ARTIFICIAL INTELLIGENCE IN CHEMISTRY & CHEMOMETRICS
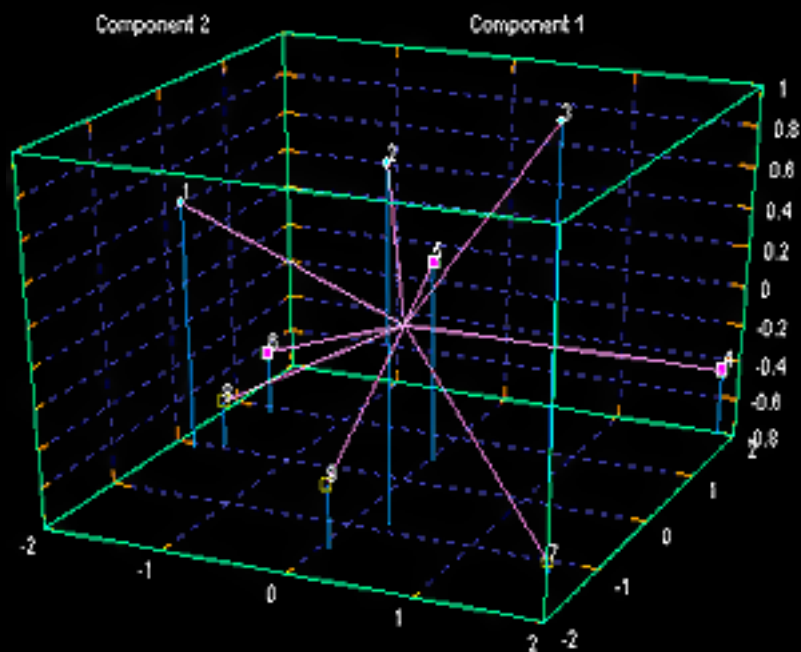
Patrycja Ciosek

CHEMOMETRICS

- Part of chemistry

- Extraction of useful information from multidimentional data

- MATHEMATICS + STATISTICS + ARTIFICIAL INTELLIGENCE

1971, Svante Vold
1974, „part of chemistry"
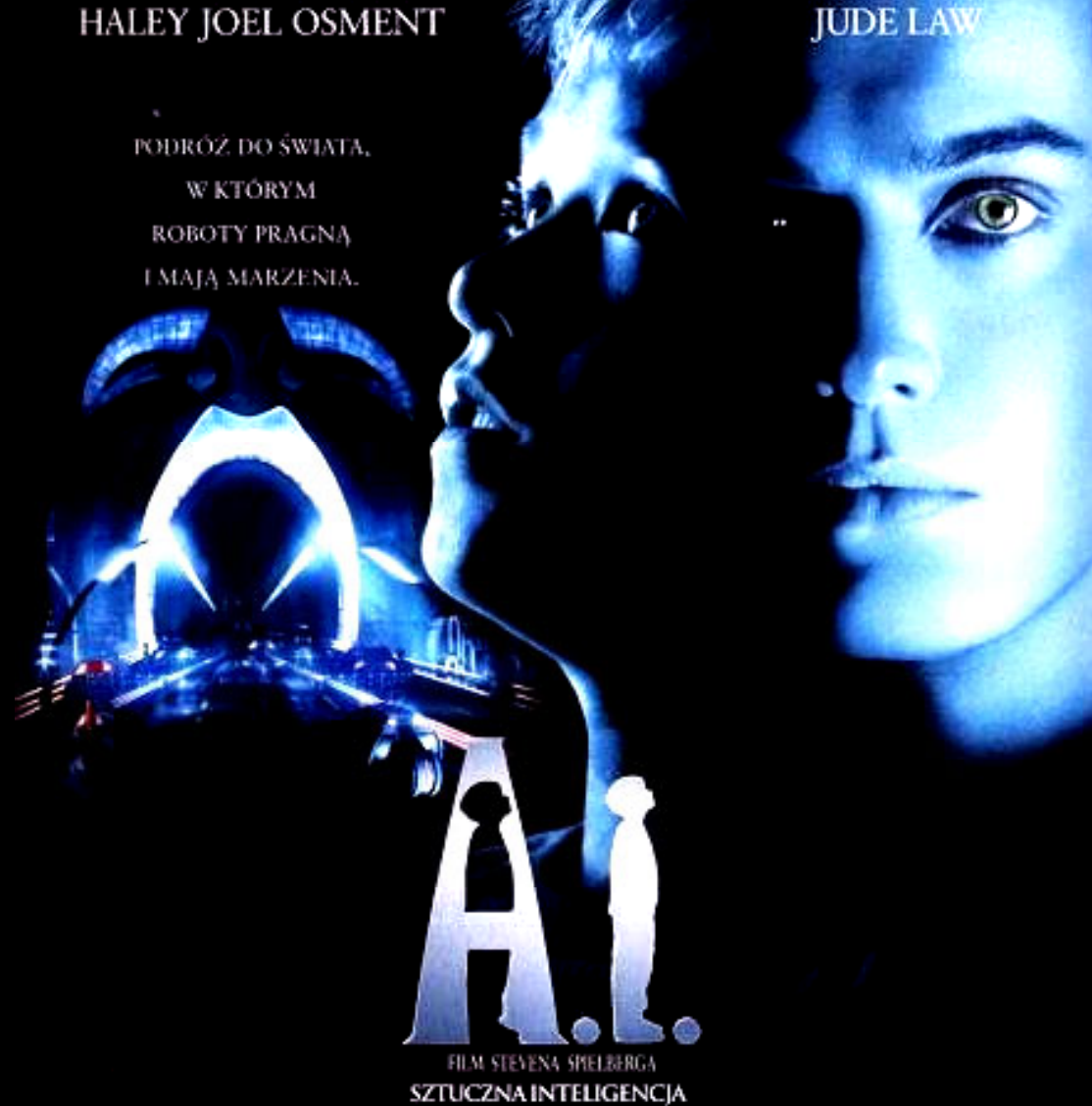
Patrycja Ciosek

