W350–W355 *Nucleic Acids Research, 2017, Vol. 45, Web Server issue* *Published online 9 May 2017*
doi: 10.1093/nar/gkx397

AMMOS2: a web server for protein–ligand–water complexes refinement via molecular mechanics

Céline M. Labbé[1,2,†], Tania Pencheva[3,†], Dessislava Jereva[3], Dimitri Desvillechabrol[1,2], Jérôme Becot[1,2], Bruno O. Villoutreix[1,2], Ilza Pajeva[3] and Maria A. Miteva[1,2,*]

[1]Université Paris Diderot, Sorbonne Paris Cité, Molécules Thérapeutiques In Silico, INSERM UMR-S 973, Paris, France, [2]INSERM, U973 Paris, France and [3]Department of QSAR and Molecular Modelling, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad G. Bonchev Str., 1113 Sofia, Bulgaria

Received February 1, 2017; Revised April 26, 2017; Editorial Decision April 26, 2017; Accepted April 28, 2017

ABSTRACT

AMMOS2 is an interactive web server for efficient computational refinement of protein–small organic molecule complexes. The AMMOS2 protocol employs atomic-level energy minimization of a large number of experimental or modeled protein–ligand complexes. The web server is based on the previously developed standalone software AMMOS (Automatic Molecular Mechanics Optimization for *in silico* Screening). AMMOS utilizes the physics-based force field AMMP sp4 and performs optimization of protein–ligand interactions at five levels of flexibility of the protein receptor. The new version 2 of AMMOS implemented in the AMMOS2 web server allows the users to include explicit water molecules and individual metal ions in the protein–ligand complexes during minimization. The web server provides comprehensive analysis of computed energies and interactive visualization of refined protein–ligand complexes. The ligands are ranked by the minimized binding energies allowing the users to perform additional analysis for drug discovery or chemical biology projects. The web server has been extensively tested on 21 diverse protein–ligand complexes. AMMOS2 minimization shows consistent improvement over the initial complex structures in terms of minimized protein–ligand binding energies and water positions optimization. The AMMOS2 web server is freely available without any registration requirement at the URL: http://drugmod.rpbs.univ-paris-diderot.fr/ammosHome.php.

INTRODUCTION

The advances in computational sciences during the last decade enabled extensive use of *in silico* methods at the interface between chemistry and biology. A modern set of approaches authorizes high-throughput screening computations, prioritization of the hit compounds and different levels of compound optimization (1–3). To date, many online tools have been developed in that direction (1). For instance, free web services for drug likeness and toxicity prediction are available, like Molinspiration (Molinspiration Cheminformatics, for RO5 computations), Aggregator Advisor (to search for molecules that aggregate, http://advisor.bkslab.org/) or FAF-Drugs3 (for compound property calculation and chemical library design) (4). In addition, several web servers performing *de novo* drug design (e.g. e-LEAD3 (5) or docking of a few ligands (e.g. SwissDock (6), CovalentDock (7)), predicting the binding affinity of a protein–small molecule complex (8), as well as some services for large-scale docking or virtual screening (like iScreen (9), DOCK Blaster (10), USR-VS (11), MTiOpenScreen (12)) have recently been reported. Yet, protein–ligand docking is not a trivial task and its performance strongly depends on the used algorithms and scoring functions, binding site definition, solvation and entropy considerations (13–15). To improve the quality of identified hits/leads and ultimately to assist the overall success of the project, additional refinement of the initially predicted docking/screening results may be required.

Here we present the new web server AMMOS2 dedicated to efficient computational refinement of protein–small organic molecule complexes and web-based visual analysis. The AMMOS2 protocol employs atomic-level energy minimization of a large number of experimental or modeled protein–ligand complex structures that can be generated via a user-chosen docking program or via virtual screening web servers as e.g. MTiOpenScreen (12). Two recent web servers, 3Drefine (16) and KoBaMIN (17), provide refinement of apo protein structures by energy minimization. The AMMOS2 web server is based on the previously developed standalone software AMMOS (Automatic Molecular Mechanics Optimization for *in silico* Screening) (18) performing

*To whom correspondence should be addressed. Tel: +33 1 57 27 83 92; Fax: +33 1 57 27 83 22; Email: maria.mitev@inserm.fr
†These authors contributed equally to the paper as first authors.

Bioorganic & Medicinal Chemistry Letters 27 (2017) 2996–3002

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

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

Synthesis, antimycobacterial activity and docking study of 2-aroyl-[1]benzopyrano[4,3-*c*]pyrazol-4(1*H*)-one derivatives and related hydrazide-hydrazones

Violina T. Angelova [a,*], Violeta Valcheva [b], Tania Pencheva [c], Yulian Voynikov [a], Nikolay Vassilev [d], Rositsa Mihaylova [a], Georgi Momekov [a], Boris Shivachev [e]

[a] *Faculty of Pharmacy, Medical University of Sofia, 2 Dunav Str., 1000 Sofia, Bulgaria*
[b] *"Stefan Angelov" Institute of Microbiology, Bulgarian Academy of Sciences, 26 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria*
[c] *Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str, 1113 Sofia, Bulgaria*
[d] *Institute of Organic Chemistry with Centre of Phytochemistry, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria*
[e] *Institute of Mineralogy and Crystallography, Bulgarian Academy of Sciences, 107 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria*

ARTICLE INFO

Article history:
Received 28 March 2017
Revised 3 May 2017
Accepted 4 May 2017
Available online 5 May 2017

Keywords:
Antimycobacterial activity
Chromeno[4,3-*c*]pyrazol-4-one
Cytotoxicity
Hydrazide-hydrazones
Molecular docking

ABSTRACT

A new convenient method for preparation of 2-aroyl-[1]benzopyrano[4,3-*c*]pyrazol-4(1*H*)-one derivatives **5b–g** and coumarin containing hydrazide-hydrazone analogues **4a–e** was presented. The antimycobacterial activity against reference strain *Mycobacterium tuberculosis* H37Rv and cytotoxicity against the human embryonic kidney cell line HEK-293 were tested *in vitro*. All compounds demonstrated significant minimum inhibitory concentrations (MIC) ranging 0.28–1.69 μM, which were comparable to those of isoniazid. The cytotoxicity (IC_{50} > 200 μM) to the "normal cell" model HEK-293T exhibited by 2-aroyl-[1]benzopyrano[4,3-*c*]pyrazol-4(1*H*)-one derivatives **5b–e**, was noticeably milder compared to that of their hydrazone analogues **4a–e** (IC_{50} 33–403 μM). Molecular docking studies on compounds **4a–e** and **5b–g** were also carried out to investigate their binding to the 2-*trans*-enoyl-ACP reductase (InhA) enzyme involved in *M. tuberculosis* cell wall biogenesis. The binding model suggested one or more hydrogen bonding and/or arene-H or arene-arene interactions between hydrazones or pyrazole-fused coumarin derivatives and InhA enzyme for all synthesized compounds.

A reemergence of tuberculosis accompanied by an increasing number of drug resistant *Mycobacterium tuberculosis* strains highlights the urgent need of searching and developing of new antitubercular drugs, capable of bypassing the resistance mechanisms. In the last decades, major advances in molecular biology have increased the knowledge of the mechanisms of resistance to the main anti-TB drugs, with the identification of specific gene mutations that are associated with drug resistance[1,2]

Isoniazid (INH), an essential antitubercular agent recommended by the WHO, is a prodrug that penetrates the tubercle bacilli by passive diffusion and is bio-activated by the bacterial anti-oxidant enzyme (KatG).[3,4] It exerts its anti-tubercular activity *via* interference with the synthesis of mycolic acids, which comprise crucial elements of the mycobacterial cell wall**.** Even with the clinical success of isoniazid, severe adverse effects, especially peripheral neuropathy and hepatotoxicity, are associated with INH-based treatment protocols; moreover its usefulness is further limited by the occurrence of resistance.[5] To overcome the resistance,[6] the drug design strategies frequently employ a combination of the INH molecule with other pharmacophores, rendering antitubercular activity. The novel INH hydrazide derivatives appeared to be promising anti-tubercular agents - more effective and less hepatotoxic than isoniazid.[7–16] In the meantime Ellis *at al.*,[17] have described the mechanism of action of the pyridoxal isonicotinoyl hydrazones (PIC) and suggested that hydrazones act as a lipophilic vehicle for the transport of its intact INH moiety into the mammalian cell and the mycobacterium (Fig. 1). The mechanism of antimycobacterial activity of INH[18] and isonicotinoyl hydrazone derivatives passes through formation of electrophilic intermediate species (i.e. a hydrazyl radical or ion) (Fig. 1). The acyl radical being coupled to NADH or NAD + seems to be crucial in yielding adduct responsible for the inhibition of 2-*trans*-enoyl-ACP reductase (InhA), and in restraining the mycobacterial cell wall synthesis. InhA catalyze the final step in the elongation cycle of the bacterial

* Corresponding author.
E-mail addresses: violina@pharmfac.net, violina_stoyanova@abv.bg (V.T. Angelova).

http://dx.doi.org/10.1016/j.bmcl.2017.05.011
0960-894X/

Med Chem Res
DOI 10.1007/s00044-017-1902-1

MEDICINAL CHEMISTRY RESEARCH

ORIGINAL RESEARCH

In vitro and in silico evaluation of chromene based aroyl hydrazones as anticonvulsant agents

Violina T. Angelova[1] · Yulian Voynikov[1] · Pavlina Andreeva-Gateva[2,3] · Slavina Surcheva[2] · Nikolay Vassilev[4] · Tania Pencheva[5] · Jana Tchekalarova[6]

Received: 3 February 2017 / Accepted: 13 April 2017

Abstract Series of aroyl hydrazones of 2*H*-chromene and coumarin carbaldehydes were synthesized and evaluated for their anticonvulsant activity and neurotoxicity. Further docking study on gamma-aminobutyric acid receptor was performed to elucidate their mechanisms of action. The highest protection was demonstrated by 2-furyl substituted 2*H*-chromene **8b** in the maximal electroshock test ($ED_{50} = 12.51\ mg\ kg^{-1}$, PI MES > 23.98) and the subcutaneous pentylenetetrazole tests ($ED_{50} = 127.10\ mg\ kg^{-1}$). Furyl-substituted derivative **4b** ($ED_{50} = 68.66\ mg\ kg^{-1}$) was the most active in the maximal electroshock test while methoxyphenyl-substituted derivate **4c** was the most active in the 6-Hz test ($ED_{50} = 94.34\ mg\ kg^{-1}$). None of the compounds displayed neurotoxicity in the rota-rod test. In silico assessment of their blood-brain barrier permeability indicated them as central nervous system active agents. The results suggest that coumarin/2*H*-chromene aroyl hydrazones scaffold deserve further evaluation in models of epilepsy and derivatization.

Keywords Coumarin · Docking · Hydrazones · 6-Hz · MES · PTZ

✉ Violina T. Angelova
violina_stoyanova@abv.bg
violina@pharmfac.net

✉ Jana Tchekalarova
janetchekalarova@gmail.com

[1] Department of Chemistry, Faculty of Pharmacy, Medical University of Sofia, Sofia, Bulgaria

[2] Department of Pharmacology and Toxicology, Medical University of Sofia, Sofia, Bulgaria

[3] Department of Pharmacology, Faculty of Medicine, Sofia University "St. Kliment Ohridsky", Sofia, Bulgaria

[4] Institute of Organic Chemistry with Centre of Phytochemistry, Bulgarian Academy of Sciences, Sofia, Bulgaria

[5] Department of QSAR and Molecular Modelling, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria

[6] Institute of Neurobiology, Bulgarian Academy of Sciences, Sofia, Bulgaria

Introduction

Epilepsy is a chronic neurologic disease with high susceptibility to seizure generation and propagation, and with a predisposition to neurobiological, cognitive, psychological, and social disturbances (Fisher et al. 2005). Epileptic seizures result from hyper synchronized discharge of neurons and their clinical manifestation depends mainly on their brain localization and spread (Terrone et al. 2016). According to a recent meta-analysis, the median lifetime epilepsy ranges from 5.8 per 1000 for developed countries, to 15.4 per 1000 for developing countries (Ngugi et al. 2010). Although more than 25 new antiepileptic drugs (AEDs) had been registered by 2015, it is well-known that they neither prevent, nor cure completely this brain disorder, and are used mostly to suppress its symptoms (Picaud et al. 2016). Moreover, it is known that up to 40% of patients with epilepsy are drug-resistant (Mohanraj and Brodie 2005). Most of AEDs exhibit adverse effects and because of the need of chronic treatment the quality of life of epilepsy patients can be aggravated. Therefore, nowadays a lot of research is focused on discovering new pharmacophores characterized with high activity and less neurotoxicity.

