Grant Agreement no:	279039 (FP7)
Project Acronym:	**COMPLEXINC**
Project full title:	NEW TECHNOLOGIES AND PRODUCTION TOOLS FOR COMPLEX PROTEIN BIOLOGICS
Funding scheme:	Collaborative Project
Period covered:	from 01/11/2011 to 31/10/2015
Project co-ordinator:	Dr Imre Berger
Organisation:	European Molecular Biology Laboratory
Tel:	+33 476 20 70 61
Fax:	+33 476 20 77 86
E-mail:	iberger@embl.fr
Project website:	www.complexinc.eu

1 Executive summary

From the early beginnings of the biologics era, production technology development has fundamentally transformed drug discovery. Technology has been the engine, but frequently also the major limiting factor in the drug discovery process. Initially substances from nature, such as plant extracts, were empirically selected for biological activities in combating disease. Increasingly sophisticated techniques in synthetic chemistry allowed for the synthesis of natural substances, or their analogues, leading to therapeutics including antibiotics and morphine. The dawn of molecular biology heralded the emergence of powerful new technologies such as receptor and enzyme biology, genomics and proteomics which have greatly accelerated the drug discovery process and transformed modern medicine and the pharmaceutical industry.

The recent explosion of biological data from genomics, proteomics and interactomics research has provided invaluable new information into how disease occurs at the DNA, protein and systems levels. However, a new and imposing technical bottleneck has also thereby been created. Because recombinant protein production tools did not match the sophistication and complexity of living organisms, the provision of the increasingly complex biological specimens identified, in assayable or druggable form, has become a serious challenge, considerably hampering progress and success in academic and industrial research and development.

The EC FP7 ComplexINC project (2011-2015) was conceived to address this vital technology gap. ComplexINC conceptualized and generated advanced tool-kits to enable high-throughput assembly of complex biologics and metabolic pathways in eukaryotic expression systems, for micro- to large scale production of high-quality protein specimens for drug discovery and as bio-therapeutics. Essential parts of these tool-kits are innovative next-generation DNA assembly pipelines for automatable multigene expression construct generation, superior synthetic baculovirus tools, genetically engineered animal and yeast cell-lines with unique capabilities with respect to protein maturation and modification, as well as advanced methods for bioprocessing and integrated quality control.

Essential objectives of ComplexINC were the provision of innovative DNA tool-kits and assembly pipelines for manufacturing eukaryotic proteins and their complexes, the engineering of synthetic nanosystems that improve the folding and functional assembly of important eukaryotic protein targets and the creation of superior new designer baculovirus genomes to overcome current bottle-necks of the application of the baculovirus expression technology for drug discovery and biologics production. ComplexINC applied the tools developed to the production of a large set of high-value protein targets including human gene regulation protein complexes, receptors, membrane proteins, virus-like particles and important new drug targets, thereby directly benefitting and driving research and development at each of the partner institutions, academic and SMEs, alike. A particular focus of activity in the ComplexINC work program was to create genetically engineered cell lines (animal and yeast) with unique capabilities in protein production, thus answering a currently neglected area of genomic development where significant improvements can be expected.

Further, ComplexINC was characterized by a strong bioprocessing component and success-fully endeavored to bridge micro-technologies such as robotics based micro-fermentation of eukaryotic cell cultures with large scale (up to pharma-scale) production of selected high value target proteins. Specific areas within the ComplexINC work program were potential vaccine candidates including enhanced influenza virus-like particles, proteins used for enzyme replacement therapy (ERT) of specific rare diseases, and novel and powerful new protein scaffolds with the potential to transform peptide and protein delivery. Putting these tools, production technologies and innovations in place generated a considerable impact on the availability of challenging proteins and protein complexes including such of human origin, with tailor-made properties specifically geared towards their application in drug discovery and in the treatment of human disease.

In four intense and exciting years, ComplexINC succeeded in creating superior and user-friendly next generation technology platforms, directly accelerating academic and industrial R&D pipelines of the partner institutions. A particular focus of ComplexINC was to commercialize the IP, protein targets and therapeutics generated. The innovations in the ComplexINC project served as a nucleus to found Geneva Biotech (www.geneva-biotech.com), a ComplexINC spinoff and already profitable new start-up enterprise which commercializes high-end biologics discovery and intelligent production tools.

2 Description of the project context and objectives

ComplexINC was conceived in response to an EU 7th Framework SME-directed Call under the theme “Health” (http://ec.europa.eu/research/health).

ComplexINC acknowledged that, in addition to their essential role as tools for drug discovery, complex biologics themselves are increasingly being employed as next-generation medicines. Indeed, the fastest growing new drug segment in the pharmaceutical industry is therapeutic proteins. They are predicted to dominate the next generation of drugs. In selected areas, such as the production of therapeutic antibodies in mammalian cells, currently available therapeutic protein production technologies are rather advanced. However, platforms used to manufacture other classes of therapeutic proteins were at a relatively primitive developmental state. For example, insect cell and yeast production platforms for biologics are presently used to produce several marketed therapeutic proteins, but, when ComplexINC was conceptulaized, were only at a development state comparable to mammalian cell antibody production platforms used in the early 1990s. In summary, sophisticated high-performance platforms for the production of most complex biologics were distinctly lacking.

Based on the leading expertise and innovative drive of the partners, ComplexINC addressed this urgent and imposing bottleneck, by providing new and sophisticated tools and technologies, to overcome the challenge of complex biologics production in drug discovery and therapy for the 21st century. The work plan of ComplexINC thereby precisely covered the activities outlined in the EC HEALTH call 1.1 to ”catalyze progress in developing new research tools for modern biology“. In particular, ComplexINC addressed HEALTH 2011.1.1-1 ”to focus on the development and improvement of high throughput research tools and technologies“. ComplexINC conceptualized and systematically generated cutting-edge tool-kits to enable high-throughput (HT) assembly of complex biologics in eukaryotic expression systems (animal and yeast) for production of high-quality protein specimens used in drug discovery and as therapeutics. Particular attention was paid to establish reliable and robust production processes from micro-scale to the large scale relevant to pharmaceutical production.

