

Protein Recovery, Separation and Purification. Selection of Optimal Techniques Using an Expert System

Eduardo W Leser*, Juan A Asenjo

Biochemical Engineering Laboratory, University of Reading, Reading RG6 2AP, England

The paper discusses the utilization of new techniques to select processes for protein recovery, separation and purification. It describes a rational approach that uses fundamental databases of protein molecules to simplify the complex problem of choosing high resolution separation methods for multi component mixtures. It examines the role of modern computer techniques to help solving these questions.

Key words: protein purification – downstream processing – artificial intelligence – expert systems

Although Biotechnology had been developed during the last half of this century, particularly with the production of antibiotics and amino-acids, the outcome of the progress in Genetic Engineering, which is known as the "New Biotechnology", has brought a completely different approach to this field. This change can be perceived in a new perspective in what was known a few years ago, as fermentation technology. Instead of giving the priority to the biological reactor, the field is now clearly divided in three parts, and what happens inside the culture system or fermenter shares importance with the upstream and downstream set of operations and activities. It is important to understand what provoked this change of focus. First, before the advent of genetic manipulation of microorganisms, the achievement of high productivities within a process was only possible with the improvement of the production organism, using traditional genetic modification and selection methods, or the optimization of the fermentation process itself, by manipulation of operating parameters. Secondly, in the former systems, based on biological processes, the concentration of products in the fermented broth was normally high (Table I) and the separation and purification of products was based on well defined unit operations and extensive equilibrium data.

Furthermore, these products have relatively low molecular weight and are normally secreted into the broth. The economics of the bioprocesses also have an important role; traditional fermented products have a low associated value

and economical feasibility was possible within a large production scale. A major exception to this was the production of items aimed to human health care, in particular vaccines. However, in this case, the guidelines for production were different: consistency and quality control for production and product. Nevertheless, this example is important due to the strict regulations that had to be followed, concerning the question of dealing with injectable substances. It creates a link between the more traditional procedures of biochemical engineering and the present challenges of production processes of human health care proteins.

TABLE I

Common yield values, in grams per litre, of some fermented products, obtained using traditional methods and genetically transformed organisms

Traditionally produced		Genetically engineered	
Product	Yield (g.l ⁻¹)	Product	Yield (g.l ⁻¹)
Ethanol	100	TPA	0.2
Glutamic acid	35	Hb Vaccine	0.1
Penicillin	12	Insulin	0.05
Vitamin B ₂	1.5	Antibodies	0.01 - 0.1

The features introduced in production after the implementation of new techniques suggest an opposing trend as the desired products have usually an important commercial value, not only due to the intrinsic value associated to the "high value - low volume" products but also because of the promising figures of an expanding market as showed in Table II and in Fig.1.

These products, usually proteins, are produced by a genetically modified group of organisms. The systems are well known and amongst them the most commonly used are *Escherichia coli*,

Work supported by CNPq (Proc. 26.0158/90.9) and European Science Foundation.

*On leave from Fundação Oswaldo Cruz, Brazil.

Received 31 May 1993

Accepted 14 December 1993

TABLE II

Estimated values for sales of biotherapeutic products in U.S.A. according to different applications in two periods (Biotech Forum Europe, vol 9, 4, April 1992, p. 178)

Application	U.S. biotherapeutic product sales (millions of US\$ 1992)	
	1992	2002
CARDIOVASCULARS		
Erythropoietin (EPO)	600	1,200
Tissue Plasminogen Activator (tPA)	180	100
Blood Factors	0	300
Superoxide Dismutase	0	200
Others	0	300
Subtotals	780	2,100
CANCER		
Colony Stimulating Factors (CSFs)	400	1,400
Interferons (IFNs)	175	700
Interleukins (ILs)	25	400
Others	0	300
Subtotals	600	2,800
HORMONES/GROWTH FACTORS		
Human Growth Hormone (hGH)	300	700
Other Growth Factors	0	100
Insulin	325	600
Others	0	400
Subtotals	625	1,800
VACCINES		
AIDS		
Herpes	0	200
Hepatitis B	125	400
Others	0	300
Subtotals	125	1,900
MONOCLONAL ANTIBODIES		
	120	600
Totals	2,250	9,200

animal cells (mainly Chinese hamster ovary cells, CHO), *Saccharomyces cerevisiae*, *Pichia pastorii*, some bacilli, and now, also, insect cells. This limits the possibilities of modifications upstream and in the fermentation process itself. In an important number of cases, there are no cell transport systems to secrete the product from the cell to the fermentation environment. Consequently, cells must be disrupted and the target product will be dispersed amongst other similar substances (contaminants). This situation makes recovery, separation and purification an extremely important task.

The present state of the art should be analyzed in terms of the existing gap between the emergence of a new process within a research laboratory and the steps conducting to its industrial implementation. It is recognised (Wheelwright & Asenjo 1993) that the problems arising during scale up of these processes are due to lack of tradition in the field, probably because the knowledge of the unit operations related to downstream processing is still in the early steps of development (Wang 1987). Scaling up is usually done by repeating the purification sequence devised in the laboratory with larger equipment. The resulting processes will eventually meet problems and, above all, do not fulfil the real aims of optimal process design: to maximize yield and quality and to minimize costs and complexity.

The production of biologicals is strictly regulated by government agencies and there are so many requirements to be followed that, in a strongly competitive market, it is important to act as quickly as possible. Once the process is approved, there are no possibilities to change production protocols and the tendency is to disregard the search for the best scheme for the production, i.e., the purification steps. The picture is even more complex due to the lack of engineering data on the most common unit operations: solid-liquid and liquid-liquid separation and this concerns not only data on transport phenomena but also data on biochemical and physico-chemical properties of products and contaminants.

PRODUCTION OF PROTEINS

The production of proteins in a large scale poses the usual problems found in common downstream processing added to the striking difficulties for the concentration and purification of large molecules. First we must consider the utilization of the protein. Conditions are remarkably different if the product is a bulk industrial raw material, as, for instance, proteases to be