Published online: 22 April 2017

Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 133 (2017) 124–134
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Original articles

Molecular dynamics simulation of the human estrogen receptor alpha: contribution to the pharmacophore of the agonists

Dessislava Jereva[a], Filip Fratev[a,b], Ivanka Tsakovska[a], Petko Alov[a], Tania Pencheva[a,*], Ilza Pajeva[a,**]

[a] *Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 105, 1113 Sofia, Bulgaria*
[b] *Micar21 Ltd., Persenk Str. 34B, 1407 Sofia, Bulgaria*

Received 14 November 2014; received in revised form 9 June 2015; accepted 8 July 2015
Available online 23 July 2015

Abstract

Human estrogen receptor alpha (ERα) is one of the most studied targets for *in silico* screening of bioactive compounds. The estrogenic activity of a vast number of chemicals has been studied for their potentially adverse effects on the hormone regulation of the endocrine system. The commonly accepted presentation of the ERα agonist pharmacophore includes terminal phenolic groups and a hydrophobic rigid backbone. In this study we report on molecular dynamics (MD) simulations of ERα to get a deeper structural insight into the agonist–receptor interactions and the pharmacophore pattern of compounds with agonistic activity. We rely on a crystallographic structure of a complex of ERα (PDB ID 2P15) with an agonist of picomolar affinity. As the X-ray structure has a mutation next to a key structural element for ERα agonistic activity (helix H12, Y537S), a series of MD simulations have been performed on the mutated and on the wild type receptor to prove the stability of the agonist–receptor interactions. No significant difference in the ligand–protein interactions has been detected between the studied proteins implying that the Y537S mutant structure can be used for refinement of the pharmacophore model of the ERα agonists. The results suggest that the pharmacophore of compounds with ERα agonistic activity can be extended by a feature that occupies a free hydrophobic region of the binding pocket. The extended pharmacophore model has been evaluated by a pharmacophore-based virtual screening of databases of ERα binders and decoys. The results also imply that MD simulations are a powerful *in silico* tool for both protein dynamics and structure investigation, especially when mutations are available that can potentially disturb the protein structure and functions.

Keywords: Human estrogen receptor alpha; Agonist; Pharmacophore; Molecular dynamics simulation

1. Introduction

The estrogen receptor alpha (ERα) belongs to the nuclear receptor superfamily of transcription factors [5]. It is, perhaps, the most studied among these receptors due to its involvement in many physiologically and pharmacologically significant phenomena. ERα is expressed in many tissues with endocrine-related functions and its dysregulation

* Corresponding author. Tel.: +359 2 979 3648.
** Corresponding author. Tel.: +359 2 979 3605.
E-mail addresses: dessislava.jereva@biomed.bas.bg (D. Jereva), fratev@biomed.bas.bg (F. Fratev), itsakovska@biomed.bas.bg (I. Tsakovska), petko.alov@biophys.bas.bg (P. Alov), tania.pencheva@biomed.bas.bg (T. Pencheva), pajeva@biomed.bas.bg (I. Pajeva).

http://dx.doi.org/10.1016/j.matcom.2015.07.003

NPC Natural Product Communications

2017
Vol. 12
No. 2
175 - 178

ADME/Tox Properties and Biochemical Interactions of Silybin Congeners: *In silico* Study

Antonia Diukendjieva, Merilin Al Sharif, Petko Alov, Tania Pencheva, Ivanka Tsakovska and Ilza Pajeva*

Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Acad. G. Bonchev Str., Block 105, 1113 Sofia, Bulgaria

pajeva@biomed.bas.bg

Received: June 1st, 2016; Accepted: November 2nd, 2016

Silymarin, the active constituent of *Silybum marianum* (milk thistle), and its main component, silybin, are products with well-known hepatoprotective, cytoprotective, antioxidant, and chemopreventative properties. Despite substantial *in vitro* and *in vivo* investigations of these flavonolignans, their mechanisms of action and potential toxic effects are not fully defined. In this study we explored important ADME/Tox properties and biochemical interactions of selected flavonolignans using *in silico* methods. A quantitative structure–activity relationship (QSAR) model based on data from a parallel artificial membrane permeability assay (PAMPA) was used to estimate bioavailability after oral administration. Toxic effects and metabolic transformations were predicted using the knowledge-based expert systems Derek Nexus and Meteor Nexus (Lhasa Ltd). Potential estrogenic activity of the studied silybin congeners was outlined. To address further the stereospecificity of this effect the stereoisomeric forms of silybin were docked into the ligand-binding domain of the human estrogen receptor alpha (ERα) (MOE software, CCG). According to our results both stereoisomers can be accommodated into the ERα active site, but different poses and interactions were observed for silybin A and silybin B.

Keywords: Silymarin, Silybins, ADME/Tox properties, Estrogen receptor.

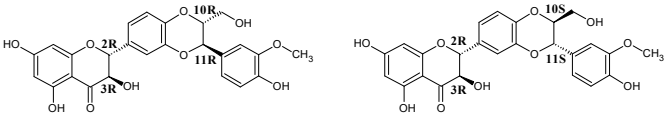
Silybum marianum (L.) Gaertn. (milk thistle) is an ancient medicinal plant that has been used for almost 2000 years for treatment of liver and gallbladder disorders of different etiologies [1,2]. The active component of this herb, silymarin, is a mixture of phenolic compounds, mainly silybin A, silybin B, but also other flavonolignans such as isosilybin A, isosilybin B, silychristin and silydianin, which are located predominantly in the fruit and seeds. The main component of silymarin is silybin, which is a quasi equimolar mixture of two diastereomers A and B (Figure 1) [3].

Today silymarin is best known for its antioxidant and chemoprotective effects on the liver [4], and is often either prescribed or self-prescribed as a complementary hepatoprotective medicine [5]. It has also gained attention due to its hypocholesterolemic, cardioprotective, neuroactive and neuroprotective properties [4]. Although silymarin is reported as nontoxic in human studies, little is known about its mechanism of action and biochemical interactions [6]. Recent works have explored inhibition and modulation of some drug transporters [7] and nuclear receptors [8] by silybin congeners as well as their biotransformation products [9]. For example, an *in vitro* study focusing on interactions of flavonolignans with the aryl hydrocarbon receptor (AhR) and estrogen receptor (ER) demonstrated that silymarin has partial estrogenic activity, with silybin B being probably responsible for it [8]. It was outlined that stereochemistry plays an important role for the investigated biological activities and there is a need for studies on the pure forms of the compounds that are otherwise therapeutically used as mixtures [10]. Another important prerequisite for broader and safer therapeutic use of flavonolignans is the better understanding of their metabolism, pharmacokinetics and potential toxic effects.

The pharmaceutical industry has used *in silico* methods for decades to search, optimize and evaluate drugs [9]. In recent years the *in silico* ADME/Tox (Absorption, Distribution, Metabolism, Excretion, and Toxicity) prediction is receiving particular attention due to the increased evidence that these pharmacokinetic properties should be considered earlier in the drug discovery process [11].

In the present study we aimed at exploring important ADME/Tox properties and biochemical interactions of selected flavonolignans (Figures 1 and 2) using *in silico* methods. For estimation of bioavailability after oral administration (gastrointestinal absorption) an in house developed quantitative structure–activity relationship (QSAR) model utilizing data from a parallel artificial membrane permeability assay (PAMPA) was used. Predictions of toxicity and metabolism were performed using knowledge-based expert systems, and molecular modelling studies were applied for investigation of the interactions of the stereoisomeric forms of silybin with the ligand-binding domain (LBD) of the human estrogen receptor alpha (ERα).

In silico estimation of gastrointestinal absorption was performed using a QSAR model for prediction of PAMPA permeability. PAMPA is a high throughput *in vitro* assay that evaluates transcellular permeation of small drug-like molecules [12]. PAMPA is used in pharmaceutical research to screen for human intestinal absorption since PAMPA permeability has been shown to correlate with both Caco-2 cell permeability and human intestinal absorption [13].



Silybin A Silybin B

Figure 1: Silybin A (2*R*, 3*R*, 10*R*, 11*R*) and silybin B (2*R*, 3*R*, 10*S*, 11*S*).

InterCriteria Analysis of Simple Genetic Algorithms Performance

Tania Pencheva and Maria Angelova

Abstract Recently developed approach of InterCriteria Analysis is here applied aiming at an assessment of the performance of such a promising stochastic optimization technique as simple genetic algorithms. Considered algorithms, as representatives of the biologically-inspired ones, are chosen as an object of investigation since they are proven as quite successful in solving of many challenging problems in the field of complex dynamic systems optimization. In this investigation simple genetic algorithms are applied for the purposes of parameter identification of a fermentation process. Altogether six simple genetic algorithms are here considered, differ from each other in the execution order of main genetic operators, namely selection, crossover and mutation. The apparatuses of index matrices and intuitionistic fuzzy sets, underlying the InterCriteria Analysis, are implemented to assess the performance of simple genetic algorithms for the parameter identification of *Saccharomyces cerevisiae* fed-batch fermentation process. The obtained results after the InterCriteria Analysis application are thoroughly analysed towards the algorithms outcomes, such as convergence time and model accuracy.

1 Introduction

In general, modelling of fermentation processes is a challenging and rather difficult to be solved problem. Logically explanation of this fact is connected with a complex structure of the considered processes, usually described by a systems of nonlinear differential equations with several specific growth rates. Thus, model parameter

T. Pencheva (✉) · M. Angelova
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
e-mail: tania.pencheva@biomed.bas.bg

M. Angelova
e-mail: maria.angelova@biomed.bas.bg

K. Georgiev et al. (eds.), *Advanced Computing in Industrial Mathematics*, Studies in Computational Intelligence 681, DOI 10.1007/978-3-319-49544-6_13

Bulgarian Chemical Communications, Volume 48, Number 4 (pp. 713 – 719) 2016

Modified multi-population genetic algorithms for parameter identification of yeast fed-batch cultivation

T. K. Pencheva*, M. K. Angelova

Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences
105 Acad. Georgi Bonchev Str., 1113 Sofia, Bulgaria

Received February 8, 2016; Revised March 12, 2016

In this investigation two new modifications of the standard multi-population genetic algorithm have been developed. Modifications differ from each other in the sequence of implementation of main genetic operators selection, crossover and mutation. The main idea of newly developed modifications is the operator selection to be executed between the operators crossover and mutation, no matter their order. Both modifications, together with the standard one multi-population genetic algorithm, have been investigated for parameter identification of yeast fed-batch cultivation. The obtained results have been compared and the newly proposed modifications have been shown to be as accurate as the standard multi-population genetic algorithms and proven to be even faster.

Keywords: Multi-population genetic algorithms, Genetic operators, Fermentation process, Parameter identification.

INTRODUCTION

Fermentation processes (FP) as representatives of biotechnological processes attract sustained interest due to the fact that they are an indigenous part of many industries such as industrial biotechnology, microbiology and the pharmaceutical industry. FP combine the dynamics of both biological and non-biological processes but their specific peculiarities are predominantly determined by the characteristics of live microorganisms. Since FP are complex dynamic systems with interdependent and time-varying process variables, their modeling, optimization and future high quality control is a real challenge. Adequate modeling of the non-linear FP significantly depends on the choice of a certain optimization procedure for model parameter identification. Conventional optimization methods usually fail in leading to a satisfactory solution [1]. This fact provokes the idea to apply stochastic algorithms, i.e. genetic algorithms (GA). GA are known as a quite promising stochastic global optimization method and have been widely applied to solve different complicated engineering problems [2-5]. Among a number of searching techniques, GA are representatives of the methods inspired by biological evolution and the principle of Darwin's theory of "survival of the fittest". GA are a feature of hard problem solving, tolerant to noise, easy to interface and hybridize. All these properties make GA convenient and more workable for different optimization problems, among them parameter identification and optimization of fermentation processes [6-9].

The standard simple genetic algorithm (SGA) [10] imitates the processes that occur in nature and searches for a global optimum solution using three main genetic operators implementing them in a sequence selection, crossover and mutation. SGA works with "chromosomes" (coded parameters) and starts with a selection of such chromosomes that represent better possible solutions according to their objective function values. Then a new offspring is formed applying the crossover operator. Finally, mutation is applied with deterministic probability, aiming to prevent the failing of all the solutions into a local optimum of the solved problem.

If there are many populations (called subpopulations), that evolve independently from each other, the single-population GA is converted to a multi-population GA (MpGA) [10]. This feature presents MpGA as more similar to nature than SGA. After the isolation time (a certain number of generations), part of the individuals "migrate" – they are distributed between the subpopulations. Similar to SGA, the standard MpGA as originally presented in [10], implements the three main genetic operators in a sequence selection, crossover and mutation. In this investigation this algorithm will be denoted as MpGA_SCM, coming from selection, crossover and mutation. According to [10, 11] the working principle of MpGA_SCM can be shortly presented as shown in Fig. 1.

