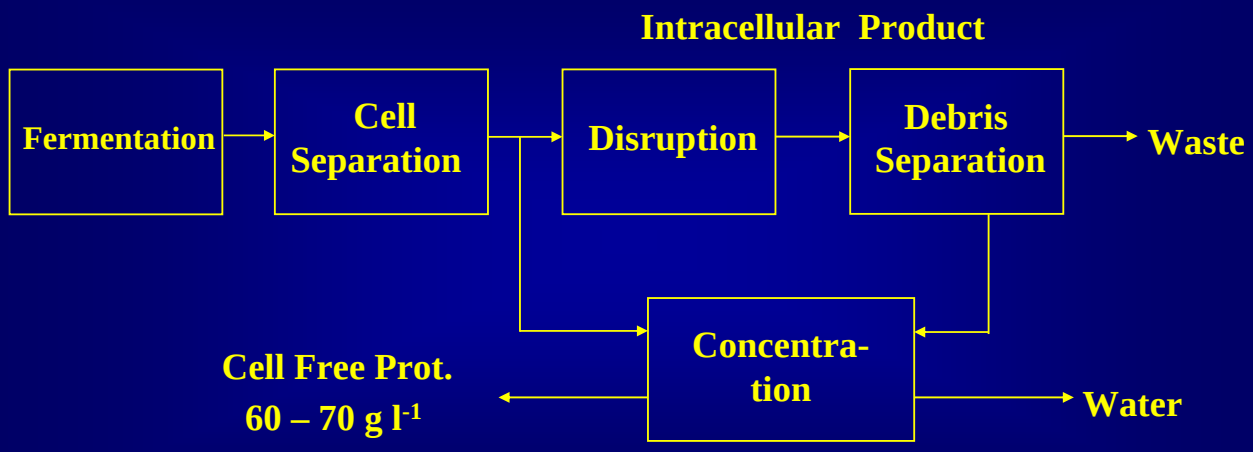
Is there a Rational Method to Purify Proteins? From Expert Systems to Proteomics

M.E.Lienqueo and J.A Asenjo
Centre for Biochemical Engineering and Biotechnology
University of Chile
mlienque@ing.uchile.cl

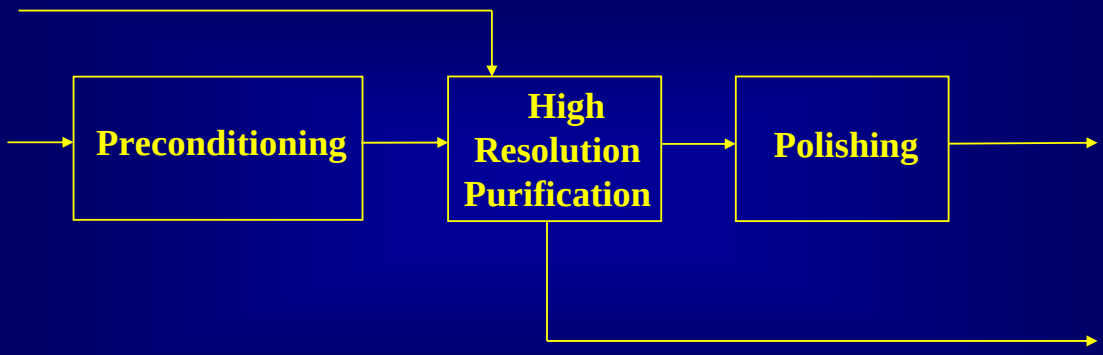
PASI 2008
Mar del Plata-Argentina

Protein Production Process

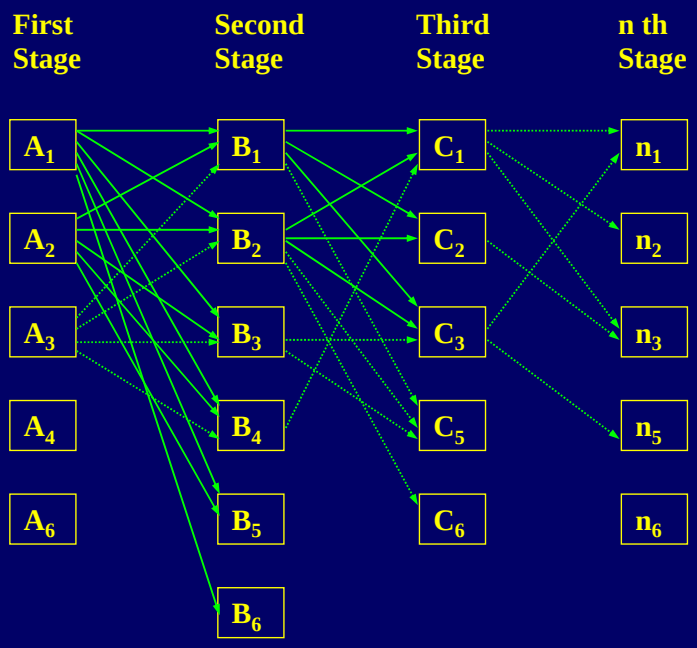
The operations involved in the *Recovery* block of a separation process.



The operations involved in the *Purification* block of a separation process.



The Combinatorial Characteristic of Choosing the Sequence of Operations for Protein Purification



1) Ion Exchange Chromatography

2) Hydrophobic Interaction Chromatography

3) Affinity Chromatography

4) Aqueous Twø Phase Separation

5) Gel Filtration

6) HPLC

Basic Information for Designing a Separation Process

1.- Defining Final Product - Final Utilization

- Final Purity level desired
- Level of production

2.- Characterisation of Starting Material

- Fermentation Source
- Cell Concentration
- Type of cultivation medium used
- Localization of the product
- **Physicochemical properties**

3.- Possible separation steps and constraints

4.- Evaluated possible process integration

Properties to be Exploited for the Separation and Purification of Different Proteins

1. **Charge (Titration Curve)**
2. **Surface Hydrophobicity**
3. **M. W. (Molecular Weight)**
4. **Biospecificity toward certain ligands (Affinity)**
5. **pI (Isoelectric Point)**
6. **Shape (Stokes Radius)**

Properties of Main Protein Contaminants in fermentation source:

- **Bacterial- *E.coli***
- **Yeast - *S. cerevisiae***
- **Mammalian cell -CHO**

Properties of Main Protein Contaminants in *E.coli*[a] *Lysate*

Band Number	Molecular Mass [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	0.13M	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 permeation.
[c]measured by hydrophobic interaction chromatography (HIC) using a Phenyl-Superose gel in an FPLC and gradient elution from 2.0 M to 0.0 M $(NH_4)_2SO_4$ in 0.1 M KH_2PO_4. Units used are the concentration of $(NH_4)_2SO_4$ at which the protein eluted.
[d]measured by isoelectric focusing using a Sephadex gel.

