Project no. LSHB-CT-2003-503017

Project acronym: DATAGENOM

Post-genomic datamining of enzymes for the synthesis of chiral pharmaceutical intermediates

Specific Targeted Project
Priority 1: Life sciences, genomics and biotechnology for health

Publishable executive summary
Reporting period 1 & 2: 1-36 months

Period covered: 1. December 2003 to 30. November 2006

Date of preparation:

May 2007.

Start date of project: 1. December 2003 Duration: 36 months

Project coordinator: Jakob K. Kristjansson,
Prokaria ltd. Gylfaflöt 5. IS-112 Reykjavik, Iceland

Publishable Executive Summary

Project objectives: The project focusses on the discovery of novel enzymes, from both public and proprietary genomes, in particular novel alcohol dehydrogenases and amino acid modifying enzymes for use in established and innovative processes for chiral synthesis. The aim is to use innovative datamining and processing of a large number of genes to ensure high flow-through in the process and rational selection of best enzyme candidates. Optimized vector-host systems are used for efficient gene expression and production of the selected enzymes. The enzymes will undergo detailed biochemical characterization with regard to potential industrial applications. Rational protein engineering or directed molecular evolution will be employed in order to obtain more robust variants, new substrate preferences or enhanced enantiomeric selectivity. Selected enzymes will be tested in existing and/or novel biocatalytic processes for production of chiral pharmaceutical intermediates with applications in therapeutic areas including AIDS, cancer and Alzheimer´s disease. Thus, the DATAGENOM project extends from genome analysis, through cloning, expression, enzyme production, screening and protein engineering, to the enzymatic production of chiral biomolecules.

Coordinator & Contractors involved

No. 1 Coordinator
Prof. Dr. Jakob K. Kristjánsson, CEO
Prokaria ehf
Gylfaflöt 5,
IS-112 Reykjavik, Iceland.
Tel +354-570-7908
Fax +354-570-7901
Email: jakob@prokaria.com

No. 2
Dr. Ian Fotheringham
Ingenza Ltd
Joseph Black Building
The King´s Building
West Mains Road
Edinburgh EH9 3JJ,
Scotland, UK
Tel: +44 131-650 7747
Fax: +44 131-650-4743
Email: ian.fotheringham@ingenza.com

No. 3
Prof. Olle Holst,
Lunds Universitet
Deptm. of Biotechnology,
Center for Chemistry and Chemical Engineering,
P.O. Box 124,
Getingevägen 60,
S-221 00 Lund, Sweden
Tel+46-46-222 9844
Fax +46-46-222 4713,
Email: olle.holst@biotek.lu.se

No. 4
Prof. Peter Schönheit,
Christian-Albrechts-Universität Kiel,
Institute fuer Allgemeine Mikrobiologie,
Biologiezentrum, An Botanischen Garten 1-9,
D-24118 Kiel, Germany
Tel +49-431-880-4330
Email: peter.schoenheit@ifam.uni-kiel.de

No. 5
Prof. Ralf Mattes
Universität Stuttgart,
Institut für Industrielle Genetik
Allmandring 31,
D-70569 Stuttgart, Germany
Telefon: 49+ 711/6856970
Fax: 49+ 711/6856973
Email: ralf.mattes@po.uni-stuttgart.de

No. 6.
Prof. Vicente Gotor
Department of Organic and Inorganic Chemistry,
University of Oviedo,
Av. Julian Claveria 6,
33006 Oviedo, Spain
Tel 34-985-103448
Fax 34-985-103448
E-mail: vgs@sauron.quimica.uniovi.es
Tel: +34 985 104061
Email: gotor@rectorado.uniovi.es

No. 7
Dr. Kai Doderer
Degussa AG
Rodenbacher Chaussee 4
D-63457 Hanau-Wolfgang, Germany
Tel: 49 6181-59-6280
Fax: 49-6181-59-2961
Email: Kai.Doderer @degussa.com

No. 8
Dr. Leszek Rychlewski
BioInfoBank Institute,
Limanowskiego 24A/16,
60-744 Poznan, Poland
Tel: +48-61-8653520,
Fax: +48-61-8643350,
E-mail: leszek@bioinfo.pl