CSRG

CHEMOMETRICS

- Design of experiments (DOE)
- Simplex optimisation
- Factor design
- Cluster analysis
- Relation modelling
- Classifiers
- PCA
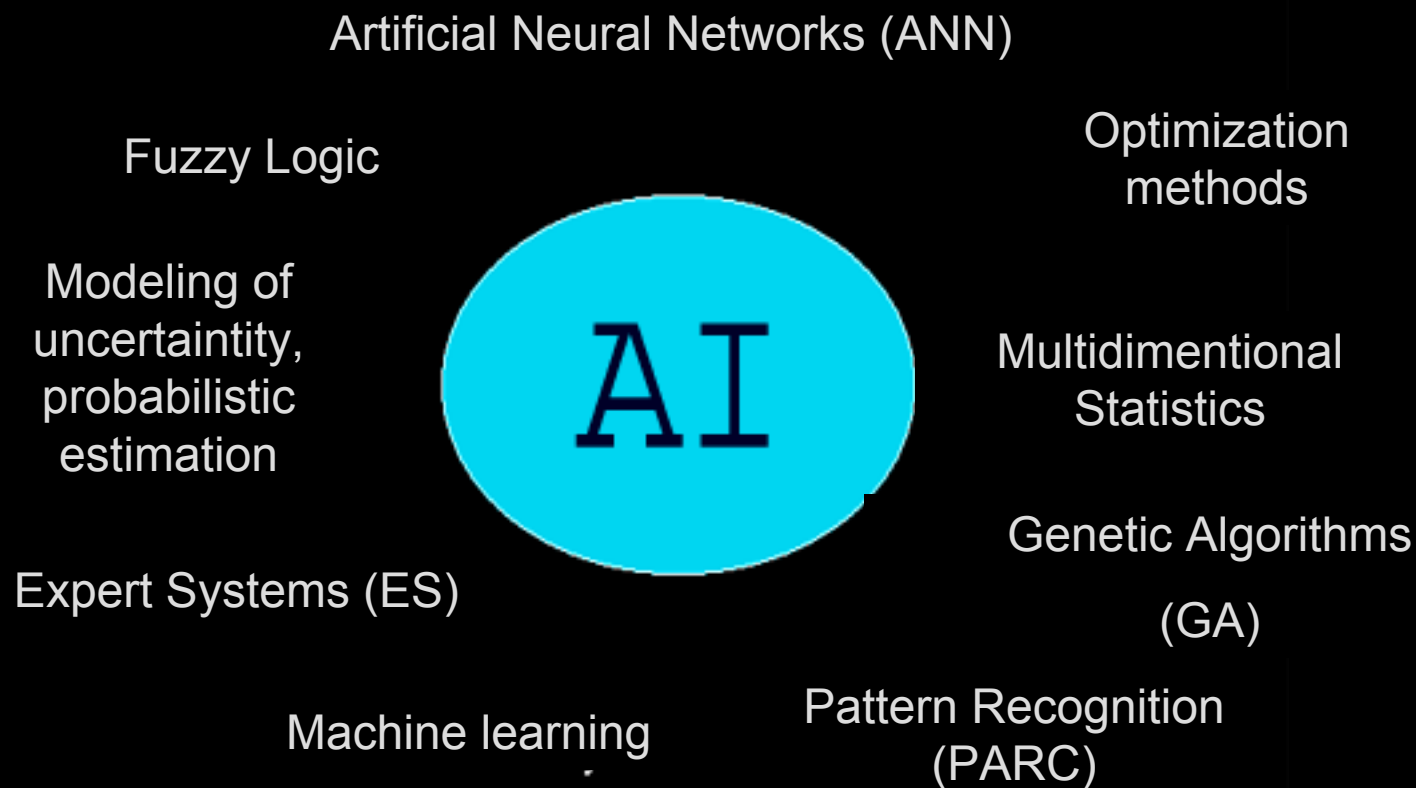- Pattern recognition