Properties of the 10 Main Protein Bands Present in *S. cerevisiae Lysate*[a]

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.
[c]measured by hydrophobic interaction chromatography (HIC) using a Octyl-Sepharose gel in an FPLC and a gradient elution from 1.5 M to 0.0 M $(NH_4)_2SO_4$ 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 deonized water.
[d]measured by isoelectric focusing using a Sephadex gel.

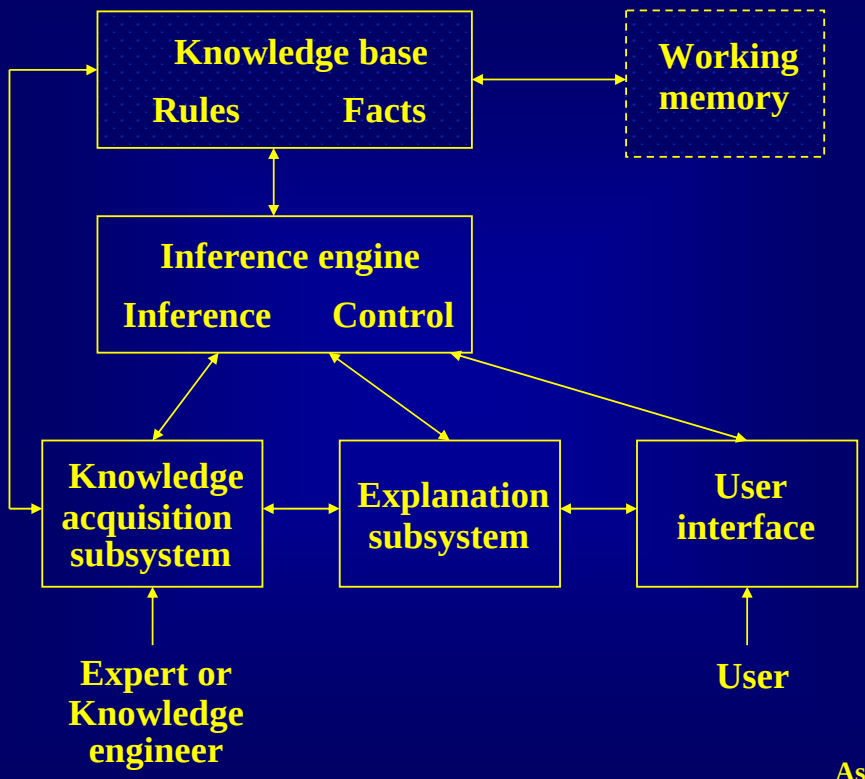
Properties of the 10 of Main Protein Bands in CHO* Culture Supernatant

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

*Chinese Hamster Ovary Cells
[a]measured by gel filtration.
[b]measured by hydrophobic interaction chromatography (HIC) using a Phenyl-Superose gel in an FPLC and a gradient elution from 1.7 M to 0.0 M $(NH_4)_2SO_4$ to avoid protein precipitation. Some protein bands still precipitated (ppt. in table).
[c]measured by isoelectric focusing using a Sephadex gel.

Expert System for selection of protein purification process

The architecture of a knowledge based expert system.



Asenjo, Herrera and Byrne, 1989

Journal of Biotechnology, 11 (1989) 275–298
Elsevier

BIOTEC 00422

Development of an expert system for selection and synthesis of protein purification processes

J.A. Asenjo, L. Herrera * and B. Byrne

Biochemical Engineering Laboratory, University of Reading, P.O. Box 226, Reading RG6 2AP, U.K.

(Received 12 May 1989; accepted 30 May 1989)

Summary

A 'second generation' of protein purification rules has been developed and implemented into a prototype expert system (ES) using a 'shell'. The work has concentrated on developing a more precise and accurate knowledge base for selection of optimal large scale protein purification sequences. Expert knowledge was obtained partly from the literature but mainly from industrial experts working on the large scale separation of therapeutic, diagnostic and analytical proteins. The knowledge was expressed in ~ 65 rules, some of which carry a degree of uncertainty. The downstream process was divided into two distinct subprocesses: a first subprocess called *recovery* after which the total protein concentration is 60–70 g l^{-1} and a second subprocess called *purification*. A limiting factor in the development of ESs for protein purification at present is the acquisition, clarification, formalization and structuring of the domain of expert knowledge. The main deficiency of accurate information was found to be in that required for the selection of high resolution purification operations on a rational basis. An expert system for selection of optimal protein separation sequences will give the user a number of alternatives chosen on the basis of extensive data back-up on proteins and unit operations. This constitutes a clear case of 'expert amplification' and not of 'expert replacement'.

Expert system; Selection; Protein purification; Purification process

Basic heuristic rules for the downstream processing design

(1) Choose the separation based on the diferent physicochemical properties.

(2) Eliminate those proteins and compounds that are found in greater percentage first.

(3) Use a high resolution step, as soon as possible.

(4) Do the most arduous purification step at the end of the process (ünal polishing).

Rules

1.- To select the initial harvesting equipment (H-EQUIPMENT).

		Certainty factor
IF	MICROORGANISM = FUNGI	
THEN	H-EQUIPMENT = Microporous membrane system	0.4
	H-EQUIPMENT = Rotary vacuum filter	0.3
	H-EQUIPMENT = Filter press	0.3
IF	MICROORGANISM = YEAST	
THEN	H-EQUIPMENT = Disc centrifuge	0.6
	H-EQUIPMENT = Microporous membrane system	0.4
IF	MICROORGANISM = BACTERIA	
THEN	H-EQUIPMENT = Microporous membrane system	
IF	MICROORGANISM = MAMMALIAN	
THEN	H-EQUIPMENT = Disc centrifuge	0.7
	H-EQUIPMENT = Microporous membrane system	0.3
IF	MICROORGANISM = UNDEFINED	
THEN	H-EQUIPMENT = Disc centrifuge	0.5
	H-EQUIPMENT = Microporous membrane system	0.5

