

# Utilization of a Bioinformatics Algorithm for the Comparison of Process Chains

F. Reichert<sup>1</sup>, A. Kunz<sup>1</sup>, C. Bender<sup>2</sup>, R. Moryson<sup>1</sup>, K. Wegener<sup>1</sup>

<sup>1</sup> Swiss Federal Institute of Technology Zurich, Raemistrasse 101, 8092 Zurich, Switzerland

<sup>2</sup> German Cancer Research Center, Division of Molecular Genome Analysis, Im Neuenheimer Feld 580, 69120 Heidelberg

## Abstract

Large manufacturing companies usually work in manufacturing networks, where they add LCL (low cost locations) manufacturing sites into their network and fill these from different HCL (high cost locations) with their production lines. These different fillings give evidence to an inhomogeneous production structure at the LCL.

To solve this problem, the method MAE-P<sup>3</sup> (MAE-P<sup>3</sup>: Machines And Equipment; Processes; Products; Planning) is introduced, which visualizes complex production structures in a two-dimensional matrix. This method is integrated into a computer program, which displays the sequence of the production lines' manufacturing processes together with the sequence of the products' manufacturing processes.

The comparison of production lines and products for relocation planning or production structure optimization is based on an algorithm for sequence alignment originally developed in the bioinformatics field.

This paper introduces the application of this algorithm to a relocation problem and gives a practical example of this method.

## Keywords:

Factory and Production Planning, Networking in International Production/Manufacturing and Logistic Networks, Algorithm of Bioinformatics

## 1 INTRODUCTION

The changing international competition conditions – as for example the reduction of production lifecycle times, increasing globalization and raising competition – have also an impact on the internationalization of many companies' production [1]. The major results of the changing competition conditions are the opening of new markets and the reduction of production costs by relocating production from high-cost-locations (HCL) to low-cost-

locations (LCL) [2]. Through internationalization and opening of new production locations, the internal production networks of these companies grow. Thus, the efficient design of global production networks becomes an increasing factor of success.

The new locations generate an additional complexity in the network. If different products from different HCL are accommodated at one single LCL, an inhomogeneous production structure may arise at the new low-cost location. This is caused by the integration of production lines of independent and often competing HCL. Differences in the lines may occur in type and sequence of resources, their components, their IT-infrastructure, or in the administration of production data. As a consequence, products from different HCL cannot be transferred to another production line easily.

The more manufacturing processes a line has to deal with and the more production lines from different HCL are relocated in the LCL, the more complex the relocation becomes from a technical point of view. The construction of new plants in the production network and the mixture of competing production units obstruct the overview of the network considerably. The fundamental problem is the lack of transparency in the world-wide production structure.

To solve the lack of transparency of the production structure, the method MAE-P<sup>3</sup> was developed [2]. The method MAE-P<sup>3</sup> visualizes complex production structures of a department, a factory, or a complete production network in a two-dimensional matrix. The method works by displaying the sequence of the manufacturing processes of a production line and the sequence of the manufacturing processes of a product. It allows comparing these sequences so that decisions about the relocation of a product within a production network can quickly be made and concepts for further production optimization can be deduced.

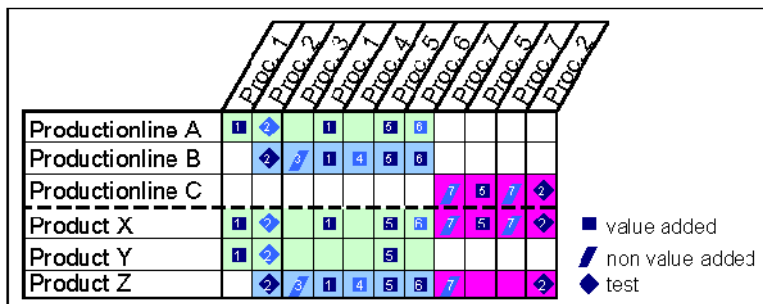


Figure 1: The MAE-P<sup>3</sup> Matrix with production lines and products.