Essential objectives of ComplexINC were:

- Provision of innovative next-generation HT DNA tool-kits and assembly pipelines for manufacturing eukaryotic proteins
- Engineering of synthetic DNA-based nanosystems for drug discovery and complex biologics production
- Creation of superior new synthetic designer baculovirus genomes for drug discovery and biologics production
- Provision of new and powerful genetically engineered cell lines (animal and yeast) with unique capabilities in protein product maturation and modification.
- Rigorous validation of all innovations and procedures by producing complex biologics and biotherapeutics from micro- to fermenter scale
- Valorisation of ground-breaking innovations by preparing launch of ComplexINC spin-off SME

The ComplexINC consortium was operational for 48 months (Nov. 2011-Oct. 2015). Operations were synergistically structured in six distinct work packages (WP01-WP06).

WP01 - Management:

Efficient management structures and procedures were implemented in Work Package Management (WP01). A Management Team consisting of Coordinator (EMBL), Deputy Coordinator (Crelux) and the Project Management Office (Eurice) was complemented by the General Assembly, a Steering Board composed of the Work Package leaders, and the Innovation Advisory board consisting of stakeholders from academia and industry. The Management Team completed the Grant and Consortium agreements. The ComplexINC consortium has convened for

Kick-Off and progress and management meetings in Munich, Oeiras, Grenoble, Barcelona, Braunschweig, and Berlin.

WP02 – New high-throughput pipeline for eukaryotic complexomics

The aim of WP02 was to create and implement new tool-kits and pipelines for producing complex protein targets in a range of eukaryotic host systems. The ACEMBL base technology which formed a core of the ComplexINC work program was consequently extended in this WP to encompass, in addition to the currently predominating baculovirus/insect cell and mammalian systems, also further eukaryotic host systems with unique properties. Notably, novel ACEMBL-based tool-kits were designed and developed for the yeasts *P. pastoris and S. cerevisiae*, providing for the first time means for efficient multiprotein co-expression in these hosts, complemented by development of a BOOST library to enhance production of difficult targets. A wide range of protein targets have been produced in the consortium using ACEMBL tool-kits, catalyzing exciting new insight in essential cellular processes such as gene regulation, cell differentiation and pluripotency, among others. Considerable progress also has been made towards engineering a metabolic pathway for ergothioneine production.

WP03 – Synthetic Biology: Creating designer baculovirus nanosystems:

Synthetic biology techniques were used in the project for creating designer baculo-viral nanosystems (SVNs) in WP03. A polyprotein approach has been successfully implemented to increase the efficiency of co-producing chaperones for difficult targets. An intensive *in silico* study was carried out to identify evolutionary conserved regions in the baculoviral genome, revealing large segments, mostly clustered in one half of the genome, for reengineering by custom-designed synthetic DNA fragments. Construction of corresponding hybrid viruses (partly wild-type, partly synthetic) has been achieved and specimens were validated with improved properties for recombinant production of protein and their complexes.

WP04 – Enabling eukaryotic expression by cell line engineering:

A particular focus of ComplexINC was the engineering of cell lines to enable eukaryotic expression (WP04). A recombination cassette-mediated exchange (RCME) technology was applied to construct eukaryotic master cell lines (mammalian and insect) displaying long-term stability of expression, which were then used to integrate multigene expression cassettes, including novel polyprotein-based chaperone modules developed in WP03. A novel yeast-based glycosylation platform (Glycomam) was put in place, and analytical methods for glycan analysis have been successfully implemented in this platform tool-kit.

WP05 – Bioprocess optimization:

In Bioprocess optimization in WP05, newly developed insect cell lines were exploited to create new production cell lines for protein targets from a range of human pathogens including viruses such as HCMV, influenza and rotavirus. Notably the HCMV target may be used for subunit vaccine design in the future. In a comparative study, production with these new insect cell lines and classical infection with recombinant baculoviruses (BEVS) was analyzed. On the metabolomics level, the fluxome of industry-relevant cell-line was scrutinized and correlated to the productivity of high-value recombinant targets including antibodies. Further achievements were the successful miniaturization of baculovirus expression using a micro-fermenter robot and the extrapolation of process parameters from micro- to fermenter scale. Moreover, yeast strains (*P. pastoris)* expressing protein products relevant for enzyme replacement therapy (ERT) were achieved.

WP06 – Dissemination, exploitation and valorisation:

Dissemination, exploitation and valorisation were covered in a dedicated work package (WP06), focused mainly on coordination of knowledge and its dissemination. The ComplexINC website

(www.complexinc.eu) was created as management and communication platform for the entire consortium as well as an interface with the outside world. Press releases have been issued announcing the launch of the project, numerous peer-reviewed articles including highest impact papers were prominently featured by press and news releases on the website. ComplexINC and its technology platforms have been presented at a large number of meetings in talks and presentations, including plenaries and keynotes, and posters at international congresses. Synergies with an FP7 communication project (CommHere) to promote dissemination were established. Moreover, a business plan was prepared in WP06 for successfully launching ComplexINC spinoff Geneva Biotech.

3 Description of main Scientific &Technological results/foregrounds

Aberrant function of our own proteins can cause severe diseases including cancer, Parkinson's, Alzheimer's and others which are among the major killers in the populations of Europe. At the same time, infectious agents such as viruses attack humans using similarly complex proteins as our own. Research and development aimed at alleviating aberrant protein function requires the availability of abundant and high-quality samples of the protein specimens causing the problem, and the therapeutic drugs that we use to combat disease are increasingly being comprised of complex proteins themselves.

The inception of recombinant protein production technologies has had a decisive impact on both life sciences and drug discovery. Through recombinant overproduction, proteins that had been elusive before can now be produced in the quantity and quality required to decipher their structure and function. This set the stage for designing and validating intervention strategies to modulate their activity, which can now be translated into medications and therapies, through an elaborate process called the drug-discovery pipeline.