Patrycja Ciosek

Artificial Intelligence (AI)

- technology and methods inspired by **informatics** and **psychology**

- Construction of machines, which way of acting can be considered as „human" (caused by „human" intelligence)

HALEY JOEL OSMENT   JUDE LAW

PODRÓŻ DO ŚWIATA,
W KTÓRYM
ROBOTY PRAGNĄ
I MAJĄ MARZENIA.

A.I.

FILM STEVENA SPIELBERGA
SZTUCZNA INTELIGENCJA

1955, John McCarthy

Artificial Intelligence

Artificial Neural Networks (ANN)

Fuzzy Logic

Optimization methods

Modeling of uncertaintity, probabilistic estimation

AI

Multidimentional Statistics

Expert Systems (ES)

Genetic Algorithms

(GA)

Machine learning

Pattern Recognition (PARC)

Artificial Intelligence – how far are we?

The progress is very hard and slow...

- Fuzzy logic – process control in industry
- Expert Systems – pharmacy and medicine
- Machine translation
- Neural Networks
- Optical recognition, speech recognition, hand writing recognition
- Deep Blue won with Gary Kasparov
- Economics – automatic systems able to estimate credit capability
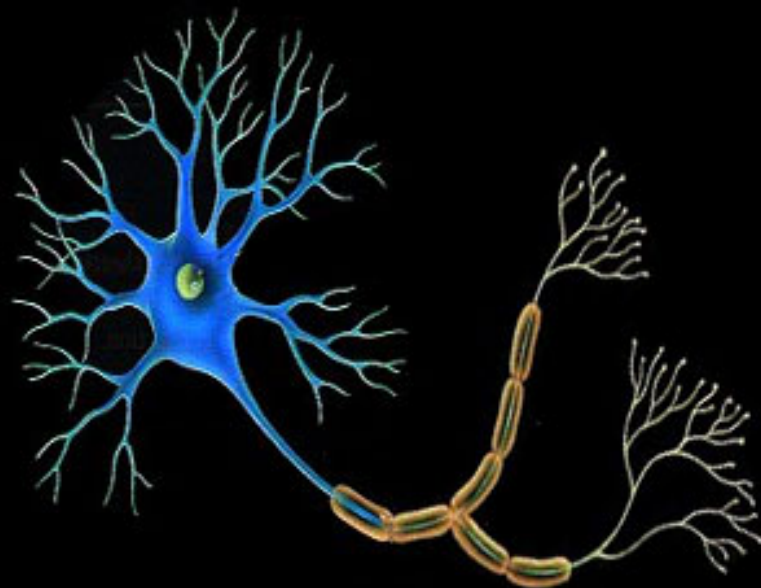
....but:

- Human conversation
- Generating profits on stock exchange
- Proper translation of literature and common parlance

Patrycja Ciosek

CSRG

Artificial Intelligence in chemistry

The main goals of AI in chemistry:

Optimization and approximation

- *Neural Networks*
- *Genetic Algorithms*
- *Expert Systems*

Patrycja Ciosek

Artificial Neural Networks (ANN)