The method's working principle is shown in Figure 1: The abscissa shows the manufacturing processes, the ordinate represents the process chains of production lines and products. A production line indicates the physical sequence of the available manufacturing processes and the process chain of a product shows the required sequence of manufacturing processes [3].

However, only identical manufacturing processes may be inserted in the same column.

The presented method MAE-P<sup>3</sup> enables to compare process chains of production lines and products, although it is currently just a two dimensional matrix. In order to compare different production lines and products and to evaluate which location has the best ability for producing a product, a flexible program is needed. It should allow inserting different production lines or different products or different production lines and products in order to compare them. To display and to compare these different process chains, an algorithm is needed to insert the aforementioned 'Gaps' at the right position.

## **2 OBJECTIVES**

The goal of this work is to apply an algorithm that has the ability to display the production structure of an internal production network. The introduced Matrix MAE-P<sup>3</sup> describes the method how to illustrate production lines and products. But an algorithm is needed to calculate the alignment of the sequences that the MAE-P<sup>3</sup> matrix can be displayed.

In an alignment similarities in the process chains of production lines and products are displayed as blocks by arranging equal processes in the same column. Both the comparison of production lines and products as well the cross comparison of lines and products are desired and part of the implementation of the algorithm.

## **3 THE WORKFLOW OF THE NEEDLEMAN-WUNSCH ALGORITHM**

The problem of comparing process chains and displaying them is solved using an algorithm originating in the bioinformatics field. Needleman and Wunsch published an article in the 'Journal of Molecular Biology', in which they developed an efficient computer-based method for finding similarities in the amino acid sequences of two proteins [4]. From these findings, it is possible to determine whether significant congruence exists between the proteins.

A Gen of an organism is stored in a DNS and will be translated via many steps in a sequence of amino acids, which display a protein. By disregarding the organization of a protein, a protein can be displayed as a sequence where every symbol correlates to an amino acid [4]. The method of finding similarities in the amino acid sequences of proteins is used in this work to detect and to compare similarities in the sequence of manufacturing processes of production lines and products.

For a comparison of sequences, regions with big congruence are searched. The similarity of two sequences is measured by a value named 'maximum match'. It is obtained by summing up the scores for pairs of symbols in both sequences. Matching symbols are given a positive score, while gap insertions are penalized via a negative score. In principle, this value shows the highest amount of similar symbols of two sequences.

The algorithm automatically arranges the symbols in a way that equal symbols are preferred to reside in the same columns, in order to maximize

the measure 'maximum match'. To be able to align symbols appropriately, gaps can be inserted in either sequence, which makes it possible to bridge over parts of sequences (see Figure 2). To determine the overall optimal maximum match, one arranges both sequences in a matrix that represents all possible alignments and computing the maximum match up to each position in this matrix.



Figure 2: Insertion of Gaps.

A diagonal step in this matrix represents the alignment of a symbol from the first sequence to a symbol of the second sequence. Proceeding in horizontal or vertical direction means that a symbol from one sequence is aligned to a gap in the second sequence. Once the matrix is completely filled, in a backtracing step the optimal path through all of the possible alignment paths is computed via following the biggest maximum matches in the matrix.

In the following paragraph the mode of operation of the algorithm is briefly explained (see Figure 3):

1. Building up a matrix with the two sequences **a** and **b**.
2. Initialization of the matrix by setting the borders to the cumulative sum of the gap penalties, since moving either vertical or horizontal corresponds to inserting a gap.
3. For the alignment of the two sequences until the actual position in the array, the maximum score has to be calculated in each field of the matrix. For the comparison we use the following score settings:

- Match of two symbols: 1 point
- Mismatch of two symbols: -1 point
- Gap Insertion: 0 points

For each field in the matrix, adjacent positions in the upper, the left or the upper left field are used to calculate the maximum score. Coming from the upper left position, the match score is added to the current score. Coming from above or left the gap score is added to the respective score. The maximum of these three proposals is filled into the current field. The procedure ends, when all fields are filled.