Recombinant protein overproduction also enables genetic engineering of complex protein specimens that themselves can be used as therapeutic drugs. For example, viruses contain protein materials, which typically enclose and protect the viral genetic information required by the virus for replication upon infecting an organism. Recombinant production of the protein shell components of a virus in the absence of its genetic material results in virus-like particles (VLPs). VLPs structurally resemble the live virus, but do not contain the genetic material which is essential for infectivity. Such recombinant VLPs are presently being used to vaccinate humans against numerous diseases including influenza, and also cancers caused by viral infection (i.e. cervical cancer caused by papilloma virus).

Research and development in academia and industry has long focused on small and isolated protein entities. Frequently, these are only the fragments or domains of a protein of interest that can be efficiently produced by the recombinant host systems available, most typically the prokaryote *Escherichia coli*. Consequently, expression in *E. coli* dominated the protein production field. For example, >90% of all entries in the protein database (PDB) were produced in *E. coli*. For therapeutic proteins, *E. coli*-based production has proven to be effective, efficient and relatively inexpensive. For example, the wide availability of *E. coli*–produced recombinant human insulin has revolutionized the treatment of patients suffering from diabetes, greatly reducing its cost and also alleviating allergic reactions hitherto occurring when insulin extracted from pigs or cows had been administered as a therapeutic.

Massive investment over the last four decades into optimizing protein expression in *E. coli* has resulted in a large collection of genetically modified strains with widely diverse properties, each tailored to particular experimental needs. An enormous and growing assortment of circular DNA molecules (plasmids) exists for expression in *E. coli*, to enable the production of heterologous proteins encoded by genes which are inserted into these plasmids. *E. coli* expression has become commonplace in virtually every molecular biology laboratory in public and private research institutions.

The explosion of biological data which became available recently thanks to breakthroughs in genomics, proteomics and interactomics research, however, has revealed that the actual actors of cellular processes, particularly in eukaryotes, are not isolated proteins but rather complex multiprotein assemblies composed of up to a dozen or more protein subunits. Many of the individual eukaryotic proteins and most of the protein complexes, including therapeutically important assemblies such as VLPs cannot be produced efficiently in *E. coli*. The failure of the prokaryote *E. coli* to produce complex human protein specimens is not surprising given that eukaryotic cells and their component proteins are in general much more complex than prokaryotes and demand different folding, processing and post-translational modifications (i.e. decorations with small molecules such as sugars or phosphate groups conferring activity) that the *E. coli* protein production machinery cannot support. Consequently, eukaryotic protein expression systems have entered centre stage. Some of the most powerful such eukaryotic multi-expression systems available today were created and developed by members of the ComplexINC consortium. Notably MultiBac, a multiprotein expression system based on baculovirus/insect cells (BEVS) which is a core technology platform of the ComplexINC work program, has catalyzed discovery in more than

a thousand laboratories world-wide since its inception in 2004.

By building on unique and world-leading know-how and expertise in the consortium, the ComplexINC project set out to decisively improve the underlying technical operation of the four major eukaryotic systems in action today (BEVS, mammalian, *S. cerevisiae*, *P. pastoris*), to substantially augment the performance of these eukaryotic systems for manufacturing complex biologics, through discoveries and innovations arrived at in the work plan of this consortium. Major advances and project breakthroughs are summarized in the following.

New high-throughput pipelines and took-kits for producing complex biologics

A major objective of the ComplexINC work programme was to generate new high-throughput pipelines and took-kits for eukaryotic complexomics. Automation is becoming a crucial prerequisite for accelerating protein research. Driven by drug discovery programs and structural genomics initiatives, there has been significant investment and progress in many laboratories in academia and industry to adapt parallelized techniques for expression construct screening, cell culture maintenance and protein purification. Considerable progress has been made in automating parallel protein expression in the host *E. coli*, mostly for single proteins and individual domains. In stark contrast, automation of eukaryotic expression, in particular of complex specimens with many subunits, had been markedly lacking.

ComplexINC took these technology developments to an entirely new level. Closely cooperating, the partners succeeded in the creation and roll-out of new and enhanced tool-kits for enabling HT complex biologics production in powerful eukaryotic production systems. New DNA reagents were created and efficient robotic scripts and protocols developed for generating tailor-made and custom designed tool-kits for baculovirus expression (BEVS), mammalian expression and yeast expression. The high-throughput routines thus developed were implemented in a robotic setup and thoroughly validated with a prioritized target list of protein specimens which were of vital interest for the partners in the consortium, including kinase complexes, enzymes for replacement therapy of diseases, eukaryotic transcription factors, viral protein complexes and a number of potential vaccine candidates.

Importantly, an array of heteromultimeric VLPs corresponding to influenza virus variants were generated in a parallelized fashion in high-throughput using this newly created platform. VLPs of papilloma serotypes, CMV VLPs, chimeric VLPs and further artificial pseudo virus-like structure forming protein specimens such as the adenovirus dodecahedron were produced, yielding new and exciting approaches to provide chimeric and multivalent biologics. These next generation biologics already attracted commercial interest as potential vaccine candidates.

A further notable foreground which may transform in particular influenza research resulted from the inception of a novel technology concept, ComplexLINK, which mimics the approach certain viruses, such as coronavirus, are exploiting to realize their proteome. Here, proteins belonging to a chosen system are produced from a long multigene construction which gives rise to a so-called polyprotein. This polyprotein is then processed into individual components by a highly specific protease which is also encoded in the polygene. ComplexLINK was successfully applied to produce influenza polymerase, which is a central, druggable target holding major promise for developing disruptive new treatments to combat flu. These studies, involving some of the leading laboratories in influenza structure research, were disseminated in *Nature*, the leading periodical of research in the life sciences. Of note, all today available near-atomic structures of complete influenza polymerases including fluA, fluB and fluC relied on ComplexINC technologies (MultiBac, ComplexLINK) to produce the specimens studied, testifying to the performance of the R&D program in this consortium. Both MultiBac and ComplexLINK technologies continue to be widely exploited to drive ambitious research programs at the partner laboratories of the ComplexINC consortium. Many protein biologics, often for the first time, were successfully produced in the consortium, including but not limited to factors of the complement system, human gene regulatory complexes implemented in neural diseases or cancer genesis, lysosomal proteins, therapeutic antibodies, metabolic sensors and novel protein and peptide display scaffolds.

