SUPPLEMENTARY MATERIAL

L-lysine production improvement: a review of the state of the art (including patents) with a focus on optimization of fermentation parameters

Fernanda Karine do Carmo Félix[1]; Pedro Gabriel Borges Bonfim[1]; Luiz Alberto Junior Letti[1]; Gilberto Vinícius de Melo Pereira[1]; Vanete Thomaz Soccol[1]; Carlos Ricardo Soccol[1a]

[1] *Department of Bioprocess Engineering and Biotechnology, Federal University of Paraná, Curitiba, Brazil.*

[a] Corresponding author - Department of Bioprocess Engineering and Biotechnology, Federal University of Paraná (UFPR), CENBAPAR building, Curitiba, Paraná 81531-980, Brazil. soccol@ufpr.br

L-lysine is an essential amino acid used in various industrial sectors but mainly in food and animal feed. Intense research has been directed toward increasing its productivity. The technologies developed in the L-lysine field are registered in documents such as patents and articles, which can be used to trace the technical-scientific and socio-geographical production profile. To study these profiles, a text-mining technique associated with carefully formulated keywords was used. Geographic analysis has shown a greater tendency for countries with industrial plants with large production capacity to submit patents or publish articles, while the social analysis reflected the close relationship between educational units and companies. Some of these companies presented a direction of effort for their physical or theoretical capacity expansion at different time intervals. The chronological analysis allows us to highlight the role of articles and patents in the registration of the technological transition. The technologies of each document were divided into (1) optimization of fermentation parameters, (2) conventional mutation, and (3) genetic engineering. Finally, the documents of the first technological class were evaluated and discussed, emphasizing the great interest in conducting continuous and fed batch processes and in the fermentative parameter controls that lead to cell metabolism in order to favor L-lysine formation. Furthermore, a descriptive flow of the substrate uptake mechanism and an amino acid metabolism and excretion mechanism were developed for the illustration of microbial physiology and molecular technologies to increase L-lysine productivity.

Keywords: L-lysine; patents; text mining; optimization; fermentation parameters; conventional mutations; genetic engineering.

Methods

Search for patents

The data presented on patents were collected from the *Derwent World Patents Index*®, a data platform that contains files from the 48 major patent-issuing authorities worldwide, until January 24, 2018. The research was carried out involving patents deposited between 1963 and 2018.

As a research strategy, the text mining technique was used, addressing key words found in the patent title. The keywords were initially composed by logical Boolean operators: "OR" for joining terms and "AND" for the terms intercession; and truncation symbols:*, to add any character after the word; and $, to add a single character after the word; and some terms were enclosed in quotation marks to search for a specific phrase or set of words . The keywords were then made up of the following terms:

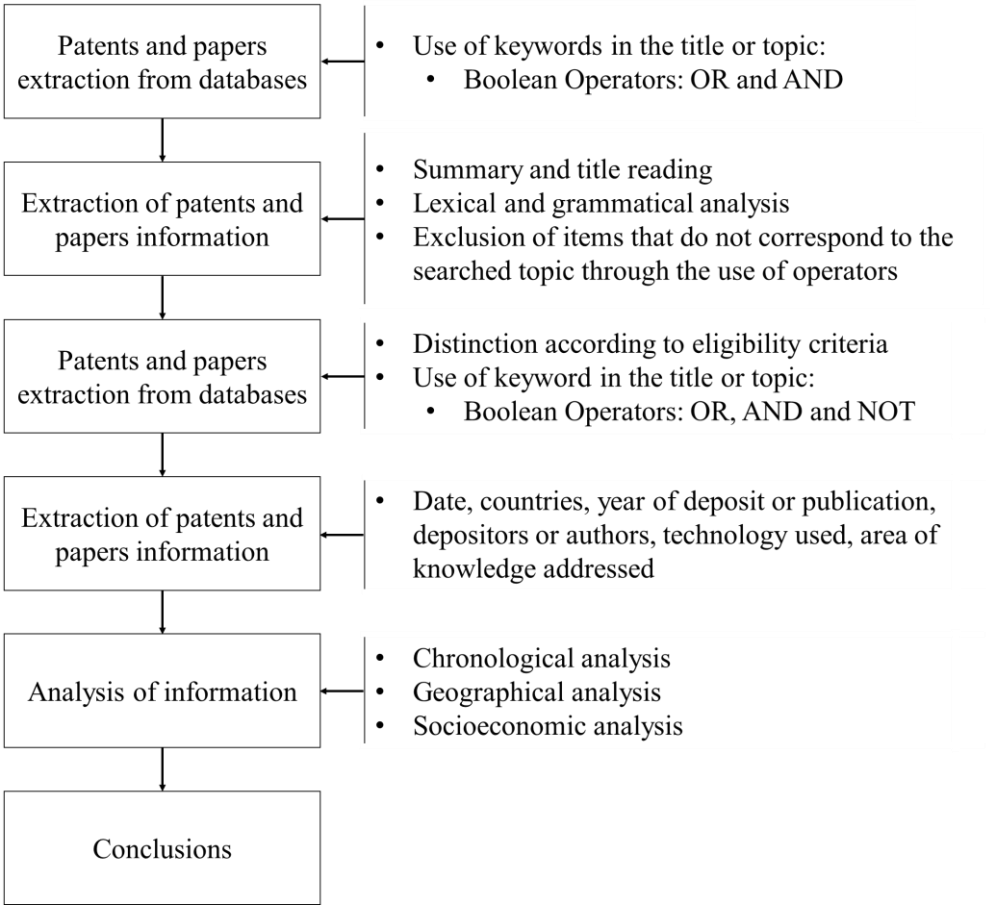
- (lysine AND (produc* OR biosynthesis OR ferment*)) AND (muta* OR gene$ OR genet* OR express* OR new)

