HTS technology paradigm

CADD

- Huge and growing number of biological targets emerging from molecular biology and genoma sequencing
- Advances in protein chemistry
- Advances in instrumentation technology (liquid handling eqiuipment, detection instruments)

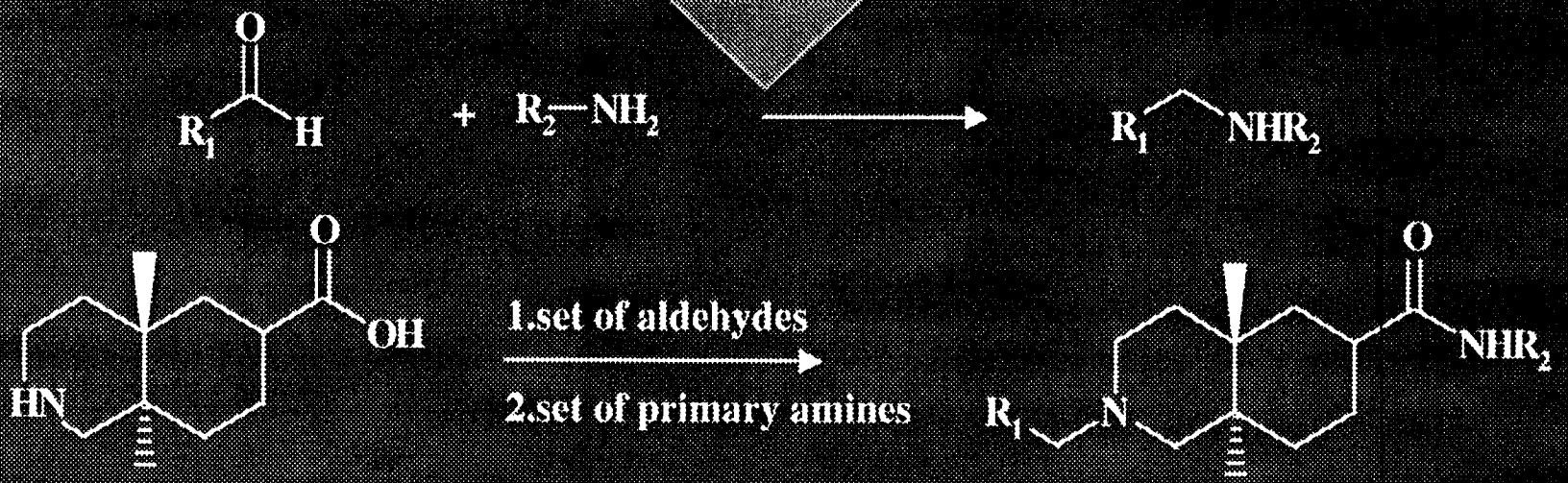
- Chemical libraries technology (Combinatorial chemistry)