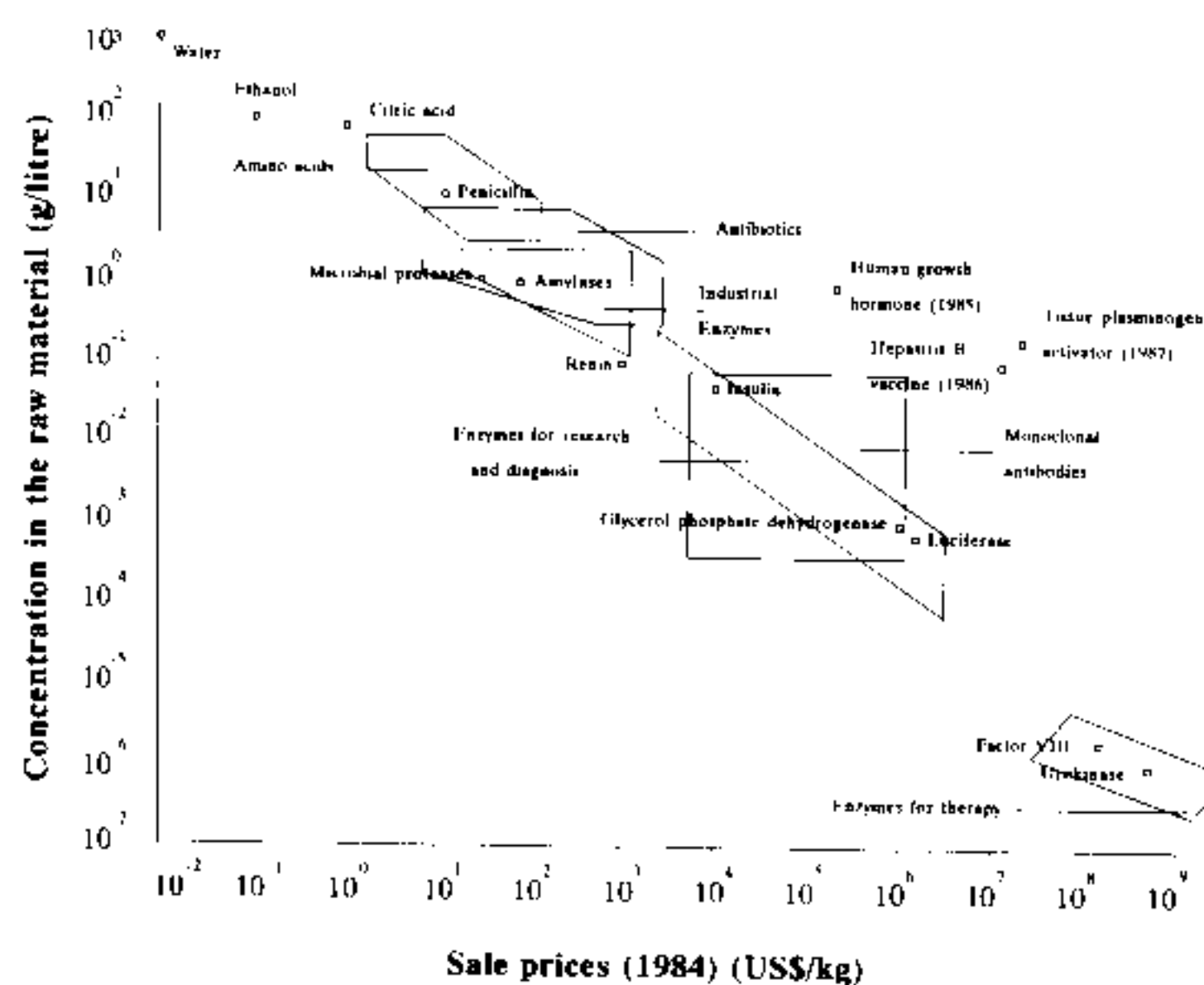


Fig. 1: sales prices (US\$.kg⁻¹), in 1984, related to the product concentration in the raw material. Separations by process scale chromatography (adapted from Dwyer, 1984).

incorporated in the formulation of powder soap, or, on the other extreme, if it is an injectable therapeutic protein, such as insulin, hepatitis B vaccine or tissue plasminogen activator (TPA). In the latter case we are trying to scale up a process that aims to produce proteins in large amounts and with a level of purity that reaches 99% or even more. If the process design is not well conceived, this can render the production not viable for technical and economical reasons. Then, if a product is supposed to be obtained in great quantities, the development of techniques should follow realistic goals from the initial stages of research. The production protocols originated at the bench scale should only consider techniques that can be realistically used in large scale, i.e., for which suitable large-scale equipment either exists or might be developed in the foreseeable future. Another goal must be the perspective, not only for each of the downstream steps but for the whole separation sequence, the maximization of yield and the minimization of the number of steps and resources (economics) (Asenjo & Patrick 1990).

Protein isolation – Isolation comprises obtaining a cell-free solution with a total protein concentration around 60 to 70 grams per litre (Pharmacia 1983, Asenjo et al. 1989). It comprises most standard unit operations normally used in separation processes, and the corresponding properties that should be considered when planning these steps; they are resumed in Table III.

TABLE III

Separation and concentration operations for large scale recovery and purification of proteins and the correspondent main physico-chemical property that drives the operation (Prokopakis & Asenjo 1990)

Operation	Physico-chemical property
Centrifugation	Sedimentation velocity
Filtration	Particle size
Microfiltration	Particle size
Homogenization	Intracellular nature (pressure gradient)
Bead milling	Intracellular nature (liquid/solid shear)
Ultrafiltration	Molecular size
Two phase extraction	Partition coefficient
Precipitation	Solubility (hydrophobic interaction)

If cell separation is necessary the most frequently used methods, at the laboratory scale, are centrifugation and filtration. This operation, concerning solid liquid separation, can pose difficulties at large scale. The smaller the size of

the particles the more difficult it is to isolate them. Large scale centrifuges are complex to operate and to maintain, specially when one must consider the operation with pathogenic sources, i.e., a contained process. On the other hand, cross flow filtration is a somewhat new alternative but there are some technological barriers to overcome. If the product location is extracellular, then the liquid part is kept; if the product is intracellular, the solid fraction of the operation is kept (Fig. 2). When a mammalian cell culture is used, the product is usually secreted by the cells. Production of monoclonal antibodies has been extensively performed using hollow fibre reactors and thus, the culture supernatant is free of particles, with exception of a small amount of cell debris that can be eliminated by gel permeation (Lee 1987).