2.- To select the operation

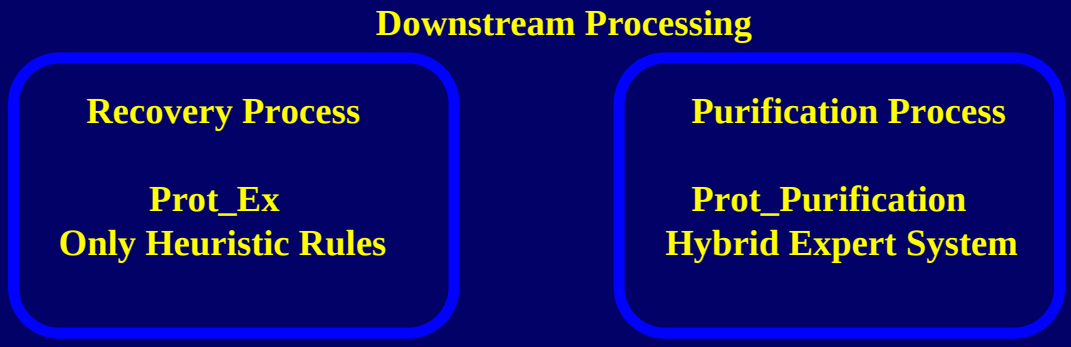
IF MICROORGANISM = MAMMALIAN
THEN PRODUCT IS EXTRACELLULAR
IF PRODUCT IS INTRACELLULAR
THEN DISRUPTION IS REQUIRED
DEBRIS SEPARATION IS REQUIRED
PRECIPITATION OF NUCLEIC ACIDS IS REQUIRED
IF PRODUCT IS EXTRACELLULAR
THEN DISRUPTION IS NOT REQUIRED
DEBRIS SEPARATION IS NOT REQUIRED
PRECIPITATION OF NUCLEIC ACIDS IS NOT REQUIRED

Design of Downstream Processing

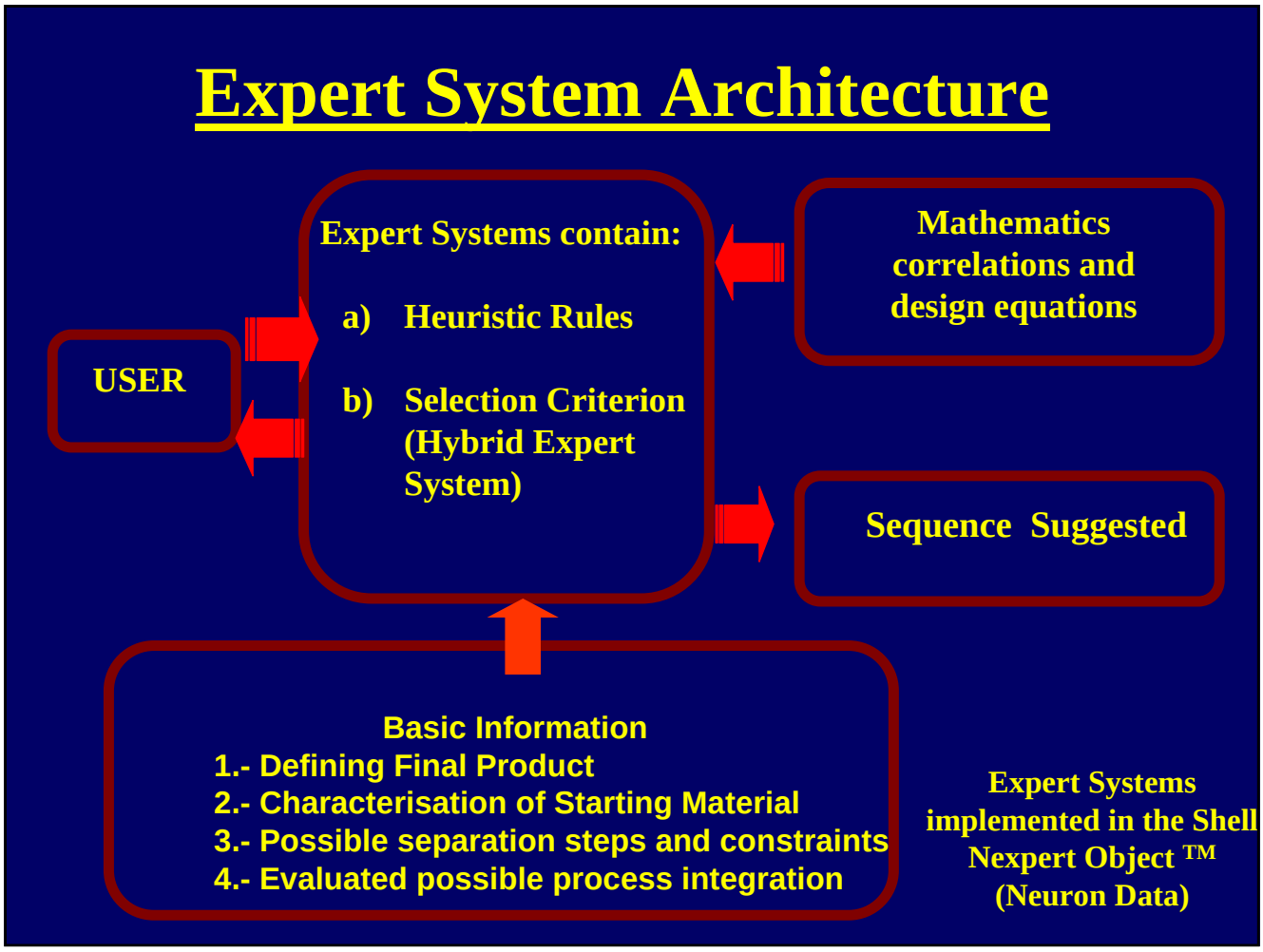
- Rigorous solution using numerical methods

Use of Artificial Intelligence techniques, Expert System

- Use of Heuristic Rules from Human Expert or/and Literature
- Use of simple mathematical correlations and strict quantitative data (Hybrid Expert System)



Expert System Architecture



Prot_Ex_Purification for Purification Process

-Choose between several chromatographic steps (more than 20)

- Use Selection Criteria defined from basic heuristic rules for separation process:

- SSC Criterion	Consider the ability of the purification operation to separate two or more proteins
- Purity Criterion	Consider the purity level obtained after a purification operation has been applied

- Use mathematics correlation for predict ability and level of purity (Hybrid Expert System)

Rational Selection Criteria

Separation Selection Coefficient Criterion

This criterion selects the best process using the SSC value calculated for each chromatographic technique and each contaminant protein.

