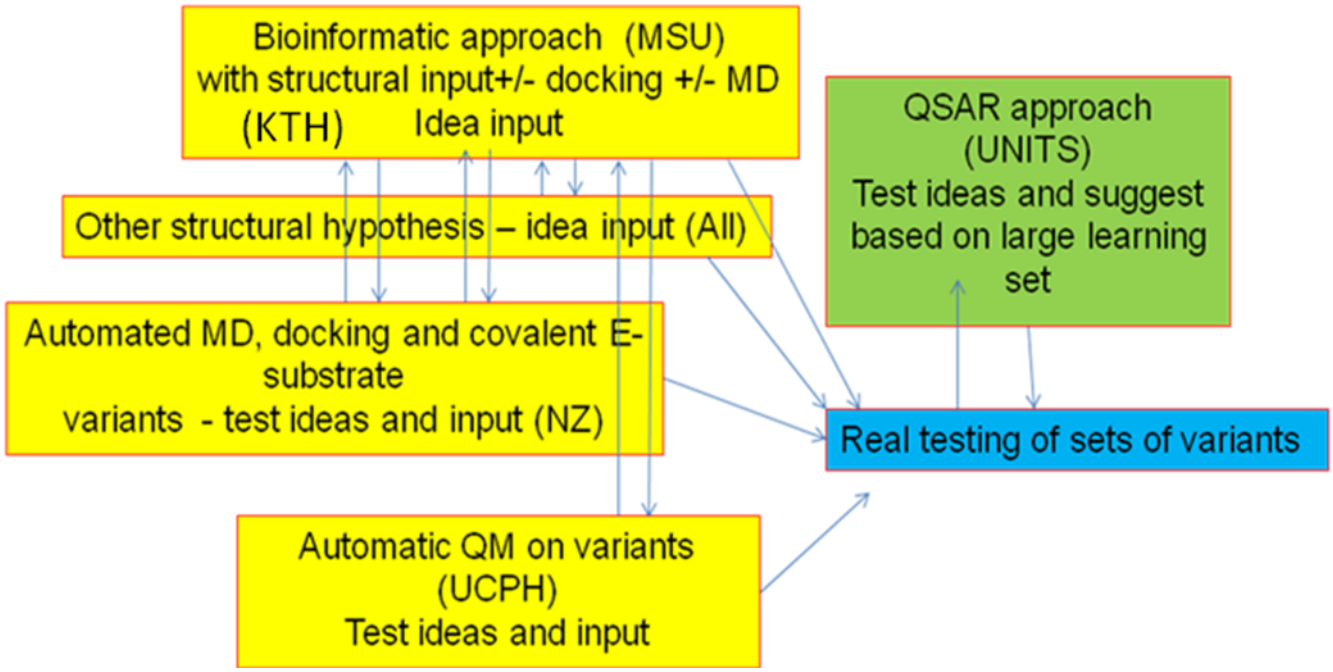
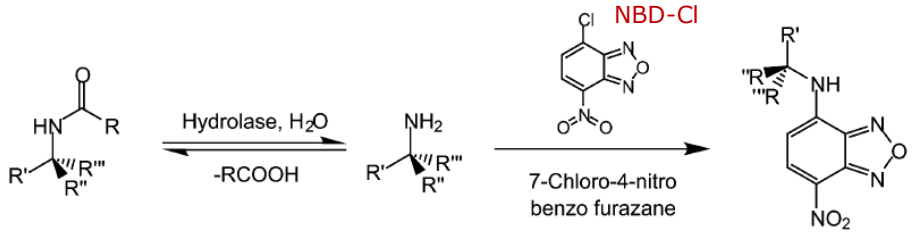


Figure 2: Schematic representation of the integration of different Partners and methodologies for the task of engineering amidase activity into lipase scaffold.

Amidase activity was engineered inside scaffold of lipase B from *Candida antarctica*. Two Partners (NZ, KTH) were involved in the production of mutants' libraries, which were evolved according to different concepts, approaches, computational methodologies. Moreover, different experimental assays were also used to achieve the objective. The mutants developed and produced by KTH and by NZ have been delivered to the Consortium as well as the necessary data for a comprehensive rational analysis of structure-function correlation (3DQSAR models). Experimental data support the conclusions and part of the results was already published. More publications will be submitted within six months from the closing of the IRENE project.

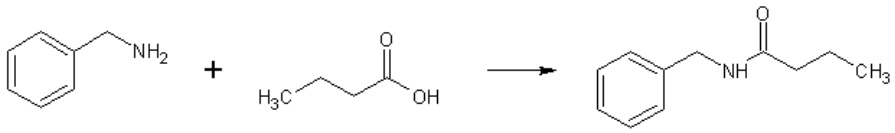
From the conceptual point of view, **Partner NZ** has addressed and investigated several enzyme substrate parameters. Even though the substrate binds nicely, the tetrahedral state (reaction intermediate) might orient incompletely for potential hydrolysis of amide bond. Parameters for an optimal placement of substrate being able to make the transition from enzyme substrate complex to transition state intermediate has been measured by UCPH. Provided the substrate is binding and orient correctly, provided that the bonding system makes the correct contacts and provided that the energetic of the reaction has low barrier energy the functionality can take place. All four measures need to be positive in order to secure a good functioning enzyme (Publication to be made). For the promiscuity of esterase to amidase function the specificity topic can be added to the above measures. The specificity being the correct binding and orientation of the substrate to make the correct bonds necessary for catalysis.