4. After filling the fields, in the bottom right corner the total maximum score is displayed, shown in Figure 3.
5. From the bottom right corner of the matrix the final path and thus the alignment is produced by following the maximum values in each adjacent field, which is called the 'trace back'. A step to the left means inserting a gap with the symbol of the column. A step up means inserting a gap with the symbol of a row. A diagonal step in the array means aligning the two symbols of the actual row and column.

		Sequence a					
		-	A	B	C	E	H
Sequence b	-	0	0	0	0	0	0
	A	0	1	1	1	1	1
	C	0	1	1	2	2	2
	F	0	1	1	2	2	2
	H	0	1	1	2	2	3
		Maximum Match: 3					

Figure 3: Example of an array of two sequences.

Following, the recurrence relation used to calculate each field depending on the adjacent fields is shown (as done in step 3):

$$M(0,0) = 0$$

$$M(i, j) = \max \left\{ \begin{array}{l} M(i-1, j-1) + w(a_i + b_j) \\ M(i-1, j) + gap \\ M(i, j-1) + gap \end{array} \right\} \quad (1)$$

For all  $1 \leq i \leq m, 1 \leq j \leq n$

- gap -1
- a, b Sequences
- m Length of Sequence (a)
- n Length of Sequence (b)
- M(i,j) Maximum match between a Prefix of (a), which ends in i and a prefix of (b) ending in j

If more than two sequences are compared, first a so called guide-tree is generated using the pair wise euclidean distances between each pair of sequences. It is used to guide the order in which the sequences are aligned pair wise to each other. Pairs with the smallest distances are grouped together and the alignments are generated for those pairs first. Sequences with larger distances are then successively aligned to the already computed pair alignments (see Figure 4).

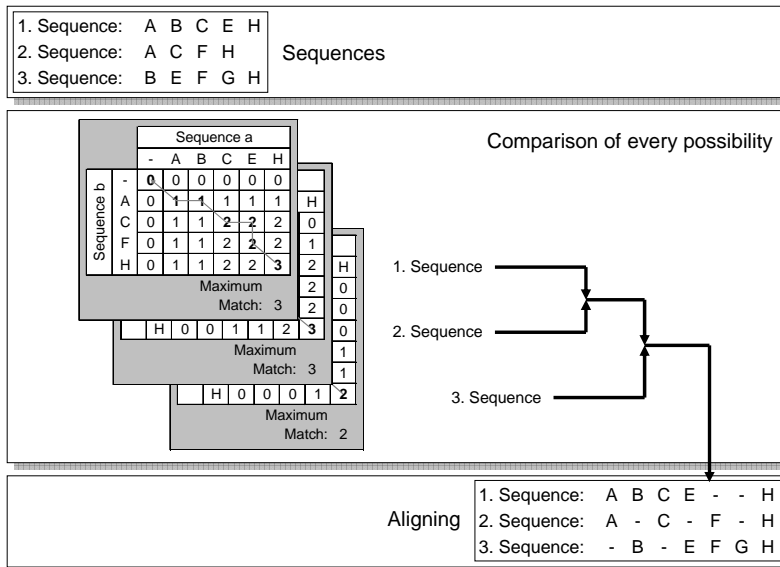


Figure 4: The tree structure of sequences with the highest maximum matches.

To use the algorithm for production planning or production optimization, a data base is needed. In this database, process chains of production lines of the different locations of the production network are stored. For a relocation planning project, the process chain of a product to be relocated will be derived from the process chain of the actual production line, where the product is currently produced.

A process chain of a production line and a process chain of a product are stored as sequences of manufacturing processes in the database. The list of a product's process chain can now be compared with different lists of process chains of production lines. The Needleman-Wunsch algorithm compares the lists and aligns the different sequences of manufacturing processes.

The result is a visualization of the product's process chain in comparison to different process chains of production lines with equal manufacturing processes in the same columns. This visualization can be analyzed on a computer or can be printed and analyzed visually on a poster. The next step is to check, which production line has the best match with the product. If there exists no production line, which fulfils the requirement of

the product, a production line can be chosen, which fulfils most of the requirements. If a production line doesn't offer all required manufacturing processes for a product, then a machine can also be added to the production line.