No. 9
Dr. Thomas Daussmann
Managing Director
JFC - Jülich Fine Chemicals GmbH,
Prof.-Rehm.Str. 1
D-52428 Jülich, Germany
+49 (2461) 980401
Fax. +49-2461-980410
Email: t.daussmann@juelich-chemicals.de

Subcontractor
Prof. Oreste Piccolo
Studio di Consulenza Scientifica
Via Bornò 5,
23896 Sirtori (LC), Italy
Tel +39-039-958985
Fax +39-039-955024
Email: orestepiccolo@tin.it

Work performed and results

The total number of scanned genomes were from 363 bacteria, 25 archea and 72 eukaryotic microbes plus 6 proprietary bacteria. Databases were prepared for 3 main enzyme families: Alcohol dehydrogenases (Short chain & medium chain ADH); Oxidases (L-Amino acid & monoamine OX); Formate dehydrogenases (D-2 hydroxyacid & molybdopterin binding domain FDH)

Annotation was done for 1170 novel genes: 750 ADH (500 from public genomes & 150 from propietary genomes); 55 L-AA Ox (43 bacterial & 12 fungal); 40 Mono AOX (25 bacterial & 15 fungal); 77 2D-OH-FDH (34 bacterial & 33 fungal); 348 Molybdo-FDH (304 bacterial & 44 fungal)

Existing strain collection was phylogenetically classified and 200 diverse strains selected for activity screening on various ADH substrates. This yielded 42 strains showing interesting ADH activity patterns that were used for retrieving new genes with the GENEMINING™ method. The improved method also yielded 40 Short chain ADH´s & 23 Medium chain ADHs from existing partial genomes. The GENEMINING™ method was improved by using ThermoPhage™ ssDNA ligase.

The ActiveBlast tool was used & improved for the selected gene families. The current version has improved statistical analysis and estimation of accuracy. It was tested for predicting effects of mutations and compared with other programs. It is available as a service at: www.cropnet.pl/ligprof. From the predicted annotated genes about 520 best candidate genes in the target families ADH, Oxidases & FDH were proposed for cloning.

Total of 98 enzymes were expressed & tested for activity, stability or immobilization: 33 SDH & 15 MDH; 13 + 2 mutants AAO/MonoAO; 5 2D-OH FDH; 15 FDH and 12 + 2 synthetic GDH.

Both D- & L-AAO were random mutated leading to 10-fold increased activity towards several substrates, including two commercially relevant unnatural L-amino acids. The mutation screen was improved by shorter time & 20-fold lower cost. Consensus sequence & loop analysis were used to resynthesise both D-AAO and GDH. Ten single or quadurple mutants of FDH were made for improved immobilazation strategy.

Total of 60 enzymes (72 variants) were screened; 25 ADH (33 clone variants) were characterized in details for enantiomeric excess & conversion %, resulting in at least 3 interesting enzymes acting on ethylacetoacetate, a typical β-ketoester. One rare thermotolerant, strictly R-selective ketoreductase was discovered. Total of 13 AAO/MonoAO were screened by moderate HTP assay on diverse range of AA targets on whole cells and cell extracts. Also enzyme variants for mutant selection resulting in 35- & 24-fold improvment on a major commercial target & improved half-life of a D-AAO. Two mutant MonoAO were tested for derazemization processes, and 3 hydrolases for synthesis of γaminoalcohols & secondary cyclic amines. Five archaeal FDHs were screened. Also 14 new GDH were screened on NADH & NADPH for relative cofactor regeneration activity & stability.

The enzymes were purified on small scale for the activity screening tests. Three putative archeal FDHs were purified and shown to be novel, acting preferentially on long-chain D-specific 2-OH acids of industrial interest.

The small-scale (up to 10 L) production protocolls using *E. coli* fed-batch systems using standard fermentation equpments with on-line measurements & controls were developed. The technology increased rutine yields from 3-5g/L to over 54g/L wet cell mass.

Some AAO enzymes were moved over to 10 L scale. Stable lysed & fractionated whole cells of D-AAO have been transerred to a pilot manufacturing site for industrial process testing. Some of the ADHs were made commercially available and therefore also prepared for this larger scale.