During the project, expression hosts and purification strategy were changed and optimized in order to improve efficiency of the enzyme engineering process. The best enzymes characterized so far show ca. 7 fold improved activity on substrate N-benzyl-2-chloroacteamide compared to wild type CalB. A patent application has been submitted (Polynucleotides encoding the variants; nucleic acid constructs, vectors, and host cells comprising the polynucleotides; methods of using the variants).



Scheme 1. A*ctivity assay used for characterization of amidase hydrolytic activity of CalB variants.*

A number of variants were also characterized for synthetic activity (UNITS Partner) in buffer medium to allow using soluble substrate while avoiding the need of enzyme immobilization.



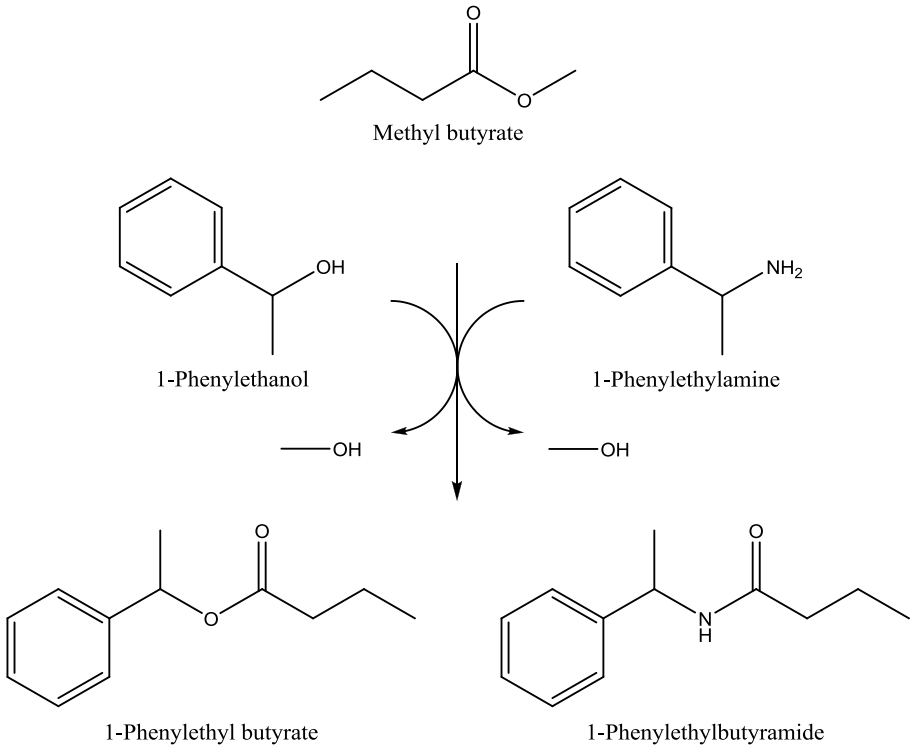
Scheme 2. A*ctivity assay used for characterization of amidase synthetic activity of CalB variants.*

The libraries of mutant produced by Partner NZ were exploited by Partner UNITS for the construction of a 3D-QSAR model that confirmed the existence of a correlation between the introduced mutations and the observed variations of activity.

The **design strategy of Partner KTH** was based on the output of an *ab initio* modelling of the reaction path of amide hydrolysis. It was demonstrated that the highest energy point was for inversion of the nitrogen of the amide bond. The calculated geometry of that transition state was by force field modelling studied in 16 proteases/amidases representing 11 different folding families and 10 different reaction mechanisms. In all these enzymes a hydrogen bond acceptor was found that could bind to the hydrogen of the nitrogen of the scissile amide bond. The position of the acceptor was such that it could facilitate nitrogen inversion and thereby reduce the energy of the highest energy point of the reaction path with up to 20 kJ/mol, which would correspond to a rate enhancement of up to 2500 times. The hydrogen bond acceptor was found either as a part of the enzyme, mostly as a back bone carbonyl oxygen, or in the substrate itself. A corresponding hydrogen bond possibility was not found in any of the studied esterases/lipases. From these results it is clear that the introduction of a hydrogen bond in transition state for nitrogen inversion is needed in the lipase structure to achieve a increase in amidase activity. This could be seen as a change in reaction mechanism as a new stabilizing interaction is present in transition state.

The strategy followed by KTH was based on the introduction of a stabilizing H-bond, as previously reported. Each mutant was characterized by titrating the active sites to have a reliable evaluation of enzyme concentration. Hydrolytic assays were carried on *p*-nitrophenyl butyrate and *p*-nitrophenylbutyramide. The best mutants display an amidase activity more than 7 folds higher than WT. In terms of amidase/estera activity, an improvement of 51 times was observed. Some of the mutants were also characterized for their synthetic activity, by working in toluene. The synthesis of 1-phenylethylbutyrate and phenylethylbutyramide were studied at the same time in the same reaction.

Ester and amide synthesis were also studied using CALB wt and selected variants that were the best ones found in hydrolysis of amides. Reactions that were catalyzed:



The synthesis was performed in dry methyl butyrate with immobilized enzyme. The results show that the selected mutants did not have higher synthetic rate than the wild-type enzyme. This is astonishing as one normally would expect the same relative behavior in the two reaction directions. One explanation could be that water is coordinated in the active site and is essential in the reaction mechanism. Under water free conditions such as during synthesis performed in methyl butyrate no water would be present and the reaction rate would be low. Two publications illustrate the whole design strategy and the conceptual basis (ChemCatChem, 5, 853-860, 2011; *ChemBioChem,* 13, 645-648, 2012.)

b.2 *Engineering Lipases to be used in polycondensation of lactide, (substrate promiscuity)*
The objective was the understanding molecular basis of how to promote substrate promiscuity, with particular emphasis on external nucleophiles in reactions catalysed by hydrolases. That would overcome obstacles that hamper applications of hydrolases in synthetic reactions. More specifically, lipases able to catalyze polylactate (polyesters) synthesis would be useful in the production of one of the most environmentally interesting compostable polymers, such as polylactic acid, other polyhydroxy alkanoic acids.
The kinetic and modeling study of Partner KTH demonstrated that the deacylation step in the propagation is rate limiting so that the rational design was based on making a larger space in both the acyl donor and the acyl acceptor pockets.