Innovating baculovirus/insect cell expression

The baculovirus insect cell expression system (BEVS) has firmly established itself as a powerful system for many applications, including production of protein biologics, notably human proteins and protein complexes. However, significant unmet technological challenges were unsolved for BEVS, including inherent folding, assembly and quality control obstacles for challenging protein targets and large complexes. Further, use of BEVS for pharmaceutical production was impeded by genetic instability of the baculovirus when production was scaled up. ComplexINC tackled these challenges by creating and validating synthetic viral nanosystems (SVNs) that decisively enhanced protein production, including novel synthetic baculoviral genomes that were created and validated in the project.

For instance, new folder modules were constructed and integrated into the baculoviral genome to create customized genomes to overcome the handicap of previously available baculovirus expression systems which relied on the folding and processing machinery available in the infected insect cells to assist maturation of heterologous protein products produced. Effective production of therapeutically important proteins such as kinases relevant for anticancer drug design crucially depends on very specific chaperone systems containing numerous protein components. ComplexINC created a highly versatile tool-box were a series of folders and helper proteins were integrated in a format fully compatible with the enhanced synthetic BEVS and HT tools developed in the consortium. SVNs that comprise a variety of HSP70 and HSP90-based chaperone complexes were combinatorially generated. These were stably integrated in a combinatorial manner into customized baculoviral genomes. Their efficiency in assisting folding was validated with numerous test cases including difficult-to-express kinases and kinase complexes. Moreover, SVNs were created which modified the glycosylation pattern of secreted proteins. In addition to SVns that provide humanized glycosylation (SweetBac), variants were developed and successfully applied which entirely remove sugars from the surface of proteins of interest. A whole range of such “factories” was successfully put in place throughout the ComplexINC project.

A particular objective in ComplexINC was to create new baculoviral genomes by applying synthetic biology approaches. Baculovirus is a highly efficient delivery system for recombinant genes into eukaryotic cells, with great impact on the production of eukaryotic proteins, including high-value drug targets for pharmaceutical development. This method is also suitable for producing vaccines against influenza, herpes, cervical cancer and others. More recently, baculovirus has emerged as a versatile tool for gene therapy.

These applications, at the forefront of modern biology, rely on a large double-stranded DNA genome derived from a natural baculovirus. This genome has been intensively researched, mainly by entomologists. Genes essential for baculovirus propagation in nature and in cell culture have been determined, as well as non-essential genes, and genes which impede applications in the laboratory. Several genetic alterations of the wild-type viral genome have been performed using classical knock-out technologies, to improve protein expression or maturation, or even to facilitate upstream and downstream processing. However, based on the large and complex nature of the baculovirus genome, even these relatively modest manipulations required tedious and labour intensive effort by specialists. Using available technologies, it was not possible to exploit the full potential of the baculovirus system.

ComplexINC proposed to reverse this approach by designing and assembling a partly or fully synthetic baculoviral genome that is functionally tailored for rapid recombinant multigene insertion, flexible reengineering, combinatorial gene delivery and optimal protein production. Building on a significant body of research in this direction including data mining and phylogenetic analyses available in the partner laboratories, a virtual minimal genome of a synthetic baculovirus was compiled. Starting from this blueprint, new, partly synthetic hybrid baculoviruses were created for optimized, streamlined, highly versatile format for gene transfer and optimal recombinant protein production. Computational biology and comparative genome bioinformatics were utilized to delineate essential and non-essential regions in coding and non-coding sequences of the genome, and to create a refined blueprint for optimally arranging essential genes and non-coding regulatory elements in the optimized minimal baculovirus genome.

Protoypes of these designer genomes were synthesized in large sub-fragments, and recombination technologies were used to assemble defined sets of synthetic DNA fragments. The

resulting novel synthetic baculoviruses (SynBac) were rigorously validated using a selection of important targets, including human proteins and complexes and biologics of pharmaceutical value Significant progress towards a fully synthetic, optimized and minimized genome has been made in ComplexINC, and the first versions of the SynBac genome have already reached the market. In addition to these new baculoviral reagents, modified insect cell lines which allow direct integration of DNA cargo into the insect cell genome were developed in the consortium, to add to the possibilities for multi-protein co-production.

BEVS miniaturization and microprocess optimization

Miniaturization is a top priority in contemporary protein science. If cell culture volumes are reduced, processes standardized, and parameters optimally controlled, then it becomes feasible to increase parameter space for bioprocessing optimization to multidimensional matrices, optimizing for reagents (synthetic gene variants, cell lines, baculoviruses), physical parameters (growth temperature, media, supplements, oxygenation etc.) and biological parameters (virus performance, protein production levels, solubility levels, etc.) concomitantly, thereby increasing speed and throughput of experiments decisively. The ComplexINC consortium successfully miniaturized explorative biologics production and developed meaningful multidimensional matrices for process control and optimization for ComplexINC's baculovirus/insect cell system technologies. A particular focus of this project was to establish meaningful parameters in micro-scale which enable predictive planning of large scale experiments. A micro-fermentation table-top robot was put in place (collaboration with TAP Biosystems, UK). Successful proof-of-concept for parameter identification in micro-scale and extrapolation to fermenter-scale was achieved and will be communicated in due course.

Innovating yeast expression

For more than two decades, *S. cerevisiae* and *P. pastoris* expression platforms have served as workhorses for recombinant protein production in academia and industry. Both organisms share many beneficial traits such as rapid and inexpensive cultivation to very high densities in shaker flasks and fermenters, high yield and high productivity of protein products using stable, durable strains, cost-effectiveness, and suitability for the production of isotopically labelled proteins.