- Adaptative structure
- Able to model complicated relations In-Out
- Generalization of obtained knowledge
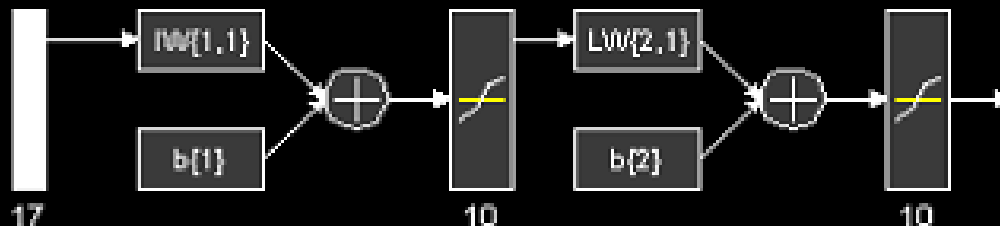- Proper processing of incomplete data
- Parallel computing

40' - Warren McCulloch,

Walter Pitts

CSRG

Patrycja Ciosek

ANN in chemistry

- Calibration of devices, sensors
- Development of new measurements methods
- Dynamic process on-line monitoring
- Signal processing
- „Shape classification"
- QSAR, QSPR

Patrycja Ciosek

CSRG

ANN in chemistry – QSAR, QSPR

↗ QSPR = Quantitative Structure-Property Relationship

↗ QSAR = Quantitative Structure-Activity Relationship
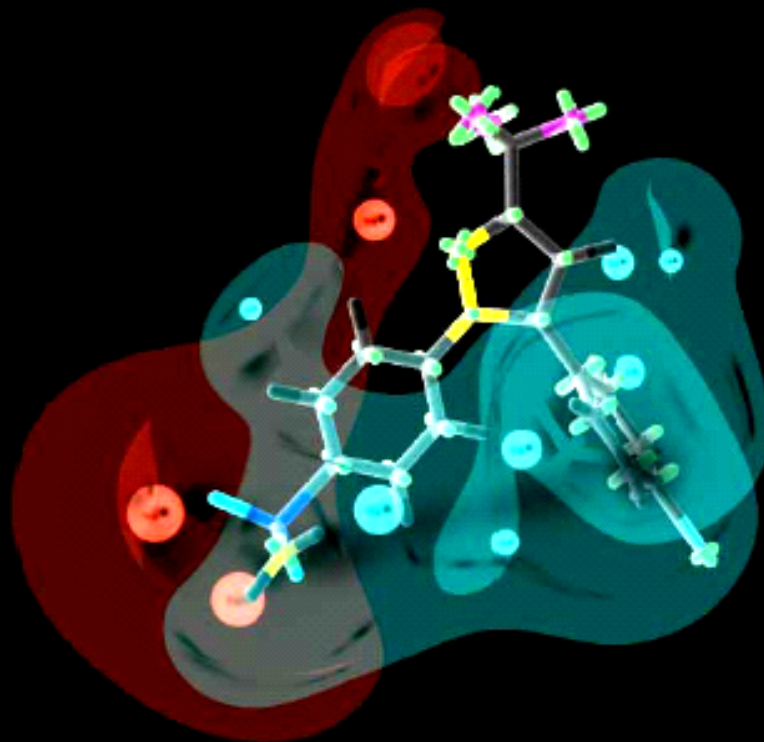
Genetic Algorithms (GA)

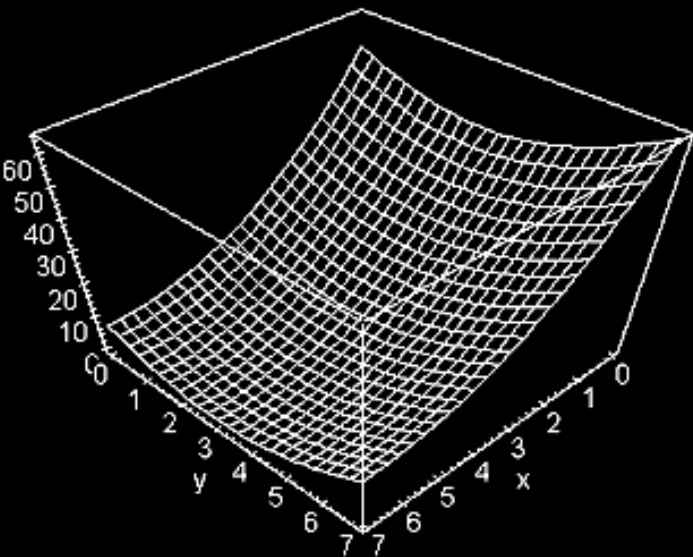↗ Search the space of alternative solutions in order to find the best one

