Pharmaceutical BIOPROCESSING

Opening the black box: Chinese hamster ovary research goes genome scale

Genome scale science for CHO production cells: genomes, -omics and big data. *acib, the Austrian Center of Industrial Biotechnology*, Vienna, Austria, 12–14 March 2014

The scientific community of Chinese hamster ovary (CHO)-cell scientists had a meeting to discuss the opportunities and requirements for a new era of genome scale science for CHO cells. With multiple genome sequences available, the challenge to the field is to use approaches of systems and synthetic biology to enhance quality and yield of recombinant therapeutics from CHO.

Nicole Borth
University of Natural Resources & Life Sciences (BOKU), Vienna, Austria
and
Austrian Center of Industrial Biotechnology (acib), Petersgasse 14, 8010 Graz, Austria
nicole.borth@boku.ac.at

The last 4 years have seen an explosion of available data on the genome of Chinese hamster ovary (CHO) cells [1] and more recently the Chinese hamster [2,3] as the most appropriate reference genome. As a result, cell biologists and engineers are now keen to dive into the molecular details of how this cell line achieves its high productivities and growth, to understand the basis for their variability and clonal heterogeneity and to increase our control of the behavior of cells [4]. The availability of genome sequence information comes with the lucky and timely coincidence of the emergence of efficient genome-editing tools, such as the CRISPR/cas technology [5], thus opening a window of opportunity during the next years to understand and control the cellular machinery and behavior of CHO cells on a gene per gene level. Hence the challenge to the field is to use approaches of systems and synthetic biology to enhance quality and yield of recombinant therapeutics from CHO.

For efficient use of the new tools and information, a concerted effort of the scientific community, both in academia and industry, is required. The workshop organized by acib together with CHOgenome.org [6] aimed at bringing together the key players in the field, to coordinate different approaches and to provide a summary of where we stand. In the following chapters the different sessions will be described along with a summary of the most important decisions reached during the pre-meeting discussions on how to bring the field forward.

Genomic sequences & what they tells us

The first session obviously was dedicated to the genomic information now available. Nathan Lewis (UCSD, CA, USA) presented data on differences between the sequenced CHO cell lines with respect to single nucleotide polymorphisms, indels and other structural variations, such as gene duplications or deletions [2]. Overall, these variants are clearly related to the respective history of the different cell lines, thus outlining their genetic 'family trees'. In a more detailed analysis for instance of pro- and anti-apoptotic genes, the expression pattern was consistently changed in CHO cell lines relative to the Chinese hamster, with a tendency towards overexpression of antiapoptotic genes and a reduction of pro-apoptosis. Clearly this is an advantage to cells in culture for their long-term survival. Nevertheless, many of the apoptosis-related genes, whether pro- or anti-apoptotic, showed duplications in several of

FUTURE SCIENCE part of fsg

the cell lines analyzed, without any distinct pattern emerging. The precise regulation appears to be very cell line specific, resulting in a balanced interplay of pro- and anti-apoptotic signals and thus also reflecting the phenotypic diversity of different CHO cell lines.

Karina Brinkrolf and Oliver Rupp (Bielefeld University, Bielefeld, Germany) then presented their results of a Chinese hamster reference genome sequenced from sorted chromosomes [3]. The knowledge about chromosome location of genes and scaffolds is especially important in view of the large number of chromosome rearrangements present in CHO cell lines: future sequencing projects of cell lines can use this information to identify the precise location of breaks and crossovers between chromosomes and to follow the emergence of new translocations during long term culture of cells.

The two Chinese hamster genomes now published are both in a similar state of assembly, consisting of several hundred thousand scaffolds each. The participants at the workshop saw the next most important step in an improvement both of the assembly and the annotation of the reference genome, which was unanimously agreed should be the Chinese hamster rather than any of the 'chromosomally re-arranged' CHO cell lines. Using the short reads generated by next-generation sequencing (NGS) technology, it is difficult to impossible to assemble regions of high repeat frequency and thus to generate a complete assembly consisting of a single scaffold per chromosome. An alternative approach here is the use of PacBio sequencing technology (Pacific Biosciences, CA, USA), which is able to generate reads up to 20 kb long. PacBio technology may not have the precision of other NGS techniques, however, in combination the highly parallel NGS results together with the long PacBio reads will enable a dramatically improved assembly. It was decided therefore to approach the scientific community and industry to contribute by crowd financing to the generation of additional long-read sequences at 20× coverage by PacBio (see CHOgenome.org for details on how to contribute). The next release of a new assembly is planned for the first half of 2015. This then will be the starting point for a highly curated annotation that in addition to genes, introns and exons also needs to contain transcription start sites, promoter regions and noncoding RNAs.