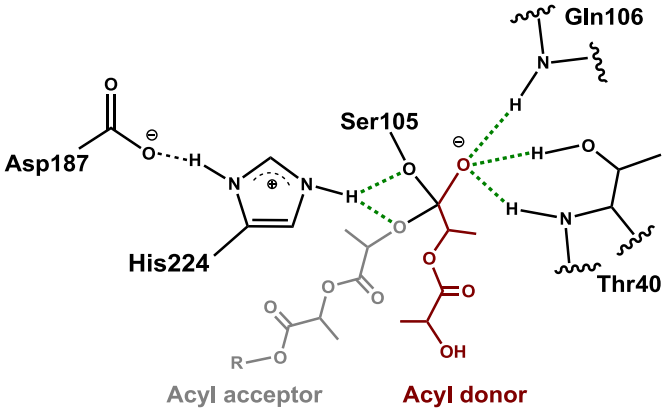


Figure 3: Model of the tetrahedral intermediate representing the propagation step where a D,D-lactide unit is the acyl donor and benzyl D,D-dilactate is the acyl acceptor.

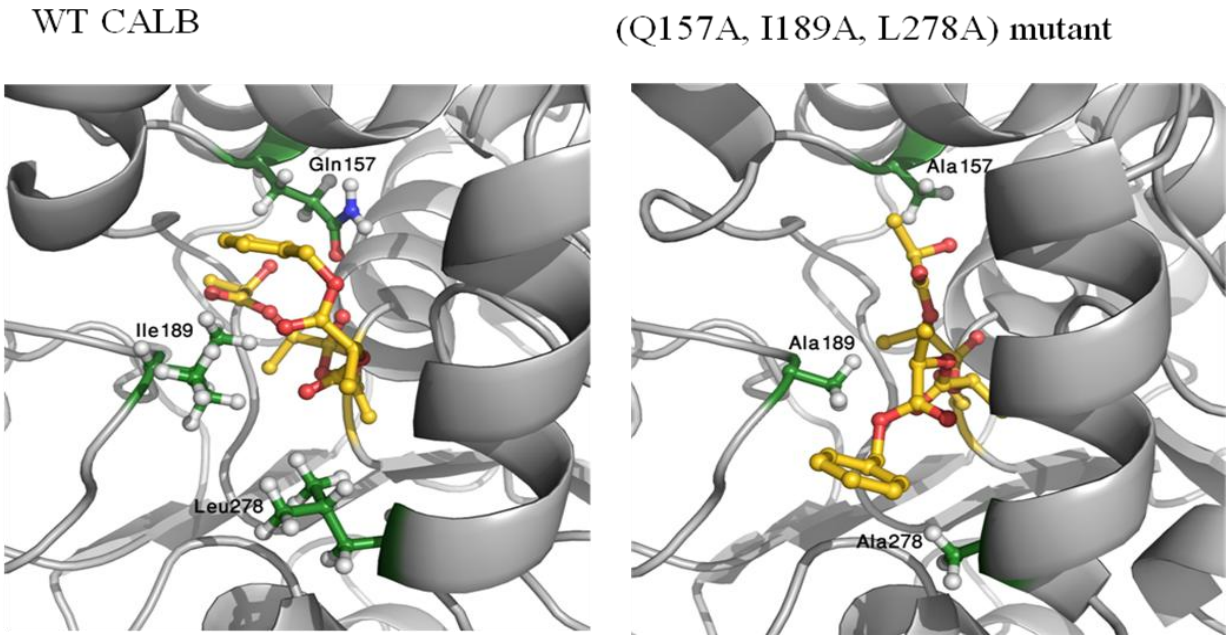


Figure 4: Molecular dynamics simulation of the tetrahedral intermediate

Site directed mutagenesis was carried out to produce 5 mutants: Gln157Ala, Ile189Ala, Leu278Ala, Ile189Ala, Leu278Ala, Gln157Ala, Ile189Ala, Leu278Ala, which were produced in *P. pastoris.* Initial screening of *Candida antarctica* lipase B mutants towards the ring opening polymerization of *D*,*D*-lactide revealed that mutants (Q157A) and (Q157A, I189A, L278A) had the highest improvements as compared to WT. These were chosen for a detailed kinetic study. Preparative ring opening polymerization of *D*,*D*-lactide were run using the WT and mutants (Q157A) and (Q157A, I189A, L278A). The reactions were allowed to run for 48 hours in D8-toluene at 60 ºC using 1-phenylethanol as initiator. The synthesized polymers were characterized by NMR, MALDITOF and SEC. The mutants (Q157A) and (Q157A, I189A, L278A) showed about 90-fold increase in activity as compared to the WT enzyme. The mutants greatly improved the rate and degree of polymerization in a preparative synthesis of poly(*D,D*-lactide).
The positive results obtained with the first generation of mutants have induced to evaluate the upper limit for possible kinetic improvements through mutagenesis. For that purpose, the kinetic investigation of tributyrin hydrolysis and ethyl octanoate transacylation was carried out on the WT and mutants with the aim of observing the maximum improvement achievable.
The results indicated that the first generation of mutants are already close to the maximum performance achievable through mutagenesis. The successful mutagenesis strategy has been reported in a publication. (ChemComm, 47, 7392–7394, 2011).

