Biotechnology Advances 34 (2016) 621–633

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow

Aydin Golabgir [a,1], Jahir M. Gutierrez [b,c,1], Hooman Hefzi [b,c], Shangzhong Li [b,c], Bernhard O. Palsson [b,c,e,f], Christoph Herwig [a,d,*], Nathan E. Lewis [c,e,f,**]

[a] Research Area Biochemical Engineering, Institute of Chemical Engineering, Vienna University of Technology, Gumpendorferstrasse 1a—166/4, A-1060, Vienna, Austria
[b] Department of Bioengineering, University of California, San Diego, CA, USA
[c] Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego School of Medicine, San Diego, CA, USA
[d] CD Laboratory on Mechanistic and Physiological Methods for Improved Bioprocesses, Vienna University of Technology, Gumpendorferstrasse 1a—166/4, A-1060, Vienna, Austria
[e] Department of Pediatrics, University of California, San Diego, CA, USA
[f] The Jacobs School of Engineering CHO Systems Biology Center, University of California, San Diego, CA, USA

ARTICLE INFO

Article history:
Received 26 September 2015
Received in revised form 21 February 2016
Accepted 28 February 2016
Available online 3 March 2016

Keywords:
Chinese hamster ovary
Bibliome
Meta-analysis
Bioprocess
Phenome
Statistical analysis

ABSTRACT

The scientific literature concerning Chinese hamster ovary (CHO) cells grows annually due to the importance of CHO cells in industrial bioprocessing of therapeutics. In an effort to start to catalogue the breadth of CHO phenotypes, or phenome, we present the CHO bibliome. This bibliographic compilation covers all published CHO cell studies from 1995 to 2015, and each study is classified by the types of phenotypic and bioprocess data contained therein. Using data from selected studies, we also present a quantitative meta-analysis of bioprocess characteristics across diverse culture conditions, yielding novel insights and addressing the validity of long held assumptions. Specifically, we show that bioprocess titers can be predicted using indicator variables derived from viable cell density, viability, and culture duration. We further identified a positive correlation between the cumulative viable cell density (VCD) and final titer, irrespective of cell line, media, and other bioprocess parameters. In addition, growth rate was negatively correlated with performance attributes, such as VCD and titer. In summary, despite assumptions that technical diversity among studies and opaque publication practices can limit research re-use in this field, we show that the statistical analysis of diverse legacy bioprocess data can provide insight into bioprocessing capabilities of CHO cell lines used in industry. The CHO bibliome can be accessed at http://lewislab.ucsd.edu/cho-bibliome/.

Contents

* Correspondence to: C. Herwig, Gumpendorferstrasse 1a—166/4, A-1060 Vienna, Austria.
** Correspondence to: N. E. Lewis, 9500 Gilman Drive, Mail Code 0760, La Jolla, CA 92093-0760, United States.
E-mail addresses: christoph.herwig@tuwien.ac.at (C. Herwig), nlewisres@ucsd.edu (N.E. Lewis).
[1] Contributed equally (ordered alphabetically).

http://dx.doi.org/10.1016/j.biotechadv.2016.02.011
0734-9750/

1. Introduction

Chinese hamster ovary (CHO) cells have been utilized for academic and industrial purposes since the 1950s (Tjio & Puck, 1958). Today CHO cells represent the preferred cellular factory for the production of important recombinant proteins and biotherapeutics (Walsh, 2014), including six of the top ten selling biotherapeutics in 2014 (Philippidis, 2015). In the early days of recombinant protein production, the complexity of desired products required a mammalian host. Thus, new CHO cell lines were developed and an entire new field of CHO bioprocessing was born (Jayapal, 2012). Throughout the history of CHO cell culturing, major technological advances have continued to expand its use in industry.

Over the past two decades, there has been a steady increase in the number of published studies on CHO cell culturing and bioprocessing. However, we still do not fully understand the factors that determine the optimal performance of CHO cells during culture. A deeper understanding of these factors may possibly be obtained from the retrospective analysis of large amounts of carefully collated and curated legacy data on CHO cells. Examples of questions that could be explored with the use of an organized repository of CHO bioprocessing data include the following. What are the main phenotypic differences across CHO cell lines (e.g., CHO-K1, DG44, DUKXB11, CHO-S, etc.)? How do the different culture conditions affect the performance of cells? Is it possible to predict titer, viable cell density (VCD) or viability over time given appropriate information on the cell line and culture conditions? Are there any significant differences between parental cell lines that ultimately translate into a sustained effect in culture performance? Answers to these and many other questions can have important implications on CHO cell bioprocessing and help improve recombinant protein quality.