The 403 data was tabulated in Microsoft Excel program tables and the titles and summaries of each patent were analyzed. The results who fled the subject, due to the ambiguity of the keywords, were excluded by adding the NOT operator. Thus a new set of key words was formed:

- (lysine AND (produc* OR biosynthesis OR ferment*)) AND (muta* OR gene$ OR genet* OR express* OR new) NOT ($epsilon OR fluorescent OR cancer OR peroxide OR "b*-lysine" OR "glycyl-L*" OR carnitine OR cadaverine OR "D-lysine" OR aminoval* OR disorder* OR catal* OR "lysine decarboxylase" OR archaea OR reactiv* OR "single-cell protein" OR "lysine residue$" OR nitric oxide OR hair OR fusion OR "arginine micro*" OR ectoine OR homoglut* OR heptan* OR "a*-oxidase" OR croton* OR LYS2 OR fission OR lactoferrin OR "e-amino" OR "poly:lysine" OR retrovirus OR "XYL A protein" OR plant$ OR spinach OR "corn seed" OR swine$ OR egg$ OR "2-mono*" OR "lysine protein" OR "L-sacch*" OR "lysine-containing")

From these keywords 277 results were obtained, which were analyzed qualitatively and quantitatively, and compared with the results of scientific articles and with the market for the production of lysine. Patents were organized according to their earliest filing date independent of the patent-issuing authorities in which they were deposited, to enable the tracking of technological trends and the identification of emerging technologies. From the number of patents deposited in each patent-issuing authorities (except World Intellectual Property Organization, United Kingdom, and European Patent Office) was carried out the mapping of patents with the help of the Microsoft Power Map tool for Excel. Patents were classified according to the knowledge areas presented by the platform, and the technology used was analyzed individually by reading.

Figure 1 summarizes the methodology used for the search of patents:



Figures S1. Methodology for searching in patents and papers.

Search for papers

The data presented on papers were collected from the *Web of Science Core Collection*, which contains files from more than 33,000 newspapers, and is known to be an accurate, objective and reliable platform. The data were collected until February 14, 2018, and selected from documents linked to the *Science Citation Index Expanded* (SCI-EXPANDED) and listed as "articles" in the categories “Biotechnology Applied Microbiology” or “Microbiology” or “Biochemistry Molecular Biology”. The research was carried out involving patents deposited between 1945 and 2018.