$$SSC_i = DF_i \cdot \eta \cdot \theta_i$$

$DF_i = | K_{D\ target\ protein} - K_{D\ contaminant\ i} |$

η: Efficiency

θ_i : Concentration Factor

The best process will be the one with the highest SSC value

K_D : Dimensionless Retention Time

$$DF_i = | K_{D\ target\ protein} - K_{D\ contaminant\ i} |$$

$$K_D = f(\text{physicochemical properties})$$

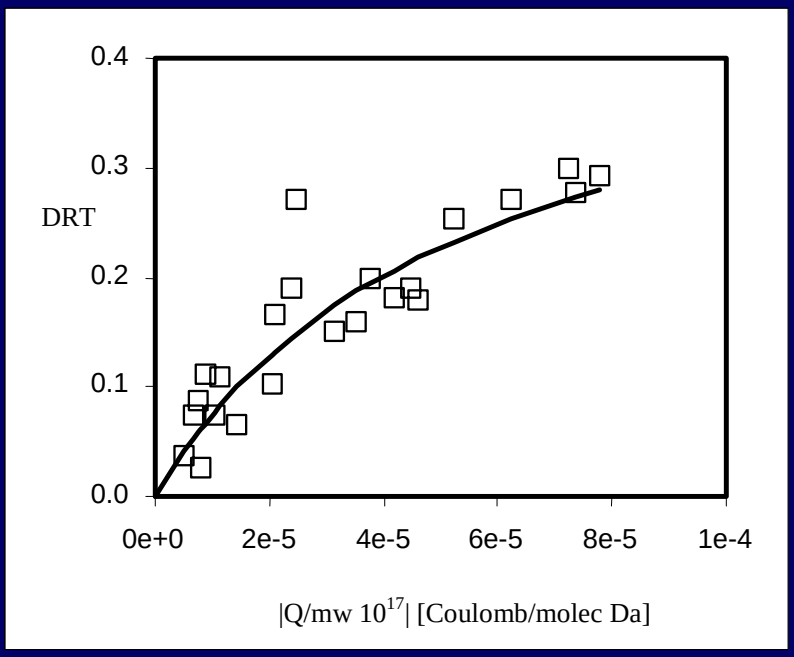
Anion and Cation Exchange : f(Q,mw)

Hydrophobic Interaction : f(ϕ)

Gel Filtration : f(mw)

Charge density as a function of retention time

$$DRT = \frac{A \cdot |\,Q/mw\,|}{1 + B \cdot |\,Q/mw\,|}$$



Charge density as a function of retention time for all pHs. Calculations were based on the results obtained for anion exchange chromatography.

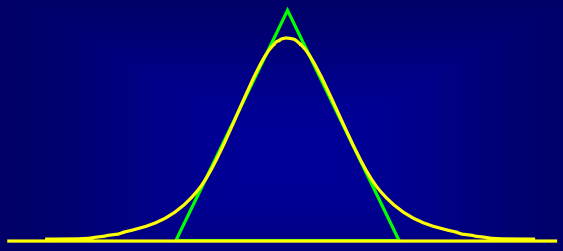
Expressions and parameters used for SSC and Purity criteria

Chromatographic Techniques	*Retention time (Kd)*	*Efficiency Factor (η)*	*Peak width (Σ)*
Anion exchange	$\frac{7383\ (Q\ 10^{26}/mw)}{1 + 15844\ (Q\ 10^{26}/mw)}$	1.00	0.15
Cation exchange	$\frac{5972\ (Q\ 10^{26}/\mathrm{mw})}{1 + 17065\ (Q\ 10^{26}/\mathrm{mw})}$	1.00	0.15
Hydrophobic interaction	ϕ	0.86	0.22
Gel filtration	2.39–0.47 log *mw*	0.66	0.46

Where: Q represents absolute value of surface charge (coulomb molecule^{-1}), *mw* represents molecular weight (Da), and ϕ represents surface hydrophobicity.

Values of Process Efficiency

Chromatographyc Process	Efficiency (η)
Ion-exchange	1.00
Hydrophobic interaction	0.86
Size exclusion	0.66

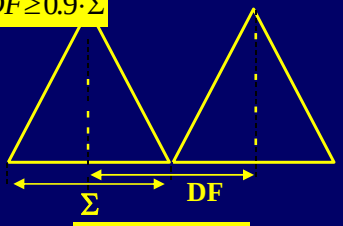


Peaks as Triangles

Concentration factor, θ

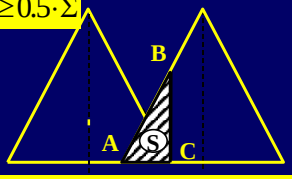
$$\theta = \frac{\text{Concentration of Contaminant Protein}}{\text{Total Concentration of Contaminant Proteins}}$$

a) $DF \geq 0.9 \cdot \Sigma$



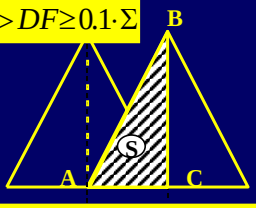
$$C_2 = C_1 \cdot 0.02$$

b) $0.9 \cdot \Sigma > DF \geq 0.5 \cdot \Sigma$



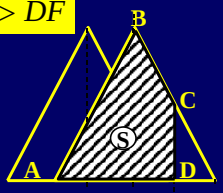
$$C_2 = \frac{C_1 \cdot 2.02 \cdot (\Sigma^2 - DF^2)}{\Sigma^2}$$

c) $0.5 \cdot \Sigma > DF \geq 0.1 \cdot \Sigma$



$$C_2 = \frac{C_1 \cdot 1.02 \cdot (\Sigma^2 - 2 \cdot DF^2)}{\Sigma^2}$$

d) $0.1 * \Sigma > DF$



$$C_2 = C_1$$

Criteria to determine the percentage of contaminant eliminated after a chromatographic step for different values of DF. Left triangle: protein product, Right triangle: protein contaminant, Σ: peak width, shaded area: contaminant left with protein after purification step