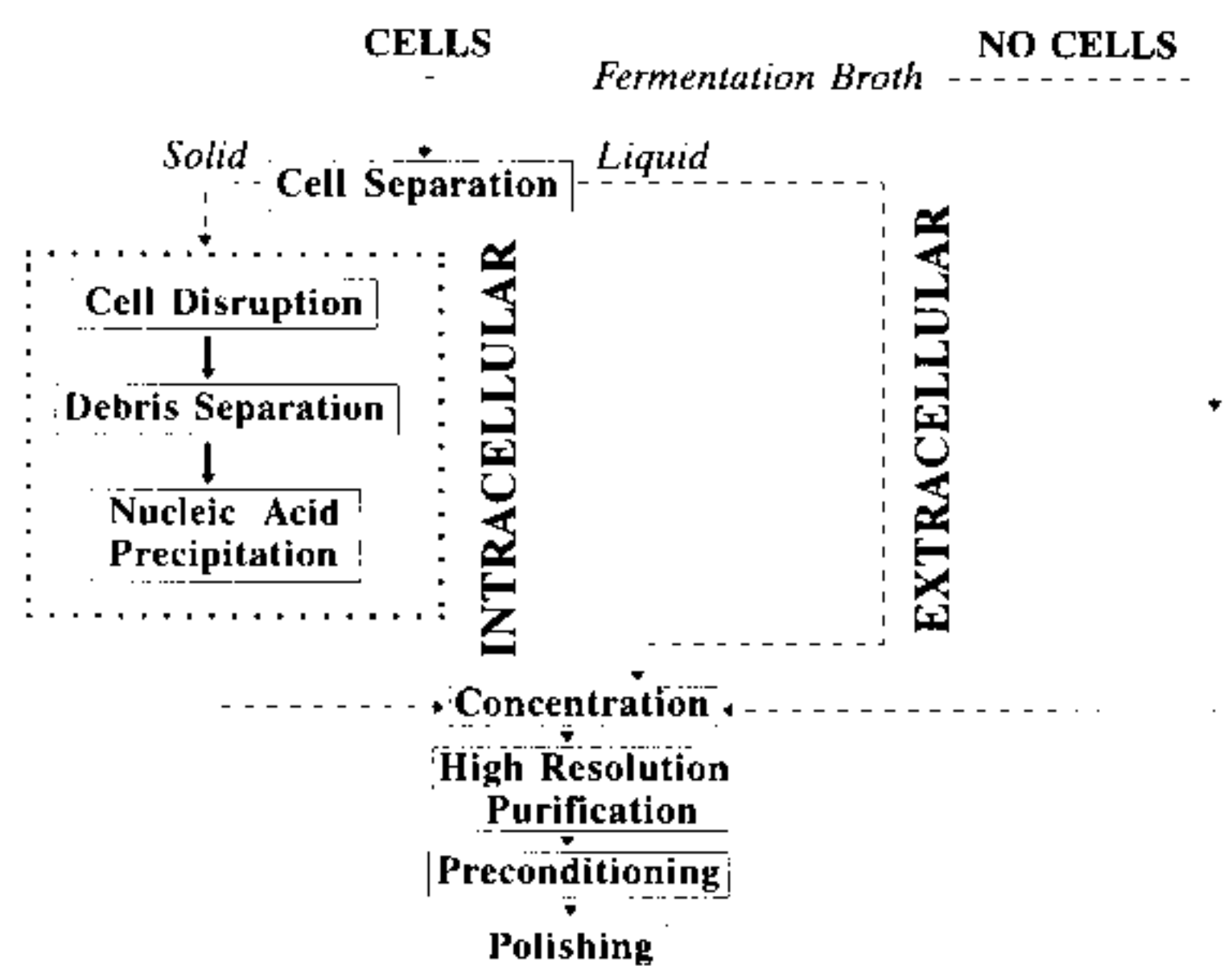


Fig. 2: general scheme for the processing steps in downstream protein purification.

Cell disruption is required when the product is intracellular. Methods and equipment are selected mainly based on the biological source and product. The choice of disruption technique determines the size of the resulting debris that in turn has an influence on subsequent operations. Typical methods used can be classified into four groups: nonmechanical, ultrasonics, high pressure homogenization, and mechanical grinding (bead milling) (Wheelwright 1991), but, for large scale purposes, only the last two categories are important at present. Bead milling is used with gram-positive bacteria and specific yeast applications; pressure homogenization for most bacteria including *E. coli* and yeast (Asenjo 1990a). Mechanical disruption releases nucleic acids that need to be precipitated. The standard method is precipitation with polyethylenimine. Separation of cell debris has to be undertaken

once the cells are disrupted and due to the small size of particles this brings extra difficulties for the large scale process. The result of this step is a solution containing the product, cell metabolites, and remaining components of the culture medium. At this point the addition of proteolytic inhibitors should be evaluated.

If the intracellular product is manufactured in *E. coli*, high expression of heterologous proteins will usually accumulate as insoluble inclusion bodies (Kane & Hartley 1988). This makes necessary the processing of the inclusion bodies into the native protein by denaturing and refolding. If the intracellular product is manufactured in yeast, often the protein is present in homogeneously particulate form, typically 30-60 nm particles such as virus-like particles (VLPs) (Muller et al. 1989). Although the processing of intracellular particulate recombinant proteins is an important aspect of downstream processing, not many satisfactory methods exist for large scale separation, denaturation, and refolding of the particulate proteins.

Protein concentration – Concentration is usually required when the protein concentration of the harvested, disrupted, and separated stream is below 60-70 g.l⁻¹, the suitable range for chromatographic purification (Pharmacia 1983, Asenjo et al. 1989, Asenjo & Patrick 1990). With some proteins it is very difficult to obtain higher concentrations without a serious increase

in viscosity, which would then impose very poor transport characteristics on the system. If a membrane (ultrafiltration) is used for concentration, the resulting flux characteristics will decide the highest possible concentration that can be obtained from the operation. If at the point where flux has dropped below an acceptable limit the concentration is below 60 g.l⁻¹, then the proteins can be precipitated (e.g., with ammonium sulphate) to increase the final concentration to an adequate level.

At this point the broth will contain proteins and other components such as lipids and/or wall or other polysaccharides, salts, and water.

Protein purification – From this point product purification begins and there will be many alternative combinations of processes (Table IV). For the recovery, resolution, and purification of a single protein, ideally one would like one step to extract virtually 100% of the target protein from the mixture with no contaminants. As this is almost impossible, two or sometimes three or four stages probably will be needed to achieve the final purity required for the particular application.

As most of the excess water has been extracted, a purification step of extremely high resolution should be used to minimize the number of stages and hence maximize yield. However, sometimes this may not be possible at this stage, as some contaminants still present may

TABLE IV

Chromatographic operations and the related physicochemical driving properties, process characteristics and applications for large scale purification of proteins (adapted from Asenjo & Patrick 1990)

Operation	Physicochemical property	Characteristic	Application
Adsorption	van der Waals forces, H bonds, polarities, dipole moments	Good to high resolution, good capacity, good to high speed	Sorption from crude feedstocks, fractionation
Ion exchange	Charge (titration curve)	High resolution, high speed, high capacity	Initial sorption, fractionation
Hydrophobic interaction	Surface hydrophobicity	Good resolution, speed and capacity can be high	Partial fractionation (when sample at high ionic strength)
Affinity chromatography	Biological affinity	Excellent resolution, high speed, high capacity	Fractionation, adsorption from feedstocks
Reversed phase liquid chromatography	Hydrophilic and hydrophobic interactions	Excellent resolution, intermediate capacity, may denature proteins	Fractionation
Gel filtration	Molecular size	Moderate resolution, low capacity, excellent for desalting	Desalting, end polishing, solvent removal

produce fouling of the affinity or high resolution ion exchange column and, consequently, shorten its life. Therefore, a first step in protein purification from other contaminants probably will be necessary. This would constitute a clean-up step of pretreatment or primary isolation. For this, a somewhat inexpensive treatment to clarify the stream from suspended materials and nonprotein contaminants besides salts should be utilized. This operation will not give a very high purity but must result in a very high recovery yield in terms of the protein product retrieved. Typical operation for this step would include inexpensive or disposable adsorption devices like a Whatman DE52 ion exchange cartridge, a hydrophobic interaction step, aqueous two-phase partitioning, or precipitation of the proteins using salt. After this procedure, a higher resolution one will most probably be used, giving a product of up to 99% (usually 95-98%) purity. Typical operations will include one or two high resolution ion-exchange chromatography steps or affinity chromatography. Although high resolution is the main concern in this stage, an adsorbent that also will give a high recovery should be chosen or designed.