->OPTIMIZATION

↗ The procedure emulates **biological evolution**
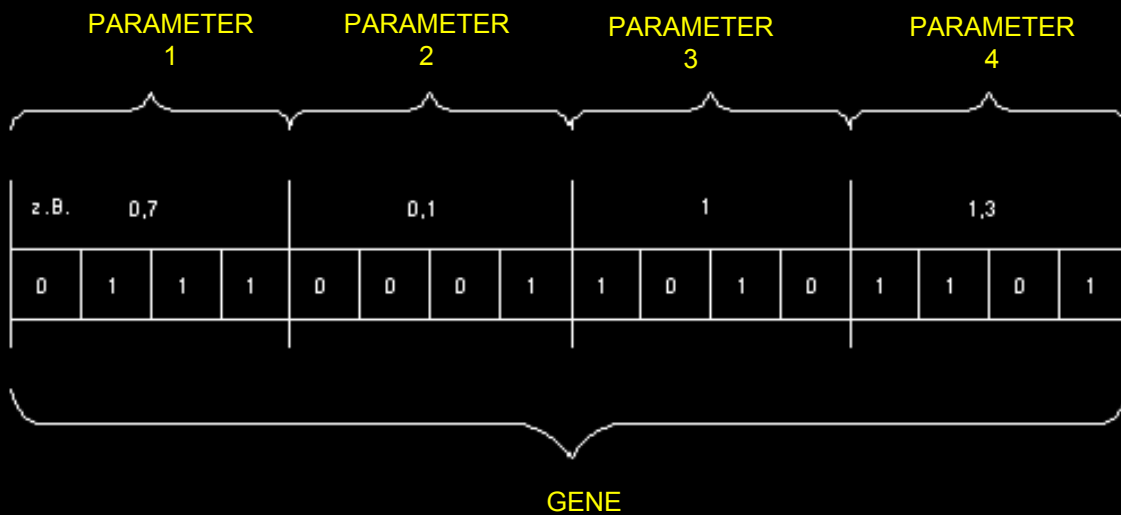
John Henry Holland

Genetic Algorithms

Natural	GA
chromosome	string
gene	feature, character or detector
allele	feature value
locus	string position
genotype	structure, or population
phenotype	parameter set, alternative solution, a decoded structure