Being eukaryotic single-celled organisms, yeasts post-translationally process proteins similarly to mammalian cells, can handle disulfide-rich proteins, assist protein folding, and can glycosylate and secrete proteins. In addition, the GRAS (generally regarded as safe) status of both yeasts streamlines the therapeutic use of its protein products.

S. cerevisiae has been successfully used for producing secreted biologics for human therapy. Examples include insulin, albumin, growth hormone and human granuloctyte colony-stimulating factor. *S. cerevisiae* is particularly well-suited for proteome-wide high-throughput screening applications based on existing homologous-recombination-based cloning methods.

P. pastoris is characterized by a respiratory metabolism that allows growth to very high cell densities and a particularly efficient secretory pathway with comparatively well-understood *N*-linked glycosylation. Recently, the first glycoengineered *P. pastoris* strains have been engineered that are capable of performing certain human-like ("humanized") glycosylations. These glycoengineered strains have considerable potential to mature into a cost-effective alternative to the significantly more expensive mammalian cell lines. To date, *P. pastoris* has been successfully used for producing biologics including hormones, antibody Fab fragments against hepatitis B virus, membrane proteins, coagulation factors, complement components and others.

Both, *S. cerevisiae* and *P. pastoris* were pronouncedly handicapped regarding the production of complex biologics that require insertion of more elaborate multigene production circuits into these organisms. As a consequence, high-throughput screening applications in *S. cerevisiae* were limited to mostly single gene expression. Disproportionately labour-intensive interventions were necessary to achieve even modest goals, such as binary systems. In *P. pastoris*, the necessity to ensure efficient linearization of the expression construct was crucial for genome integration and

posed an unmet challenge for large, multigene constructs containing complex multi-expression genetic circuits.

Therefore, despite all its attractive features and benefits, expression in *S. cerevisiae* and *P. pastoris* was limited by the lack of appropriate tools to facilitate rapid and flexible assembly of multicomponent genetic constructions in a HT format. These versatile expression systems had not yet realized their full potential for biologics production.

The ComplexINC consortium addressed this challenge and created these tools. ComplexINC partners designed, implemented and validated HT compatible tool-kits containing appropriate gene regulatory elements and selective markers for *S. cerevisiae* (MultiYeast system) and *P. pastoris* (MultiPichia system). The performance of *S. cerevisiae* and *P. pastoris* expression platforms were decisively enhanced and thus are now competitive for a large number of applications including human therapeutic protein production, from micro- to fermenter scale.

MultiYeast retained all previous advantages of *S. cerevisiae* expression systems, such as fast *in vivo* cloning of single subunit constructions by homologous recombination. Concurrently, MultiYeast enhances productivity and scope of this system by alleviating sequence constraints and facilitating the rapid and flexible generation of multi-gene expression constructs. MultiYeast thus supersedes previous approaches to implement co-expression strategies in *S. cerevisiae* in the context of structural genomics which were obstructed by a large and unwieldy effort (use of several plasmids with many different selectable markers) to achieve comparatively modest goals (co-expressing binary complexes). ComplexINC addressed with MultiYeast essential areas including production of medically important human membrane protein assemblies and, notably, human complement factors.

ComplexINC further created the foreground MultiPichia for *P. pastoris*. Implementing and rolling out MultiPichia to the larger life science community and the pharma/biotech industry is now poised to place *P. pastoris* as a viable and cost-effective alternative expression host system for a large number of therapeutics, especially complex glycosylated proteins which are currently produced by significantly more expensive technologies. Specifically, a series of novel glycoengineered *P. pastoris* yeast strains capable of performing the complex glycosylations found in lysosomal proteins were engineered with MultiPichia. This innovation is now exploited by an SME partner (Bioingenium) on an industrial scale.

By creating MultiYeast and MultiPichia, ComplexINC aimed to boost the performance of these yeast expression platforms to optimally capitalize on their established beneficial characteristics for industrial production, also for important complex biologics that are now entering centre stage. We anticipate that the now readily available MultiYeast and MultiPichia systems will have major impact in areas of production where *S. cerevisiae* and *P. pastoris* have been traditionally strong (e.g., therapeutic antibodies and hormones), while at the same time opening entirely new avenues for exploiting these systems.

Biologics production by metabolic engineering in *S. cerevisiae*

Natural products have played a very prominent role in the treatment of human disease throughout history. Often, natural products were the starting point for discovery and development of efficient therapeutics. However, production of these compounds at pharmaceutical scale can encounter insurmountable obstacles caused by their chemical complexity. Low yield and high production costs are a consequence. An alternative is isolation of these products from their natural source. However, this likewise is often hampered by low yield, associated high cost and in many cases a harmful effect on the environment.

Synthetic biology and metabolic engineering of whole pathways for the heterologous production of natural products can provide solutions to these problems, and methods to develop recombinant microorganisms capable of producing chemically challenging biological materials are evolving. Reconstruction of heterologous pathways in genetically and metabolically well-characterized organisms can permit large-scale production by fermentation of highly complex small-molecule scaffolds, starting from inexpensive carbon and nitrogen supplies, through (large-scale) fermentation processes.

The utility of implementing pathway engineering is evident form a humane, pragmatic, and economic point of view. For example, it would take six 100-year-old Pacific yew trees to provide enough Taxol (paclitaxel) to treat one single cancer patient. Metabolic engineering can contribute to forestall emergencies such as the breakdown in supply of the antimalarial drug artemisinin in 2004 due to massive seed shortages, catastrophically impacting third-world nations as no alternative sources of this drug were available at that time. The market volume of drugs that could be manufactured by metabolic engineering exceeds 30 billion €. Furthermore, the progressive depletion of fossil energy sources has created a shortage of molecular building blocks for organic chemistry synthesis and modification of natural products, rendering the development of microbial factories of small organic molecules all the more relevant.