After the high resolution step a polishing step is frequently necessary to obtain ultra high purity. This will depend on the final use of the protein, and sometimes it is probably the most difficult task to accomplish. If another physicochemical property cannot be exploited, gel permeation will be used, which can separate dimers of the product (due to aggregation phenomena) or its hydrolysis products (due to action of proteases) solely based on their different molecular weights. HPLC also can be used for polishing but this is an expensive technique for preparative purposes. It gives extremely high resolution but it may denature larger proteins.

THE SELECTION OF A PROTEIN PURIFICATION

The core principle of engineering design is the gathering of ideas to generate and evaluate methods for achieving a specified task. For this goal the procedure is normally to divide the work into three stages: synthesis, analysis and evaluation (Dym & Levitt 1991). The first stage consists in the translation of design heuristics into real methods and their assembly. To select and design an optimal sequence in a multistep separation and purification process the designer must choose between alternative operations and seek the optimal sequence with maximum yield and minimum number of steps. At this point the engineer must keep in mind the specifications and constraints. These are not only the common

physicochemical parameters but the strong limitations of quality control and the major restrictions imposed by regulatory legislation. The analysis task evaluates by calculation or comparison between similar purification schemes to see whether they satisfy other important constraints and conditions. The final stage, evaluation, creates the trial-and-error background that reveals the true nature of design.

The synthesis in process design uses, as intensively as possible, mathematical models and equations that provide the necessary amount of information to show if a particular piece of equipment is adequate for the proposed operation. However, the designer is responsible for the final decision because this kind of judgment is almost exclusively based on the designer's expertise. The knowledge for designing a process must include the characteristics of the product, extensively defined, the transport parameters in the process -here included energy, and mass balances- and for complex mixtures from which the product shall be isolated, properties of contaminants and how they interact with the product when subjected to particular processing conditions.

Thus, protein production downstream processes (DSP) contain two classes of actions: (i) protein recovery/isolation and (ii) protein purification (Asenjo et al. 1991). The design of operations belonging to the first group follows, in general lines, the traditional reasoning found in most fermentation systems. Substantial difficulties appear when attempting to solve the problems presented in the second group (purification). There is a lack of design equations and mathematical models as well as thermodynamic databases on the materials to be separated to choose the adequate optimal sequence with a minimum number of steps (chromatographic). Otherwise other approaches would lead to facing the reality of an almost infinite number of combinations to evaluate, as shown in Fig. 3.

ARTIFICIAL INTELLIGENCE AND EXPERT SYSTEMS

To solve the problem of the combinatorial nature of the selection of multi step purification where a number of choices exist at each step, as shown in Fig. 3, and to help the scientists and project engineers several authors propose the utilization of more refined tools, resorting to recent achievements of Artificial Intelligence (AI). This is one of the most contemporary developments of computer science and comprises areas of interest like robotics, natural language interpretation, image processing and expert sys-

tems, that are not only a sub area of AI but a tool for every other area. Computer systems that utilise AI techniques should be able to relate information intelligently, make inferences, and justify these inferences (and also the final result). The use of AI methods represents the transition from data processing to knowledge based processing. Expert Systems (ES) are intelligent computer programs that aims to give advice or to solve problems that are difficult enough to require significant human expertise for their solution, by logical deductions, in a simulation of human expert's reasoning. They can be used, within the above concept, in a broad range of activities like process monitoring and control, design, diagnosis, planning, consultation services, training, and many others in distinctive areas like science, industry and commerce, public sectors, banking, biology, molecular biology and engineering (Nebendahl 1988).

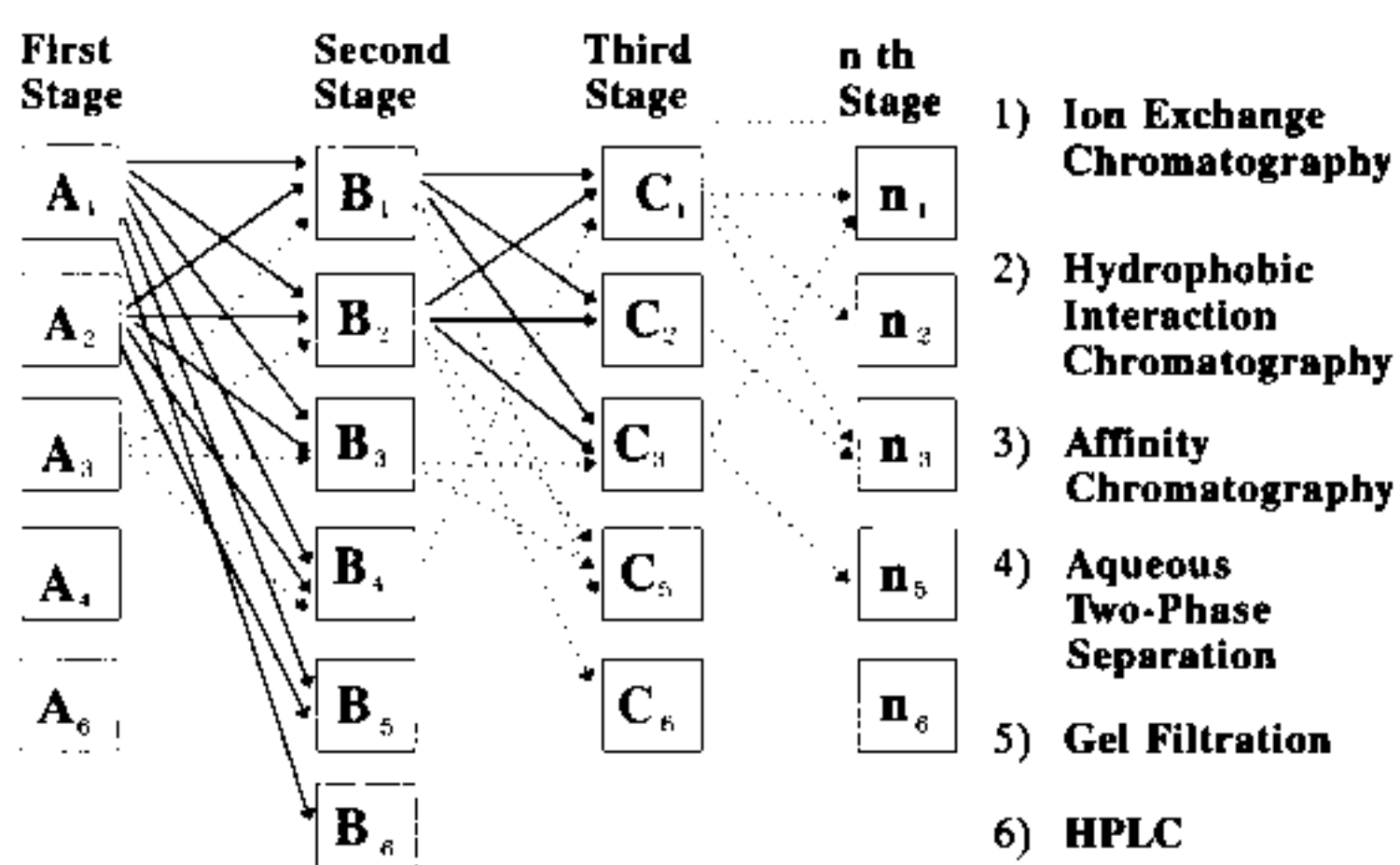


Fig. 3: the combinatorial characteristic of choosing the sequence of operations for protein purification.

