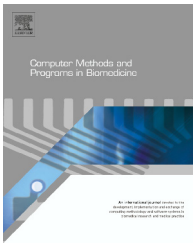
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 111 (2013) 228–248

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



On-line prediction of the feeding phase in high-cell density cultivation of *rE. coli* using constructive neural networks

M.C. Nicoletti[a,b,*], *J.R. Bertini Jr.*[c], *M.M. Tanizaki*[d], *T.C. Zangirolami*[e], *V.M. Gonçalves*[d], *A.C.L. Horta*[e], *R.C. Giordano*[e]

[a] *Depto. de Computação, UFSCar, S. Carlos, SP, Brazil*
[b] *FACCAMP, C.L. Paulista, SP, Brazil*
[c] *Depto. de Computação, ICMC, USP, S. Carlos, SP, Brazil*
[d] *Centro de Biotecnologia, Instituto Butantan, S. Paulo, SP, Brazil*
[e] *Depto. de Engenharia Química, UFSCar, S. Carlos, SP, Brazil*

ARTICLE INFO

Article history:
Received 24 September 2011
Received in revised form
18 December 2012
Accepted 11 March 2013

Keywords:
Immunology
Vaccine production
High-cell density cultivations
Bioprocess control
Constructive neural networks

ABSTRACT

Streptococcus pneumoniae (pneumococcus) is a bacterium responsible for a wide spectrum of illnesses. The surface of the bacterium consists of three distinctive membranes: plasmatic, cellular and the polysaccharide (PS) capsule. PS capsules may mediate several biological processes, particularly invasive infections of human beings. Prevention against pneumococcal related illnesses can be provided by vaccines. There is a sound investment worldwide in the investigation of a proteic antigen as a possible alternative to pneumococcal vaccines based exclusively on PS. A few proteins which are part of the membrane of the pneumococcus seem to have antigen potential to be part of a vaccine, particularly the PspA. A vital aspect in the production of the intended conjugate pneumococcal vaccine is the efficient production (in industrial scale) of both, the chosen PS serotypes as well as the PspA protein. Growing recombinant *Escherichia coli* (*rE. coli*) in high-cell density cultures (HCDC) under a fed-batch regime requires a refined continuous control over various process variables where the on-line prediction of the feeding phase is of particular relevance and one of the focuses of this paper. The viability of an on-line monitoring software system, based on constructive neural networks (CoNN), for automatically detecting the time to start the fed-phase of a HCDC of *rE. coli* that contains a plasmid used for PspA expression is investigated. The paper describes the data and methodology used for training five different types of CoNNs, four of them suitable for classification tasks and one suitable for regression tasks, aiming at comparatively investigate both approaches. Results of software simulations implementing five CoNN algorithms as well as conventional neural networks (FFNN), decision trees (DT) and support vector machines (SVM) are also presented and discussed. A modified CasCor algorithm, implementing a data softening process, has shown to be an efficient candidate to be part of an on-line HCDC monitoring system for detecting the feeding phase of the HCDC process.

* *Corresponding author at*: Depto. de Computação, UFSCar, S. Carlos, SP, Brazil. Tel.: +55 16 33518232; fax: +55 16 3351 8233.
E-mail addresses: carmo@dc.ufscar.br, carmo.nicoletti@pq.cnpq.br (M.C. Nicoletti).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.03.005

1. Introduction

Streptococcus pneumoniae (pneumococcus) is a Gram-positive bacterium responsible for a wide spectrum of illnesses varying from non-invasive, such as otitis and sinusitis, to severe invasive diseases, such as meningitis [11,48,52,71].

The surface of the bacterium consists of three distinctive membranes: plasmatic, cellular and the polysaccharide (PS) capsule. PS capsules may mediate several biological processes, particularly invasive infections of human beings [80]. Based on the structure and antigenic properties of the polysaccharide, *S. pneumoniae* strains can be divided into more than 90 structurally and antigenically distinct types, which have been classified into serogroups (designated by numbers) and further subdivided into serotypes (designated by letters) according to the Danish designation system – serotypes that are structurally and immunologically closely related belong to the same serogroup [39]. Serotypes differ from each other in relation to virulence, predominance and drug resistance [6]. Differences in the virulence characteristics between pneumococcal serotypes are important since available vaccines target only a few of the known serotypes. The epidemiological profile of a pneumococcal disease is strongly related to the prevalence of the corresponding serotype and varies with time, geographic region and individual's age [46].