b.3 Amidases with increased synthetic efficiency and improved regioselectivity for more cost effective enzymatic synthesis of beta- lactam antibiotics (substrate promiscuity)
The objective of the present task was engineering the enzyme Penicillin acylase/amidase (PA) leading to higher synthesis/hydrolysis ratio in beta-lactam antibiotic synthesis to promote the shifting from chemical to enzymatic penicillin synthesis in industry. The objective has been achieved by Partner MSU by rationally designing PA mutants endowed with: a) increased nucleophile affinity and reactivity; b) decreased unproductive acyl donor hydrolysis; c) improved selectivity to make more cost effective enzymatic routes to semi-synthetic beta-lactam antibiotics.
The rational design of the mutants was based on the modelling of molecular mechanism of the enzyme action, by means of hybrid QM/MM methods (MSU, MLT) and by modelling nucleophile

binding by native enzyme and its mutant forms using molecular docking and molecular dynamics methods, also thanks to the novel software developed (MLT, MSU).
The MD simulation of acylenzyme-nucleophile complexes was used to characterize nucleophile reactivity by applying structural filters and collecting statistics of the so-called near attack conformations from the MD trajectories. Comparison of computationally modeled and experimentally measured nucleophile reactivity parameters confirmed validity of the suggested computational approach for evaluation of the synthetic potential and nucleophile specificity of wild type penicillin acylase and its mutants.
Seven penicillin acylase mutants (five single point mutants and two double mutants) have been designed and produced according to modelling indications. Thermodynamic and kinetic characterization of mutants was carried out by MSU and BIOTIR). The mutants possess higher catalytic activity in hydrolysis of specific colorimetric substrate and a mutation was identified as able to change properties of the acyl group binding subsite and improve enzyme specificity to D-phenylglycine derivatives, which are key acyl donors in enzymatic synthesis of the most important semisynthetic penicillins and cephalosporins such as ampicillin and cephalexin.
Synthetic properties of the mutants were studied in penicillin acylase-catalyzed ampicillin synthesis All designed mutants demonstrated higher synthesis/hydrolysis ratio compared to wild type enzyme.

b.4 Glycoside hydrolases with enhanced synthetic efficiency for the production of glycoconjugates of relevance and to be used in the biotransformation of "reluctant" oligosaccharides in food industry (substrate promiscuity)

The objective of the present task was the optimization of glycoside hydrolases for a) enhancing the synthetic efficiency of glycoside hydrolase in the production of glycoconjugates of relevance b) reducing the current cost of biologically important galactosaccharides in comparison with available synthetic methods; c) opening the perspectives of biotransformations of "reluctant" oligosaccharides which are relevant for the food industry. The aim was achieved by Partner PNPI by using molecular modeling as a guideline for tuning the desired orientation of glycosyl acceptor in the active of hydrolases.
Methods of rigid and flexible docking were used to find preferable orientations of oligosaccharides within the active center of the enzyme. This enables determination of the aminoacid residues involved into the 'enzyme-ligand' interaction. Methods of molecular mechanics were used for mutual orientation of glycosyl-enzyme intermediate and acceptor of glycosidic linkage (lactose). To select *in silico* mutants able to synthesize 1→4 glycosidic linkage preferably, two criteria were applied: (i) weakening of interactions and prevention of lactose binding in active conformations advantageous for formation of 1→3 and 1→6 linkages; (ii) strengthening of interactions contributing in binding of lactose in active conformations advantageous for formation of 1→4 linkage. As a result, array of the promising α-galactosidase mutants in order to alter regiospecificty of α-galactosidase in transglycosylation reaction was generated.
The predictions on the base of *in silico* modeling were used for introducing mutations in α-galactosidase from *Thermotoga maritima* which would lead to an increasing transglycosylation ability of the enzyme. A single mutation was shown to lead to the significant alteration of regiospecificity and to increasing of total yield of transglycosylation products. The corresponding mutant is able to synthesize α1,4PNP-digalactosides with 13-times higher yield in comparison with wild-type enzyme while the production of other region-isomers was markedly decreased.
The successfull approach in the synthesis of PNP1,4-galactosides was unfortunately invalid for the synthesis of globotriose where lactose was used as acceptor instead of PNPG. To solve this problem the inactivation of nucleophile catalytic residue of TmGalA was accomplished and □-galactosyl azide was used as a donor compound. On the basis of the previous findings a set of double mutants was developed and tested in the synthesis of globotriose. As assessed by NMR spectroscopy, reaction mixture produced by one of generated double mutants contains about 60% of globotriose.