Rational Selection Criteria

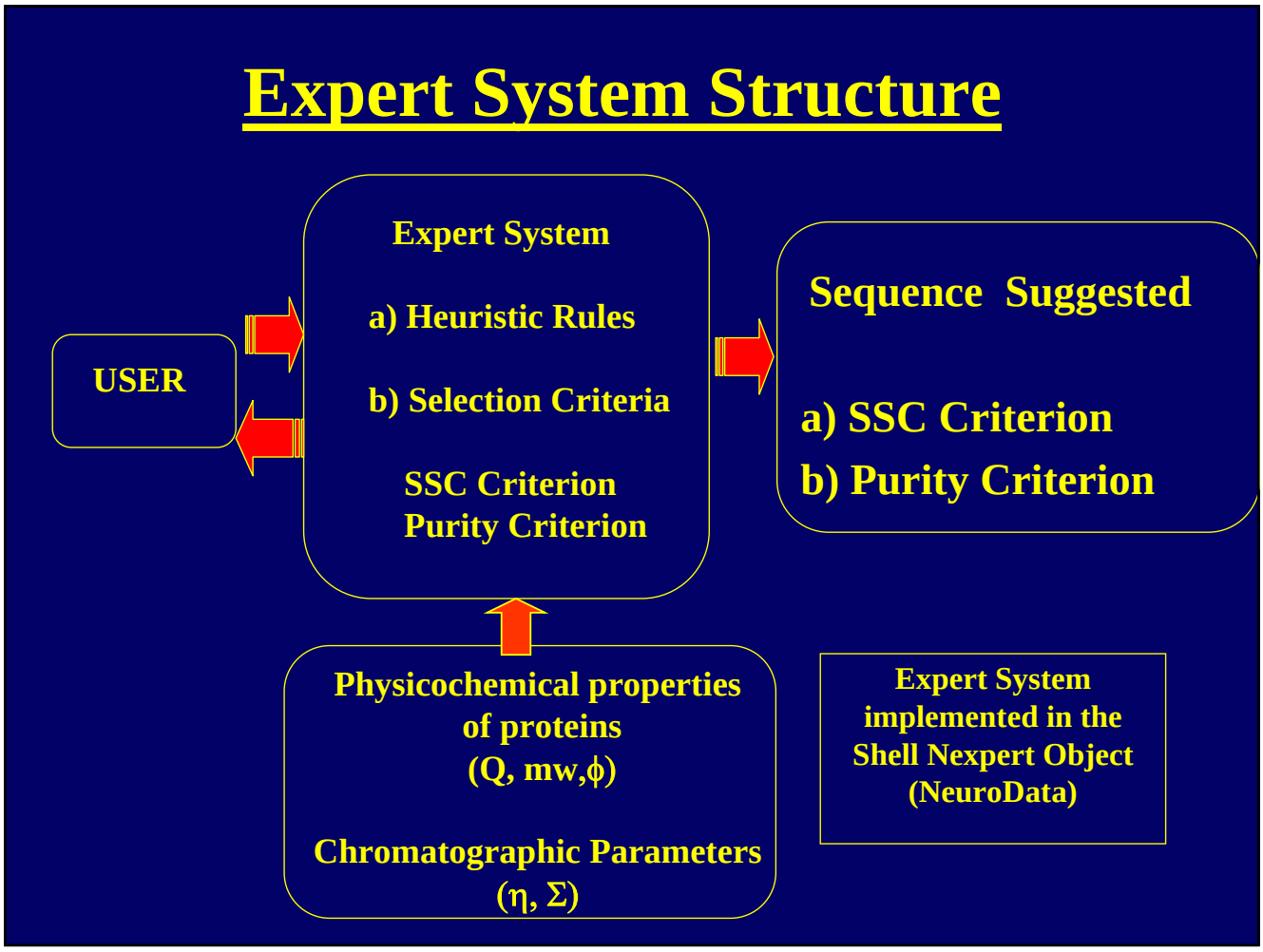
Purity Criterion

This criterion compares the final purity level obtained after a particular chromatographic technique has been applied.

$$\text{Purity Level} = \frac{[\text{Target Protein}]}{\Sigma[\text{All Proteins}]}$$

The best process will be the one with the highest purity level.

Expert System Structure



Examples

An Expert System for the selection and synthesis of multistep protein separation processes

M.E.Lienqueo, E.W. Leser and J.A. Asenjo

Computers & Chemical Engineering ,24: 2339 – 2350, 2000.

<u>Validation :</u> Recovery of Somatotropin from *E.coli*

Pergamon

Computers chem. Engng Vol. 20, Suppl., pp. S189–S194, 1996

S0098-1354(96)00042-7
0098-1354/96 $15.00+0.00

AN EXPERT SYSTEM FOR THE SELECTION AND SYNTHESIS OF MULTISTEP PROTEIN SEPARATION PROCESSES

M.E. LIENQUEO, E.W. LESER* AND J.A. ASENJO

Centre for Biochemical Engineering and Biotechnology
Department of Chemical Engineering, Universidad de Chile
Beauchef 861, Santiago, Chile
*Instituto Oswaldo Cruz, Rio de Janeiro, Brazil

Abstract - A hybrid Expert System that combines expert rules and mathematical correlations to manipulate databases for selecting the sequence of operations for the downstream purification of proteins was built. This paper describes how the physicochemical data on the protein product and the other proteins present (contaminants) was used to select a sequence of operations with a minimum number of steps to achieve a defined level of purity. An algorithm was developed to model the amount of each protein contaminat eliminated after each step. The system was tested in a practical example and it was found that process synthesis was sufficiently sensitive to important changes in the physicochemical parameters of the product and rather robust to small variations or errors in the measurement of such parameters (e.g. 10%).

Concentration, molecular weight, hydrophobicity and charge at different pHs, for the main proteins ("contaminants" of the product) in *Escherichia coli*. Data from Woolston (1994)

					Charge[4] (Coulomb per molecule x 1E25)									
		g/litre	Da	*	pH 4	pH 4,5	pH 5	pH 5,5	pH 6	pH 6,5	pH 7	pH 7,5	pH 8	pH 8,5
Contaminant	pI [1]	weight	Mol wt [2]	hydroph [3]	q A	q B	q C	q D	q E	q F	q G	q H	q I	q J
Cont_1	4.67	11.29	18,370	0.71	1.94	0.25	-0.80	-1.41	-1.76	-1.97	-2.15	-2.33	-2.45	-2.67
Cont_2	4.72	7.06	85,570	0.48	2.35	0.29	-1.17	-2.17	-2.83	-3.24	-3.50	-3.63	-3.68	-3.64
Cont_3	4.85	4.63	53,660	0.76	1.83	0.67	0.04	-0.30	-0.49	-0.65	-0.85	-1.90	-1.34	-1.50
Cont_4	4.92	5.58	120,000	1.50	3.29	1.38	-0.03	-0.69	-1.07	-1.34	-1.73	-2.30	-2.85	-2.75
Cont_5	5.01	4.83	203,000	0.36	4.08	1.83	0.04	-1.17	-1.92	-2.46	-3.07	-3.90	-4.98	-5.65
Cont_6	5.16	2.48	69,380	0.36	5.22	3.17	1.02	-0.72	-1.90	-2.60	-3.05	-3.46	-3.90	-4.24
Cont_7	5.29	7.70	48,320	0.48	3.96	3.16	1.12	-0.58	-1.36	-1.34	-1.00	-0.95	-1.59	-2.84
Cont_8	5.57	6.80	93,380	0.93	10.90	5.81	2.78	0.77	-0.81	-2.18	-3.32	-4.12	-4.45	-4.31
Cont_9	5.65	7.53	69,380		1.09	0.55	0.26	0.10	-0.03	-0.12	-0.21	-0.28	-0.32	-0.32
Cont_10	6.02	6.05	114,450	0.63	10.40	5.94	3.15	1.51	0.56	-0.05	-0.53	-0.99	-1.43	-1.72
Cont_11	7.57	3.89	198,000	0.06	0.33	0.03	0.05	0.05	0.05	0.05	0.05	-0.69	-0.97	-1.57
Cont_12	8.29	1.48	30,400		5.17	4.22	3.20	2.25	1.46	0.87	0.50	0.30	0.20	0.08
Cont_13	8.83	0.83	94,670		11.70	7.94	5.39	3.73	2.66	1.97	1.50	1.13	0.80	0.51

[*] Hydrophobicity expressed as the concentration (M) of ammonium sulphate at which the protein eluted. (Higher values represent lower hydrophobicity).