To imitate the mechanics of natural selection and genetics is enshrined in the "philosophy" of GA. Thus one can make an analogy with the processes occurring in nature and to speculate that

* To whom all correspondence should be sent: E-mail: tania.pencheva@biomed.bas.bg

InterCriteria Analysis by Pairs and Triples of Genetic Algorithms Application for Models Identification

Olympia Roeva, Tania Pencheva, Maria Angelova and Peter Vassilev

Abstract In this investigation the InterCriteria Analysis (ICrA) approach is applied. The apparatuses of index matrices and intuitionistic fuzzy sets are at the core of ICrA. They are used to examine the influences of two main genetic algorithms (GA) parameters—the rates of crossover (*xovr*) and mutation (*mutr*). A series of parameter identification procedures for *S. cerevisiae* and *E. coli* fermentation process models is fulfilled. Twenty GA with different *xovr* and *mutr* values are applied. Relations between ICrA criteria—GA parameters and outcomes, on the one hand, and fermentation process model parameters, on the other hand, are investigated. The ICrA approach is applied by pairs, as well as by triples. The obtained results are thoroughly analysed towards computation time and model accuracy and some conclusions about the derived criteria interactions are reported.

Keywords InterCriteria analysis · Index matrices · Intuitionistic fuzzy sets · Genetic algorithm · Parameter identification · Algorithm performance · *S. cerevisiae* · *E. coli* · Fermentation

1 Introduction

InterCriteria Analysis (ICrA) [11] is a contemporary approach for multi-criteria decision making. ICrA implements the apparatuses of index matrices (IM) and intuitionistic fuzzy sets (IFS) in order to compare some criteria reflecting the behaviour of

O. Roeva (✉) · T. Pencheva · M. Angelova · P. Vassilev
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria
e-mail: olympia@biomed.bas.bg

T. Pencheva
e-mail: tania.pencheva@biomed.bas.bg

M. Angelova
e-mail: maria.angelova@biomed.bas.bg

P. Vassilev
e-mail: peter.vassilev@gmail.com

S. Fidanova (ed.), *Recent Advances in Computational Optimization*,
Studies in Computational Intelligence 655, DOI 10.1007/978-3-319-40132-4_12

InterCriteria Analysis Approach to Parameter Identification of a Fermentation Process Model

Tania Pencheva, Maria Angelova, Peter Vassilev and Olympia Roeva

Abstract In this investigation recently developed InterCriteria Analysis (ICA) is applied aiming at examination of the influence of a genetic algorithm (GA) parameter in the procedure of a parameter identification of a fermentation process model. Proven as the most sensitive GA parameter, generation gap is in the focus of this investigation. The apparatuses of index matrices and intuitionistic fuzzy sets, laid in the ICA core, are implemented to establish the relations between investigated here generation gap, from one side, and model parameters of fed-batch fermentation process of *Saccharomyces cerevisiae*, from the other side. The obtained results after ICA application are analysed towards convergence time and model accuracy and some conclusions about observed interactions are derived.

Keywords InterCriteria analysis · Genetic algorithms · Generation gap · Parameter identification · *S. cerevisiae*

1 Introduction

InterCriteria Analysis (ICA), given in details in [2] is a contemporary approach for multicriteria decision making. ICA implements the apparatuses of index matrices (IM) and intuitionistic fuzzy sets (IFS) in order to compare some criteria or estimated by them objects. In [16] ICA has been applied for the first time in the field

T. Pencheva (✉) · M. Angelova · P. Vassilev · O. Roeva
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences,
105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
e-mail: tania.pencheva@biomed.bas.bg

M. Angelova
e-mail: maria.angelova@biomed.bas.bg

P. Vassilev
e-mail: peter.vassilev@gmail.com

O. Roeva
e-mail: olympia@biomed.bas.bg

K.T. Atanassov et al. (eds.), *Novel Developments in Uncertainty Representation and Processing*, Advances in Intelligent Systems and Computing 401,
DOI 10.1007/978-3-319-26211-6_33

2016 IEEE 8th International Conference on Intelligent Systems

Comparison of Different Algorithms for InterCriteria Relations Calculation

Olympia Roeva, Peter Vassilev, Maria Angelova, Tania Pencheva
Institute of Biophysics and Biomedical Engineering
Bulgarian Academy of Sciences
Email: olympia@biomed.bas.bg

Jun Su
School of Computer Science,
Hubei University of Technology, Hubei, China

***Abstract*—In this investigation different algorithms for InterCriteria relations calculation are proposed. The algorithms are investigated by exploring the influence of genetic parameters on algorithm performance during the model parameter identification of *E. coli* fermentation process. Four different algorithms performing InterCriteria Analysis (ICrA), namely μ-biased, balanced, ν-biased and unbiased, are applied. Proposed ICrA algorithms are compared based on real experimental data set of an *E. coli* MC4110 fed-batch fermentation process. The obtained results show that for considered here case study the most reliable algorithm is the μ-biased one.**

Keywords:* InterCriteria analysis; Intuitionistic fuzzy sets; Genetic algorithms; Fermentation process; *E. coli

I. Introduction

InterCriteria Analysis (ICrA) is an approach going beyond the nature of the criteria involved in a process of evaluation of multiple objects against multiple criteria [3]. As such, it is expected to discover any existing correlations between the criteria themselves. So far, ICrA has been applied in different areas of science and practice [7], [9], [14]. ICrA approach has been proved as very useful in the case of modeling of fermentation processes (FP). ICrA is applied for establishing the relations between genetic algorithm (GA) parameter *generation gap* on the one hand, and convergence time, model accuracy and model parameters on the other hand in case of *E. coli* FP [15] and *S. cerevisiae* FP [16]. In [17] ICrA is applied to establish the dependencies of considered parameters based on different criteria referred to various metaheuristic algorithms, namely hybrid schemes using GA and Ant Colony Optimization.

The main purpose of this research is to propose and investigate four different algorithms for InterCriteria relations calculation in order to explore the correlations between model and optimization algorithm parameters. The proposed here ICrA algorithms are realized based on the ideas presented in [8]. In [8] several rules for defining the ways of estimating the degrees of "agreement" and "disagreement", with respect to the type of data, are proposed. It is shown that it is necessary to specify the algorithms for determining the degrees of "agreement" and "disagreement".

Using GA model parameter identification of an *E. coli* fed-batch FP is considered here. Bacteria *E. coli* is among the widely used model organisms in genetic engineering and cell biology due to its well known metabolic pathways [11] and as such it has been chosen as representative one with numerous applications in food and pharmaceutical industry.

GAs [10] are chosen as an optimization technique as they are among the most widely used among the biologically inspired approaches for global search. In [1], [13] GA are successfully applied to identification of various models of FP.

Based on ICrA the influence of two of the main GA parameters, namely the number of individuals ($nind$) and the number of generations (gen), is investigated. Moreover, the results of identification of *E. coli* model parameters are used to obtain the relations between model parameters, model accuracy and GA convergence time.

II. Problem formulation

A. Case study: E. coli fed-batch fermentation process

The mathematical model of the considered here *E. coli* fed-batch process is presented by the following non-linear differential equations system [13]:

$$\frac{dX}{dt} = \mu_{max}\frac{S}{k_S + S}X - \frac{F_{in}}{V}X \quad (1)$$

$$\frac{dS}{dt} = -\frac{1}{Y_{S/X}}\mu_{max}\frac{S}{k_S + S}X + \frac{F_{in}}{V}(S_{in} - S) \quad (2)$$

$$\frac{dV}{dt} = F_{in} \quad (3)$$

where X is the biomass (*E. coli*) concentration, [g/l];

- S – substrate (glucose) concentration, [g/l];
- F_{in} – feeding rate, [l/h];
- V – bioreactor volume, [l];
- μ_{max} – maximum value of the specific growth rate, [1/h];
- k_S – saturation constant, [g/l],
- S_{in} – substrate concentration in feeding solution, [g/l];
- $Y_{S/X}$ – yield coefficient, [-].

The parameter vector that should be identified for the model (Eqs. (1)-(3)) is $p = [\mu_{max}\ k_S\ Y_{S/X}]$.

Identification procedures of model parameters of an *E. coli* MC4110 fed-batch FP are performed based on experimental data for biomass and glucose concentration. The detailed description of the processes conditions and experimental data set can be found in [13].

INTERCRITERIA ANALYSIS OF WASTEWATER TREATMENT QUALITY

Vania Georgieva[1], Nora Angelova[2, 3], Olympia Roeva[2], Tania Pencheva[2]

[1]"Prof. Dr. Assen Zlatarov" University, 1 Prof. Yakimov Str., Bourgas, Bulgaria

[2]Institute of Biophysics and Biomedical Engineering – Bulgarian Academy of Sciences, Acad. G. Bonchev Str., bl. 105, 1113 Sofia, Bulgaria

[3]Faculty of Mathematics and Informatics, Sofia University "St. Kliment Ohridski", 5 J. Bourchier Str., Sofia, Bulgaria

Abstract

Recently developed InterCriteria Analysis (ICrA) is here applied to assess the quality of the processes of water purification in a typical wastewater treatment plant (WTP). Various parameters dependences characterizing the quality of WTP processes such as water quantity, pH, COD, petrol, different mechanical admixtures, PO_4P, NH_4N, etc., are going to be analyzed based on real experimental data. Degrees of "agreement" and degrees of "disagreement" between preliminary determined criteria have been established implementing ICrA. This is expected to result in an additional knowledge acquisition about the WTP process itself and in establishment of some relations between ICrA criteria, especially in terms of time (seasonal) dependences.

Key words: *wastewater, treatment, intercriteria, analysis*

1. INTRODUCTION

The desire for a higher standard of living leads to the increase of requirements towards the quality of water both for its everyday usage by human beings, as well as for its application to domestic settings. When used in industrial application, different admixtures, varied in character and dispersion, which occur into the water, thus turning it into so called "wastewater". Logically, the treatment of large quantities of wastewater before their return to the environment has become an indispensable necessity for the contemporary industry and manner of life. This fact determines the continuous development of the technologies for wastewater treatment plant (WTP) processes. As such, there is a necessity of WTP processes analysis aiming at optimization of the apparatuses exploitation and facilities, included in the technological schemes. On the other hand, rendering all methods and factors that influent the quality of wastewater treatment makes modeling process of WTP technological scheme a real challenge to the researchers.

In this paper the InterCriteria Analysis (ICrA) is applied to establish the relations between process parameters, which reflect the wastewater treatment quality, such as water quantity, pH, COD, petrol, different mechanical admixtures, PO_4P, NH_4N. The existing relations are identified based on real experimental data from a typical WTP during a year.

ICrA is an approach aiming to go beyond the nature of the criteria involved in a process of evaluation of multiple objects against multiple criteria, and, on this basis, to discover any existing correlations between the criteria themselves [Atanassov et al., 2014]. ICrA has been further applied in different areas of science and practice [Atanassova et al., 2014; Krawczak et al., 2016; Pencheva et al., 2016]. ICrA approach has been also implemented to the Mesta river pollution modelling [Ilkova and Petrov, 2015]. Results obtained from this application prove that in the case of some ecological problems ICrA approach could be very useful. Additional knowledge for existing process parameters correlations derived after the application of ICrA might be useful in further analysis and optimization of WTP processes.