As a strategy, the text mining technique was used, addressing key words found in the paper topic. Logical Boolean operators, as explained at the patent session, composed these keywords, which initially were the same as those used at the beginning of the patent search:

- (lysine AND (produc* OR biosynthesis OR ferment*)) AND (muta* OR gene$ OR genet* OR express* OR new)

The 4.122 data was tabulated in Microsoft Excel program tables and the titles and summaries of each patent were analyzed. The results who fled the subject, due to the

ambiguity of the keywords, were excluded by adding the NOT operator. Thus a new set of key words was formed:

- ((lysine AND (produc* OR biosynthesis OR ferment*)) AND (muta* OR gene$ OR genet* OR express* OR new) NOT ("HP prod*" OR BrnFE OR "toxin prod*" OR "substrate bind*" OR BAP3 OR "change$ of lys*" OR "fluo* signal" OR RNAT OR "fusion protein$" OR "SCP prod*" OR PGK OR FACS OR "beet pulp" OR "roles of pyr*" OR $epsilon OR antibiotic* OR human* OR glutathione OR histone OR disease$ OR cadaverine OR acetylglucosaminidase OR "proton from" OR hydroxylase OR spor* OR acetylat* OR "replaced by $ lys*" OR alcaligin OR "replac* meth*" OR SILAC OR "succinate produc*" OR plant$ OR residue$ OR enterolysin OR"fatty acid$" OR "putrescine prod*" OR "gluconate kinase" OR "lysine hydroxyl*" OR "aminobutyric acid" OR ectoine OR nitrite OR pipecolic OR ubiquitin* OR "ribulose monophosphate" OR "methionine overproduct*" OR grain$ OR azido OR cyanotoxin$ OR cyanophycin OR plasma OR bioethanol OR chick$ OR oxoglut* OR erythritol OR aminoval* OR piperid* OR "lactic acid production" OR "isoleucine yield" OR "lysine decarbox*" OR antibod* OR "mycosporine-l*" OR cancer$ OR "L-serine product*" OR triticum OR "xylitol production" OR hamster OR nanostructur* OR phaseolotoxin$ OR immun* OR "isoleucine over*" OR virulence OR "butanol product*" OR ferrodoxin OR nitrosobenza* OR photosynthesis OR poly-L* OR diabet* OR EG OR phospholipase OR "twin-arginin*" OR antimicrobi* OR trypsin OR uridine OR microevolution* OR "purine biosynth*" OR "propanediol produc*" OR thiouridine OR "val*-producing" OR PDO OR "glycerol product*" OR "produc* of succ*" OR "lysine as byproduc*" OR "Caulobacter" OR aldolase OR phosphinothricin OR aminomutase OR "tyr* or lys*" OR riboflavin OR "biotinyl*" OR resveratrol OR monoamin* OR "ethanol yield" OR "guanosine up" OR napht* OR "pyruvate to l-iso*" OR "isoleucine biosyn*" OR homocitrate OR "arginine producer" OR streptolydigin OR "arginine biosyn*" OR "glutamate producer" OR virus OR "diaminopentane producer" OR "propionate metabolism" OR osteo* OR isoenzyme* or "DHA produc*" OR mycoflora OR "V lysine" OR "lys* in posit*" OR "rich loop$" OR "nitric oxide" OR tubulysin OR "acetyl lysine" OR pyrrolysine OR "valine biosynt*" OR polymer$ OR zinc OR soymilk OR "lysine-bound$" OR therapy OR "protein profi*" OR "isobutanol titer" OR "lysine-rich" OR "lysine prot*" OR "D-amino acid" OR "produce cinnamic" OR "D-lysine" OR mammal* OR "bind glutamine" OR flavin OR "arginine transport" OR acetolactate OR "lysine bind*" OR "aminohexanoate" OR "L-valine synt*" OR "productivity of glut*" OR "xylose lys*" OR "wine yeast" OR "DL-lys*" OR antigen OR "aromatic amino" OR cyclase OR "threonine production by" OR hexanoic OR "arg* to lys*" OR CEF OR archaea OR lysidine OR queuosine OR citrulline OR "alanine racemase" OR "glycogen metabolism" OR "yields of ethan*" OR "arginine concentration" OR "threonine concentration" OR polyadenyl* OR "SAM bind*" OR "HA produc*" OR mannose OR "nucleotide product*" OR root$ OR "translational bypassing" OR "prod* of