Patrycja Ciosek

CSRG

The Coding

No.	STRING	
1	100001	x=4;y=1
2	001100	X=1;y=4
3	110010	X=6;y=2
4	000100	X=0;y=4

PARAMETER 1   PARAMETER 2   PARAMETER 3   PARAMETER 4

z.B.	0,7				0,1				1				1,3		
0	1	1	1	0	0	0	1	1	0	1	0	1	1	0	1

GENE

Binary Representation
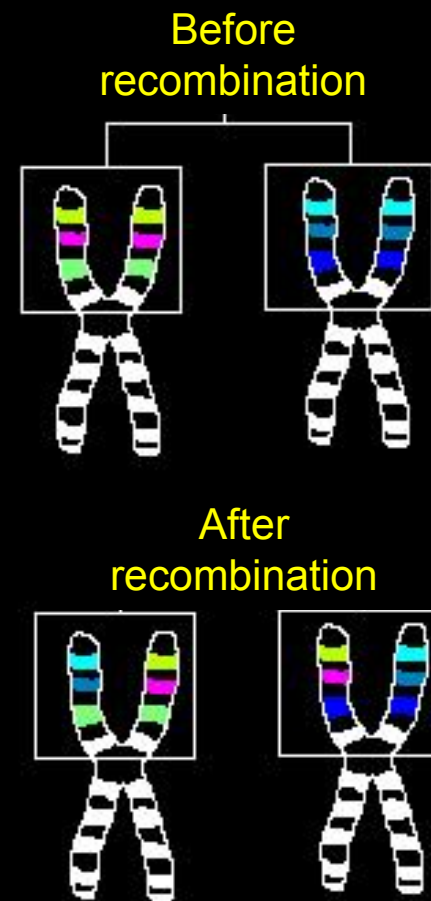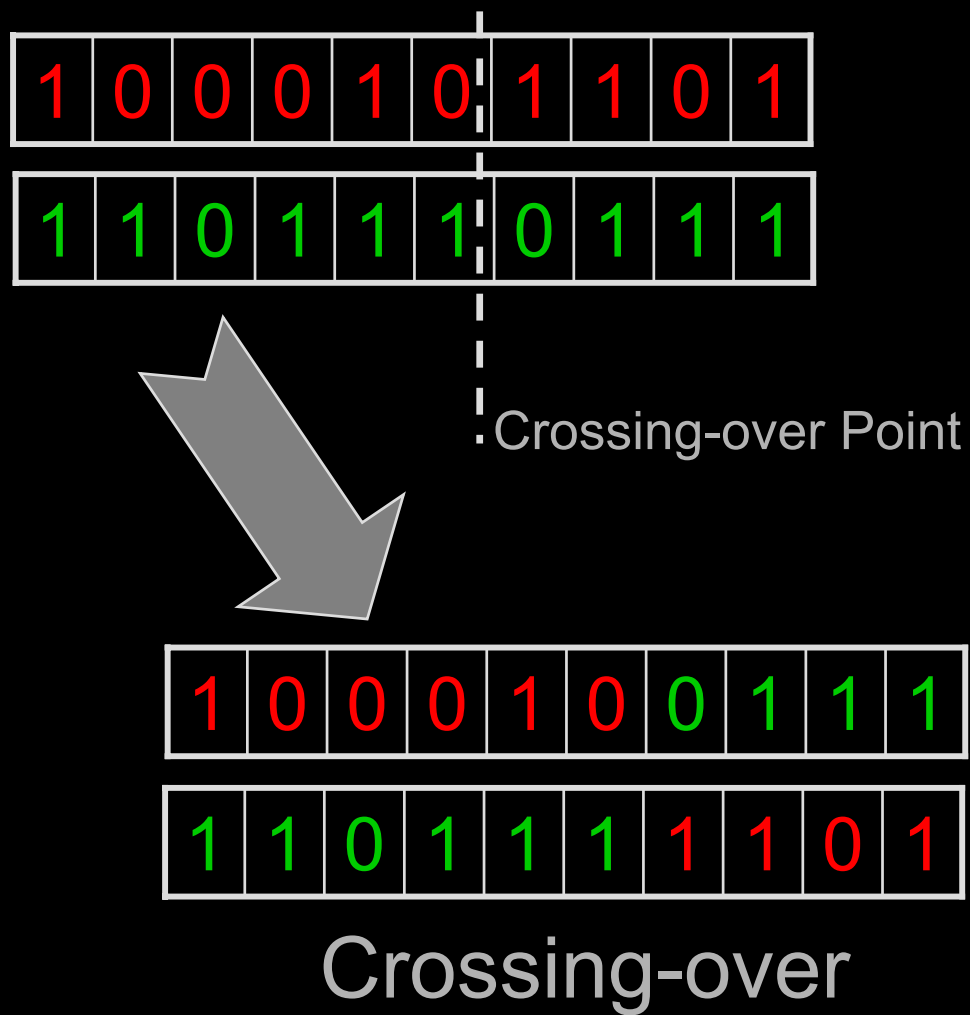(the most frequently used)

Selection

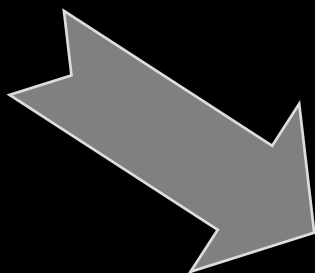**The main tool of evolution ;-)**

- Selection of individuals according to strict, determined criterion

- Criterion = **evaluation function** determined by the user
  - High value -> reproduction of the individual
  - Low value -> the individual dies

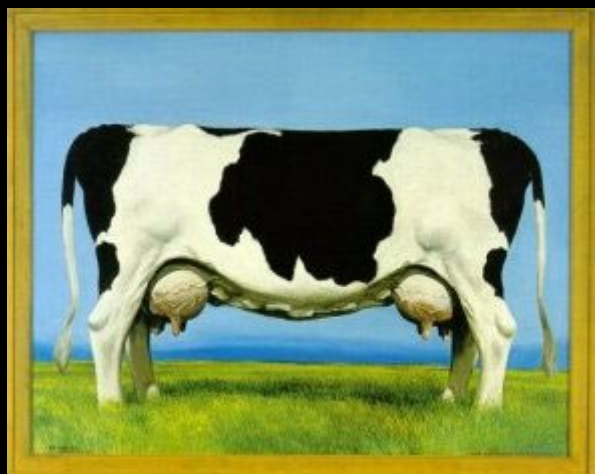- Following generations are **more and more adapted**