[1] Measured by isoelectric focusing using homogeneous poolyacrylamide gel in Phast System.

[2] Molecular weight was measured by SDS-PAGE with PhastGel media in Phast System.

[3] Hydrophobicity was measured by hydrophobic interaction chromatography using a phenyl-superose gel in an FPLC and a gradient elution from 2.0 M to 0.0 M $(NH_4)_2SO_4$ in 20 mM Tris buffer.

[4] Charge was measured by electrophoretic titration curve analysis with PhastGel IEF 3-9 in a Phast System.

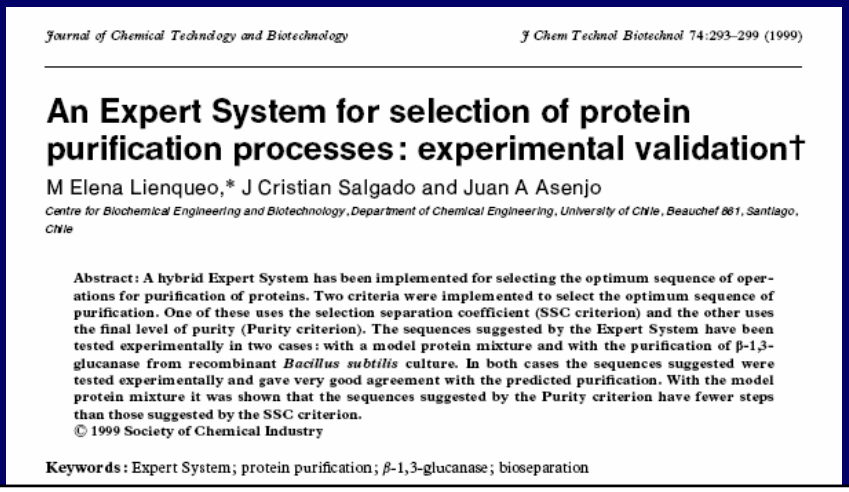
The downstream purification process of Somatotropin (Bovine Growth Hormone)

Published Process (98% purity)	PROT_EX (98,2% purity)
Centrifugation	Crossflow microfiltration
High-pressure homogenization	High-pressure homogenization
Pellet wash	Disk centrifugation
Solubilization	Solubilization
Renaturation	Renaturation
Microfiltration	Ultrafiltration
Concentration and diafiltration	
Anion exchange chromatography	Anion exchange chromatography
Hydrophobic interaction chromatography	Hydrophobic interaction chromatography

An Expert System for selection of protein purification processes: experimental validation

M.E.Lienqueo, J.C. Salgado and J.A. Asenjo

J Chem Technol Biotechnol, 74: 293-299 (1999)



Journal of Chemical Technology and Biotechnology *J Chem Technol Biotechnol* 74:293–299 (1999)

An Expert System for selection of protein purification processes: experimental validation†

M Elena Lienqueo,* J Cristian Salgado and Juan A Asenjo

Centre for Biochemical Engineering and Biotechnology, Department of Chemical Engineering, University of Chile, Beauchef 861, Santiago, Chile

Abstract: A hybrid Expert System has been implemented for selecting the optimum sequence of operations for purification of proteins. Two criteria were implemented to select the optimum sequence of purification. One of these uses the selection separation coefficient (SSC criterion) and the other uses the final level of purity (Purity criterion). The sequences suggested by the Expert System have been tested experimentally in two cases: with a model protein mixture and with the purification of β-1,3-glucanase from recombinant *Bacillus subtilis* culture. In both cases the sequences suggested were tested experimentally and gave very good agreement with the predicted purification. With the model protein mixture it was shown that the sequences suggested by the Purity criterion have fewer steps than those suggested by the SSC criterion.

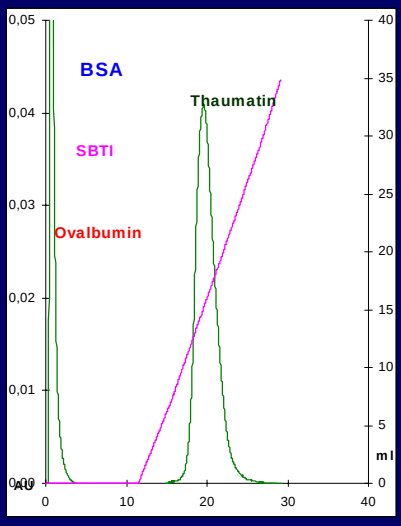
Keywords: Expert System; protein purification; β-1,3-glucanase; bioseparation

Purification of BSA

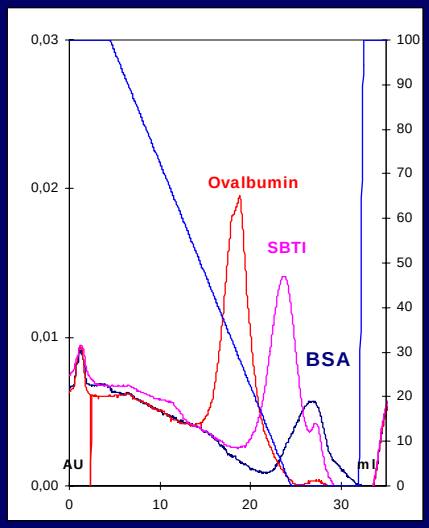
Physicochemical properties of protein mixture

Proteins	Initial Concentration (mg/ cm $^{-3}$)	Molecular weight (Da)	Hydrophobicity [$(NH_4)_2SO_4$]	Charge [Coulomb/ molecule] 10^{-25}				
				pH 4,0	pH 5,0	pH 6,0	pH 7,0	pH 8,0
BSA	2	67,000	0.86	1.03	-0.14	-1.16	-1.68	-2.05
Ovalbumin	2	43,800	0.54	1.40	-0.76	-1.65	-2.20	-2.36
SBTI	2	24,500	0.90	1.22	-0.76	-1.54	-2.17	-2.13
Thaumatin	2	22,200	0.89	1.94	1.90	1.98	1.87	0.91