Invasive diseases, such as meningitis, happen mainly during the first months after birth; a large number of pneumococcal diseases in children are associated with a restrict number of serotypes, which varies according to the geographical region. In spite of the existing over 90 different serotypes, recent data suggests that 11 serotypes are responsible for approximately 75% of worldwide *S. pneumoniae* infections. In USA and Europe the most prevalent serotypes in children are 6, 14, 19 and 23 [51] while in adults the most prevalent are 3, 4, 7, 8 and 14 [29]. In developing countries it is common to find serotypes 1 and 5 in adults [64]. In Brazil, 45% of pneumococcal diseases are caused by serotypes 1, 6B and 14. Serotypes 1 and 6B are prevalent in all ages; serotype 14 is responsible for a greater number of infections in children while serotypes 3 and 4 are more common in adults [13–15]. Prevention against pneumococcal related illnesses can be provided by vaccines. Currently all the existing pneumococcal-related vaccines are based on the free occurrence of the capsular polysaccharide or then have the PS conjugated to a protein aiming at enhancing protection.

One of the major difficulties in developing and commercializing a pneumococcal conjugate vaccine in Brazil is the large number of existing pathogenic serotypes in the country, namely 1, 3, 4, 5, 6A, 6B, 9V, 14, 18C, 19A, 19F, 23F, corresponding to approximately 80% of the clinically isolated strains [14]. It is well known that the process of producing a conjugated vaccine is highly elaborated and complex, involving many variables and whose final product yield is around 30%. The culture of the microorganism is a difficult process since not all serotypes grow in a similar way in the same medium. These difficulties, among many others, reflect on the final cost of the vaccine which makes almost impossible its broad use in developing countries.

There is a sound investment worldwide in the investigation of a proteic antigen as a possible alternative to vaccines based exclusively on PS [5,7,11,71,76]. A few proteins which are part of the membrane of the pneumococcus seem to have antigen potential to be part of a vaccine. Particularly the PspA has presented itself as one of the most promising proteins, mainly because it is externally exposed to the polysaccharide capsule of the *S. pneumoniae* [17,49,50].

The work described in this paper is part of a broader project that investigates the viability of the design of a new pneumococcal conjugate vaccine customized to Brazilian most prevalent PS serotypes. Three laboratories from the Biotechnology Center of Butantan Institute (S. Paulo – SP) have already been working for a few years on subjacent areas to the project, such as the design of efficient methods of bacterial polysaccharide production and purification [41–43,60] and the implications of considering as the carrier protein a surface protein from the *S. pneumoniae* capable of inducing protection [27,30,67,72,85]. One of the proteins of choice was the PspA. A vital aspect in the production of the intended conjugate pneumococcal vaccine is the efficient production (in industrial scale) of both, the chosen PS serotypes as well as the PspA protein. This paper particularly investigates the viability of an on-line monitoring system, based on constructive neural networks, for automatically detecting the time to start the fed-phase of a high-cell density culture (HCDC) of the recombinant *Escherichia coli* (*rE. coli*) that contains a plasmid used for PspA expression.

The layout of the paper is as follows. Section 2 presents the main aspects related to *S. pneumoniae* focusing on pneumococcal virulence factors and their role in colonization and disease. Section 3 describes a recent history of the various vaccines against the pneumococcus, their main characteristics and effectiveness, particularly taking into account the Brazilian population. Section 4 focuses on the fermentation process for growing *rE. coli* aiming at the production of the PspA protein for vaccine purposes. Section 4.1 approaches some of the technicalities involved in high cell density cultures (HCDC) of recombinant organisms, particularly the control of the feeding process and Section 4.2 describes the composition and operational conditions of four *rE. coli* cultivations, whose data are used in the monitoring experiments described in Section 6. Section 5 first presents a brief motivation for using constructive neural networks (CoNNs) and then describes the general characteristics of five CoNN algorithms used in the experiments described in Section 6. Section 6 presents the data and the methodology used for training the CoNNs as the online monitoring software for identifying the start of the feeding phase in HCDC cultures of *rE. coli*. Results of software simulations implementing the five CoNN algorithms as well as conventional neural networks (FFNN), decision trees (DT) and support vector machines (SVM) are presented and discussed. Finally Section 7 summarizes the work done highlighting its main conclusions and contributions.

2. *S. pneumoniae* – main characteristics, resistance and virulence factors

S. pneumoniae is commonly found in the respiratory tract colonizing the nasopharyngeal cavity of the human host