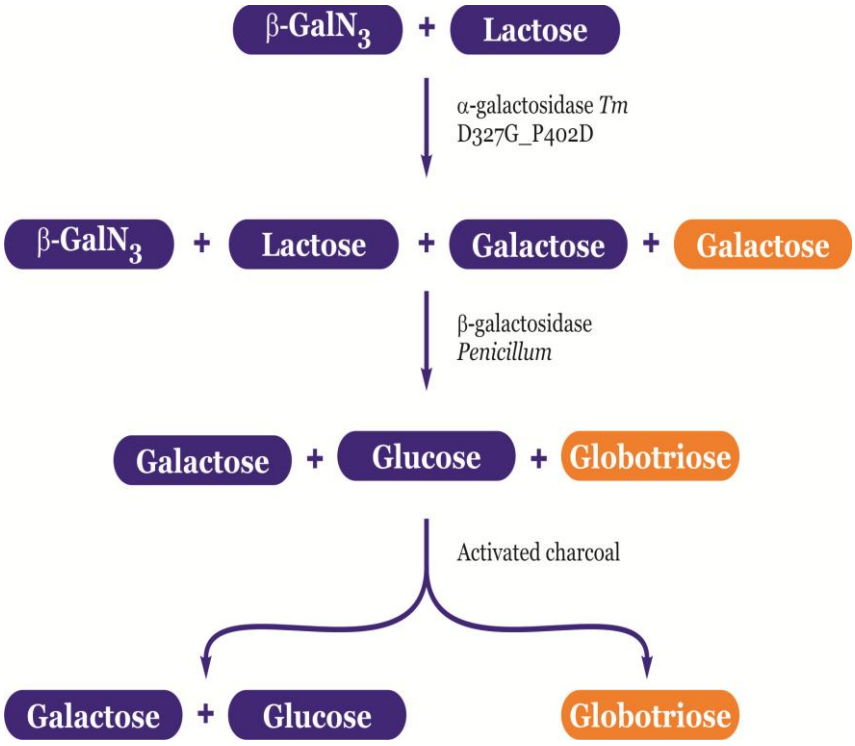


Fig. 5. Scheme of the enzymatic synthesis of globotriose

Mutants and synthetic process were patented.

b.5 *Enteropeptidases with higher activity and higher selectivity for biological applications (substrate promiscuity)*

Mutant variants of human enteropeptidases engineered and tested able a) to make available an enzyme more stable, active and inexpensive than the bovine enzyme, b) to minimize possible unspecific cleavage.

The rational for the engineering strategy has been developed firstly by modeling human and bovine enzymes using homology on the basis of crystal structure of the last one. Then the catalytic subunit of human enteropeptidase (L-HEP) were engineered using the bovine enzyme (L-BEP) X-ray structure. As a matter of fact, L-BEP possesses lower activity in comparison with L-HEP, but higher specificity. Catalytic subunit of human enteropeptidase has a high homology with other serine proteinases, its folding is similar to other members of this family of enzymes for which the crystal structure is obtained, – chymotrypsin and thrombin. The catalytic mechanism and the S_1-site recognizing P_1 are the same as in others chymotrypsin-like serine proteinases, but the residues of Asp in positions P_2-P_4 in inhibitor are basically coordinated by ionic interactions with unique sites on enzyme surface. The analysis of amino acids residues L-HEP and L-BEP, making the nearest environment of a substrate (10 Å) has been performed.

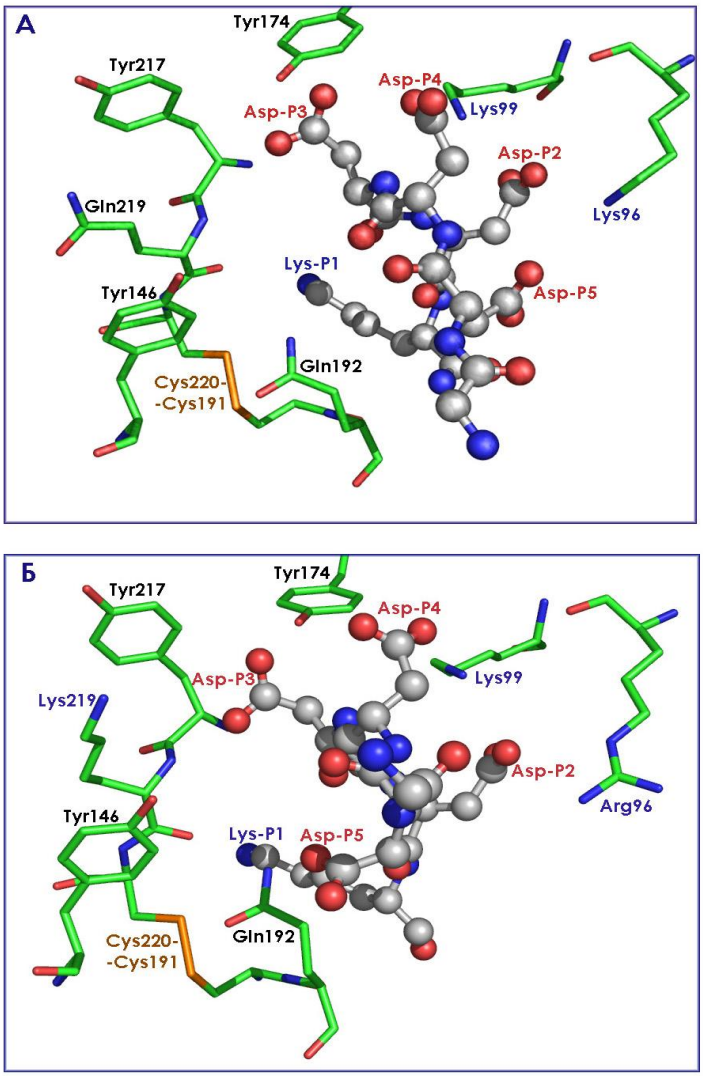


Fig. 6.Comparison of L-BEP (А) and L-HEP (Б) interaction with the substrate VD_4K-cm. The L-HEP 3-D structure has been engineered on the basis of L-BEP crystal structure (1EKB).

Structural differences of human and bovine enzymes, capable to affect binding to a substrate and, thus, to change the specificity, were revealed. Comparing interactions of L-BEP (Fig. 2А) and L-HEP (Fig. 2Б) with the substrate residues P_1-P_5 one can see the differences caused by presence of positively charged lysine residue in position 219 in L-HEP and replacement of the 96 lysine residue to more voluminous Arg residue. Therefore, the conclusion has been drawn that these residues and their nearest environment can be mainly responsible for the difference of catalytic constants of human and bovine enzymes. The mutant genes Arg96Lys, Lys219Gln and the double mutant gene have been obtained.