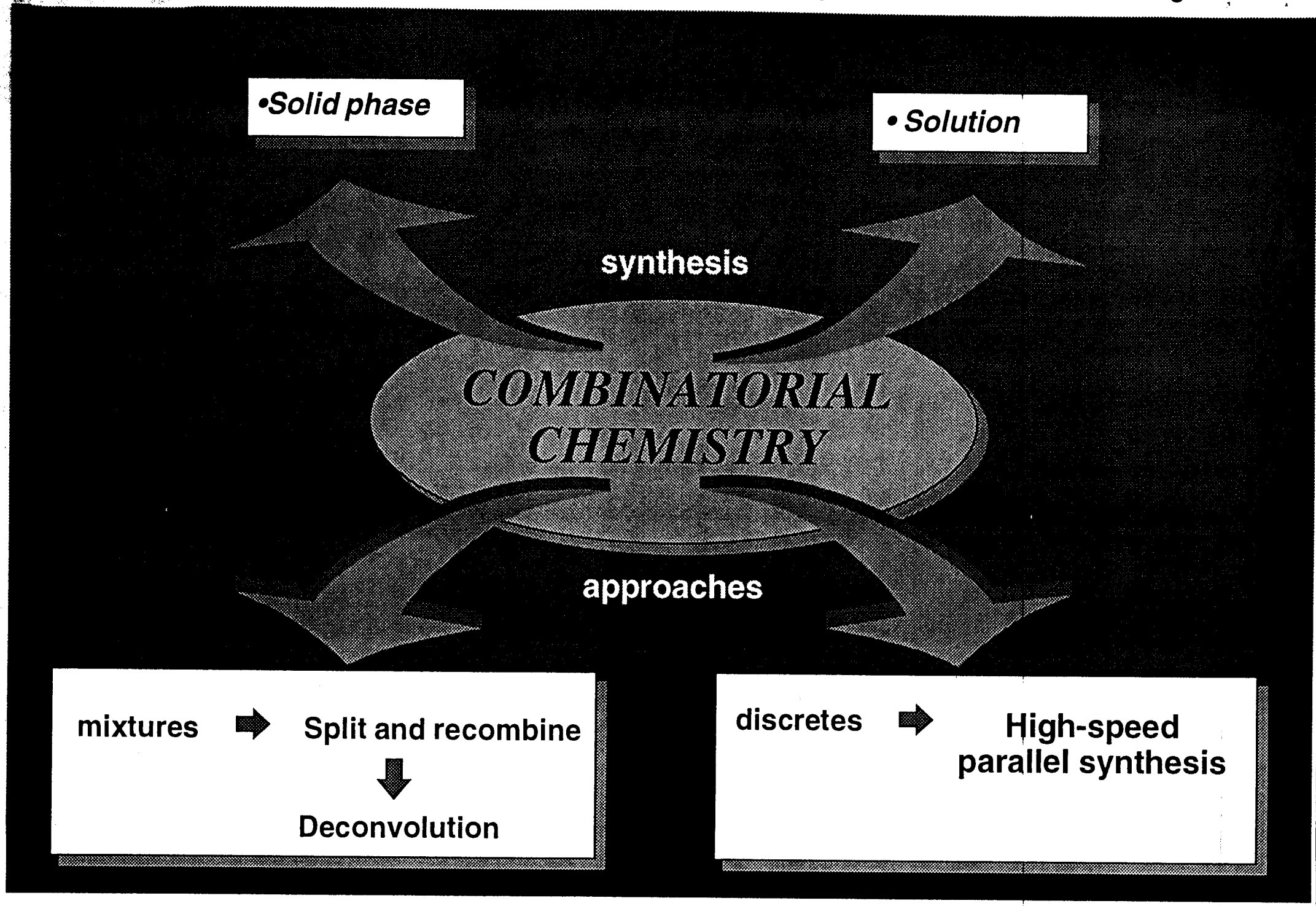
COMBINATORIAL CHEMISTRY

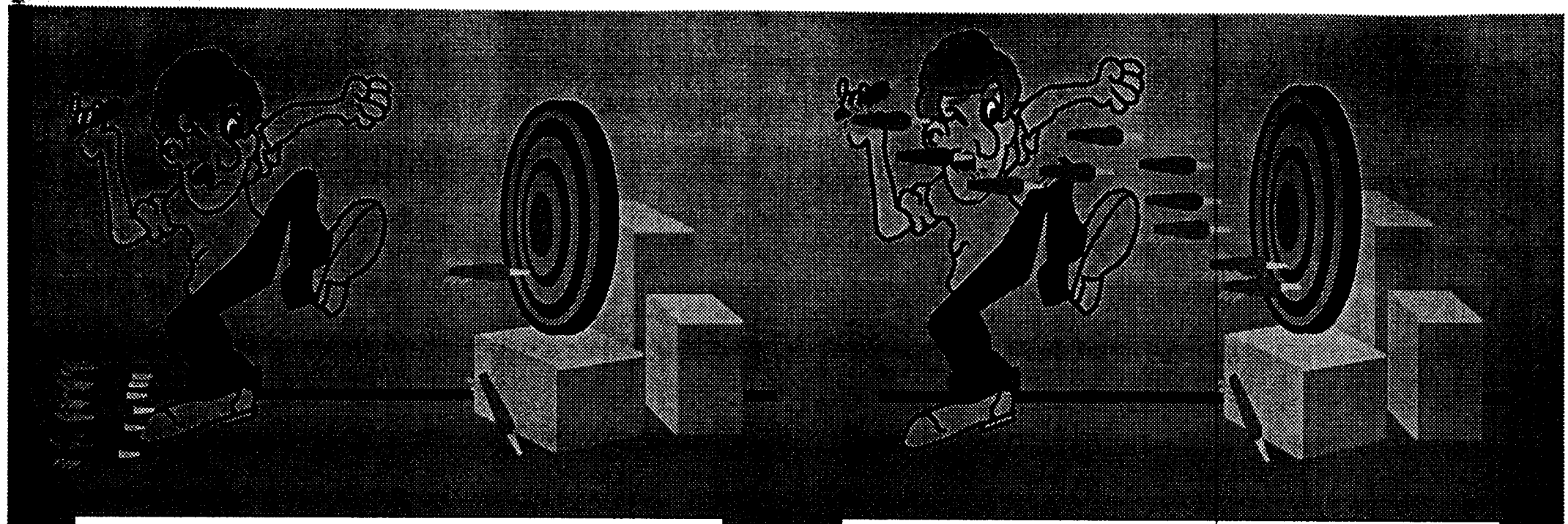
A synthetic strategy capable of producing large chemical libraries by covalently combining a basis set of modular components

what is a chemical library?