A thighly regulated L-rhamnose-inducible *E. coli* system for expression was improved (by *T7gene10* leader sequence), stability (by *cer*-region deletion), copy number (by deletion of *rop*-gene) & size. The optimized vector was used by great success by Partners 2 & 3. By using a plasmid from *Bordetella bronchiseptica*, a broad host range system, functioning in many Gram-negative bacteria was made. A small plasmid gave best results & was very stable at 30°C but less at 37°C. By deleting the *dadX* & *alr* alanine racemase genes, an antibiotic-resistance free vector was made. By cloning an endonuclease gene it was shown that the rhamnose system was thight enough for cloning & expression of toxic genes in *E. coli*. Temperature regulated expression vectors were constructed, giving great improvements and better scale-up.

A *Thermus thermophilus* thermophilic system was improved by introducing new features such as polylinkers, tags and promoter sequences. Another thermophilic *Rhodothermus marinus* shuttle vector expression system was established with a stable tryptophan marker & different promoters and reporter genes were tested.

The kinetic resolution of optically enriched 1,3-aminoalcohols through aminolysis processes in organic solvents has been achieved with reasonable success. Enzymatic resolution of γ-aminoalcohols, secondary cyclic amines, and derivatives from 4-(N,N-dimethylamino)pyridine have been performed using lipases. Use of oxidoreductases for the bioreduction of ketones has been studied.

Synthesis of (S)-Dapoxetine has been developed with good overall yield. Kinetic resolution of chiral building blocks that are possible intermediates of interesting drugs has been successfully achieved.

Statistical reaction design & optimised reactions using oxidase biocatalysts yielded specific benefits to the reaction efficiency to prepare L-2-aminobutyric acid. It was applied to oxidase biotransformations to prepare two further unnatural amino acids. Amine oxidases and lipases and the influence of different enzymatic parameters in organic solvents as well as scale up parameters & the recycling of immobilised enzymes, were studied in order to lower the operating cost of the bioprocesses.

Formulation of novel and improved oxidases, as semi-pure cell fractions in biotransformations, lowered the overall biomass use by 8-20 fold, giving greatly improved productivity and simpler work process. This manufacturing process for L-2-aminobutyric acid and two further unnatural L-amino acids has been evaluated at 100 L pilot scale and is moving to 1000 L scale on rute to full commercial manufacturing.

A general method of process development & optimised bioreduction reactions using whole cell biocatalysts was devised. This work yielded no specific benefits for the tested reaction to produce tert-leucine of 2-(p-chlorophenyl) ethanol. Nevertheless the intrinsic properties of the newly discovered cofactor regenerating enzymes expanded the possibilities of the reduction biocatalysts. Novel and improved cofactor regeneration enzymes were formulated as whole cell biocatalysts and have been used in biotransformations operated at 2 L pilot scale in preparation for full manufacturing scale. Three new industrially interesting ADHs were found and one has been produced in high cell density fermentation and is already being sold commercially for bioproduction of speciality compunds on 5-10L scale.

Plan for using and disseminating the knowledge

The dissemination at this stage of the projects has been primarily within the project at meetings and bilateral discussions and confidential reports. Several publication from the work have been published, are in preparation or will be published later. This also applies to student reports and Ph.D. thesis. Particular care is taken to evaluate the commercial potential of the results and the partners intend to protect their innovation with patenting where appropriate. This can be seen by the fact that 3 patent applications have already been filed for innovations made in this projects and enzymes and processes have already been used commercially. This project is ideally composed of 5 academic participants and 3 highly focussed SME´s plus 1 large industrial company. This ensures on one hand dissemination of knowledge to students and the general scientific community through open litterature and conferences but on the other hand maximum exploitation by the industrial partners. Where appropriate, patents have and will be filed on methods of genome mining and activity screening and on matrials composed of genes and enzymes backbones discovered or improved by directed mutations, as well as chemical processes and entities developed. Web page for the project is: http://datagenom.bravehost.com

The main dissemination activities of the partners in the project are listed in the following overview table.