Patrycja Ciosek

Genetic Operators

1	0	0	0	1	0	1	1	0	1

1	1	0	1	1	1	0	1	1	1

Crossing-over Point

1	0	0	0	1	0	0	1	1	1

1	1	0	1	1	1	1	1	0	1

Crossing-over

Before recombination

After recombination

Patrycja Ciosek

CSRG

Genetic Operators

1	0	0	0	1	0	1	1	0	1

1	0	0	0	0	1	1	1	0	1

Mutation

Patrycja Ciosek

GA in Chemistry

- Curve fitting (IR spectra)

- feature selection – multicomponent calibration

- Determination of the configuration of some systems (for example C60)

- The composition of complex materials (for example composites)

- Molecular structure optimization

- Protein folding (3D structure of proteins)

- Protein-ligand docking
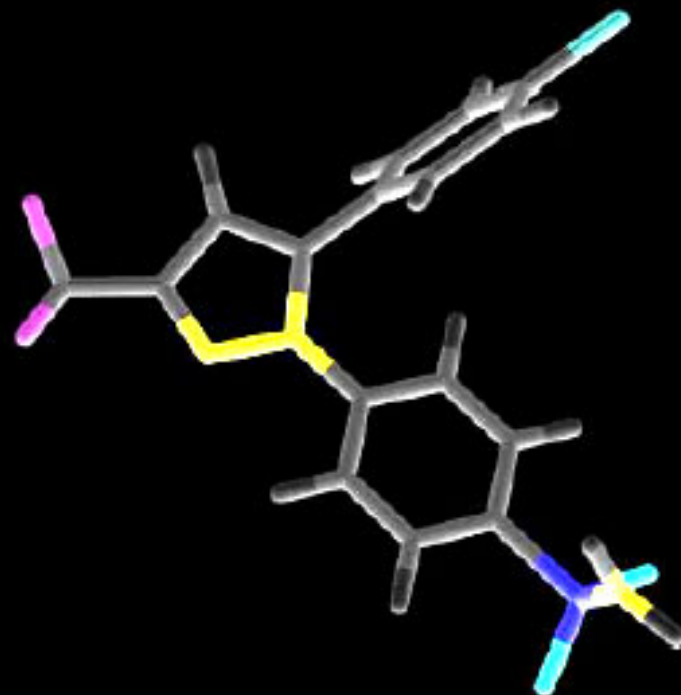
Expert Systems (ES)

↗ Program / set of programs

↗ It aims to recall the use of knowledge and making decisions

Why?

- Costs
- No experts in many cases
- Work more fast
- Not get tired
- Consequent
- Objective
- Precise
- The analysis of huge amounts of data demands the use of a computer