Functional State Modelling of Cultivation Processes: Dissolved Oxygen Limitation State

Olympia Roeva[1#], Tania Pencheva[1#*], Stoyan Tzonkov[1], Bernd Hitzmann[2]

[1]*Institute of Biophysics and Biomedical Engineering*
Bulgarian Academy of Sciences
Sofia, Bulgaria
E-mails: olympia@biomed.bas.bg, tania.pencheva@biomed.bas.bg, tzonkov@biomed.bas.bg

[2]*Institute of Food Science and Biotechnology*
University of Hohenheim
Stuttgart, Germany
E-mail: bernd.hitzmann@uni-hohenheim.de

[*]*Corresponding author*
[#]*Authors contributed equally*

Received: August 01, 2014 **Accepted: November 11, 2014**

Published: April 20, 2015

Abstract: *A new functional state, namely dissolved oxygen limitation state for both bacteria Escherichia coli and yeast Saccharomyces cerevisiae fed-batch cultivation processes is presented in this study. Functional state modelling approach is applied to cultivation processes in order to overcome the main disadvantages of using global process model, namely complex model structure and a big number of model parameters. Alongwith the newly introduced dissolved oxygen limitation state, second acetate production state and first acetate production state are recognized during the fed-batch cultivation of E. coli, while mixed oxidative state and first ethanol production state are recognized during the fed-batch cultivation of S. cerevisiae. For all mentioned above functional states both structural and parameter identification is here performed based on experimental data of E. coli and S. cerevisiae fed-batch cultivations.*

Keywords: *Functional state modelling, Dissolved oxygen limitation state, Fed-batch cultivation, Escherichia coli, Saccharomyces cerevisiae.*

Introduction

Cultivation processes, as representatives of processes in living nature, are characterized by a complicated structure of organization and interdependent characteristics, which determine their non-linearity and non-stationary properties. These processes are known to be very complex and their modelling may be a rather time consuming task. Many mathematical models of cultivation processes have been proposed but just a few have been used to optimize industrial plants. The common modelling approach is the development of an overall non-linear process model that performs satisfactorily through the entire operating range. Unfortunately, this approach has a lot of disadvantages. Complex global models of cultivation processes are characterized with a big number of parameters [5, 16, 20, 21], which complicate the model identification and simulation. Moreover, the global model is not able to describe the metabolic changes during the entire operating range and the parameter non-stationary. As an alternative, an increasingly popular multiple-model approach [2, 4, 6, 7, 14, 15], and in particular – functional state modelling approach can be applied to cope with strongly non-linear and time-varying systems [22, 23]. Using this approach complicated problems are

Asian Journal of Physics Vol. 23, No 5 (2014) 745-756

Post-docking optimization of protein-ligand interactions involving water molecules

Dessislava Jereva[1], Tania Pencheva[1], David Lagorce[2,3], Dimitri Desvillechabrol[2,3], Ilza Pajeva[1], Maria A Miteva[2,3]

[1]*Institute of Biophysics and Biomedical Engineering – Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria*

[2]*INSERM, UMR-S 973, MTi, 75013 Paris, France*

[3] *University Paris Diderot, Sorbonne Paris Cité, UMR-S 973, MTi, 75013 Paris, France*

Water is an important component in biological systems mediating the interactions of macromolecules with various partners. Water molecules accommodated in binding sites of proteins are of particular importance for drug design. Taking water molecules into account in computational approaches to drug design, however, is an extremely challenging task, especially for a large number of ligands usually used for structure-based virtual screening. We previously developed the approach AMMOS to improve the quality of protein-ligand interaction predictions for *in silico* screening by using molecular mechanics optimization of docked protein-ligand complexes. In this study, we included explicit water molecules mediating protein-ligand interactions in AMMOS molecular mechanics optimization. The impact of considering water molecules in the protein binding site during the minimization on prediction of the ligand binding modes and binding energies was assessed and reported here.

Keywords: AMMOS, post-docking optimization, binding site, protein flexibility, protein-ligand interactions, water molecules

1 Introduction

Water is an important component in biological systems mediating the interactions of macromolecules with various partners. The dynamics of water interactions play a key role in molecular recognition, protein folding, and stability [1, 2]. The presence of water molecules at protein-ligand [3], protein-protein, and protein-DNA interfaces highlights their role as direct hydrogen bonds or salt bridges for polar interactions and stability of such complexes [4–6]. The thermodynamics of buried water clusters at protein-ligand interfaces has already been studied (for example [3]). **In fact, the displacement of waters upon association is dependent on the decrease of free energy of binding** through the entropy gain [7, 8]. Interestingly, Amadasi and co-workers classified water molecules with 87% accuracy into 2 categories: conserved/ functionally displaced and sterically displaced/missing [9]. By applying a double-decoupling method to calculate the binding energy of waters in protein-ligand complexes, it has been suggested that, on average, water molecules that can be displaced by a ligand are less strongly bound (by 2.5 kcal/mol) than conserved waters are [10].

Understanding the role of water molecules accommodated in binding sites of proteins is of particular importance for structure-based drug design [11]. However, considering water molecules in computational approaches is a challenging task, especially for a large number of ligands usually used for virtual screening experiments [12, 13]. Water molecules are often ignored in protein-ligand docking. In order to incorporate waters in protein-ligand interaction predictions, a lot of efforts have been made during the last decade. Several protein-ligand docking programs (Flex-X [14], Autodock [15], GOLD [16], and rDOCK [17], among others) **allow for the inclusion of structural waters for ligand docking**. It has been shown in many cases that explicit water molecules improve the binding prediction accuracy [18–20]. Furthermore, several methods are available to predict positions of interfacial waters important for protein-ligand interactions (for example, GRID [21], AQUARIUS [22], CS-Map [23], Fold-X [24], WaterMap [25], etc.). Molecular dynamics or Monte Carlo simulations can also be used to predict water sites at protein surfaces [26]. A more recently available tool, WaterDock [27], predicts binding sites of water molecules by using docking of explicit waters with AutoDock

Corresponding author :

e-mail: pajeva@biomed.bas.bg (Ilza Pajeva); *maria.miteva@paris7.jussieu.fr* (Maria A Miteva);

19th Int. Workshop on IFSs, Burgas, 4–6 June 2015
Notes on Intuitionistic Fuzzy Sets
ISSN 1310–4926
Vol. 21, 2015, No. 2, 99–110

InterCriteria analysis of genetic algorithm parameters in parameter identification

Tania Pencheva, Maria Angelova, Vassia Atanassova and Olympia Roeva

Institute of Biophysics and Biomedical Engineering
Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str., Sofia 1113, Bulgaria
e-mails: tania.pencheva@biomed.bas.bg, maria.angelova@biomed.bas.bg,
vassia.atanassova@gmail.com, olympia@biomed.bas.bg

Abstract: An application of InterCriteria Analysis (ICA) – recently proposed approach for multicriteria decision support – is here presented. The apparata of Index Matrices and Intuitionistic Fuzzy Sets are in the grounds of ICA. In this investigation, ICA is applied to examine the influences of genetic algorithms parameters during the model parameter identification of *E. coli* MC4110 and *S. cerevisiae* fermentation processes. The impact of two of the main genetic algorithms parameters, namely number of individuals and number of generations, is here studied. The obtained results after ICA application are discussed towards convergence time and model accuracy. Some conclusions about existing relations and dependencies between genetic algorithms parameters, from one side, and fermentation process model parameters from the other side, are derived.
Keywords: InterCriteria analysis, Intuitionistic fuzzy sets, Genetic algorithms, Fermentation process, *E. coli*, *S. cerevisiae*.
AMS Classification: 03E72.

1 Introduction

InterCriteria Analysis (ICA) is a recently developed approach [3] aiming to go beyond the nature of the criteria involved in a process of evaluation of multiple objects against multiple criteria, and, on this basis, to discover any existing correlations between the criteria themselves. Given in details in [3], ICA has been further applied for the purposes of temporal, threshold and trends analyses of an economic case-study of European Union member states' competitiveness [8, 9, 10].

Biotechnology & Biotechnological Equipment, 2014
Vol. 28, No. 5, 968−974, http://dx.doi.org/10.1080/13102818.2014.934550

Taylor & Francis
Taylor & Francis Group

ARTICLE; BIOINFORMATICS

Functional state modelling approach validation for yeast and bacteria cultivations

Olympia Roeva[a]* and Tania Pencheva[b]

[a]*Bioinformatics and Mathematical Modelling, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria;* [b]*QSAR and Molecular Modelling, Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Sofia, Bulgaria*

(*Received 6 February 2014; accepted 17 April 2014*)

In this paper, the functional state modelling approach is validated for modelling of the cultivation of two different microorganisms: yeast (*Saccharomyces cerevisiae*) and bacteria (*Escherichia coli*). Based on the available experimental data for these fed-batch cultivation processes, three different functional states are distinguished, namely primary product synthesis state, mixed oxidative state and secondary product synthesis state. Parameter identification procedures for different local models are performed using genetic algorithms. The simulation results show high degree of adequacy of the models describing these functional states for both *S. cerevisiae* and *E. coli* cultivations. Thus, the local models are validated for the cultivation of both microorganisms. This fact is a strong structure model verification of the functional state modelling theory not only for a set of yeast cultivations, but also for bacteria cultivation. As such, the obtained results demonstrate the efficiency and efficacy of the functional state modelling approach.

Keywords: functional state; modelling; *E. coli*; *S. cerevisiae*; cultivation

Introduction

Biotechnological processes, and especially cultivation processes, have enjoyed enormous advances in recent years. Due to their multidisciplinary nature, cultivation processes have attracted the interest of microbiologists, molecular biologists, bio- and chemical engineering, food and pharmaceutical chemists, etc. These complex processes are characterized with properties like non-linearity and non-stationarity. Thus, the development of accurate mathematical models essential for the design, optimization and high-quality control is still a challenging task.

When modelling such processes, the common approach is to develop a global non-linear model valid over the entire operation range. The main disadvantages of the global model are its very complex structure, inability to reflect possible metabolic changes that might occur during the process, as well as the non-stationarity of the parameters. To overcome these global model disadvantages, an alternative approach based on a multiple-model framework could be considered. The multiple-model approach allows some real phenomena or events to be reflected, leading to process description with simpler local models; and offers possibilities for direct incorporation of high-level and qualitative plant knowledge into the model.[1]

Considering the applications of the multiple-model approach for biotechnological processes, the one considered more convenient for further process control is the functional state modelling (FSM) approach. The FSM approach was originally developed by Zhang et al. [2,3] for aerobic yeast growth processes and several works have already shown its benefits.[4,5] Its applicability to the mathematical modelling of *Saccharomyces cerevisiae* CEN.PK haploid batch and *S. cerevisiae* DY 7221 fed-batch cultivations has been shown.[6,7] This approach can be used not only for process predictions, but also for early stabilization of process,[7] robust control,[8] model-based control [9] and design of multiple-model non-linear adaptive control algorithms.[10] This provoked us to attempt to validate the theory of the FSM approach not only for yeast cultivations, but also for mathematical modelling of bacteria cultivations.

Although the concept of FSM was originally developed for yeast cultivation processes,[2,3] it could be applied for the modelling of *Escherichia coli* cultivations as well.[11] Based on [12] it is demonstrated that there is an analogy between the *S. cerevisiae* and *E. coli* growth curves.[5,11]

Yeasts are one of the most important microorganisms with various applications in food and bread production, beer and wine fermentation, etc. In addition, yeasts include some of the most widely used model organisms in genetic engineering and cell biology due to the fact that their metabolic pathways are well known. That is why yeast cultivations are often used as a test process for new methods or ideas.

*Corresponding author. Email: olympia@biomed.bas.bg

Genetic Operators Significance Assessment in Simple Genetic Algorithm

Maria Angelova(✉) and Tania Pencheva

Institute of Biophysics and Biomedical Engineering,
Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
{maria.angelova,tania.pencheva}@biomed.bas.bg

Abstract. Genetic algorithms, proved as successful alternative to conventional optimization methods for the purposes of parameter identification of fermentation process models, search for a global optimal solution via three main genetic operators, namely selection, crossover, and mutation. In order to determine their importance for finding the solution, a procedure for significance assessment of genetic algorithms operators has been developed. The workability of newly elaborated procedure has been tested when simple genetic algorithm is applied to parameter identification of *S. cerevisiae* fed-batch cultivation. According to obtained results the most significant genetic operator has been distinguished and its influence for finding the global optimal solution has been evaluated.

Keywords: Simple genetic algorithm · Genetic operators · Parameter identification · *S. cerevisiae* fed-batch cultivation

1 Introduction

Genetic algorithms (GA) [1] are a metaheuristic method based on biological evolution. Some properties such as hard problems solving, noise tolerance, easiness to interface and hybridize make GA a suitable and quite workable tool especially for incompletely determined tasks. Such a task and a real challenge for researchers is the parameter identification of fermentation processes (FP) models [2–6]. FP are known as complex, dynamic systems with interdependent and time-varying process variables, and their modeling is a specific task, rather difficult to be solved. Failure of conventional optimization methods to reach to a satisfactory solution for parameters identification of FP models [6] provokes idea as an alternative technique genetic algorithms to be tested.

Inspired by natural genetics, Goldberg [1] initially presents the standard single-population genetic algorithm (SGA) that searches a global optimal solution using three main genetic operators in a sequence selection, crossover and mutation. When GA are applied for the purposes of model parameter identification, there are many operators, functions, parameters and settings that may vary depending on the considered problems [1,7]. In [7] three of the main GA parameters, namely generation gap (GGAP), crossover (XOVR) and mutation

I. Lirkov et al. (Eds.): LSSC 2013, LNCS 8353, pp. 223–231, 2014.
DOI: 10.1007/978-3-662-43880-0_24,

Global Journal of Technology & Optimization

Pencheva and Angelova. Global J Technol Optim 2014, 5:1
http://dx.doi.org/10.4172/2229-8711.1000164

Research Article Open Access

Purposeful Model Parameters Genesis in Multi-population Genetic Algorithm

Tania Pencheva* and Maria Angelova

Unstitute of Biophysics and Biomedical Engineering – Bulgarian Academy of Sciences, 105 Acad. Georgi Bonchev Str., 1113 Sofia, Bulgaria

Abstract

IIn this paper recently proposed procedure for purposeful model parameters genesis, originally developed for simple genetic algorithm, has been validated for multi-population genetic algorithm when it is applied for the purposes of parameter identification of S. cerevisiae fed-batch cultivation. Proposed procedure aims to improve the algorithm effectiveness in respect to the convergence time and model accuracy treating intervals of variations of model parameters. Obtained results after the procedure application show more than 12% improvement of multi-population genetic algorithm convergence time while saving and even slightly meliorating the model accuracy.