When "significant human expertise" is mentioned in the above definition this implies, mainly, the application of *heuristics* for finding a solution to a special problem. Heuristics or rules-of-thumb are the knowledge on when to apply hypotheses to facts, or a mean for controlling the use of deduction. They are rules that order the choice of asserted knowledge to deduction. They do not lead to the absolutely optimal conclusion but they provide advice of a good assessment based on incomplete and sometimes ambiguous information.

The first large ES to operate as an emulation of an expert, providing users with an explanation of its reasoning method, was developed at Stanford University: it was called MYCIN and was designed to aid physicians in the diagnosis and treatment of hospital acquired infections (Buchanan & Shortliffe 1984). Although used only on a research base, its success proved that this new developed technology could be used on

commercial applications. The number of applications is growing steadily. Hushon (1990) shows that the number of ES for environmental applications grew from 20 to 80 between 1986 and 1989. Another relevant aspect is that this technology evolved from networks to PC's and now there are more than thirty commercial programmes, software designed for running in every possible hardware platform.

Figure 4 shows a diagram of an ES and its components. The core of the scheme is the inference engine but its operation will only be triggered if there is a knowledge base that contains facts (or objects) and rules to organize and define how the reasoning will be conducted. The role of the knowledge engineer is to translate the expertise into the system and to create the knowledge base. This is the most arduous problem to solve because it is done by someone not familiar with the subject and the installation can take a long time to be accomplished. Sometimes this task is done by the expert him or herself and unless he or she holds a strong base on computer science the implementation time also will be the bottleneck of the process.

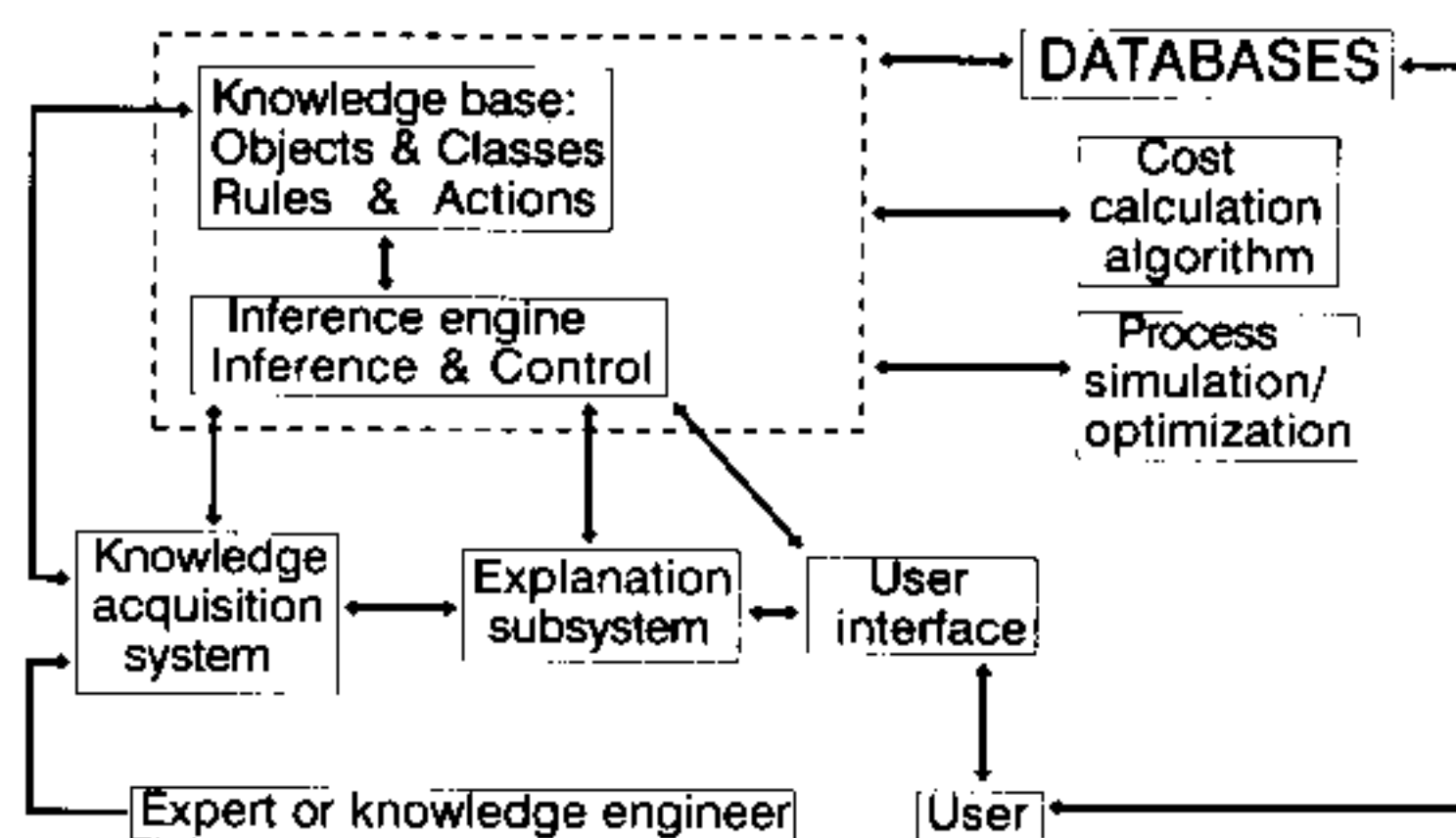


Fig. 4: the architecture of an expert system and its links with external devices (adapted from Harmon & King 1985).

The operation of the system is done by the so called user's interface. Usually the system is self explanatory and the user can seize upon help in any moment during consultation. Contemporary shells, or software, present graphical capabilities that give a much more clear vision of the solution search path inside the problem space.

EXPERT SYSTEMS AND PROTEIN PURIFICATION

The use of ES for Chemical Engineering problems is very well established (Stephanopoulos & Mavrouniotis 1988) and their application to biotechnology problems is becoming more frequent (Clapp & Ruel 1991). The development of systems to deal with the question of

protein purification started some years ago (Wacks 1987, Wang 1987, Wheelwright & Asenjo 1993). The use of commercial software proved to be suitable too (Purves 1990, Asenjo 1990a, Asenjo & Maugeri 1992).

When analysing the general problems of the separation of multi-component mixtures, the engineer's expertise can be translated by four classes of heuristics (Nadgir & Liu 1983):

- method heuristics { rules for choosing amongst a set of separation operations;
- design heuristics { rules for defining the order of the steps;
- species heuristics { rules based on the properties of the components;
- composition heuristics { rules related to the influence of the feed and product composition on the separation costs.