Kinetic parameters of cleavage of low molecular mass non-specific (Z-Lys-SBzl) and specific (GD_4K-na) substrates by the mutant variants of enteropeptidase in comparison with the wild type enzyme (L-HEP) indicate that all enteropeptidase variants have similar kinetic characteristics on non-specific substrates. On the other hand, the GD_4Kna cleavage K_m increased 2.2 folds with K219Q mutation, while R96K mutation practically did not influence to the kinetic parameters. Structural analysis of the mutant variants of enteropeptidase catalytic subunit showed that residue 96 influenced substrate selectivity via interactions with the residue P2 and its shielding from solvent, while the residue 219 can set ionic and hydrogen bonds with the residues P3 and P5.

Therefore, substitution of the Lys residue for Arg was suggested in the enteropeptidase cleavage sequence in fusion proteins to enhance large scale protein production in fusion expression systems. Also the improvement of enzyme affinity will diminish the amount of

nonspecifically cleaved sites and as a result will improve the yield of target proteins in fusion systems.

b.6 *Penicillin acylase mutants with improved enantioselectivity to be applied in specific enantio-resolutions and cascade reaction (reengineering enantioselectivity)*

Penicillin acylase (PA) has been rationally engineered in order to increase its enantioselectivity towards an array of natural and non-natural amino acids, amino alcohols, aliphatic and aromatic amines which are fundamental building blocks in chemical and pharmaceutical industry. The ultimate aim is to allow the full exploitation of this well known, stable and industrially largely employed enzyme.

To generate a library of *in silico* mutants and to evaluate them in a rational way we have proposed, created and tested unique knowledge based rational prediction pipeline based on bioinformatics. It consists of four primary steps: bioinformatic analysis of target enzyme family, rational prediction of positions for point mutations based on this analysis, library generation, docking evaluation of mutant structures.

The first stage was to perform a bioinformatic sequence and structural analysis on the target enzyme family and to identify the members with clear indications of relationship. To perform this step the classical PSI-BLAST approach was extended by previously reported idea that a true homolog identified in a later iteration should have already revealed its identity as a homolog in the second round (the one using the first profile generated by PSI-BLAST and least likely corrupted).

The second stage was to identify potential sites for point mutagenesis to convert the activities. By definition, orthologous and paralogous proteins have a common ancestor and thus almost always have the same general biochemical function. Orthologs, which diverge after speciation, normally have the same specificity. Thus a protein family can be divided into ortholog groups. Proteins from one group are considered to have the same functional specificity, whereas different groups generally have different specificities. A “specificity group” in this case is a group of orthologous proteins having the same functional feature. Specificity of some groups may coincide or be unknown. The union of the derived groups should not necessarily cover the entire protein family. A set of positions of the MSA, which can best discriminate between these specificity groups, is called Specificity Determining Position. To perform the analysis we have extended out ZEBRA program that was originaly developed in our group in Lomonosov Moscow State University to analyze protein conservation. Advantageously, ZEBRA implements explicit calculation of physicochemical properties of residues located in one alignment column to benefit the cases with high physicochemical similarity inside the group and at the same time low physicochemical similarity between the groups.

It is important to note that no *a priori* knowledge about “specificity” groups is required to run the program though availability of such knowledge could accelerate the prediction.