Keywords: Multi-population genetic algorithm; Purposeful genesis; Model parameters; Fermentation process; *Saccharomyces cerevisiae*

Introduction

Wide application of fermentation processes (FP) in different branches of industry maintains contemporary scientific interest to investigate in this area. FP combine the dynamic of two fundamental components – non-biological and biological, but their specific peculiarities are ultimately determined from characteristics of live microorganisms. As complex, dynamic systems with interdependent and time-varying process variables, FP constitute a serious challenge for investigators working in the field of their modeling, optimization and control. Adequate modeling of FP mostly depends on the choice of a certain optimization procedure for model parameters identification. Genetic algorithms (GA) [1] are a good alternative to conventional optimization methods for solving such a complex problem [2-4].

The effectiveness of a certain optimization technique can be evaluated by the model accuracy achieved and the convergence time needed. Obtained promising results when purposeful model parameter genesis was originally developed and successfully applied for simple genetic algorithm [5], provokes the idea such a procedure to be tested for another kind of genetic algorithm. Thus, the aim of the study is to apply and validate a procedure for purposeful genesis for parameter identification of *S. cerevisiae* fed-batch cultivation when using multi-population genetic algorithm.

Materials and Methods

Multi-population genetic algorithms

Standard genetic algorithms, originally presented in [1], search a global optimal solution using three main genetic operators in a sequence selection, crossover and mutation over the individuals in the population. While simple genetic algorithms work over one population at time, multi-population genetic algorithms (MpGA) is more similar to the nature since in it many populations, called subpopulations, evolve independently from each other. After a certain number of generations, a part of individuals migrates between the subpopulations.

Procedure for purposeful model parameters genesis

Due to the stochastic nature of genetic algorithms, a great number of algorithm runs have to be executed in order to obtain reliable results in parameter identification of a fermentation process model. When results from many algorithms executions were accumulated and analyzed, they show that the values of model parameters can be assembled and predefined boundaries could be restricted. That provoked the idea resulted in purposeful model parameters genesis (PMPG) [5] for shrinking variation boundaries of model parameters values, aiming to decrease convergence time while improve or at least save model accuracy.

The procedure for PMPG has been originally developed for single-population genetic algorithm [5] and consists of six steps, as shown in the right-side of the flowchart in Figure 1. Left side of the flowchart in Figure 1 presents the working principle of standard MpGA according to [1,6].

The stepwise procedure of PMPG passes through all the six steps described in Figure 1, not omitting any of them and without cycles.

Results and Discussion

In this investigation the PMPG procedure has been tested for the purposes of parameter identification of *S. cerevisiae* fed-batch cultivation when using multi-population genetic algorithm.

Experimental data of *S. cerevisiae* fed-batch cultivation is obtained in the Institute of Technical Chemistry – University of Hannover, Germany [7]. The cultivation of the yeast *S. cerevisiae* is performed in a 1.5 l reactor, using a Schatzmann medium. Glucose in feeding solution is 50 g/l. The temperature was controlled at 30°C, the pH at 5.7. The stirrer speed was set to 500 rpm.

Considered here fed-batch cultivation of *S. cerevisiae* is characterized by keeping glucose concentration equal to or below its critical level (S_{crit}=0.05 g/l), sufficient dissolved oxygen $O_2 \geq O_{2crit}$ (O_{2crit} = 18%) and availability of ethanol in the broth. This state corresponds

***Corresponding author:** Tania Pencheva, Institute of Biophysics and Biomedical Engineering – Bulgarian Academy of Sciences, 105 Acad. Georgi Bonchev Str., 1113 Sofia, Bulgaria, Tel: (3592) 9793648; E-mail: tania.pencheva@biomed.bas.bg

Received September 08, 2014; **Accepted** October 04, 2014; **Published** October 14, 2014

Citation: Pencheva T, Angelova M (2014) Purposeful Model Parameters Genesis in Multi-population Genetic Algorithm. Global J Technol Optim 5: 164. doi: 10.4172/2229-8711.1000164

Genetic operators' significance assessment in multi-population genetic algorithms

Maria Angelova* and Tania Pencheva

Institute of Biophysics and Biomedical Engineering,
Bulgarian Academy of Sciences,
105 Acad. G. Bonchev Str.,
1113 Sofia, Bulgaria
E-mail: maria.angelova@biomed.bas.bg
E-mail: tania.pencheva@biomed.bas.bg
*Corresponding author

Abstract: Genetic algorithms are widely applied bioinspired optimisation technique that search for a global optimal solution via three main genetic operators, namely selection, crossover, and mutation. In order to determine the operators importance when multi-population genetic algorithm is applied for parameter identification of *S. cerevisiae* fed-batch cultivation, recently presented procedure for significance assessment has been implemented. Based on obtained results the most significant genetic operator has been distinguished and its influence on finding the global optimal solution has been evaluated as well.

Keywords: multi-population genetic algorithm; genetic operators; parameter identification; *S. cerevisiae* fed-batch cultivation.

Reference to this paper should be made as follows: Angelova, M. and Pencheva, T. (2014) 'Genetic operators' significance assessment in multi-population genetic algorithms', *Int. J. Metaheuristics*, Vol. 3, No. 2, pp.162–173.

Biographical notes: Maria Angelova received her MSc in Bioelectroengineering from the Technical University, Sofia in 2006. Since 2009, she has been a PhD student at the Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences (BAS), Sofia. She has published more than 20 publications. Her scientific interests are in the fields of modelling and optimisation of biotechnological end ecological processes, genetic algorithms and ant colony optimisation.

Tania Pencheva received her MSc in Bioengineering (1994) and MSc in Applied Mathematics and Informatics (1998) from the Technical University, Sofia and PhD (2003) from the Centre of Biomedical Engineering, Bulgarian Academy of Sciences (BAS), Sofia. Currently, she is an Associate Professor at the Institute of Biophysics and Biomedical Engineering (BAS). She is a Deputy Chief Editor of *International Journal Bioautomation* and a member of many programme and organisation committees of national and international conferences. She has published more than 150 publications including eight books and book chapters with about 200 known citations. She has been awarded by the Union of the Scientists in Bulgaria for young scientist's

Post-Docking Optimization and Analysis of Protein-Ligand Interactions of Estrogen Receptor Alpha using AMMOS Software

Tania Pencheva[1,§], Dessislava Jereva[1,§], Maria A. Miteva[*,2,3] and Ilza Pajeva[*,1]

[1]*Institute of Biophysics and Biomedical Engineering – Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria*

[2]*INSERM, UMR-S 973, MTi, 75013 Paris, France*

[3]*University Paris Diderot, Sorbonne Paris Cité, UMR-S 973, MTi, 75013 Paris, France*

Abstract: Understanding protein-ligand interactions is a critical step in rational drug design/virtual ligand screening. In this work we applied the AMMOS_ProtLig software for post-docking optimization of estrogen receptor alpha complexes generated after virtual ligand screening protocol. Using MOE software we identified the ligand-receptor interactions in the optimized complexes at different levels of protein flexibility and compared them to the experimentally observed interactions. We analyzed in details the binding sites of three X-ray complexes of the same receptor and identified the key residues for the protein-ligand interactions. The complexes were further processed with AMMOS_ProtLig and the interactions in the predicted poses were compared to those observed in the X-ray structures. The effect of employing different levels of flexibility was analyzed. The results confirmed the AMMOS_ProtLig applicability as a helpful post-docking optimization tool for virtual ligand screening of estrogen receptors.

Keywords: AMMOS, binding site, estrogen receptor, *in silico* screening, protein flexibility, protein-ligand interactions.

INTRODUCTION

In silico approaches have gained immense significance in the recent years and have become an integral part of research in both, academic institutions and different industry branches, thus assisting drug design and discovery. Among these methods the virtual ligand screening (VLS) of bioactive compounds has been established as an effective approach to handle large sets of compounds improving in this way the "hit-rate" of drug discovery programs [1-3]. Structure-based VLS (SB-VLS) relies on the three dimensional structure of the biological target obtained either through experimental methods such as X-ray crystallography or NMR spectroscopy [4], or predicted by homology modeling. Most of the SB-VLS protocols [5-7] employ docking of all compounds in an appropriate chemical library into the binding pocket of the selected target and evaluate the fit between the molecules *via* scoring functions. Critical for docking accuracy is the treatment of the ligand flexibility and, in many cases, the receptor flexibility [8-10], the last increasing considerably the computational time. It is impractical to perform SB-VLS using docking on a fully flexible protein receptor for a large number of ligands. It has been shown that post-docking optimization, either after conventional docking-scoring procedures or after hierarchical VLS protocols [6, 11, 12] helps in improving both, the docking pose and the scoring, and thus the overall efficiency of SB-VLS. This view is supported by a number of examples of binding pose prediction and enrichment improvements after post-docking energy minimization [13-17].

The effectiveness of the post-docking optimization is often estimated on the achieved final enrichment. It is also critical to analyze at atomic level how the protein interacts with the most potent ligands and to validate such analysis based on experimental structures of protein-ligand complexes. Considering the X-ray complex as "the best docking pose" one can compare the number and the type of the protein-ligand interactions in the optimized post-docking complexes to the X-ray ones.

Recently, we reported the software platform AMMOS [18, 19] that performs an automatic procedure for structural refinement of compound collections (AMMOS_SmallMols package) and energy minimization of protein-ligand complexes (AMMOS_ProtLig package) making use of the AMMP (Another Molecular Mechanics Program) force field [20-22]. Employing molecular mechanics optimization AMMOS_ProtLig allows five levels of protein flexibility – from fully flexible to rigid protein, while ligands are always kept flexible. The tool can be applied on a large number of protein-ligand complexes pre-generated with user-selected docking programs.

In this work we focused on the analysis of protein-ligand interactions for the estrogen receptor alpha (ERα) ligand binding domain complexes using AMMOS_ProtLig. ERα is one of the most studied targets for *in silico* screening [23-25]. In the recent years this receptor received a tremendous amount of attention due to the potentially adverse effects of

*Address correspondence to these authors at the (MAM) University Paris Diderot, Sorbonne Paris Cité, UMR-S 973, MTi, 75013 Paris, France; Tel: +331 57 27 83 92; Fax: +331 57 27 83 72; E-mail: maria.miteva@univ-paris-diderot.fr and (IP) Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria; Tel: +359 2 9793605; Fax: +359 2 8723787;
E-mail: pajeva@biomed.bas.bg

§Contributed equally.

1875-6697/13 $58.00+.00

Chapter 1

Improvement of Multi-population Genetic Algorithms Convergence Time

Maria Angelova and Tania Pencheva

Abstract. Different kinds of multi-population genetic algorithms have been investigated for a parameter identification of a fermentation process. Altogether six realizations of multi-population genetic algorithms have been proposed, four of them with a different sequence of implementation of main genetic operators selection, crossover and mutation, and another two without mutation. A comparison of the considered six kinds of genetic algorithms is presented for a parameter identification of a fed-batch cultivation of *S. cerevisiae.* The influence of the most important genetic algorithm parameters, namely generation gap and rates of crossover, mutation, insertion, and migration, have also been investigated. All kinds of considered multi-population genetic algorithms lead to similar values of the optimization criterion. Among those with three genetic operators, the algorithm with a sequence of selection, crossover, and mutation is significantly faster than the others. When mutation is avoided, the genetic algorithm with a sequence of selection and crossover is faster than the other one.

Keywords. Multi-population Genetic Algorithms, Parameter Identification, Fermentation Process, Saccharomyces Cerevisiae.

Mathematics Subject Classification 2010. 37N25, 49J15, 90C31, 97M60.

1.1 Introduction

Investigation of fermentation processes (FP) is a question of present interest due to the fact that different branches of industry use them in the production of pharmaceuticals, chemicals and enzymes, yeast, foods, and beverages. Taking into account that FP are complex, nonlinear, dynamic systems with interdependent and time-varying process variables, their examining as modelling and control objects can present a serious challenge. An important step for adequate modeling of FP is the choice of a certain optimization procedure for model parameter identification. Genetic algorithms (GA) are a quite promising stochastic global optimization method which is widely applied for solving different complicated problems [2–5, 7–9]. GA, inspired by Darwin's theory of "survival of the fittest." [3], are one of the methods based on biological evolution. Properties like hard problems solving, noise tolerance, easy to interface and hybridize make genetic algorithms suitable and more workable for a parameter identification of fermentation models [1, 2, 4, 5, 7–9].