More specifically related to protein purification, there are five rules-of-thumb (heuristics) that guide the choice of the downstream processes and its sequence (Asenjo & Maugeri 1992): (1) choose the separation processes based on different physical, chemical or biochemical properties; (2) separate the most plentiful impurities first; (3) choose those processes that will exploit the differences in the physicochemical properties of the product and impurities in the most efficient manner; (4) use a high resolution step when possible; (5) do the most arduous step last.

The purpose of this work is to build an ES, using a defined software (usually called shell), which contains knowledge on protein separation processes, organized and structured following the heuristics stated above. The system will provide suggestions on schemes for separation and purification of the defined product. To reason on a real design basis, the economics of the project should be analyzed through algorithms containing elements of cost associated with calculations of material and energy balances. Alternative solutions and the utilization of other unit operations also should be considered as options to minimize the cost.

EXPERT SYSTEMS PROTOTYPES

The implementation of ES for protein purification can be the result of the development of specific programs to handle the knowledge and the inference unit (Wacks 1987, Siletti 1989). The use of available software proved to be suitable too (Purves 1990, Asenjo 1990a, Asenjo & Maugeri 1992). The first attempt to build an ES prototype was carried out using a program from

Texas Instruments Inc., called PcPlus. This was an experimental program but it could run the first prototype that contained 65 rules. The knowledge base structure comprised two main sections of operations: recovery/isolation and purification. It did not take in consideration problems like the processing of a recombinant protein present in inclusion bodies or other particles. The suggested process involves a sequence of operations to seize the declared design purpose. To find different arrangements to accomplish the same objective the expert could assign certainty factors and the ES would carry them into the proposed solution. Because the prototype did not contain data on product and contaminant properties, the only way to solve the second part of the problem, that is, the choice of a sequence to purify the product, was very empirical. The only available method was to allocate different degrees of certainty for the high resolution separation techniques to obtain the expected level of purification. The evaluation of solutions for the recovery phase was good but, concerning the purification stage, results were very vague, and the implementation of the knowledge base was not successful. This reflected the lack of information on properties of substances involved in the process, which is vital for the selection of operations in the best possible order according to their relative efficiencies (Asenjo 1990b, Asenjo et al. 1991, Asenjo & Maugeri 1992, Leser & Asenjo 1992). The same shortage of data leads to an unreliable prototype to find the solution for the final polishing operations to remove minor impurities.

Information on thermodynamic properties of protein product and main contaminant proteins should be possible to use to predict the performance of a particular separation. In order to do this it is necessary to find the differences in physicochemical properties that determine the separation behaviour of the different proteins in the system. The calculation of a deviation factor (DF), that takes into account the differences between the product and each one of the contaminants for a particular thermodynamic property (e.g. molecular weight, hydrophobicity or charge) is shown in Table V. An efficiency factor (η) for the separation process in exploiting this difference in a specific property has to be included in the evaluation. It is possible, then, to define a 'separation coefficient' (SC) that can be used to characterize the ability of the separation operation to separate two or more proteins (Asenjo 1990b).

$$SC = f(DF, \eta)$$

A term for considering the concentration has also been included in the selection. This will affect the selection criteria, since the contaminants in higher concentrations have to be removed first (Heuristics: Rule 2). But, as concentration does not appear intrinsically to influence the actual separation coefficient, the suggestion of using the term 'Separation Selection Coefficient'(SSC), when including the concentration term θ has been preferred

$$SSC = DF \cdot \eta \cdot \theta$$

When using chromatography there are differences in the cost of the matrices used (e.g., protein A affinity chromatography uses a more expensive matrix than CM-Sepharose ion-exchange). Also different adsorption capacities and flow characteristics of the matrices will result in columns of different size. Yet, column hardware cost is only a fraction of the total cost hence the total hardware cost of a chromatographic step is somewhat constant. Differences in the cost of a purification operation are considered by using a cost factor (CF) giving an expression for the term economic separation coefficient (ESC)

$$ESC = f(SC, CF)$$

The values of the parameters in these two expressions should range between $0 < \eta \leq 1$ and $0 \leq DF \leq 1$. As such values are relative, the maximum value for DF for individual properties is defined within this range and the value for η given to a particular operation also will depend on the range (or maximum possible value for the deviation of a specific protein property) that must be standardized for different operations. The value of the cost factor (CF) will be $CF < 1$ or $CF \geq 1$ and a standard operation (such as Ion-exchange using CM-Sepharose) has been given a value of 1. We have made a first attempt to define both a separation coefficient, SC, and an economic separation coefficient, ESC. They are shown in Table V. It should be noted that the two parameters η and the cost factor, CF, are thus far empirical and subjective. A more rigorous estimate is presently under study in our group (Watanabe et al. 1993). The cost factor is not based on a rigorous economic evaluation, such as has recently been carried out (Kosti 1989), but on an 'approximate', preliminary evaluation of the cost involved using such an operation. There are many elements apart from the direct variable and capital costs that affect the choice of process and therefore the 'approximate' evaluation of cost and thus CF (e.g., availability of matrix in the pilot plant, reliability, robustness with variation in feedstock, speed of

process implementation or quality control). This role of other elements is partly related because the cost of production of a therapeutic or diagnostic protein is still only a small fraction of the final price. Therefore the cost differences found in a rigorous economic evaluation are much more marked than those shown in the expression in Table V. All values shown in Table V will be subjected to modifications as the rationale proposed is tested in real cases (Watanabe et al. 1993).

TABLE V

Separation Coefficient (SC), Selection Separation Coefficient (SSC) and Economic Separation Coefficient (ESC), their definitions and hypothetical values associated to some operations (Leser & Asenjo 1992)

Separation coefficients					
$SC = DF \cdot \eta$					
DF =	Deviation factor for hydrophobicity, molecular weight, and pl				
DF =	$\frac{ \text{Protein Value} - \text{Contaminant Value} }{\text{Max.} (\text{Protein Value} , \text{Contaminant Value})}$				
DF = 1.0 for Affinity Chromatography					
$\eta =$ Efficiency =	<table border="0"> <tr><td>1.00 for Affinity Chromatography</td></tr> <tr><td>0.70 for Ion Exchange</td></tr> <tr><td>0.35 for Hydrophobic Interaction Chromatography</td></tr> <tr><td>0.20 for Gel Filtration</td></tr> </table>	1.00 for Affinity Chromatography	0.70 for Ion Exchange	0.35 for Hydrophobic Interaction Chromatography	0.20 for Gel Filtration
1.00 for Affinity Chromatography					
0.70 for Ion Exchange					
0.35 for Hydrophobic Interaction Chromatography					
0.20 for Gel Filtration					
$SSC = DF \cdot \eta \cdot \theta$					
$\theta =$ Concentration factor					
$\theta =$	$\frac{\text{Concentration of Contaminant Protein}}{\text{Total Concentration of Contaminant Proteins}}$				
$ESC = \frac{SSC}{CF}$					
CF = Cost factor =	<table border="0"> <tr><td>1.0 for Affinity Chromatography</td></tr> <tr><td>0.6 for Gel Filtration</td></tr> <tr><td>0.3 for Ion Exchange</td></tr> <tr><td>0.3 for Hydrophobic Interaction Chromatography</td></tr> </table>	1.0 for Affinity Chromatography	0.6 for Gel Filtration	0.3 for Ion Exchange	0.3 for Hydrophobic Interaction Chromatography
1.0 for Affinity Chromatography					
0.6 for Gel Filtration					
0.3 for Ion Exchange					
0.3 for Hydrophobic Interaction Chromatography					