S. cerevisiae is at the forefront of this emerging field. It is a genetically well characterized and tractable organism into which reconstruction of complete biosynthetic pathways is likely feasible and significant progress has been made. Successful examples include yeasts that produce taxol, artemisinin, and zeaxanthin. In all reported cases, the assembly of complete pathways relied on essentially linear and inflexible conventional cloning methods, or on more efficient but ad hoc strategies such as the DNA assembler method, which relies on simultaneous construction of a multi-gene plasmid by yeast homologous recombination. DNA assembler, arguably one of the most efficient methods for pathway engineering to date, is severely compromised by the requirements to avoid plasmid instability derived from the recombination of multiple direct repeats in yeast, thus limiting multiple use of efficient regulatory elements in the same artificial multigene expression circuit defining the pathway.

ComplexINC provided to the community tools to accomplish multigene plasmid assembly with unprecedented ease and maximum flexibility. Rapid and combinatorial construction of entire pathways is now feasible by using the MultiYeast tool-kit in combination with the high-throughput DNA assembly protocols honed in the consortium. This now allows facile construction of pathways and concomitantly provides the option of exploratory dissection of the pathway by establishing, in a combinatorial manner, the effects of each pathway component and their combinations on product yield. Metabolic processes can then be optimized by permutation or replacement of individual pathway components and rewiring those into the circuit, enhancing speed and accuracy to test hypotheses and implement modifications to optimize a target metabolic pathway of choice.

The MultiYeast tool-kit substantially augments existing options for pathway engineering in *S. cerevisiae*. Pathway engineering will also make available the option to screen or rationally add any desired genetically encoded functionality to the genes for the pathway of choice. For example, genes encoding factors required for activating a pathway component can be incorporated in this way. The possibility to supply gain-of-function genes will be an invaluable addition to the established loss-of-function approach (gene deletion) in *S. cerevisiae*, by which interfering genes are eliminated from the yeast background.

As a proof-of-concept, ComplexINC implemented a MultiYeast approach to metabolic engineering by constructing a pathway for the ergothioneine pathway. Ergothioneine is a histidine betaine derivative with a thiol group attached to the C2 atom of the imidazole ring. It plays an important role in the control of intracellular redox potential and metal homeostasis, and in inactivating electrophilic toxins. Understanding its chemistry is essential for the development of therapies against infectious and inflammatory diseases and cancer. Neither humans nor yeast can synthesize ergothioneine. Humans absorb it from the diet and accumulate it in specific tissues, such as liver, kidney, blood cells and the central nervous system. The complete ergothioneine pathway consists of five genes as demonstrated *in vitro* with purified enzymes from *M. smegmatis*. Using MultiYeast, ComplexINC partners successfully engineered a novel *S. cerevisiae* strain capable of producing essential components of the ergothioneine pathway.

A combinatorial BOOST library to further enhance biologics production

It is widely recognized that expression of mammalian proteins in prokaryotic hosts often fails because of the absence of eukaryotic chaperones or necessary enzymatic modifications. Both *S. cerevisiae* and *P. pastoris* are capable of supplying many of these folding and stabilizing protein factors and modifications (e.g., phosphorylations, glycosylations). However, mammalian proteins may require specific chaperonins, more sophisticated modifications, or simply exhaust the

available cellular complements when over-expressed.

The ComplexINC consortium assembled a definitive solution to overcome this burdensome limitation by creating and exploiting a combinatorial enhanced expression boost library (compatible with MultiYeast and MultiPichia), comprising eukaryotic chaperones, folding co-adjuvants and modification enzymes. This BOOST library now allows systematic co-expression screening of multiple fungal and mammalian genes for improved production of the target biologic of interest. The library is designed such that it will be fully compatible with the MultiYeast and MultiPichia tool-kits carrying the gene(s) of interest which were established in the consortium.

Expression screening of the resultant combinatorial fusion constructs permits the optimization of the production parameters of a particular target protein of interest through co-expressing these library members in a combinatorial fashion in small-scale in high-throughput. Identification of successful library members enhancing target protein production will then allow up-scaling to the production volumes desired.

Finding the appropriate complement of folding and stabilizing helper molecules may prove absolutely crucial for efficiently producing human protein complexes as therapeutics, since higher eukaryotic proteins typically require specific modifications or chaperones which are not naturally present in *S. cerevisiae* or *P. pastoris*. The future success of these microorganisms as modern factories for the production of human biologics will depend on the possibility to adapt to the needs of the more challenging next generation products ahead.

Areas that will profit are glycoengineering, which may be considerably accelerated and holds promise of substantial health improvement and economic revenue; secretion of heavily modified proteins that may require co-expression with specific proteases; and therapeutic antibody production, whose chains, glycosylation, and optimum level of production could be tuned for efficacy by using the BOOST library.

The combinatorial approach will be instrumental for broadly enhancing the production of challenging human proteins, biologics and therapeutics that have specific folding and post-translational modification requirements, in both *S. cerevisiae* and *P. pastoris*.

Innovating mammalian cell expression

Protein therapeutics has proven to be very effective at treating a number of human diseases, and continues to be the fastest growing sector of the pharmaceutical industry. Animal cell expression systems are particularly suited for the production of large, complex protein biologics requiring post-translational modification for their biological activity. Chinese hamster ovary (CHO) cells are among the most widely used mammalian hosts for the production of biotherapeutics. Cultivation of these cells in industrial fermenter scale is well established and the list of successful therapeutic protein drugs produced in CHO cells is continuously growing. Cell lines of human origin are also being pursued and offer advantages regarding the glycosylation pattern of their products.

CHO production cell lines with stable high-level performance contain the recombinant transgene at a genetically stable hot spot for transcription. Extensive screening for isolating such cells lines can considerably slow down the drug development process. Isolating highly productive cell lines for antibodies or Fc-fusions is relatively straightforward, but can be quite difficult for other protein classes. A significant challenge for producing any protein with mammalian cells remained in the difficulty to insert recombinant genes in a well-producing spot in the host genome. Only a small number of cell lines were commercially available that offer choice over sites for genomic insertion. An extensive selection of mammalian cell lines, ideally tailored to the specific requirements of a particular protein target to be produced, was largely lacking. This was partly due to a lack of tools for rapidly generating cell lines containing defined and well-producing spots in their genome for integrating one or more genes of interest.