Planned/ actual dates	Title	Type	Type of audience	Countries adressed	Size of audience	Partner respon-sible
8-9 May 2004	Datagenom kick off meeting Hanau, Germany	Conference meeting	Internal	Internal	Internal	all
24-25 Oct.2004	Datagenom, Oviedo, Spain,	Conference meeting	Internal	Internal	Internal	all
20-21 June 2005	Datagenom Reykjavik, Iceland	Conference meeting	Internal	Internal	Internal	all
September 2006	Datagenom, Edinburgh Scotland	Conference meeting	Internal	Internal	Internal	all
5. Nov 2004	Compatible solutes, Witten, Germany	Conference	Research	Europe	50	1
12-13 Oct 2004	Extremophiles 2004, Baltimore, USA	Conference	Research	Global	300	1
2004-2005	The work has been disseminated in Japan, USA, Singapore, Australia,	Conference	Research	Global	100-200	2
2004	Jurjen N. van Bolhuis. 2004. Cloning and functional expression of oxido-reductase genes from extremophiles; characterization of recombinant oxido-reductases. Master Thesis	Publication	Higher education	Internal CO	Prokaria/ University of Utrecht, holland	1
2004	A THERMOPHILIC HOST-VECTOR SYSTEM. International Patent Application No.: PCT/IS2004/000013 Publication No.	Patent application	Industry	Global	Global	1
2005	Anke Hummel, biochemistry student at the Greifswald University, Germany. 2005. Internal reports on alcohol dehydrogenases done at Prokaria in Iceland in first half 2005.	Publication	Higher education	Internal CO	Prokaria/ Greifswald University, Germany	1
2006	*Rhodothermus marinus*: physiology and molecular biology. Extremophiles 10:1, 1-16.	Publication	Research	Global	Global	1
2007	Cloning and expression of heterologous genes in *Rhodothermus marinus*. Extremophiles 11, 283-293	Publication	Research	Global	Global	2
2004	Enzyme Catalysed Deracemisation and Dynamic Kinetic Resolution Reactions. *Curr Opin Chem Biol*, 8, 114-119.	Publication	Research	Global	Global	2
2004	Novel Biocatalyst Technology for the Preparation of Chiral Amines. Innov. Pharm. Technol. 4, 114-122.	Publication	Research	Global	Global	2
2005	New platform bioprocesses for chiral amines and unnatural amino acids. Chemical Industry Magazine	Publication	Industry	Global	Global	2
2006	Preparative deracemization of unnatural amino acids. Biochem. Soc. Trans. 34, 287-290	Publication	Industry	Global	Global	2

2006	Compositions of Variant Biocatalysts and Method of Preparing Enantioselective Amino Acids and Amines Using a Variant Biocatalyst. (covers new oxidase biocatalysts evolved towards novel amino acids) US Provisional	Patent application	Research Industry	Global	Global	2
2006	Stereoinversion of Amino acids (covers improvements to the overall process of stereoinversion of amino acids) US Provisonal	Patent application	Research Industry	Global	Global	2
2005	Effect of postinduction nutrient feed composition and use of lactose as inducer during production of thermostable xylanase in Escherichia coli glucose-limited fed-batch cultivations. J. Biosci. Bioeng. 99, 477-484	Publication	Research Industry	Global	Global	3
2005	A feeding strategy for E. coli fedbatch cultivations operating close to the maximum oxygen transfer capacity of the reactor. Biotechnol. Letters, 27, 983 - 990	Publication	Research Industry	Global	Global	3
2005	Optimized expression of soluble cyclomaltodextrinase of thermophilic origin in E.coli by using a soluble fusion-tag and by tuning of inducer concentration. Protein Exp. Purif, 39, 54-60	Publication	Research Industry	Global	Global	3
2005	The methylotrophic yeast *Pichia pastoris* as a host for the expression and production of thermostable xylanase from *Rhodothermus marinus*. FEMS Yeast Res, 5, 839-850	Publication	Research Industry	Global	Global	3
2006	Production of a lipolytic enzyme originating from *Bacillus halodurans* LBB2 in the methylotrophic yeast *Pichia pastoris*. Appl. Microbiol. Biotechnol. 71, 463-472	Publication	Research Industry	Global	Global	3
2006	Glycolysis in hyperthermophiles. Annual Meeting of the New Zealand Microbiological Socoiety	Key note lecture	Research	Inter-national	250	4
2007	Characterization of novel D-isomer specific 2-hydoxyacid dehydrogenases from hyperthermophilic archaea	Publication, In preparation	Research Industry	Global	Global	4
2007	Building Blocks of Plasmid Functions: Characterization of Genetic Modules implemented in a rhaP Expression vector.	Publication, In preparation	Research Industry	Global	Global	5
2007	Construction of a rhamnose inducible selfselecting host-vector system.	Publication, In preparation	Research Industry	Global	Global	5