Chapter 11
Genetic Algorithms Quality Assessment Implementing Intuitionistic Fuzzy Logic

Tania Pencheva
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Bulgaria

Maria Angelova
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Bulgaria

Krassimir Atanassov
Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, Bulgaria

ABSTRACT

Intuitionistic fuzzy logic has been implemented in this investigation aiming to derive intuitionistic fuzzy estimations of model parameters of yeast fed-batch cultivation. Considered here are standard simple and multi-population genetic algorithms as well as their modifications differ from each other in execution order of main genetic operators (selection, crossover, and mutation). All are applied for the purpose of parameter identification of S. cerevisiae fed-batch cultivation. Performances of the examined algorithms have been assessed before and after the application of a procedure for narrowing the range of model parameters variation. Behavior of standard simple genetic algorithm has been also examined for different values of proof as the most sensitive genetic algorithms parameter toward convergence time, namely, generation gap (GGAP). Results obtained after the intuitionistic fuzzy logic implementation for assessment of genetic algorithms performance have been compared. As a result, the most reliable algorithm/value of GGAP ensuring the fastest and the most valuable solution is distinguished.

DOI: 10.4018/978-1-4666-4450-2.ch011

Computers and Mathematics with Applications 64 (2012) 266–271

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

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

Data extraction module—A supplementary tool for the AMMOS_ProtLig software package

Dessislava Jereva, Ilza Pajeva, Tania Pencheva *

Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 105, Acad. Georgi Bonchev Str., 1113 Sofia, Bulgaria

ARTICLE INFO

Keywords:
Virtual ligand screening
Post-docking optimization
AMMOS_ProtLig
Data extraction
Protein–ligand complex

ABSTRACT

Data Extraction Module is a supplementary tool for the software package *AMMOS_ProtLig* (Automated Molecular Mechanics Optimization tool for *in silico* Screening of Protein–Ligand interactions). The purpose of the module is to facilitate further analyses of the results after *AMMOS ProtLig* application. *Data Extraction Module* automatically processes the output files after post-docking optimization with *AMMOS_ProtLig*. The module retrieves the ligand and the protein data files of a protein–ligand complex, specified according to user preferences. *Data Extraction Module* is developed as an open-source graphical application for *Microsoft Windows* and *Linux* operating systems. The applicability of the module has been tested on five protein–ligand complexes with different physicochemical properties and topology. *Data Extraction Module* reduces the user's involvement and accelerates the process of data extraction by 5 to 10, depending on the size of the protein–ligand complex.

1. Introduction

Nowadays, *in silico* approaches have become an integral part of research in both academic institutions and branches of industry, aiming at directing optimal drug design and discovery. Among these methods, virtual screening of bioactive compounds has become an established technique, whose ultimate goal is the identification of potential drugs among thousands of chemical structures [1–4]. A crucial step in the virtual screening is docking. Once the knowledge of the target is available, rapid docking algorithms are used to place the available candidate compounds within the active site of the biochemical target of interest [1]. Recently, it has been suggested that post-docking optimization, either after conventional docking–scoring procedures or after hierarchical virtual screening protocols, may help to further improve both the docking pose and the scoring, and in this way the overall efficiency of virtual screening experiments [5]. For this purpose, the software package *AMMOS_ProtLig* (Automated Molecular Mechanics Optimization tool for *in silico* Screening of Protein–Ligand interactions) has been developed [5] for energy minimization of pre-docked protein–ligand complexes. *AMMOS_ProtLig* is based on molecular mechanics, and it provides different levels of flexibility of the protein atoms; the ligand is always flexible. As a result of the optimization, two of the *AMMOS_ProtLig* output files contain the coordinates of the minimized ligands and the active atoms (atoms of the protein that can move depending on the selected level of flexibility). Other *AMMOS_ProtLig* output file stores the energies of the minimized protein–ligand complexes ranked by increasing energy values (corresponding to decreasing protein–ligand binding affinity). The stored data can be further used for generation of enrichment curves in the virtual ligand screening protocols, analysis of protein–ligand interactions in the minimized complexes, and other applications.

So far, treatment of *AMMOS_ProtLig* output files for the purposes of further analysis has been performed manually by the user. Thus the process is time consuming and prone to errors. In this paper, we present *Data Extraction Module*, which aims

* Corresponding author.
E-mail addresses: dessislava.jereva@biomed.bas.bg (D. Jereva), pajeva@biomed.bas.bg (I. Pajeva), tania.pencheva@biomed.bas.bg, tania.pencheva@clbme.bas.bg (T. Pencheva).

0898-1221/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2012.01.060

Computers and Mathematics with Applications 64 (2012) 221–228

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

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

Purposeful model parameters genesis in simple genetic algorithms

Maria Angelova, Krassimir Atanassov, Tania Pencheva *

Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad. Georgi Bonchev Str., 1113 Sofia, Bulgaria

ARTICLE INFO

Keywords:
Genetic algorithms
Purposeful genesis
Model parameters
Fermentation process
Saccharomyces cerevisiae

ABSTRACT

Simple genetic algorithms have been investigated aiming to improve the algorithm convergence time. Because of the stochastic nature of genetic algorithms, several runs have to be performed in order to achieve representative results. A procedure for purposeful genesis concerning intervals of variations of model parameters is proposed for a standard simple genetic algorithm, aiming to improve significantly the algorithm effectiveness. Such a stepwise methodology is applied to parameter identification of fed-batch cultivation of *S. cerevisiae*. The procedure is further validated to a modified simple genetic algorithm with changed sequence of main genetic algorithm operators, namely mutation, crossover and selection, proven to be faster than the standard one. Results obtained from both applications show significant improvement of the algorithm convergence time while saving the model accuracy.

1. Introduction

Application of fermentation processes (FPs) in different branches of industry makes their investigation a very topical problem. Modeling and further optimal control of FPs is a non-trivial task since they are complex, dynamic systems with interdependent and time-varying process variables. An important step for FP adequate modeling is the choice of a certain optimization procedure for model parameter identification. Inability of conventional optimization methods such as Nelder–Mead's minimization, sequential quadratic programming, quasi-Newton algorithms (i.e. Broyden, Fletcher, Goldfarb and Shanno), etc., to reach a satisfactory solution [1,2] provokes an idea that some stochastic algorithms should be applied. As an alternative for solving such a complex problem, evolutionary algorithms can be taken into consideration. Among these, the concept of genetic algorithm (GA) [3], inspired by Darwin's theory of "survival of the fittest", is a stochastic global optimization technique with applications in different areas [4–11]. Some properties of GA, such as the ability of solving hard problems, noise tolerance, easiness to interface and hybridize, make them a suitable and quite workable technique for parameter identification of fermentation models [2,7–11]. Promising results obtained by the utilization of GA encourage their future investigation.

The effectiveness of a certain optimization technique can be evaluated by the model accuracy achieved and the convergence time needed. Due to the stochastic nature of GA obtained results might be quite diverse. That is why several runs have to be performed in order to achieve representative results. The accumulation of data from different runs provoked the idea for purposeful genesis concerning intervals of variations of model parameters. Such an idea is going to be elaborated for standard simple genetic algorithms and further promptly applied to a modified one, previously proven to be faster [9].

Standard simple genetic algorithm (SGA), originally presented in [3], is here denoted with the abbreviation SGA-SCM, derived from the sequential execution of the main genetic operators: *selection*, *crossover*, *mutation*. In SGA-SCM, chromosomes (a coded parameter set) representing better possible solutions according to their own objective function

* Corresponding author.
E-mail addresses: maria.angelova@biomed.bas.bg (M. Angelova), krat@bas.bg (K. Atanassov), tania.pencheva@biomed.bas.bg, tania.pencheva@clbme.bas.bg (T. Pencheva).

0898-1221/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2012.01.047

Доклади на Българската академия на науките
Comptes rendus de l'Académie bulgare des Sciences
Tome 65, No 3, 2012

MATHEMATIQUES
Modèles mathématiques

ALGORITHMS IMPROVING CONVERGENCE TIME IN PARAMETER IDENTIFICATION OF FED-BATCH CULTIVATION

Maria Angelova, Tania Pencheva

(*Submitted by Academician I. Popchev on October 25, 2011*)

Abstract

Fermentation processes are characterized with non-linear and time-dependent parameters that makes their parameter identification non-trivial task. Failure of conventional optimization methods to lead to a satisfied solution provokes an idea some stochastic algorithms to be applied. As such, different modifications of simple genetic algorithms (SGA) have been investigated aiming to improve the model accuracy and the algorithm convergence time. Altogether six modifications of SGA have been proposed, four of them with a different sequence of implementation of selection, crossover and mutation, and another two – without mutation, firstly presented here. A comparison of six considered modifications of SGA is performed for parameter identification of a fed-batch cultivation of *S. cerevisiae*. The influence of the most important genetic algorithm parameters, namely generation gap, crossover and mutation rates, has been investigated too. Almost all considered SGA lead to similar values of the optimization criterion but the algorithm with an operators sequence of mutation, crossover and selection, and especially those without mutation, are significantly faster than the others. Among the considered GA parameters, generation gap influences most significantly the algorithm convergence time, saving up to 40% of the time without affecting the model accuracy.

Key words: genetic algorithms, genetic operators, genetic algorithm parameters, parameter identification, fermentation process, *Saccharomyces cerevisiae*

This work is partially supported by the European Social Fund and Bulgarian Ministry of Education, Youth and Science under Operative Programme "Human Resources Development", Grant BG051PO001-3.3.04/40 and National Science Fund of Bulgaria, Grant DID 02-29 "Modelling Processes with Fixed Development Rules".

Chapter 9

AMMOS Software: Method and Application

Tania Pencheva, David Lagorce, Ilza Pajeva, Bruno O. Villoutreix, and Maria A. Miteva

Abstract

Recent advances in computational sciences enabled extensive use of in silico methods in projects at the interface between chemistry and biology. Among them virtual ligand screening, a modern set of approaches, facilitates hit identification and lead optimization in drug discovery programs. Most of these approaches require the preparation of the libraries containing small organic molecules to be screened or a refinement of the virtual screening results. Here we present an overview of the open source AMMOS software, which is a platform performing an automatic procedure that allows for a structural generation and optimization of drug-like molecules in compound collections, as well as a structural refinement of protein-ligand complexes to assist in silico screening exercises.

Key words: 3D structure generation, Structure refinement, Virtual screening, AMMOS, AMMP, Open source/free software

1. Introduction

Recent advances in computational sciences enabled extensive use of *in silico* methods in projects at the interface between chemistry and biology. Among them virtual ligand screening, a modern set of approaches, facilitates hit identification and lead optimization in drug discovery programs (1–3). Nowadays various in silico methods can be employed for such purposes, i.e., drug-like properties' predictions (4, 5), ligand-based virtual screening (i.e., chemical similarity search (6–8), pharmacophore search (9)), or structure-based virtual screening employing docking and scoring techniques (10–14). Most of these approaches require preparation of the libraries containing small organic molecules to be screened (15, 16) or refinement of the virtual screening results (17, 18).

Riccardo Baron (ed.), *Computational Drug Discovery and Design*, Methods in Molecular Biology, vol. 819,
DOI 10.1007/978-1-61779-465-0_9,

Modified Simple Genetic Algorithms Improving Convergence Time for the Purposes of Fermentation Process Parameter Identification

MARIA ANGELOVA
Institute of Biophysics and Biomedical Engineering
Bulgarian Academy of Sciences
105 Acad. Georgi Bonchev Str., 1113 Sofia
BULGARIA
maria.angelova@biomed.bas.bg

PEDRO MELO-PINTO
Centre for the Research and Technology of Agro-Environmental and Biological Sciences
University of Trás-os-Montes and Alto Douro
Quinta de Prados, Apartado 1013, 5001-801 Vila Real
PORTUGAL
pmelo@utad.pt

TANIA PENCHEVA
Institute of Biophysics and Biomedical Engineering
Bulgarian Academy of Sciences
105 Acad. Georgi Bonchev Str., 1113 Sofia
BULGARIA
tania.pencheva@biomed.bas.bg

Abstract: - Fermentation processes are characterized with non-linear and time-dependent parameters that make their parameter identification non-trivial task. Failure of conventional optimization methods to yield a satisfactory solution provokes the idea for some stochastic algorithms to be applied. As such, different modifications of simple genetic algorithms (SGA) have been investigated aiming to improve the model accuracy and the algorithm convergence time. For that purpose two new modifications of SGA are developed here. SGA realizations differ from each other in the sequence of implementation of the main genetic operators selection, crossover and mutation. A comparison of the herewith developed two modifications of SGA and standard SGA towards algorithm convergence time and model accuracy is presented for parameter identification of *S. cerevisiae* fed-batch cultivation. The influence of the most important genetic algorithm parameters, namely generation gap, crossover and mutation rates has been investigated, too. Both proposed modifications of SGA produce similar values of the optimization criterion, meanwhile being significantly faster than the standard SGA. Among the considered genetic algorithms parameters, generation gap influences the algorithm calculation time most significantly, saving up to 53% of the time without affecting the model accuracy.

Key-Words: - Genetic algorithms, genetic operators, genetic algorithm parameters, parameter identification, fed-batch fermentation process.