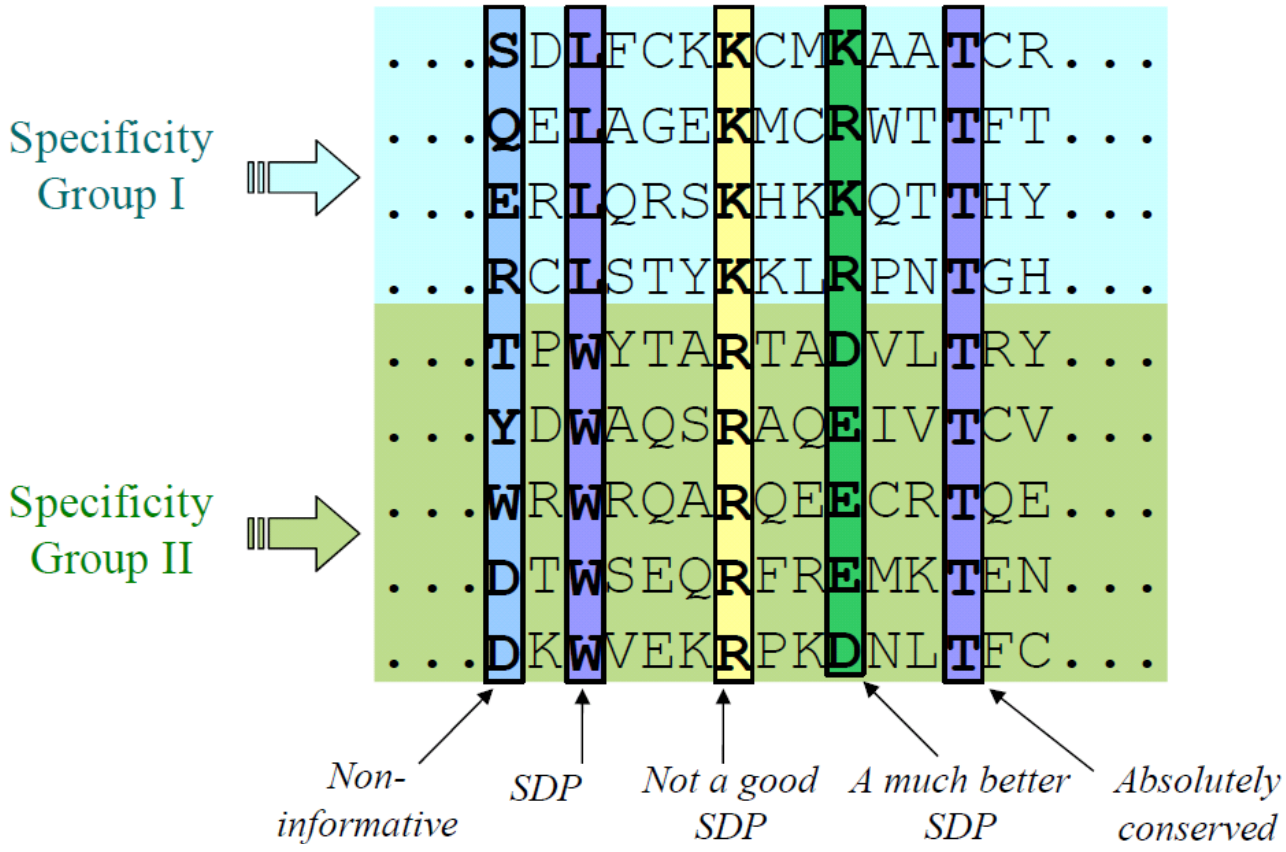


Fig.7. ZEBRA is a program for effective prediction of aminoacids crucial for protein structure and function. ZEBRA v. 3.1 implements explicit calculation of physicochemical properties of residues located in one alignment column to benefit the cases with high physicochemical similarity inside the group and at the same time low physicochemical similarity between the groups.

ZEBRA could work with sequence and structure-based alignments. Updated ZEBRA v.3.1 now can perform advanced group-specific positions search, it has the capabilities to analyze big (MSA of hundreds of sequences) datasets, it is fast and highly scalable for multi-core CPUs.

Third stage was to generate the library based on acquired predictions. For modeling mutations in protein structures we used protocol based on a scoring function that consists of several types of physical potential energy terms and homology-derived restraints to predict the first generation coordinates of mutated side chains.

To perform wide-scale *in silico* screening with AutoDock4 software we have developed an Autodock Operation Environment (AOE) program package that automatically performs the following activities: protonation of input mutant and ligand structures based on predicted pKas, optimization of mutant structure in case of local spatial clashes, optimization of hydrogen bond network to ensure proper protonation of mutant, calculation of grid nets, docking of specified substrates into specified mutants and post processing of acquired docking results – that is energy ranking and filtering out the structures with incorrect orientation of the substrate in enzyme-substrate complex (this requires certain a priori knowledge of binding mechanism). To perform specified tasks AOE integrates Autodock, Pdb2pqr and Propka software products with all intermediate and final results stored in a database controlled by PostGreSQL database manager. AOE gave an extreme opportunity not only to perform potentially infinite screening, but also to store and analyze the results in a rational way. It is limited only to HDD capacity and time to perform docking operations. AOE was implemented on Lomonosov Moscow State University Supercomputer complex SKIF (www.parallel.ru/cluster) and used for *in silico* evaluation of penicillin acylase, lipase and hydroxynitrile lyase mutant libraries using different screening methods. The most promising penicillin acylase mutants with modified enantioselectivity have been recommended for experimental evaluation.

A range of mutants has been prepared, isolated and purified. Substrate specificity of the prepared mutants has been studied (Partner BIOTIR) as well as their ability to catalyze enantioselective

conversions. Remarkably increased catalytic activity (up to 36 fold increase) and enantioselectivity (more than 6 fold increase) was validated experimentally.
The results are object of a patent application (RU2009142994 A "Method to improve catalytic properties of penicillin acylase", Guranda D.T., Jamskova O.V., Panin N.V., Svedas V.K.)

Conclusions:
Despite the physiological difficulties encountered in a few Tasks of the projects, new advanced computational software and methodologies were developed. They allowed the eengineering of enzymes with promiscuous activities and the understanding of the structural and molecular basis of the phnomenom. New enzymes were developed within all Wok Packages devoted to protein engineering (reaction promiscuity, substrate promiscuity, redesign enantioselectivity).
Experimental and conceptual difficulties were encountered in the Tasks concerning the engineering of nitrilase and HNL enzymes. Notwithstanding the efforts of Partner TUD in developing and optimizing cloning strategies, the resulting HNL enzymes were not endowed of activity, although retain their folded conformation. However, the methodological routes drawn for the successful examples listed above represent a powerful tool for planning effecting strategies for the solution of problems faced with these two enzymes.

4.1.4: Potential impact
Scientific and technological impact
The output of IRENE project comprises:

a) **software/computational methods**
b) **engineered enzymes.**

The platform of **computational tools** developed by IRENE project is directed to scientists and industrial operators for being applied to:

- rational design of efficient biocatalysts to be produced through engineering
- fast and efficient *in silico* screening of available enzymes/mutants to fully exploit catalytic potential of existing biocatalysts and providing quantitative parameters describing enzyme's efficiency and for correlating enzyme structure-function
- fast substrate-screening and rational substrate engineering
- understanding molecular basis of biocatalyst' action and properties

Most of the **software** produced inside the consortium is "open-source" and readily accessible to the scientific community and to end-users, which typically are large companies working on biomass processing or synthetic biology.
Design and screening methodologies have been set up relying on both commercial and open sources software. In some cases the use of commercial software allowed the exploitation of technical support for the refining of the computational strategies and to achieve "the proof of concept" in a shorter time frame. However, in each case, specific alternative solutions are available in terms of "open source" products.
The method for automatic in-silico design and screening of enzymes represents an innovative instrument for multi-tasks optimization of enzyme design, that can be tailored on the basis of specific needs and requirements. Therefore the present project represents the first example of application of this methodology in the Biotechnology sector and has positively assessed its general validity, applicability and versatility.
The expected impact will be in terms of "improving the effective exploitation and integration of structural and functional information.
The scientific expertise of Partners along with the actual cooperation within the consortium, has assured the high standard of the scientific know-how and its transfer to scientific community, society, SMEs and industries. As a matter of fact, the Consortium has already published part of the project results in more than 15 publications on peer reviewed journals. So far, four patent applications have been filed and at least ten further scientific publications are in the pipeline.
The experimental work of five PhD thesis was built up inside the IRENE project. Globally, more than 20 young researchers were hired, trained and actively involved in the international cooperation inside the project. This has strengthened the links between European and Russian

research and has promoted the cohesion of the scientific community at an European and international level.