Patrycja Ciosek

CSRG

Expert System - a scheme

Patrycja Ciosek
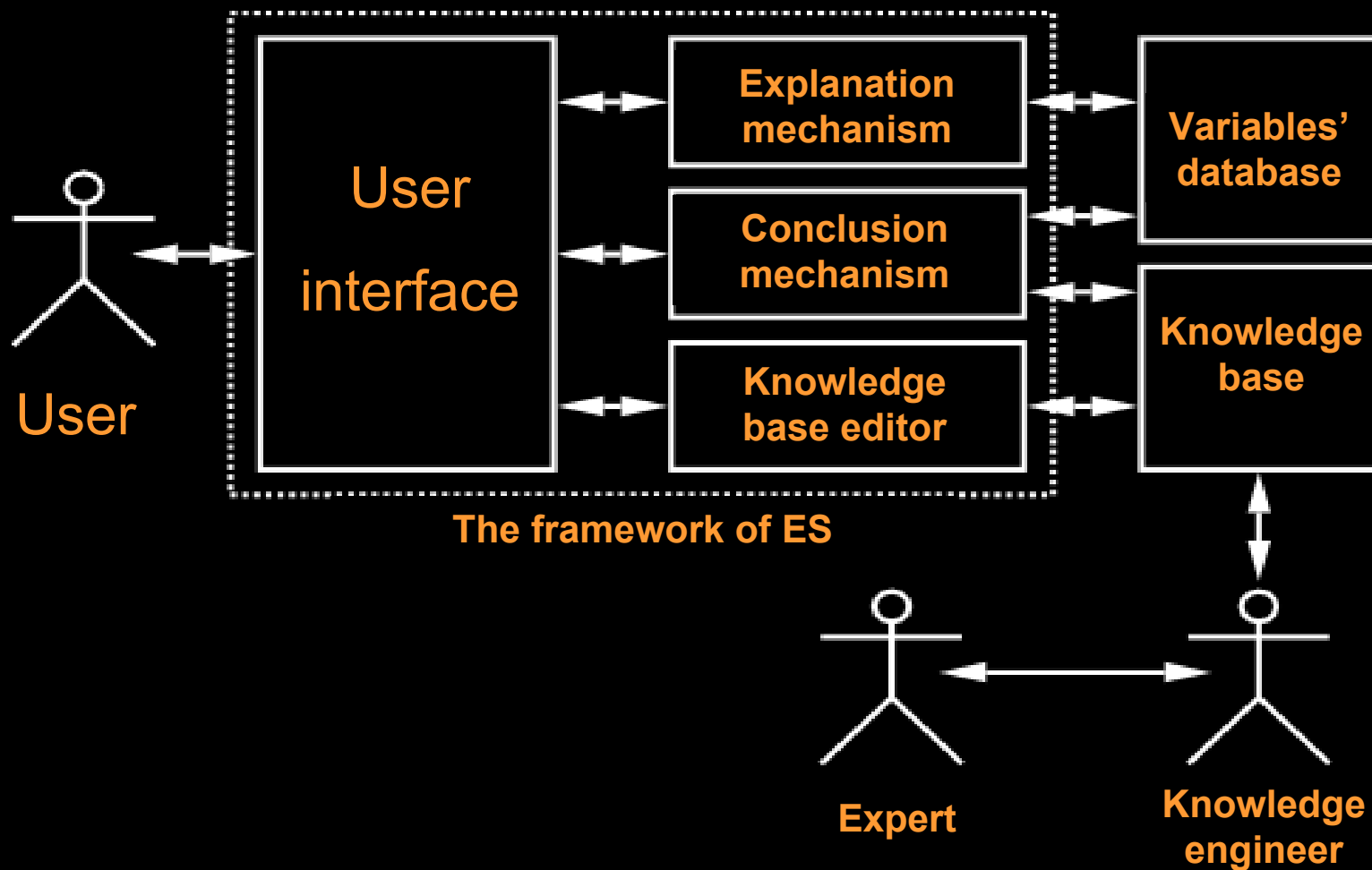
ES in Chemistry - DENDRAL

- Generation of chemical structures according to data obtained from MS, NMR, IR, UV
- From 1969 used in chemistry, many problems solved:
    - structure of organic esters,
    - hormones,
    - antibiotics,
    - impurities in chemical substances.

The results for mixtures are better than those obtained by experts!
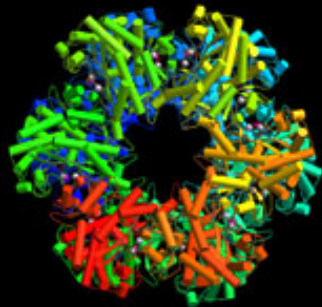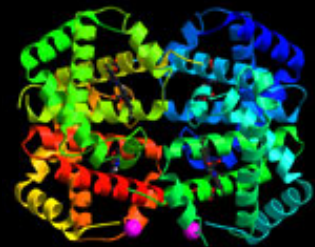
Patrycja Ciosek

CSRG

ES in Chemistry - CRYSTALIS
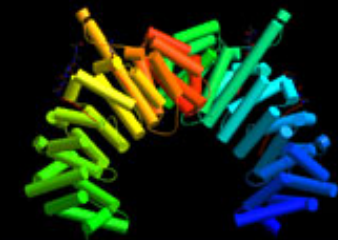
- Program, which elaborates crystallographic data of proteins

  (costs of classical analysis – even 1.000.000 $).

- Interpretation of electron density maps and other data

ES in Chemistry - SYNTHESIS

- **CASD** - Computer Aided Synthesis Design

- 20 atoms - 10^{18} **various molecules!**

- Number of basic substrates in synthesis ~500, in industry even more (~2000).

- Database – about 500 reactions, more than 100.000 concrete ones.

Patrycja Ciosek

Summary

- Chemometrcs & Chemistry <-> AI
- Analysis of multidimensional, complex data
- Optimization, approximation
- ANN, GA, ES

MORE ➡

- Chemometrics and Intelligent Laboratory Systems
- Journal of Chemometrics
- Environmetrics
- Analytical Chemistry
- Analytical Letters
- Analytica Chimica Acta