ComplexINC exploited a superior new method for generating stably modified master cell lines which contain an integration site at a validated, well-producing hot-spot which is engineered to accept foreign genes of interest into a locus occupied by a fluorescent protein marker in the master cell line. High-level production master cell lines are isolated by preparative sorting of cells expressing green fluorescent protein from the site of integration on the genome. Different master

cell lines are selected for high, medium and low level expression of the fluorescent marker, and thoroughly validated by means of monitoring fluorescent protein expression over generations. Then, the green fluorescent protein gene can be rapidly exchanged for a heterologous gene of interest by using recombinase-mediated cassette exchange (RMCE). In this way, novel production cell lines can be quickly and easily obtained from the resulting master cells. Genome engineering by RMCE thus allows inserting transgenes of interest precisely into a defined expression hot-spot of the host cell genome. Glycosylation mutant CHO Lec cell lines for suspension culture were successfully established with this system for production of glycoprotein suitable for crystallization and structure determination by X-ray diffraction methods.

In the ComplexINC work plan, FACS and RMCE were used to establish an assortment of production cell lines tailored for secreted target protein biologics. Glycosylation mutant and normal cell lines producing lysosomal enzymes for enzyme replacement therapy (ERT) were established. For the first time, this allowed correlation of the activity and targeting into recipient cells of enzymes with complex or high-mannose type glycans. In the consortium, the resulting information was used to guide the targeted development of *P. pastoris* strains for production of lysosomal enzymes. This provided a unique opportunity to compare the integrity, activity and bioavailability of these therapeutically important lysosomal enzymes that were produced in *P. pastoris* or CHO cells, respectively.

Co-expression of accessory factors that modify intracellular pathways in CHO cells can be instrumental for obtaining better product yields and for improved post-translational modifications. Cellular pathways including apoptosis, unfolded protein response and secretion have been targeted successfully. Co-expression of a recombinant antibody with three accessory factors (two cell cycle regulatory proteins and a growth factor), maximized product yield upon transient transfection of a human cell line. For stably transfected cell lines, such co-expression of several factors would be very difficult to establish by previous technology.

Bioprocess optimization and quality control - the industrial perspective

Scaling-up expression is a major focus of biologics production, especially in pharma and biotech industries where protein biologics are often required in large quantities, for example for use as biotherapeutics. For complex eukaryotic biologics production, reliable methods are currently lacking that would allow for meaningful extrapolation from micro-scale, matrix-based optimization experiments and convert these results into a reliable predictor for large (pharmaceutical) scale production.

An eminent focus of ComplexINC was the validation of its newly developed technologies in the context of large-scale production. Large-scale production of a prioritized selection of biologics in ComplexINC was carried out to demonstrate the utility of our innovations for the pharmaceutical market. Of particular interest was the question whether selected parameters gained from controlled experiments in the laboratory, preferentially at small or very small scale, and determined in a highly parallelized setup, can be related to outcomes or provide useful hints when moving to large-scale production. A particular focus of this project was to establish meaningful parameters at micro-scale to enable predictive planning of large scale experiments. This was successfully carried out with the production technology innovations developed in ComplexINC by judiciously choosing particularly promising biologics as production targets and taking advantage of the large know-how and infrastructure available in the consortium for pharmaceutical scale production, and in particular the process optimization expertise.

Large-scale production of biologics using new technologies developed in ComplexINC was a paramount aim of this consortium. SME Partner IBET has an outstanding track record for controlled large-scale production of biologics including GMP level production. IBET has accumulated know-how and developed tools for process design based on hybrid modelling (mechanistic fermentation plus fuzzy logics) and intelligent downstream processing. ComplexINC exploited this know-how to thoroughly validate new cell lines and baculovirus reagents using selected protein targets of pharmaceutical interest. Optimization strategies for cost-efficient large-scale production included producing, extracting and organizing process development data using bottom-up systems biology tools and integrative ways of screening for industrially robust production. Highly predictive space-frame definitions for quality by design (QbD) implementation

will be combined with deducing models for larger scale process control, which were validated at larger (5 to 50L) pilot plant scale.

By using cutting-edge bioprocessing approaches, and capitalizing on unique competence in this area, ComplexINC thoroughly scrutinized next generation tools and innovations achieved in the work plan, on a scale relevant to pharmaceutical production. By providing next generation biologics and biotherapeutics, ComplexINC compellingly validated its technology platforms in an industrial setting.

Accelerating Enzyme Replacement Therapy (ERT) for Lysosomal Storage Disease

Lysosomal storage diseases (LSDs) are a group of approximately 40 rare inherited metabolic disorders that result from defects in lysosomal function. LSDs occur when the lysosome, a specific organelle, malfunctions. Fabry disease, Tay-Sachs disease, Gaucher disease and Pompe disease are examples for LSDs. The discovery of the lysosome as the organelle for intracellular digestion and recycling of macromolecules in the 1960s was the breakthrough that led to an understanding of the physiological basis of LSDs. These diseases are typically caused by deficiency of a single enzyme required for the metabolism of lipids, glycoproteins or mucopolysaccharides.

LSDs are comparatively rare but grave genetic diseases, affecting mostly children who often die at a young and unpredictable age if untreated. Diverse and often severe symptoms are associated with LSDs. For example Fabry disease, an LSD characterized by deficiency in α-galactosidase, is associated with pain crises, fever, fatigue and conditions such as ventricular hypertrophy. Fabry disease has a prevalence of roughly 1 in 50,000 in the European Union with associated costs of currently 200,000 € per patient per year, totalling more than 2 billion € annual health expenditures. There are no cures for LSDs currently available. A significant advance for the treatment of LSDs came through enzyme replacement therapy (ERT). In ERT, the enzyme lacking in the patients is replaced by intrevenous infusion containing high-quality preparations of active enzyme.