penicillin" OR "tryptophan bios*" OR "lysine group$" OR "threonine at" OR "lysine side" OR ribozyme$ OR respirometric OR "penicillin $-produc*" OR "instationary label*" OR "lys* to ala*" OR MaGe OR "NAC activ*" OR "conser* tyr*" OR ildD OR "pept* targ*" OR "containing $-lys*" OR "wine fermen*" OR KAPA OR "pore-form*" OR "methionine export" OR "P-alanine" OR seawater OR dietary OR "beer product*" OR Amadori OR "cis-encoded" OR "regulatory hypot*" OR denticola OR drosophila OR OSBS OR "pyridoxine bios*" OR "peptide$ to lys*" OR sea OR "subs* of a lys*" OR XAR OR heterodimer* OR "threonine accum*" OR antimutagenic OR "TOR sign*" OR "tRNA-lys*" OR "electronic struct*" OR "arg* product* path*" OR "lys*-alpha-oxidase" OR "protein purif*" OR "methionine biosyn* path*" OR "produc* of beer" OR polyketide OR K199 OR "lysine-cysteine" OR "SV-produc*" OR "cysteine exp*" OR "lys* to arg*" OR "aminoadipic path*" OR gossypii OR "lys* and biotin uptake" OR "glutam* produc* process" OR "isoleucine path*" OR "leucine concent*" OR "lys* at position$" OR "bind* protein$" OR clonixinate OR "pool$ of lys*" OR "chang* amino acid" OR "threonine biosyn*" OR "heme biosyn*" OR "guanidino group" OR "lys*-$-ami*" OR "synthetic protein" OR "freeze-tolerant" OR "chain of lys*" OR Maillard OR "lys*-sepharose" OR "sequence of lys*" OR zipper OR "lys* substitution$" OR "mutant protein$" OR aminochelin OR indole OR asparagine OR "hut operon" OR "absolute glut*" OR "the cation" OR pigment$ OR "beta-ly*" OR cefotaxime OR "methionine product*" OR "methylation site" OR "alcoh* ferm*" OR hydroxybenz* OR "mutation of lys*" OR thioether OR "produc* beta-gal*" OR "mM threonine" OR "polymerase produc*" OR K30L OR actinomycin OR "replac* of lys*" OR "proteolytic produc*" OR "alleles" OR "trehalose-ferm*" OR msgB OR AB1157 OR polytopic OR "lys* to glut*" OR conidia OR neuro* OR vasopress* OR "peptide site$" OR "protein product*" OR swine OR "production of L-thr*" OR "lysine mutase"))

From these keywords 299 results were obtained, which were analyzed qualitatively and quantitatively, and compared with the results of papers and with the market for the production of lysine. The articles were organized in relation to the year of publication in periodicals and compared with the data collected from the patents. The articles were also ordered according to the countries where the authors were affiliated and the branch in which they worked, and this information was mapped with the help of the Microsoft Power Map tool for Excel. Figure X summarizes the methodology used for the search of papers.

Analysis of technologies

The technology used was analyzed individually by reading both articles and patents and classified in: "optimization of fermentation parameters", "conventional mutation" and "genetic engineering". Due to the search terms used, the files related to the "**optimization of fermentation parameters**" included only processes with genetically

modified strains (whether by means of conventional modification techniques or by genetic engineering techniques). We classify these processes in batch, fed batch or continuous, and the techniques used for optimization and the main parameters evaluated and controlled in these documents were registered and reviewed.

In relation to **conventional mutation techniques**, we highlight the mutagenic agents used (physical or chemical) and the characteristics adopted by the strains after the mutation. **Genetic engineering** strategies were organized as cited by Li et al. [1] in “enter”, “flow”, “modarate”, “block” and “exit”, with appropriate adaptations.

In the "**enter**" technology were grouped the patents and the articles that presented/displayed genetic engineering techniques that modified the microorganism with the aim of enabling it to consume unconventional carbon sources or techniques that optimize the consumption rate of usual substrates.

The documents grouped in the "**flow**" technology are aimed at obtaining a greater flow of metabolites in the route of L-lysine synthesis, by releasing the feedback inhibition/repression or by overexpression of genes that allow an acceleration of that flow.

The documents grouped as "**moderate**" strategy use as technique the overexpression of genes that are located at branch points of the metabolic pathway, with the objective of slowing down the competing metabolic branches. Already the files that encompass target gene silencing techniques of competing metabolic branches or that disrupt genes composing the L-lysine degradation pathway are grouped as “block” technology.

Finally, patents and articles grouped as “**exit**” strategy report the use of genetic engineering to facilitate efflux of L-lysine across the membrane and bacterial cell wall.

Reference

[1] Li Y, Wei H, Wang T, et al. Current status on metabolic engineering for the production of L-aspartate family amino acids and derivatives. Bioresour. Technol. 2017;245:1588–1602.