1 Introduction

Fermentation processes (FP) underlie the production of pharmaceuticals, chemicals and enzymes, yeast, foods, beverages, etc. in various industry branches. That is why FP modeling and future optimal control are questions of continued interest. Meanwhile, the modeling and control of FP pose serious challenges to their researchers as they are complex, nonlinear dynamic systems with interdependent and time-varying process parameters. An important step for adequate modeling of non-linear models of FP is the choice of a certain optimization procedure for model parameter identification. Failure of conventional optimization methods such as Nelder-Mead's minimization, sequential quadratic programming, quasi-Newton algorithms (i.e. Broyden, Fletcher, Goldfarb and Shanno), etc. to yield a satisfactory solution [1, 2] provokes the idea for some stochastic

Hindawi Publishing Corporation
International Journal of Chemical Engineering
Volume 2011, Article ID 646917, 7 pages
doi:10.1155/2011/646917

Research Article

Tuning Genetic Algorithm Parameters to Improve Convergence Time

Maria Angelova and Tania Pencheva

Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Street, 1113 Sofia, Bulgaria

Correspondence should be addressed to Tania Pencheva, tania.pencheva@clbme.bas.bg

Received 31 March 2011; Revised 17 June 2011; Accepted 5 July 2011

Academic Editor: Ahmet N. Palazoglu

Fermentation processes by nature are complex, time-varying, and highly nonlinear. As dynamic systems their modeling and further high-quality control are a serious challenge. The conventional optimization methods cannot overcome the fermentation processes peculiarities and do not lead to a satisfying solution. As an alternative, genetic algorithms as a stochastic global optimization method can be applied. For the purpose of parameter identification of a fed-batch cultivation of *S. cerevisiae* altogether four kinds of simple and four kinds of multipopulation genetic algorithms have been considered. Each of them is characterized with a different sequence of implementation of main genetic operators, namely, selection, crossover, and mutation. The influence of the most important genetic algorithm parameters—generation gap, crossover, and mutation rates has—been investigated too. Among the considered genetic algorithm parameters, generation gap influences most significantly the algorithm convergence time, saving up to 40% of time without affecting the model accuracy.

1. Introduction

Fermentation processes (FP) are preferred and widely used in different branches of industry. The modeling and control of FP pose serious challenges as FP are complex, nonlinear dynamic systems with interdependent and time-varying process parameters. An important step for adequate modeling of nonlinear models of FP is the choice of a certain optimization procedure for model parameter identification. Different metaheuristics methods have been applied to surmount the parameter estimation difficulties [1–3]. Since the conventional optimization methods cannot overcome the limitations of FP [4], genetic algorithms (GAs), as a stochastic global optimization method, are quite promising. Among a number of searching tools, the genetic algorithms are one of the methods based on biological evolution and inspired by Darwin's theory of "survival of the fittest" [5]. GAs are directed random search techniques, based on the mechanics of natural selection and natural genetics. GAs find the global optimal solution in complex multidimensional search spaces simultaneously evaluating many points in the parameter space. They require only information concerning the quality of the solution and do not require linearity in the parameters. GAs have been successfully applied in a variety of areas to solve many engineering and optimization problems [6–8]. Properties such as noise tolerance and ease of interfacing and hybridization make GA a suitable method for the identification of parameters in fermentation models [9–13].

Simple genetic algorithm (SGA) presented initially in Goldberg [5] searches for a global optimal solution using three main genetic operators in a sequence selection, crossover, and mutation. GAs work with a population of coded parameter set called "chromosome." Each of these artificial chromosomes is composed of binary strings (or genes) of certain length (number of binary digits). Each gene contains information for the corresponding parameter. Through selection chromosomes representing better possible solutions according to their own objective function values are chosen from the population. After the reproduction, crossover proceeds in order to form new offspring. Mutation is then applied with determinate probability. Even though selection and crossover effectively work, occasionally, a GA may lose some potentially useful information. That is why

TOPICS IN CHEMISTRY AND MATERIAL SCIENCE, Vol. 5 (2011) pp. 43–51
Current Issues in Organic Chemistry 2, edited by R.D. Nikolova, S. Simova,
P. Denkova, G.N. Vayssilov

Post-Docking Optimization and Analysis of Protein-Ligand Interactions with AMMOS Free Software

T. Pencheva[1], **P. Alov**[1], **D. Jereva**[1], **I. Pajeva**[1], **M.A. Miteva**[2]
[1]Institute of Biophysics and Biomedical Engineering, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria
[2]MTI, INSERM U973 – University Paris Diderot, 75205 Paris Cedex 13, France

Abstract
AMMOS (Automated Molecular Mechanics Optimization tool for *in silico* Screening) is a software tool for structural refinement of compound collections and energy minimization of protein-ligand complexes. It performs an automatic procedure for energy minimization, based on molecular mechanics (AMMP force-field), at five different levels of protein receptor flexibility. AMMOS has been shown to improve the enrichment in virtual screening experiments after docking: 40% to 60% of the actives were found in the top 3% to 5% of the entire screened compound collection for the tests proteins. In this work a further validation of AMMOS is reported in relation to enrichment improvement and protein-ligand interactions in several protein-ligand complexes. Binding sites of different topology and physico-chemical properties have been analyzed using Ligand Interaction Application module in MOE software. The results show that AMMOS refines the protein-ligand complexes and, depending on the level of flexibility, restores to a different extent the interactions identified in the experimental structures of the protein-ligand complexes studied.

1 Introduction

Virtual ligand screening (VLS), or *in silico* screening, has been established as an effective approach to handle large sets of compounds and to improve the "hit-rate" of drug discovery programs [1]. For SB-VLS (structure-based VLS) methods, it is assumed that the 3D structure of the target is known either by X-ray crystallography or NMR experiments, or predicted by homology modeling. The aim here is to dock all the ligands present in a database into the binding pocket of the selected target and evaluate the fit between the molecules [2]. While SB-VLS is known to give valuable information for selection of new hits among chemical libraries of millions of compounds by using common docking-scoring algorithms [3], it is also costly in terms of computational time. The extensive treatment of the ligand, and, in some cases, the receptor flexibility in the docking process is critical for docking accuracy but increases significantly the duration

European Journal of Medicinal Chemistry 45 (2010) 2622–2628

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Short communication

Post-docking virtual screening of diverse binding pockets: Comparative study using DOCK, AMMOS, X-Score and FRED scoring functions

Tania Pencheva [a], Oumarou Samna Soumana [b], Ilza Pajeva [a], Maria A. Miteva [b,*]

[a] Centre of Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria
[b] MTI, INSERM U973, University Paris Diderot, 5 rue Marie-Andrée Lagroua, 75205 Paris Cedex 13, France

ARTICLE INFO

Article history:
Received 7 July 2009
Received in revised form
8 December 2009
Accepted 10 December 2009
Available online 4 January 2010

Keywords:
Scoring functions
DOCK
AMMOS
GBSA
X-Score
FRED

ABSTRACT

Most of the benchmark studies on docking-scoring methods reported in the last decade conclude that no single scoring function performs well across different protein targets. In this study a comparison of thirteen commonly used force field and empirical scoring functions as implemented in DOCK, AMMOS, X-Score and FRED is carried out on five proteins with diverse binding pockets. The performance is analyzed in relation to the physicochemical properties of the binding sites. The solvation effects are considered via the Generalized Born/Surface Area (GBSA) solvation method for one of the assessed scoring functions. We examined the ability of these scoring functions to discriminate between active and inactive compounds over receptor-based focused libraries. Our results demonstrated that the employed here empirical scoring functions were more appropriate for the pocket of predominant hydrophobic nature while the force field scoring functions performed better on the mixed or polar pockets.

1. Introduction

Structure-based virtual screening (SBVS) is successfully applied today to discover new hit molecules and it has been demonstrated to increase the efficiency of drug design projects [1–4]. A common strategy of SBVS is to select potential actives through evaluation of ligand binding affinities by applying docking-scoring methods [5–7] to a 3D receptor structure (most often a protein) and a huge number of small molecule candidates. Despite of development of rigorous methods such as the free energy perturbation [8] or the linear interaction energy (LIE) [9] the fast and accurate calculation of the free ligand binding energy still remains a great challenge in SBVS [9]. Over the past 15 years different approximations, the so called scoring functions, have been proposed for faster estimation of protein-ligand interactions. The scoring functions aim to guide the positioning (docking) of candidate compounds to receptor binding sites, to decide on the probable binding modes, and finally to discriminate between potential active and inactive molecules. A comprehensive analysis of scoring functions advantages and shortcomings can be found in [5–7].

Current scoring functions can be divided into three groups, force field based, empirical, and knowledge-based. The force field scoring functions estimate the binding free energy as a sum of independent molecular mechanics force fields potentials. Solvation and entropy contributions can also be considered, e.g. free energy scoring methods in terms of Molecular Mechanics – Poisson-Boltzmann Surface Area/Generalized Born Surface Area (MM-PBSA/GBSA) can be successful in improving the binding affinity prediction and ranking the actives [10,11]. The empirical approximation represents the binding free energy as a weighted sum of protein-ligand interaction terms by fitting the scoring function to experimental binding affinity data. Recently new empirical scoring functions have been proposed, involving, for instance, improved models for hydrophobic interactions [12], quantum chemical energy binding terms [13] or NMR chemical shift perturbation prediction [14]. The knowledge-based scoring functions [15,16] are exclusively derived from statistical analyses of atom-pairs frequencies in protein-ligand complexes with known 3D structures.

In the last decade a number of scoring functions have been benchmarked. A common conclusion from different comparative studies is that no single scoring function performs well across different protein targets [5,16–19]. Assessment reports show that the size, shape, polarity and flexibility of the binding sites can dramatically affect the performance.

* Corresponding author. Tel.: +331 57 27 83 92; fax: +331 57 27 83 72.
E-mail addresses: tania.pencheva@clbme.bas.bg (T. Pencheva), oumarou.samna@univ-paris-diderot.fr (O.S. Soumana), pajeva@bio.bas.bg (I. Pajeva), maria.miteva@univ-paris-diderot.fr (M.A. Miteva).

0223-5234/$ – see front matter © 2009 Published by Elsevier Masson SAS.
doi:10.1016/j.ejmech.2009.12.025

BMC Chemical Biology

Software

Open Access

DG-AMMOS: A New tool to generate 3D conformation of small molecules using *D*istance *G*eometry and *A*utomated *M*olecular *M*echanics *O*ptimization for *in silico* *S*creening

David Lagorce[†1], Tania Pencheva[†2], Bruno O Villoutreix[1] and Maria A Miteva*[1]

Address: [1]MTI, INSERM U973 - University Paris Diderot, 5 rue Marie-Andrée Lagroua, 75205 Paris Cedex 13, France and [2]Centre of Biomedical Engineering, Bulgarian Academy of Sciences, 105 Acad. G. Bonchev Str., 1113 Sofia, Bulgaria

Email: David Lagorce - david.lagorce@univ-paris-diderot.fr; Tania Pencheva - tania.pencheva@clbme.bas.bg; Bruno O Villoutreix - bruno.villoutreix@univ-paris-diderot.fr; Maria A Miteva* - maria.miteva@univ-paris-diderot.fr

* Corresponding author †Equal contributors

Published: 13 November 2009

BMC Chemical Biology 2009, **9**:6 doi:10.1186/1472-6769-9-6

Received: 17 July 2009
Accepted: 13 November 2009

This article is available from: http://www.biomedcentral.com/1472-6769/9/6

Abstract

Background: Discovery of new bioactive molecules that could enter drug discovery programs or that could serve as chemical probes is a very complex and costly endeavor. Structure-based and ligand-based *in silico* screening approaches are nowadays extensively used to complement experimental screening approaches in order to increase the effectiveness of the process and facilitating the screening of thousands or millions of small molecules against a biomolecular target. Both *in silico* screening methods require as input a suitable chemical compound collection and most often the 3D structure of the small molecules has to be generated since compounds are usually delivered in 1D SMILES, CANSMILES or in 2D SDF formats.

Results: Here, we describe the new open source program DG-AMMOS which allows the generation of the 3D conformation of small molecules using Distance Geometry and their energy minimization via Automated Molecular Mechanics Optimization. The program is validated on the Astex dataset, the ChemBridge Diversity database and on a number of small molecules with known crystal structures extracted from the Cambridge Structural Database. A comparison with the free program Balloon and the well-known commercial program Omega generating the 3D of small molecules is carried out. The results show that the new free program DG-AMMOS is a very efficient 3D structure generator engine.

Conclusion: DG-AMMOS provides fast, automated and reliable access to the generation of 3D conformation of small molecules and facilitates the preparation of a compound collection prior to high-throughput virtual screening computations. The validation of DG-AMMOS on several different datasets proves that generated structures are generally of equal quality or sometimes better than structures obtained by other tested methods.