In an initial step to explore such questions, we compiled and curated the literature between January 1995 and June 2015 and identified studies containing biotech-relevant data on CHO cells. In addition, we classified each article based on the type of data it contains. Next, we extracted the detailed experimental data from a sample of 74 articles (Keen & Rapson, 1995; Schumpp-Vonach et al., 1995; Zang et al., 1995; Matsuzawa et al., 1997; Mastrangelo et al., 2000; Nishijima et al., 2000; Takagi et al., 2000; Yamamoto et al., 2000; Kim et al., 2002, 2005, 2006, 2007, 2009; Kim & Lee, 2002a, 2002b, 2007, 2009; Oettl et al., 2003; Yoon et al., 2003; Fox et al., 2004, 2005; Yoon et al., 2006, 2007; Zhang et al., 2006; Choi et al., 2007; Chun et al., 2007; Figueroa et al., 2007; Fujiwara et al., 2007; Hong et al., 2007; Liu et al., 2007; Liu & Chen, 2007; Muller et al., 2007; Takuma et al., 2007; Yoon & Ahn, 2007; Yoon & Jeong, 2007; Ahn et al., 2008; Allen et al., 2008; Astley & Al-Rubeai, 2008; Baik et al., 2008; Becker et al., 2008; Hwang et al., 2008; Lee et al., 2008a, 2008b, 2008c; Majors et al., 2008a, 2008b, 2009; Ohya et al., 2008; Omasa et al., 2008; Tan et al., 2008; Zhao et al., 2008; Florin et al., 2009; Gigout et al., 2009; Hwang & Lee, 2009; Ju et al., 2009; Mohan & Lee, 2009; Nam et al., 2009; Peng & Fussenegger, 2009; Yee et al., 2009; Chong et al., 2010; Cost et al., 2010; Fan et al., 2010; Han et al., 2010; Kaneko et al., 2010; Kantardjieff et al., 2010; Malphettes et al., 2010; Rodrigues et al., 2010; Altamirano et al., 2000; Prentice et al., 2007; Shen et al., 2010; Sung et al., 2007; Sunley et al., 2008; Tharmalingam et al., 2008; Wulhfard et al., 2008) that contain relevant data on CHO cell phenotype, culture performance and production characteristics. Through several statistical analyses, we identified significant trends across bioprocesses corresponding to specific attributes, such as parental cell lines, culture conditions, growth rates, production capabilities, and other research parameters. While it is often assumed that technical variation and opaque publication practices limit research re-use in this field, here we successfully integrate data from diverse studies to quantitatively validate long-held assumptions in bioprocessing. Thus, the collation and analysis of the ever-increasing data on CHO bioprocessing can provide valuable insights for future bioprocessing efforts.

2. Methods

The methodology for realizing the presented quantitative review involved two main phases: 1) bibliographic compilation of scientific literature on CHO, or "bibliome", along with the extraction and digitization of the metadata to be used for 2) statistical analysis. Fig. 1 illustrates the step-by-step workflow and detailed descriptions about each step are provided in the following sections.

2.1. Identification and selection of publications

Thomson Reuters Web of Science™ was queried to search for all research articles published between January 1995 and June 2015 that contained the keywords "CHO cells" and/or "Chinese hamster ovary" in the title or abstract. Although the first mention of CHO cells in the scientific literature dates back to 1958 (Tjio & Puck, 1958) we focused here on studies published within the last 20 years to focus more on CHO cell bioprocesses that employ current technologies. This initial set of articles was then manually filtered by removing any study involving characterization of a recombinant protein expressed in CHO for basic science purposes (e.g., localization, interaction within the cell, effects of mutations or consequences of exposure to UV light or radiation).

2.2. Extraction of metadata

Most articles in our bibliome utilize graphs and time course plots to present the results. Thus, WebPlotDigitizer (Rohatgi, 2014) was used to digitize the data contained in the corresponding articles of our sample (see Supplementary File 1 for details of WebPlotDigitizer validation). From here on, the data extracted from these figures and the associated meta-features will be referred as the *Phenotype and Production Characteristics dataset*. In order to make the proposed analysis comprehensive, we manually annotated each article and figure with experimental details that may influence cell phenotypes of interest (see Table 1).

Data series were grouped based on their associated metadata to facilitate subsequent analyses. To do this, we assigned a bioprocess identification number (bioprocess ID) to each data series corresponding to the same experiment. That is, a bioprocess ID was assigned to each set of data series with the same values in each of the metadata features such as *cell line*, *culture media* and *culture conditions*. Many articles contain multiple bioprocess IDs since there can be more than one bioprocess in a single study (e.g. when a study tests the performance of two cell lines under same culture conditions). The raw data from these extracted bioprocesses are provided in Supplementary Files 2–3, hosted at Synapse (http://dx.doi.org/10.7303/syn5570798).