3-8 July, 2005	Enzymatic resolution of 1,3-amino alcohols using lipases in organic solvents. Poster presentation P321 at Biotrans 2005 Symposium.	Poster, Delft, Holland	Research	Global	Global	6
2006	Lipase-catalyzed Resolution of Chiral 1,3-Aminoalcohols. Application in the Asymmetric Synthesis of (S)-Dapoxetine. Tetrahedron: Asym. 17, 860-866	Publication	Research Industry	Global	Global	6
2006	Enzymatic Chemoenzymatic Preparation of Optically Active Secondary Amines: a New Efficient Route to Enantiomerically Pure Indolines. Tetrahedron: Asym. 17, 2558-2564	Publication	Research Industry	Global	Global	6
2006	Biocatalytic Preparation of Optically Active 4-(N,N-Dimethylamino) pyridines for Application in Chemical Asymmetric Catalysis. Tetrahedron: Asym. 17, 1007-1016	Publication	Research Industry	Global	Global	6
2006	Enantioselective Kinetic Resolution of 4-Chloro-2-(1-hydroxyalkyl) pyridines using *Pseudomonas cepacia* Lipase. Nature Protocols. 1, 2061-2067	Publication	Research Industry	Global	Global	6
2006	*Candida Antarctica* Lipase B: An Ideal Biocatalyst for the Preparation of Nitrogenated Organic Compounds. Adv. Synt. Cat. 348, 797-812	Publication	Research Industry	Global	Global	6
2006	Lipases: Useful Biocatalysts for the Preparation of Pharmaceuticals. J. Mol. Cat. B: Enzymatic. 40, 111-120	Publication	Research Industry	Global	Global	6
2006	Enzymatic Aminolysis and Ammonolysis Processes in the Preparation of Chiral Nitrogenated Compounds. Curr. Org. Chem. 10, 1125-1143	Publication	Research Industry	Global	Global	6
2007	Preparation of Chiral Pharmaceuticals through Enzymatic Acylation of Alcohols and Amines. In Biocatalysis in the Pharmaceutical and Biotechnology Industries. Chapter 7. (Ed. R. M. Patel. Dekker): Taylor and Francis. New York	Publication Book chapter	Research Industry	Global	Global	6
2007	Enantioselective Synthesis of 4-(N,N-Dimethylamino) pyridines through a Chemical Oxidation-Enzymatic Reduction Sequence. Application in Asymmetric Catalysis. Adv. Synt. Cat.	Publication In Press	Research Industry	Global	Global	6

2006	PDB-UF: database of predicted enzymatic functions for unannotated protein structures from structural genomics. BMC Bioinformatics. 7, 53	Publication	Research Industry	Global	Global	8
2006	Eukaryotic Domain of Unknown Function DUF738 Belongs to Gcn5-related N-acetyltransferase Superfamily. Cell Cycle 5: 2927-2930	Publication	Research Industry	Global	Global	8
2007	LigProf: A simple tool for in silico prediction of ligand-binding sites. J. Mol. Model	Publication In Press	Research Industry	Global	Global	8
2007	http://www.cropnet.pl/ligprof	Website	Research	Global	Global	8
2007	http://www.bioinfo.pl/PDB-UF	Website	Research	Global	Global	8
2004	http://datagenom.bravehost.com	Website	All	Global	Global	1
2007-2008	Planned publications from various partners, depends on ultimate results, and IPR measures done	Publication	Research	Global	Global	all

If not specified as already published, patent filed or commercialised, most of the exploitable results at this stage are still mosely for internal dissemination as they are beeing actively used and prepared for publications and exploitation. At the end of the project several results have already been IP protected or even commercialised. Other results are already or being prepared for scientific puplications. It is expected that more patent filing, publications and commercializations based on this project will follow in the next two years.