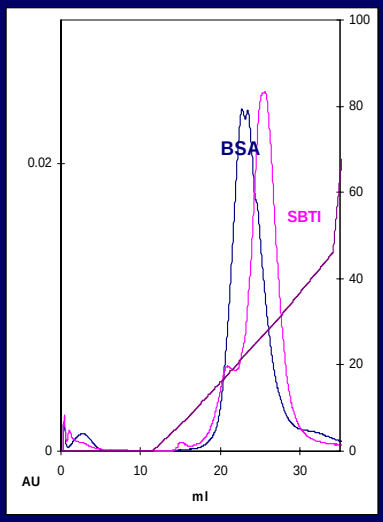
Sequence Suggested by SSC Criterion



First step: CationExchange Chromatography at pH 6.0

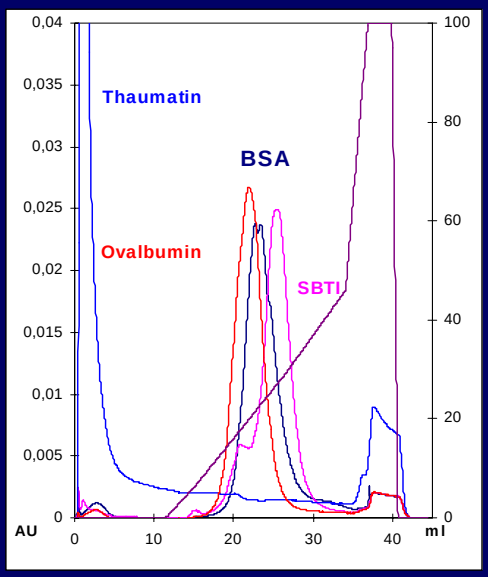


Second step : Hydrophobic Interaction Chromatography

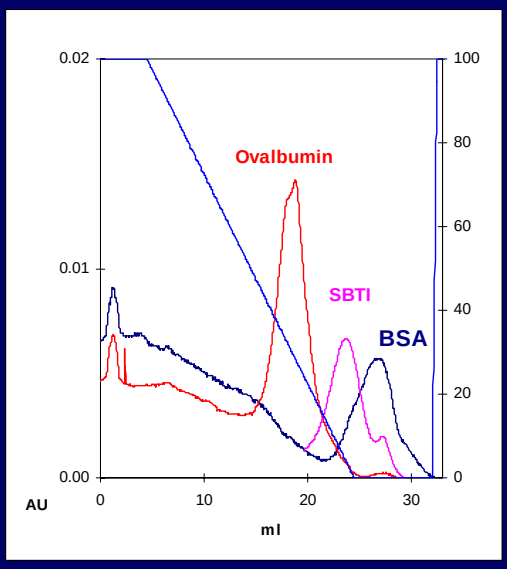


Third step : Anion Exchange Chromatography at pH 7.0

Sequence Suggested by Purity Criterion



First step : Anion Exchange Chromatography at pH 7.0



Second step : Hydrophobic Interaction Chromatography

Sequence Suggested by Expert System to Obtain a Purity Superior to 94% in the Purification

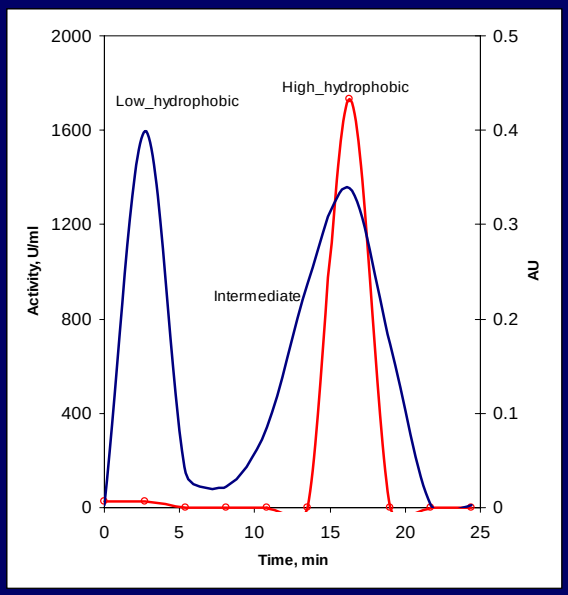
SSC Criterion Chromatography steps	Purity	Purity Criterion Chromatography steps	Purity
Cation Exchange at pH 6.0	33.1 %	Anion Exchange at pH 7.0	63.7 %
Hydrophobic Interaction	49.5 %	Hydrophobic Interaction	94.5 %
Anion Exchange at pH 7.0	97.0 %		

Purification of a recombinant beta-glucanase from a supernatant of *Bacillus subtilis*

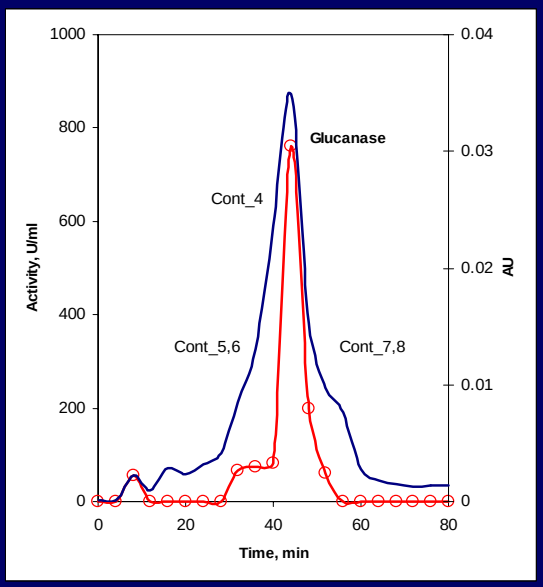
Physicochemical Properties and Concentration for the main proteins in *B.subtilis* ToC46 (pFF1) Culture

	Initial Concentration (mg/ ml)	Molecular weight (Da)	Hydrophobicity $[(NH_4)_2SO_4]$	Charge [Coulomb/ molecule] 10^{-25}				
				pH 4,0	pH 5,0	pH 6,0	pH 7,0	pH 8,0
-1,3-glucanase	0.60	31000	0.00	1.46	-0.62	-1.02	-2.33	-2.52
Contaminants								
Low hydrophobic								
Contaminant_1	2.74	41000	1.50		0.26	-0.87	-1.65	-2.04
Contaminant_2	2.74	32900	1.50		0.00	-2.70	-3.51	-3.51
Medium hydrophobic								
Contaminant_3	0.25	35500	0.20		-0.55	-0.22	-0.73	-1.82
High Hydrophobic								
Contaminant_4	0.42	62500	0.00		-1.06	-1.17	-2.79	-3.32
Contaminant_5	0.25	40600	0.00		-0.55	-0.22	-0.73	-1.82
Contaminant_6	0.25	69600	0.00		-0.55	-0.22	-0.73	-1.82
Contaminant_7	0.09	40600	0.00		1.46	-0.47	-1.06	-1.04
Contaminant_8	0.09	69600	0.00		1.46	-0.47	-1.06	-1.04