The -omics: tools & applications

The available genomic sequences now enable the development of better analysis tools to study different CHO phenotypes. Mike Betenbaugh and Amit Kumar from Johns Hopkins University (MD, USA) presented the proteomics database [7] now available through CHOgenome.org, which lists more than 6000 proteins expressed by CHO under various culture conditions, while Jong Youn Baik, now at the University of Delaware (DE, USA), presented his work performed at the Korean Institute of Science and Technology (KAIST; Seoul, South Korea) in Gyun Min Lee's group on the response of CHO cells under stressful conditions. Differentially expressed genes were identified during hyperosmosis, reduced temperature and butyrate addition, all of which influence specific productivity. Based on the gained knowledge, targeted engineering of cells enables improvement of cellular phenotypes. This was demonstrated by overexpression of HSP60 and HSP70, which enabled more rapid adaptation of cells to grow in serum-free conditions. Diethard Mattanovich (acib, Vienna, Austria) presented a comparative study of the behavior of *Pichia pastoris* and CHO cells producing two model proteins, HSA and a single-chain antibody, using transcriptomics, proteomics and image analysis. Samples were taken from chemostat cultures, an approach that improves the reliability and consistency of gene-expression patterns as the cellular environment is stable and consistent. It was shown that in addition to a response to increased productivity, both cell types also show protein-specific stress responses.

Big data

The generation of large data sets in systems biology presents the researcher with a new set of problems that require innovative tools for data-set handling, not the least of which is comparability and standardization.

Colin Clarke discussed the major statistical problems encountered with typical cell culture approaches, such as the low number of samples and replicates (usually two to three) in relation to the large number of analytes (e.g., 15,000 expressed genes) [8]. This results in low statistical significance of results leading either to a high false-negative or false-positive rate. Two possible approaches can reduce this limitation: one is to increase available data sets by uploading them to searchable databases containing information on analyzed -omics data linked to relevant phenotypic data; alternatively the biological significance of results can be increased by analyzing in parallel multiple systems levels (such as transcriptome, microRNome and proteome together). Removal of irrelevant data from the statistical analyses, such as genes that are not expressed, will also increase significance.

Another approach is the development of genome-scale models and their integration with multiple -omics data sets to enable better understanding and prediction of cell behavior, as presented by Nathan Lewis, as well as by Dong-Yup Lee from Singapore and Michael Hanscho from acib (Vienna, Austria), in

collaboration with Lars K Nielsen from the University of Queensland, Australia. A positive outcome of the meeting is a collaboration of these groups to merge their individual models into a consolidated version, which will be published in early 2015.

Dissection of the different molecular steps of protein production and growth also opens the possibility to redesign cell factories using approaches of synthetic biology. As succinctly outlined by David James from the University of Sheffield (Sheffield, UK), next-generation cell factories will have various properties, as befits the respective product and its quality attributes, with the ability to fine-tune precise expression levels of genes engineered to assist the cell in its specific production task [9].

The future

Kelvin H Lee from the University of Delaware (DE, USA) finally presented his vision of the future of biopharmaceutical production [10]. The interest in the genome sequence and systems biology data of CHO is caused by our new ability to understand the detailed biology of this cell which makes it such a successful producer of therapeutic proteins. The main focus nowadays is on product quality, as yields have increased over the last years to a level that causes many to consider CHO bioprocessing a mature technology. Nevertheless, the impact of detailed genetic and phenotypic changes on product quality and its variation are still not understood and require further research. The data and tools that should be available on CHOgenome.org for this purpose, both for academics and the industry, will contribute largely to the speed of developments in this field.

The general problems currently encountered in -omics studies are inconsistencies in gene IDs between the different databases available and the lack of curated annotation for CHO. While the latter will improve with time (see above), the problem that most online analysis tools are set up for human and mouse and not for CHO remains. To establish systems biology approaches as standard tools for cell line development and bioprocessing, CHOgenome.org will have to develop to provide the infrastructure required, including tools to compare multiple genomes and transcriptomes, hosting or at least linking to relevant publications and data sets, providing online analysis tools for CHO and also serving as a repository that enables researchers interested in a specific gene to immediately look for previous reports on whether and under which conditions this genes was expressed and at what level. For this reason, acib is supporting the effort of developing the website into a multifunctional port. First versions will be out for beta-testing toward the end of 2015, with bi-annual updates expected over the next 3 years. Feedback from scientists and users will be highly appreciated to make the website as functional and user-friendly as possible.

Financial & competing interests disclosure

The author has no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1 Xu X, Nagarajan H, Lewis NE *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* 29, 735–741 (2011).

2 Lewis N, Liu X, Li Y *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nat. Biotechnol.* 31, 759–765 (2013).

3 Brinkrolf K, Rupp O, Laux H *et al.* Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.* 31, 694–695 (2013).

4 Kildegaard HF, Baycin-Hizal D, Lewis NE, Betenbaugh MJ. The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. *Curr. Opin. Biotechnol.* 24, 1102–1107 (2013).

5 Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347–350 (2014).

6 Hammond S, Kaplarevic M, Borth N, Betenbaugh MJ, Lee KH. Chinese Hamster Genome Database: an online resource for the CHO community. *Biotechnol. Bioeng.* 109(6), 1353–1356 (2012).

7 Baycin-Hizal D, Tabb DL, Chaerkady R *et al.* Proteomic analysis of Chinese hamster ovary cells. *J. Proteome Res.* 11, 5265–5276 (2012).

8 Clarke C,Henry M, Doolan P *et al.* Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics* 13, 656 (2012).

9 Grainger RK, James DC. CHO cell line specific prediction and control of recombinant monoclonal antibody *N*-glycosylation. *Biotechnol. Bioeng.* 110(11), 2970–2983 (2013).

10 Baik JY, Lee KH. Toward product attribute control: developments from genome sequencing. *Curr. Opin. Biotechnol.* 30, 40–44 (2014).