Recombinant expression of biotherapeutics in high quality is critical for the treatment of LSDs by ERT. Overproduction in particular in yeast combines the eukaryotic capacity of performing post-translational modifications, notably glycosylation, with the microbial ability to grow to high cell densities in inexpensive media. However, many therapeutic proteins require specific glycosylation patterns. A correct glycosylation pattern is a key parameter to ensure correct folding and function of a given protein biotherapeutic and its capacity to remain circulating in the bloodstream, as well as for interacting with the appropriate receptors. Yeast glycosylation is mainly mannose type (in contrast to mammalian-like complex glycans). Therefore, proteins produced in yeast can cause a strong immunogenic response and rapid clearance from the bloodstream.

In ComplexINC, an innovative platform, GLYCOMAM, was developed containing a series of strains of the yeast *P. pastoris* with modified glycosylation patterns as compared to wild-type by exploiting the innovative MultiPichia approach. A key part of biotherapeutics production with GLYCOMAM was rigorous quality control by glycan content analysis as well as *in vitro* and *in vivo* biological activity assays. The strains developed within the scope of the GLYCOMAM platform were optimized for cost-efficient large-scale production.

By implementing the GLYCOMAM platform in conjunction with large-scale production of selected proteins, ComplexINC set the stage to produce sufficient amounts of suitable protein biotherapeutics for further tests and (pre)clinical studies towards treatment by ERT of the LSD Fabry disease.

A ComplexINC high-tech spin-off SME, Geneva Biotech

Importantly, ComplexINC had, as a major objective, to bundle advances and discoveries from the work program and accelerate their commercialization by founding a new high-tech start-up to market intelligent protein production technologies. This was successfully set in motion by the creation of the ComplexINC spin-off company Geneva Biotech (http://geneva-biotech.com/) which was subsequently successfully capitalized and today already employs 10 high-skilled FTEs and has an in-house R&D program for next-generation therapeutics.

The remarkable success of the ComplexINC consortium was readily acknowledged as the consortium and research outcomes received external expert evaluation when the 2nd Periodic Project Report submitted to the European Commission was reviewed. In the review, ComplexINC was classified as ‘exceptionally productive’, a ‘highlight’, a ‘success case’, and of ‘significant R&D breakthrough character’. A large number of peer reviewed publications including papers in highest-impact periodicals, several patent applications, a large number of news and press releases compellingly underscore the success of the consortium.

4 Description of the potential impact and the main dissemination activities and exploitation of results

The immediate impact of the ComplexINC consortium's achievements is recognized and already extensively exploited by experts from both academia and industry on a global level. By making available cutting edge expression tools for a large range of well-established expression hosts, a number of key requirements are addressed for many labs that need to tackle a multitude of challenging protein expression projects. Using the next generation of enhanced baculovirus genomes for example puts the global research community in a position to express and characterize a large number of high value disease related proteins for the first time that were hitherto next to impossible to be recombinantly produced in an active form. Along the same lines, the availability of fast and efficient tools such as MulitBac, MultiYeast and MultiPichia enable the international research community to tackle ever larger and more complicated recombinant protein production challenges. By directly addressing the very needs of researchers in academia and industry, ComplexINC developments not only speed up protein production projects, but also directly address the growing need in high value proteins that represent human drug targets, in particular ever larger multi-protein complexes for basic research and drug discovery in academia and industry. It is immediately apparent, that these cutting-edge enabling technologies thereby contribute significantly to provide our society with better drugs, proteins for enzyme replacement therapies and vaccines, faster and more efficiently.

Wide-ranging and highly effective dissemination activities were an integral part of the ComplexINC concept and significantly contributed to the impressive speed of implementing the consortium's achievements on a global scale in academia and industry. The ComplexINC website was active from the onset of the four year project and constantly updated to serve as a central and readily accessible information hub not only for internal communication but also and likely more importantly for the public community. Results were and still are being communicated during invited talks at key academic institutions as well as by oral and poster presentations at international scientific meetings to directly inform the scientific community about the latest results and the major progress that has been achieved.

Along the same lines, key decision makers in the global biotech and pharmaceutical industry were introduced to ComplexINC and its achievements during hundreds of presentations during face-to-face meetings and at partnering conferences. As an immediate outcome, beyond raising the awareness of the ComplexINC consortium, SME consortium members Crelux and Geneva Biotech could secure a large number of licensing and fee-for-service research contracts. The constant flow of contract research deals that are at least based in part on ComplexINC innovations resulted in a large number of permanent positions being generated for highly skilled personnel at the SMEs.

Further dissemination activities involved a large number of press releases issued by all consortium members, mailing of newsletters by Crelux to the global decision makers and distribution of project flyers.

A large number of publications in high profile peer reviewed scientific publications furthermore contributed significantly to the dissemination of the achievements and their adaption. Additional publications are planned for the future with a notable highlight in 2016 of a book being published by Springer Scientific "Advanced Technologies for Protein Complex Production" that was edited by the Madrid group of Cristina Vega. The publication contains, among others, contributions by most ComplexINC group members and will further enhance the visibility of ComplexINC and contribute to further dissemination of the results.

During the entire project duration and beyond, monitoring of novel IP was a cornerstone of the project and a large number of patent applications have and will be submitted to the patent organizations. The successful spin-out of a significant share of the IP has helped to generate the new SME Geneva Biotech. Conceiving the SME, writing the business plan and supporting the financing were an integral part of the ComplexINC project and significantly contributed to the generation of currently 10 permanent positions for highly trained FTEs.

At the final project meeting, all consortium members were informed during an exploitation workshop about how to best valorise the significant achievements of all ComplexINC members even more in the immediate future. It is anticipated that the already highly successful ComplexINC

achievements will be further enhanced and exploited by additional commercialization of the cutting-edge results.

THE COMPLEXINC CONSORTIUM:

EMBL - European Molecular Biology Laboratory Germany

Crelux – Crelux GmbH

HZI – Helmholtz-Zentrum fuer Infektionsforschung GmbH

iBET – Instituto de Biologia Experimental e Tecnologica

CSIC – Agencia Estatal Consejo Superior de Investigaciones Cientificas

Bioingenium – Bioingenium SL

Eurice - European Research and Project Office GmbH Germany

Geneva Biotech Sàrl