Sequence suggested for purifying glucanase



First step : Hydrophobic InteractionChromatography

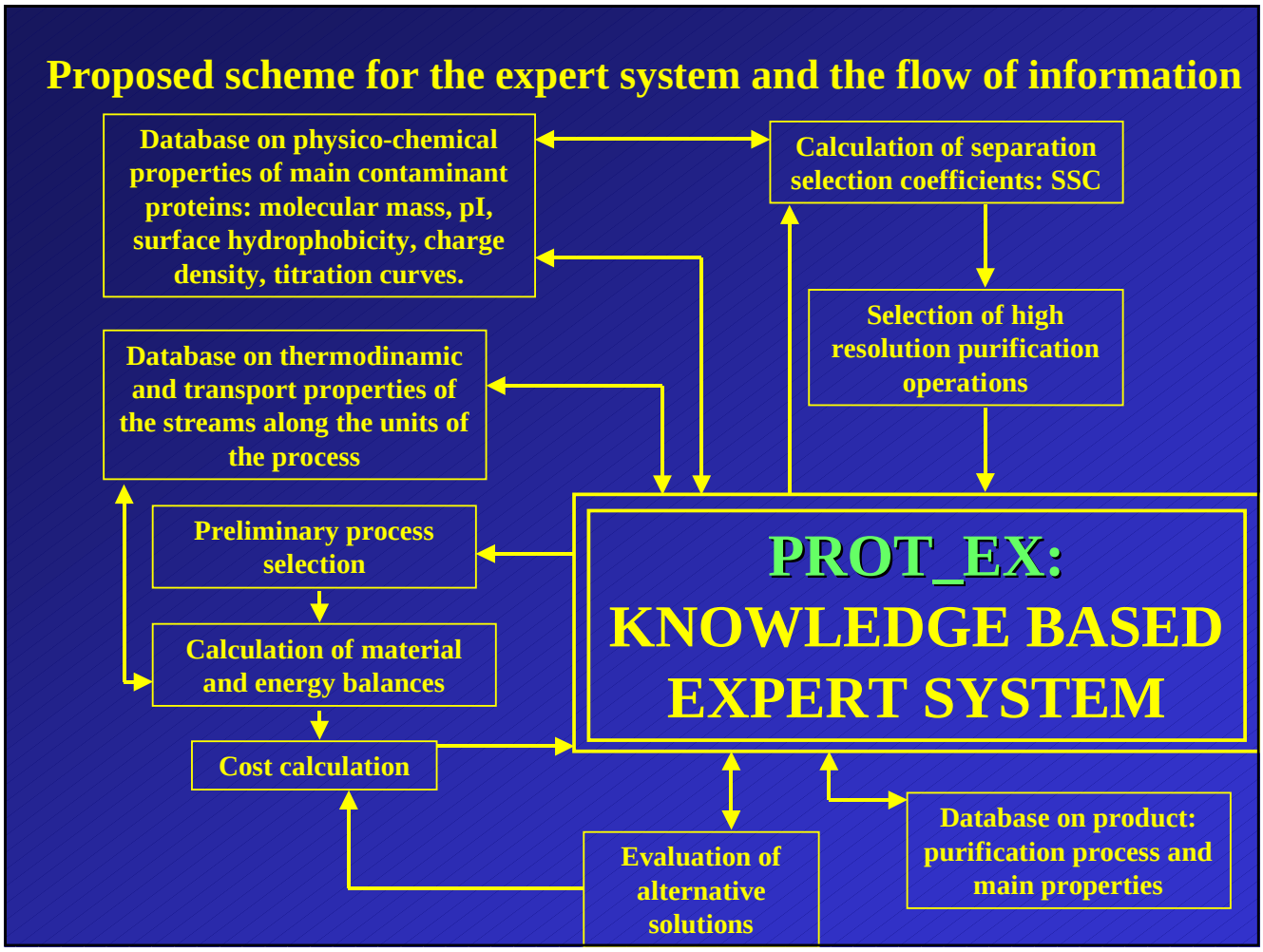


Second step :AnionExchange Chromatography at pH 6.5

Sequence Suggested by Expert System for SSC Criterion and for Purity Criterion

SSC and Purity Criterion Chromatography steps	Purity	Experimental Validation Chromatography steps	Purity
Hydrophobic Interaction	32.7 %	Hydrophobic Interaction	33 - 38 %
Anion Exchange at pH 6.5	70.3 %	Anion Exchange at pH 6.5	65 - 70 %

Next step


Proposed scheme for the expert system and the flow of information
Database on physico-chemical properties of main contaminant proteins: molecular mass, pI, surface hydrophobicity, charge density, titration curves.
Calculation of separation selection coefficients: SSC
Selection of high resolution purification operations
Database on thermodinamic and transport properties of the streams along the units of the process
Preliminary process selection
Calculation of material and energy balances
Cost calculation
PROT_EX:
KNOWLEDGE BASED
EXPERT SYSTEM
Evaluation of alternative solutions
Database on product: purification process and main properties

Reference

- J.A Asenjo, L. Herrera and Byrne, “Development of an expert system for selection and synthesis of protein purification processes “ J. Biotechnol., 11,275-298 (1989).
- J.A. Asenjo, B. A. Andrews “Is there a rational method to purify proteins? From expert systems to proteomics” J. Mol. Recognit. 2004; 17: 236–247
- E.W Leser,. M.E.Lienqueo, J.A Asenjo, "Implementation in an Expert System of Selection Rationale for Purification Processes for Recombinant Proteins.", Ann. N.Y.Acad of Sci., 782: 441-455, 1996.
- E.W Leser,. J.A Asenjo “The rational selection of purification processes for proteins: an expert system for downstream processing design” Ann N Y Acad Sci. 1994 May 2;721:337-47.
- M.E. Lienqueo, J.A. Asenjo “Use of expert systems for the synthesis of downstream protein processes” Computers & Chemical Engineering ,24: 2339 – 2350, 2000.
- M.E Lienqueo,J.C. Salgado, J.A. Asenjo “An Expert System for selection of protein purification processes: Experimental Validation", J.Chem.Technol Biotechnol, 74: 293-299, 1999.