The prototype was further improved with the implementation of a rationale for selection of high resolution purification operations. This was done by interfacing a program in PASCAL in which the main molecular properties of a target product protein were compared with those of the main protein contaminants and then used to select

the most appropriate high resolution purification operations (Purves 1990, Asenjo et al. 1991). The 'expanded' prototype with access to the databases and the more rational selection of high resolution separation operations resulted in an ES with approximately 130 rules besides the Pascal interface that implements the database and the rationale. For its implementation and testing, the data shown in Tables VI, VII and VIII were used. In these tables, *Band Number* refers to the identification of individual proteins with reference to their position in protein band profiles by polyacrylamide gel electrophoresis (PAGE) of the crude material.

For our prototype presently being developed (work in progress) we are using a newer shell. Recent developments in AI point towards use of two important concepts, namely the use of an open architecture and object-oriented programming for ES. NEXPERT Object, from Neuron Data, which incorporates these features proves to be extremely well suited to the improvement of the previous prototypes. The use of these attributes enables a more integrated solution, making easier to interface the system with external sources to access banks of data and spreadsheets.

TABLE VI

Properties of the 10 main protein bands in *Escherichia coli* lysate^a (Andrews et al. 1993)

Band number	Molecular weight ^b	Hydrophobicity Φ^c	Isoelectric point ^d
1	90,000	0.02 M	4.8
2	145,000	1.12 M	4.8
3	80,000	0.13 M	4.9
4	200,000	1.02 M, 0.13 M	4.8
5	12,800	0.64 M	5.1
6	25,000	0.26 M	4.5
7	45,000	0.13 M	5.4
8	40,000	0.64 M	4.6
9	44,000	0.13 M	4.3
10	120,000	0.02 M	5.4
11	80,000	0.13 M	4.6

a: cell lysate was prepared by bead milling.

b: measured by gel filtration using Sephacryl S-200 in 0.05 M sodium phosphate buffer, at pH 7.0.

c: measured by hydrophobic interaction chromatography (HIC) using a Phenyl-Superose gel in an FPLC and a gradient elution from 2.0 M to 0.0 M (NH₄)₂SO₄ in 0.1 M KH₂PO₄. Units used are the concentration of (NH₄)₂SO₄ at which the protein eluted.

d: measured by isoelectric focusing using either pH 4.0-6.5 or pH 3-10 range Sephadex gel.

TABLE VII

Properties of the 10 main protein bands present in *Saccharomyces cerevisiae* lysate^a (Andrews et al. 1993)

Band number	Molecular weight ^b	Hydrophobicity Φ^c	Isoelectric point ^d
1	80,000	0.50 M	6.6
2	44,000	0.60 M, etOH	6.4
3	22,000	0.25 M	5.6
4	80,000	etOH	6.6, 8.8
5	49,000	ppt.	5.5
6	71,000	0.30 M	5.7
7	170,000	0.40 M	5.7, 6.9
8	12,000	ppt.	7.1
9	170,000	0.15 M	5.7
10	65,000	0.65 M	6.0, 7.7

a: cell lysate was prepared by bead milling.

b: measured by gel filtration using Sephacryl S-200 in 0.05 M sodium phosphate buffer, at pH 7.0.

c: measured by hydrophobic interaction chromatography (HIC) using Octil-Sepharose gel in an FPLC and a gradient elution from 1.5 M to 0.0 M (NH₄)₂SO₄ to avoid protein precipitation. Some protein bands still precipitated (ppt. in table) etOH means tightly bound band that needed to be eluted with 24% ethanol in deionized water.

d: measured by isoelectric focusing using either pH 4.0-6.5 or pH 3-10 range Sephadex gel.

TABLE VIII

Properties of the 10 main protein bands in Chinese hamster ovary cells culture supernatant (Andrews et al. 1993)

Band number	Molecular weight ^a	Hydrophobicity Φ^b	Isoelectric point ^c
1	66,000	0.83 M	5.0
2	140,000-205,000	0.83 M, ppt.	5.4, 8.7
3	295,000	0.83 M	6.0
4	72,000	0.70 M	5.4
5	53,000	1.25 M	5.2
6	72,000	0.70 M	5.4
7	170,000	1.10 M	4.6
8	3,000	1.25 M	5.4
9	6,000	0.02 M	4.0
10	170,000	0.71 M	5.7

a: measured by gel filtration using Sephacryl S-200 in 0.05 M sodium phosphate buffer, at pH 7.0.

b: measured by hydrophobic interaction chromatography (HIC) using Phenyl-Superose gel in an FPLC and a gradient elution from 1.7 M to 0.0 M (NH₄)₂SO₄ to avoid protein precipitation. Some protein bands still precipitated (ppt. in table).

c: measured by isoelectric focusing using either pH 4.0-6.5 or pH 3-10 range Sephadex gel.

PROPERTY PROTEINS	CONCENTRATION	MOLECULAR WEIGHT	ISOELECTRIC POINT	HYDROPHO- BICITY	CHARGE				
					pH 4.0	pH 4.5	pH 9.5	pH 10.0
PRODUCT									
CONTAMINANT 1									
CONTAMINANT 2									
CONTAMINANT 3									
CONTAMINANT 4									
CONTAMINANT 5									
⋮									
CONTAMINANT N									

Fig. 5: the structure of the database components for main proteins and contaminants in one of the production streams to be used in the selection of optimal separation operations.

The present work changes the concept of creating branches in the downstream process and reasons in terms of a linear flow of operations. The separation amongst "frames" is avoided. There are improvements in terms of a more precise definition of work conditions: it follows more closely the GMP (Good Manufacturing Practice) standards for bioprocessing. Within a more accurate designer's point of view, there are features enclosed in the knowledge base that can create a documented reasoning when the user must choose, e.g. between centrifugation and microfiltration for cell separation. The analysis of processes conditions associated with the expert reasoning (Kroner et al. 1984, Bowden 1985, Mackay & Salysbury 1988, Lee 1989) leads to a proposed solution for the cell separation operations. Although these changes are being carried out in the core of the knowledge base, due to the necessity of a different approach and reasoning, the most important modification is the one occurring in the improvement of bridges to link the system to databases and spreadsheets. Secondly, the source to calculate the separation coefficients relies on real data on production streams. The determination of physicochemical properties of contaminants in a systematic generation of data banks for process selection is being carried out now (Woolston et al. 1992) besides a refinement and the validation of the rationale for selecting operations (Tsoka 1992, Watanabe et al. 1993). Now the prototype focuses on three of the most common production microorganisms. However its open architecture will enable the addition of other categories. Recently it has been demonstrated that charge density (net charge/molecular weight), determined over a range of Ph using an electrophoretic technique (titration curve), is the major factor

affecting protein behaviour in ion exchange chromatography. With minor deviations this parameter can be used to predict the chromatographic behaviour of a protein (Watanabe et al. 1993).