With this method, a location can be found in the network, where the production structure is already present to produce a certain product. After the possible relocations are clear, the capacity of these lines and the costs for the production on this line can be analyzed. Based on these findings, a substantiated decision where to produce a product can be made.

#### **4 PRODUCTION STRUCTURE OPTIMIZATION OF RELOCATED PRODUCTION LINES**

Besides the planning of relocation, also a production structure optimization can be performed. In this work, a project is introduced where the MAE-P<sup>3</sup> matrix with the Needleman-Wunsch algorithm was used for optimization of the production structure.

In an LCL, an automotive supplier runs a production department, where the range of products was constantly increased by product relocation activities. However, the production structure was not adequately adjusted. From 800 product variants, at least 95% were divided into 14 product families and were produced on various preliminary assembly lines, final assembly lines, and on single resources. The consequence was a very complicated meshwork of product chains through the production structure. The aim of the project was to readjust the resources to the necessary sequence of the product chains' manufacturing processes.

The first step was to collect the data of the production structure. The process chains of the four preassembly lines, 10 final assembly lines and the unconnected single resources were collected in lists and stored in a database. The next step was to collect the process chains of the 14 products. Using the Needleman-Wunsch algorithm, the production structure of these production lines and products were aligned and the same manufacturing processes were displayed in the same columns.

Based on the aligned production structure by the algorithm, the MAE-P<sup>3</sup> matrix was printed as a poster and used in a workshop to analyze the production structure. Out of this poster with the now structured visualization, cluster of similar sequences of manufacturing processes were searched.

The result of this analysis was a production area with defined preliminary assembly lines, defined intermediate storages (supermarkets), and product family specific final assembly cells. Furthermore, the workload of the resources could be raised, the quality improved, and the number of required employees could be reduced.

#### **5 CONCLUSION**

The MAE-P<sup>3</sup> matrix allows to compare sequences of manufacturing processes of production lines and products so that decisions on the

relocation of a product within a production network can be made on an objective basis. Beside this, the proposed method can also be used for production optimization.

For a flexible comparison of these sequences, a database with an algorithm is needed, which can compare production lines, products or production lines and products. For this purpose, the Needleman-Wunsch algorithm was applied, originally invented for comparing amino acid sequences. In this work the algorithm is used to compare process chains of manufacturing processes.

The Needleman-Wunsch algorithm gives the ability to exploit very fast and structurally how similar process chains of production lines and products are which are intended to be relocated. By comparing the sequences of manufacturing processes of the chosen production lines to the manufacturing processes required for the product, the procedure supports the decision-making of the relocation from a technical point of view. The main relevant technical information is now centrally available.

Many parameters affect the decision of production relocation. Parameters like the degree of education, labor costs or political stability are just a cutout of the variety of parameters affecting such a decision. The method MAE-P<sup>3</sup> with the integrated algorithm is a contribution to the technical aspects of a relocation decision. Also, the MAE-P<sup>3</sup> matrix can be used to visualize and in a second step to optimize inhomogeneous production structures of newly built low cost location plants.

## 6 REFERENCES

- [1] Abele, E, Kluge, J., Näher, U, 2006, Handbuch globale Produktion, Carl Hanser Verlag, München, Wien.
- [2] Reichert, F., Kunz, A., Moryson, R., 2008, MAE-P<sup>3</sup> - A System to Gain Transparency of Production Structure as a Basis for Production Relocation Planning. In: icseng, 19th International Conference on Systems Engineering, 458-463.
- [3] Rother, M., Shook, J., 2003, Learning to see: value-stream mapping to create value and eliminate muda – Version 1.3, The Lean Enterprise Institute Cambridge, Mass.
- [4] Needleman, S. B., Wunsch, C. D., 1970: A general method applicable to the search for similarities in the amino acid sequence, in: Journal of Molecular Biology, Volume 48: 433 - 453.