BMC Bioinformatics

Methodology article

AMMOS: Automated Molecular Mechanics Optimization tool for *in silico* Screening

Tania Pencheva[1,2], David Lagorce[1], Ilza Pajeva[2], Bruno O Villoutreix[1] and Maria A Miteva*[1]

Address: [1]INSERM U648, Bioinformatics-MTI University Paris Diderot, 5 rue Marie-Andrée Lagroua, 75205 Paris Cedex 13, France and [2]Centre of Biomedical Engineering "Prof. Ivan Daskalov", Bulgarian Academy of Sciences, 105, Akad. Georgi Bonchev Str., 1113 Sofia, Bulgaria

Email: Tania Pencheva - tania.pencheva@clbme.bas.bg; David Lagorce - david.lagorce@univ-paris-diderot.fr; Ilza Pajeva - pajeva@bio.bas.bg; Bruno O Villoutreix - bruno.villoutreix@univ-paris-diderot.fr; Maria A Miteva* - maria.miteva@univ-paris-diderot.fr

* Corresponding author

Published: 16 October 2008

BMC Bioinformatics 2008, **9**:438 doi:10.1186/1471-2105-9-438

Received: 18 April 2008
Accepted: 16 October 2008

This article is available from: http://www.biomedcentral.com/1471-2105/9/438

Abstract

Background: Virtual or *in silico* ligand screening combined with other computational methods is one of the most promising methods to search for new lead compounds, thereby greatly assisting the drug discovery process. Despite considerable progresses made in virtual screening methodologies, available computer programs do not easily address problems such as: structural optimization of compounds in a screening library, receptor flexibility/induced-fit, and accurate prediction of protein-ligand interactions. It has been shown that structural optimization of chemical compounds and that post-docking optimization in multi-step structure-based virtual screening approaches help to further improve the overall efficiency of the methods. To address some of these points, we developed the program AMMOS for refining both, the 3D structures of the small molecules present in chemical libraries and the predicted receptor-ligand complexes through allowing partial to full atom flexibility through molecular mechanics optimization.

Results: The program AMMOS carries out an automatic procedure that allows for the structural refinement of compound collections and energy minimization of protein-ligand complexes using the open source program AMMP. The performance of our package was evaluated by comparing the structures of small chemical entities minimized by AMMOS with those minimized with the Tripos and MMFF94s force fields. Next, AMMOS was used for full flexible minimization of protein-ligands complexes obtained from a mutli-step virtual screening. Enrichment studies of the selected pre-docked complexes containing 60% of the initially added inhibitors were carried out with or without final AMMOS minimization on two protein targets having different binding pocket properties. AMMOS was able to improve the enrichment after the pre-docking stage with 40 to 60% of the initially added active compounds found in the top 3% to 5% of the entire compound collection.

Conclusion: The open source AMMOS program can be helpful in a broad range of *in silico* drug design studies such as optimization of small molecules or energy minimization of pre-docked protein-ligand complexes. Our enrichment study suggests that AMMOS, designed to minimize a large number of ligands pre-docked in a protein target, can successfully be applied in a final post-processing step and that it can take into account some receptor flexibility within the binding site area.

Electronic Journal of Biotechnology ISSN: 0717-3458
Vol.10 No.4, Issue of October 15, 2007

Received November 16, 2006 / Accepted May 7, 2007
DOI: 10.2225/vol10-issue4-fulltext-5
RESEARCH ARTICLE

Multiple model approach to modelling of *Escherichia coli* fed-batch cultivation extracellular production of bacterial phytase

Olympia Roeva*
Centre of Biomedical Engineering Prof. Ivan Daskalov
Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str.
Sofia 1113, Bulgaria
Tel: 359 2 979 36 48
E-mail: olympia@clbme.bas.bg

Tania Pencheva
Centre of Biomedical Engineering Prof. Ivan Daskalov
Bulgarian Academy of Sciences
105 Acad. G. Bonchev Str.
Sofia 1113, Bulgaria
Tel: 359 2 979 36 08
E-mail: tania.pencheva@clbme.bas.bg

Stoyan Tzonkov
Centre of Biomedical Engineering Prof. Ivan Daskalov
Bulgarian Academy of Sciences105 Acad. G. Bonchev Str.
Sofia 1113, Bulgaria
Tel: 359 2 979 36 48
E-mail: tzonkov@clbme.bas.bg

Michael Arndt
Institute of Technical Chemistry
University of Hannover
LeibnizUniversity of Hannover
3 Callinstrasse, 30167
Hannover, Germany
E-mail: mail@michaelarndt.net

Bernd Hitzmann
Institute of Technical Chemistry
University of Hannover
LeibnizUniversity of Hannover
3 Callinstrasse, 30167
Hannover, Germany
E-mail: hitzmann@iftc.uni-hannover.de

Sofia Kleist
Chair of Fermentation Technique
Technical Faculty
University of Bielefeld
25 Universitätsstr., 33615
Bielefeld, Germany
E-mail: sok@fermtech.techfak.uni-bielefeld.de

Gerchard Miksch
Chair of Fermentation Technique
Technical Faculty
University of Bielefeld
25 Universitätsstr., 33615
Bielefeld, Germany
E-mail: gmi@fermtech.techfak.uni-bielefeld.de

Karl Friehs
Chair of Fermentation Technique
Technical Faculty
University of Bielefeld
25 Universitätsstr., 33615
Bielefeld, Germany
E-mail: kfr@fermtech.techfak.uni-bielefeld.de

Erwin Flaschel
Chair of Fermentation Technique
Technical Faculty
University of Bielefeld
25 Universitätsstr., 33615
Bielefeld, Germany
E-mail: efl@fermtech.techfak.uni-bielefeld.de

Financial support: This work is partially supported from the National Science Fund Project No MI - 1505/2005**.**

Keywords: genetic algorithms, modelling, multiple model approach.

Abbreviations: GA: genetic algorithms

The paper presents the implementation of multiple model approach to modelling of *Escherichia coli BL21(DE3)pPhyt109* fed-batch cultivation processes for an extracellular production of bacterial phytase. Due to the complex metabolic pathways of microorganisms, the accurate modelling of bioprocesses is rather difficult. Multiple model approach is an alternative concept which helps in modelling and control of complex processes. The main idea is the development of a model based on simple submodels for the purposes of further high quality process control. The presented simulations of *E. coli* fed-batch cultivation show how the process could be divided into different functional states and how the model parameters could be obtained easily using genetic algorithms. The obtained results and model verification demonstrate the effectiveness of the applied concept of multiple model approach and of the proposed identification scheme.

Bioprocesses, and particularly cultivation processes, are characterized by a complicated structure of organization and independent characteristics, which determine their nonlinearity and non-stationary. Model formulation for a bioprocess is traditionally performed under conditions of a well-defined medium with single-substrate limitations, conditions that are not applied to most industrial cultivations, typically running in a complex medium. In many cases, the globally valid conventional numeric models, which describe the overall process behavior cannot be used in on-line monitoring and control, either because they do not describe the process well enough or contain too many poorly known parameters. Simple unstructured models, which account for key process variables (biomass, substrate and product concentrations) do not reflect metabolic changes and are unsuitable for many tasks (Zhang et al. 1994; Tartakovsky et al. 1997; Feng and Glassey, 2000; Venkat et al. 2003). Model predictions could be improved using structured models, but these models incorporate too many equations and unknown parameters and provide a qualitative, rather than quantitative description of the process. The structured model of a bioprocess is normally so complicated that it is difficult to be used for industrial scale production. Therefore, some alternative modelling methods for the purpose of monitoring and control of bioprocesses have to be searched for.

Table 1. Cultivation parameters.

Parameter	t_0	$\gamma_X(t_0)$	$\gamma_S(t_0)$	$V(t_0)$	$\gamma_{S_{in}}$	$\gamma_{S_{sp}}$
cultivation 1	4.30 h	3.20 g/l	0.78 g/l	2.70 l	500 g/l	0.2 g/l
cultivation 2	3.10 h	3.20 g/l	0.5 g/l	2.7 g/l	500 g/l	0.1 g/l

The multiple model approach is an alternative concept, which helps in modelling and control of complex processes such as bioprocesses. The state of the approaches to modelling and control problems arising working with systems of ever-increasing complexity and associated nonlinearity is presented by (Tartakovsky et al. 1997). In this work the authors describe an approach which embraces a wide range of methods by developing complex models and controllers based on multiple submodels.

The functional state concept could be use to describe and analyze the current biological state of bioprocesses, and could be applied in expert system-based fault diagnosis and in control of bioprocesses (Zhang et al. 1994). The main idea is to use a two-level hierarchy where at the first level the process is divided into macrostates, called functional states, according to behavioural equivalence. In each functional state the process is described by a conventional type of model, called a local model, which is valid only in this functional state. In each functional state certain metabolic pathways are active enough to dominate the overall behaviour of the process. The biological behaviour is quite similar during each functional state. At the second hierarchical level some numeric detection algorithms and/or rules based on expert knowledge can be used for the

*Corresponding author

T. PENCHEVA O. ROEVA
I. HRISTOZOV

FUNCTIONAL STATE APPROACH TO FERMENTATION PROCESSES MODELLING

EDITED BY

STOYAN TZONKOV
BERND HITZMANN

Prof. Marin Drinov Academic Publishing House
Sofia · 2006

Preface

Biotechnological processes, and especially fermentation processes, have the enormous advances in recent years. Due to their multidisciplinary nature, fermentation processes have attracted a significant interest of microbiologists, biochemists, molecular biologists, bioengineers, chemical engineers, and food and pharmaceutical chemists, etc. These processes are known to be very complex, with complicated structure of organization and interdependent characteristics, which determine their non-linearity and non-stationary properties. Therefore accurate mathematical models of the underlying processes are essential for design, optimization and high-quality control.

When modelling complex processes, such as fermentation processes, the common approach is the development of a global nonlinear model that performs satisfactorily through the entire operating range. The global model is characterized with a complex structure, moreover it is not able to describe the possible metabolic changes during the entire operating range, as well as the parameter non-stationary. An alternative approach based on multiple model framework could be considered in order to overcome mentioned above global model disadvantages. The application of multiple model approach allows some really existing phenomena or events to be reflected, leads to process description with simpler local models and gives possibilities for direct incorporation of high-level and qualitative plant knowledge into the model.

This monograph is a result of long standing international collaboration between *Institute of Technical Chemistry*, *University of Hannover*, *Germany* and *Centre of Biomedical Engineering* "*Prof. Ivan Daskalov*", *Bulgarian Academy of Sciences*, *Bulgaria*. This successful collaboration has started in 1996, passed through two international contracts granted by *DFG* (2000-2001 and 2003-2005) and has now continued with common project № MI – 1505/2005, granted by National Science Fund, Bulgaria.

Chapter 1 describes the model construction technology as an overview of mathematical modelling development. Some basic remarks regarding the microorganisms growth kinetics for different modes of fermentation processes are listed. Classification of fermentation processes models, with examples for each considered model type, is presented.

Chapter 2 presents the theoretical background of functional state modelling approach, starts with description of idea of multiple-model framework. A survey of different applications of multiple-model approach to fermentation processes is here summarised. Some basic knowledge about functional state modelling approach for yeast growth processes are here demonstrated, as well as preconditions of implementation of functional state modelling approach for bacteria growth processes are here grounded. This chapter is accomplished with short review of *Genetic algorithms*, which are one of the most powerful optimization techniques nowadays.

Chapter 3 gives up a presentation of experimental conditions of nine cultivations of yeast and bacteria. Three batch and three fed-batch yeast cultivations, as well one fed-batch bacteria cultivation, are performed in the *Institute of Technical Chemistry*, *University of Hannover*, *Germany* by Michael Arndt, Dirk Hüll, Egbert Stärk, Thomas Scheper and Bernd Hitzmann. Another two fed-batch bacteria cultivations are performed in the *Department of Fermentation Engineering*, *Faculty of Technology*, *University of Bielefeld*, *Germany* by Michael Arndt, Sofia Kleist, Gerhard Miksch, Karl Friehs, Erwin Flaschel and Bernd Hitzmann. Experimental data

from the cultivation processes are used in the following two chapters for functional state modelling approach application.

Chapter 4 presents an application of functional state modelling approach for batch and fed-batch cultivations of yeast aerobic growth process. Different types of local structure models are determined, as well as model parameter estimation using *Genetic algorithms* are performed. Validation of the developed local models is presented in order to demonstrate the effectiveness of the functional state modelling approach application.

Chapter 5 demonstrates the first known attempt of the application of functional state modelling approach to bacteria growth process. New local model structures for each defined functional state are developed. Model parameter identification is again performed using *Genetic algorithms*. Based on available experimental data a validation of the proposed local models is presented.

Chapter 6 summarizes the basic steps, analyses and decisions in implementation of functional state modelling approach to yeast and bacteria growth processes. The advantages of this approach are confirmed by the obtained results for structures and parameter identification of local models for each recognized functional state.

The recommendations of the reviewers, *Prof. U. Viesturs* (Latvia) and *Acad. I. Popchev* (Bulgaria), are gratefully taken into consideration. The reviewers constructive criticism has contributed to the improved presentation.

October, 2006 Authors
Sofia