The following list of **biocatalysts has been rationally designed and produced** inside the IRENE project because of their potential industrial impact:

- amide forming enzymes with higher efficiency and different specificity as compared to the known proteases to be used in fine and pharma chemistry, (patented)
- Lipases to be used in polycondensation of lactide, for the production of bio-based and compostable polymers (published)
- amidases with increased synthetic efficiency and improved regioselectivity for more cost effective enzymatic synthesis of beta- lactam antibiotics (patented)
- glycoside hydrolases with enhanced synthetic efficiency for the production of glycoconjugates of relevance and to be used in the biotransformation of "reluctant" oligosaccharides in food industry (patented both enzyme and process)
- enteropeptidases with higher activity and higher selectivity for biological applications (published)
- Penicillin acylase mutants with improved enantioselectivity to be applied in specific enantio-resolutions and cascade reaction (patented)

Industrial and economic impact

At present European companies supply about 70% of the world enzymes. To maintain and strengthen the European leadership in the biocatalysis sector, different routes must be pursued to make enzymes readily available for practical applications, thus making biocatalysis competitive as compared to conventional organic chemistry**.**

Biotec products are established in higher value business segments but there is an increasing interest of chemical industries world-wide and in EU (both high and low-end market). EU must get ready for providing adequate answers in terms of technological innovation.

The establishment of new or more **economically competitive activities in the Industrial Biotechnology EU sector** is seen as one of the primary goals, along with the prompt transfer to industry of adequate technological tools for accelerating innovation and shortening the "time to market" of biotechnological products.

Political and social impact

The results coming from the project contribute to the implementation and evolution of the European policies for the construction of a **"Knowledge based"** society by promoting the following criteria:

- Attracting and **training young researchers**, in particular women, to form a new skilled generation with an improved awareness towards science and corresponding societal benefits, trained in an international environment
- Enhancing the **competitiveness of European SMEs and industry** and accelerating the implementation of the European **bio-based economy**
- Improving the **environmental sustainability** of productive processes as well as the safety aspects associated to workers' health in chemical industry
- Boosting **international co-operation** with Russia and Uzbekistan in crucial scientific and technological areas

Cross-thematic impact

Because the computational methods studied in the present project as well as the biocatalysts have an interdisciplinary character, the knowledge and products developed in the project will contribute to different industrial sectors and policy objectives included in Themes such as Health, Environment, Energy, Materials.

Since the project aimed at "being strongly contaminated" by computational strategies used in different disciplines, in the areas of computer science, drug-design and life science more in general. Positive sinergies have been already identified during the works of "International

Conference on in silico enzyme design and screening", which was specifically organized as closing event of the IRENE project to promote the dialogue and cooperation with complementary sectors (e.g. computational science, enzymology, drug-design and protein chemistry).

4.1.5 Project web-site:

www.irene-fp7.eu

Further information on the activities of the IRENE consortium available at the following URL:

http://www.irene-fp7.eu/UserFiles/Book_of_abstract.pdf

http://www.irene-fp7.eu/UserFiles/Taiwan-NCP%20Lucia%20Gardossi.pdf

http://www.irene-fp7.eu/UserFiles/Il%20Piccolo%20article%20UNITS.pdf

http://www.irene-fp7.eu/UserFiles/UNITS%20-%20ANSA%20article%20and%20webpages.pdf

http://www.irene-fp7.eu/UserFiles/IRENE%20conference%20poster%281%29.pdf

http://www.irene-fp7.eu/UserFiles/IRENE%20conference%20flyer%281%29.pdf

http://www.irene-fp7.eu/UserFiles/IRENE%20conference%20flyer%281%29.pdf

http://www.irene-p7.eu/UserFiles/IRENE%20conference%20plenary%20speakers%281%29.pdf

http://www.irene-fp7.eu/UserFiles/VF_Irene.pdf

4.1.6: List of beneficiaries and contacts names

1.Università degli Studi di Trieste (UNITS)
Lucia Gardossi

2. School of Biotechnology/Biochemistry, Kungliga Tekniska Högskolan (KTH)
Karl Hult

3. University of Copenhagen (UCPH)
Jan Halborg Jensen

4. Technische Universiteit Delft (TUD)
Ulf Hanefeld

5. Novozymes A/S (NZ)
Allan Svendsen

6. The National University of Uzbekistan (NUU)
Mirzaatkham Mirzakhakimovich Rakhimov

7. Belozersky Institute of Physicochemical Biology Lomonosov Moscow State University (MSU)
Vytas Svedas

8. Petersburg Nuclear Physics Institute Russian Academy of Sciences (PNPI)
Anna A. Kulminskaya

9. Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences (IBCH)
Dmitry Dolgikh

10. Molecular Technologies, Ltd (MLT)
Ghermes Chilov

11. Bio:Technologies, Innovations, Research, Ltd (BIOTIR)
Maxim Ilich Yushko