The data banks referred in Fig. 4 are being built for *E. coli*, hybridoma and Chinese hamster ovary cell supernatant and insect cells. A model of a table (database) is shown in Fig. 5.

CONCLUSIONS

The complex problem of selection of an optimal sequence of processes to purify proteins from multi component mixtures can be simplified. The utilization of a rational approach based on the exploitation of fundamental physicochemical properties of protein molecules should lead to coherent and optimized decisions. To help accomplish this task the use of contemporary computers techniques supported by symbolic computation proves to be an extreme helpful tool. These techniques are developed aiming at optimized solutions for large scale production applications, but also for finding optimal and simplified separation schemes in the research laboratory. When the final aim of an investigation is production of a protein in large amounts, the utilization of feasible methods for large scale operation should be an essential criterion to choose the separation processes at the research laboratory.

REFERENCES

- Andrews AT, Noble I, Keeratipibul S, Asenjo JA 1993. Physico chemical properties of the main matrix proteins of three important culture vehicles. *Biotech Bioeng* (in press).
- Asenjo JA 1990a. Cell disruption and removal of insolubles, p. 11-20. In DL Pyle *Separations for biotechnology II*. Elsevier, New York.

- Asenjo JA 1990b. Selection of operations in separation processes, p. 3-16. In JA Asenjo, *Separation processes in biotechnology*. Marcel Dekker, New York.
- Asenjo JA, Herrera L, Byrne B 1989. Development of an expert system for selection and synthesis of protein purification processes. *J Biotechnol* 11: 275-298.
- Asenjo JA, Maugeri F 1992. An expert system for selection and synthesis of protein purification processes, p. 358-379. In P Todd, SK Sikdar, M Bier (eds) *Frontiers in bioprocessing II*. American Chemical Society, Washington.
- Asenjo JA, Parrado J, Andrews BA 1991. The rational design of purification processes for recombinant proteins. *Ann NY Acad Sci* 646: 334-356.
- Asenjo JA, Patrick I 1990. Large-scale protein purification, p. 1-28. In ELV Harris, S Angal (eds) *Protein purification applications: a practical approach*. IRL press, Oxford.
- Bowden CP 1985. Recovery of microorganisms from fermented broth. *The Chemical Engineer* 415: 50-53.
- Buchanan BG, Shortliffe EH 1984. *Rule-Based Expert Systems*. Addison-Wesley, Reading, UK, 748 pp.
- Clapp KP, Ruel GJ 1991. Expert systems in bioprocessing. *BioPharm (February)*: 23-35.
- Dwyer JL 1984. Scaling up bioproduct separation with high performance liquid chromatography. *Bio/Technol* 2: 957.
- Dym CL, Levitt RE 1991. *Knowledge based systems in engineering*. McGraw-Hill, Singapore, 404 pp.
- Harmon P, King D 1985. *Expert systems. Artificial intelligence in business*. John Wiley, New York, 283 pp.
- Hushon JM 1990. Overview of environmental expert systems, p. 1-24. In JM Hushon *Expert systems for environmental applications*. American Chemical Society, Washington.
- Kane JF, Hartley DL 1988. Formation of recombinant protein inclusion bodies in *Escherichia coli*. *Trends in Biotechnol* 6: 95-101.
- Kosti R 1989. *Economic evaluation of large scale protein purification operations*. M.Sc. thesis, University of Reading, U.K., 96 pp.
- Kroner KH, Schutte H, Hustedt H, Kula MR 1984. Cross-flow filtration in the downstream processing of enzymes. *Proc Biochem* 19: 67-74.
- Lee S-M 1987. Affinity purification of monoclonal antibody from tissue culture supernatant using protein A-Sepharose CL-4B, p. 199-216. In SS Seaver *Commercial production of monoclonal antibodies: a guide for scale-up*. Marcel Dekker, New York.
- Lee S-M 1989. The primary stages of protein recovery. *J Biotechnol* 11: 103-118.
- Leser EW, Asenjo JA 1992. Rational design of purification process for recombinant proteins. *J Chrom* 11: 103-118.
- Mackay D, Salysbury T 1988. Choosing between centrifugation and cross flow microfiltration. *Chem Eng* 447: 45-50.
- Müller F, Brühl K, Freidel K, Kowallik KV, Ciriacy M 1989. Processing of TY1 proteins and formation of Ty1 virus-like particles in *Saccharomyces cerevisiae*. *Mol and Gen Genetic* 207: 421-429.
- Nadgir VM, Liu YA 1983. Studies in chemical process design and synthesis. *Am Inst Chem Eng J* 29: 926-934.
- Nebendahl D 1988. *Expert Systems. Introduction to the technology and applications*. John Wiley & Sons, London, 211 pp.
- Pharmacia 1983. *Scale-up to process chromatography. Guide to design*. Uppsala, Sweden.
- Prokopakis G, Asenjo JA 1990. Synthesis of Downstream Processes, p. 571-601. In JA Asenjo *Separation Processes in Biotechnology*. Marcel Dekker, New York.
- Purves IJ 1990. *The testing and development of an expert system for selection and synthesis of protein purification processes*. M.Sc. thesis, City University, London, 202 pp.
- Siletti CA 1989. *Computer aided design of protein recovery processes*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, USA, 379 pp.
- Stephanopoulos G, Mavrovouniotis M 1988. Artificial intelligence in chemical engineering research and development. *Comp Chem Eng* 12(9/10): V-VI.
- Tsoka S 1992. *Selection of chromatographic protein purification operations based on physicochemical properties*. M.Sc. thesis, University of Reading, UK, 80 pp.
- Wacks S 1987. *Design of protein separation sequences and downstream process in biotechnology*. M.Sc. thesis, Columbia University, New York, USA, 39 pp.
- Wang DIC 1987. Separations for Biotechnology, p. 17-46. In MS Verral, MJ Hudson (eds) *Separations for Biotechnology*. Ellis Horwood Ltd., Chichester, UK.
- Watanabe E, Tsoka S, Asenjo JA 1993. Selection of chromatographic protein purification operations based on physicochemical properties. *Ann NY Acad Sci* (in press).
- Wheelwright SM 1991. *Protein purification: design and scale up of downstream processing*. Carl Hanser, München. 228pp.
- Wheelwright SM, Asenjo JA 1993. Process Design. In KH Kroner, N Papamicahael (eds) *Downstream processing, recovery and purification of proteins. Handbook of principles and practice*. Carl Hanser, München (In press).
- Woolston P, Wharam MJ, Kearns MJ, Asenjo JA 1992. A database for the selection of large scale protein purification processes. Presented at the *Recovery of Biological Products VI*. Interlaken, Switzerland, 20-25 September.