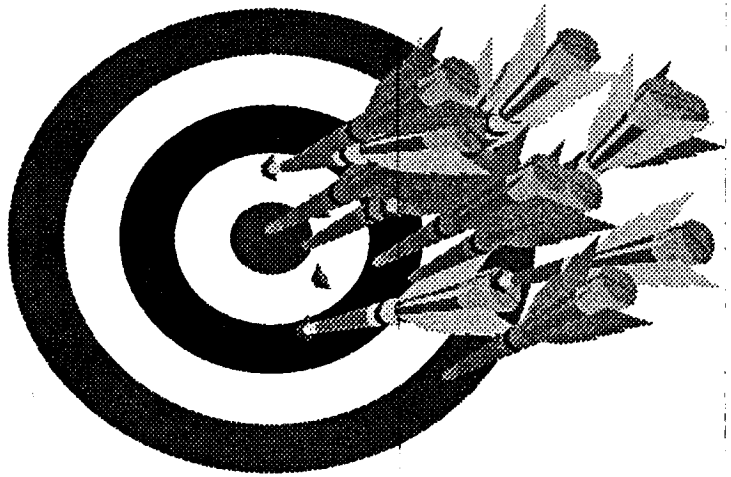
A large collection of compounds produced by reacting sets of monomers with one another sequentially, or with a template (scaffold)




•Solid phase
• Solution
synthesis
COMBINATORIAL CHEMISTRY
approaches
mixtures
Split and recombine
Deconvolution
discretes
High-speed parallel synthesis



traditional approach



combinatorial chemistry approach

COMBINATORIAL CHEMISTRY

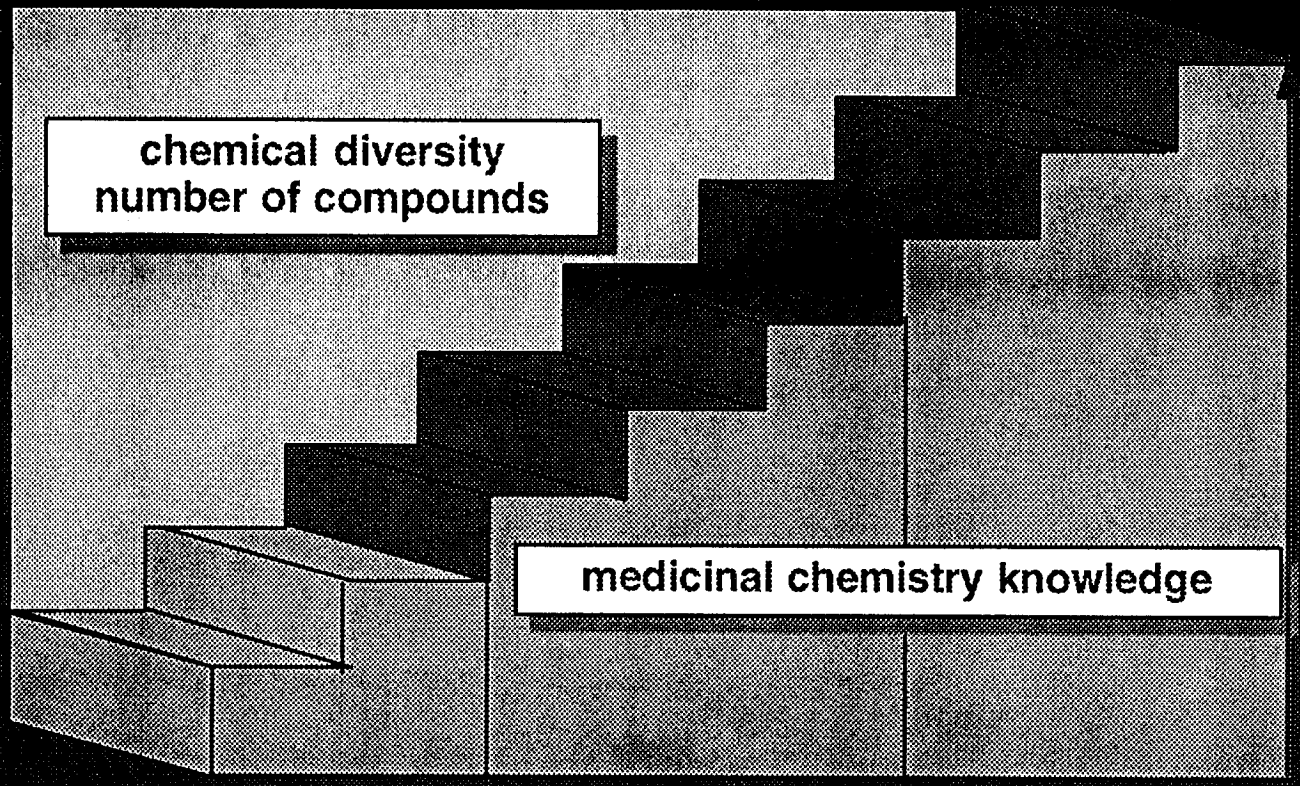
...the ability to play with a large number of molecules...

Mario Geysen

...a collection of strategies, technologies and instrumentation blended to suit specific objectives

Eric Gordon

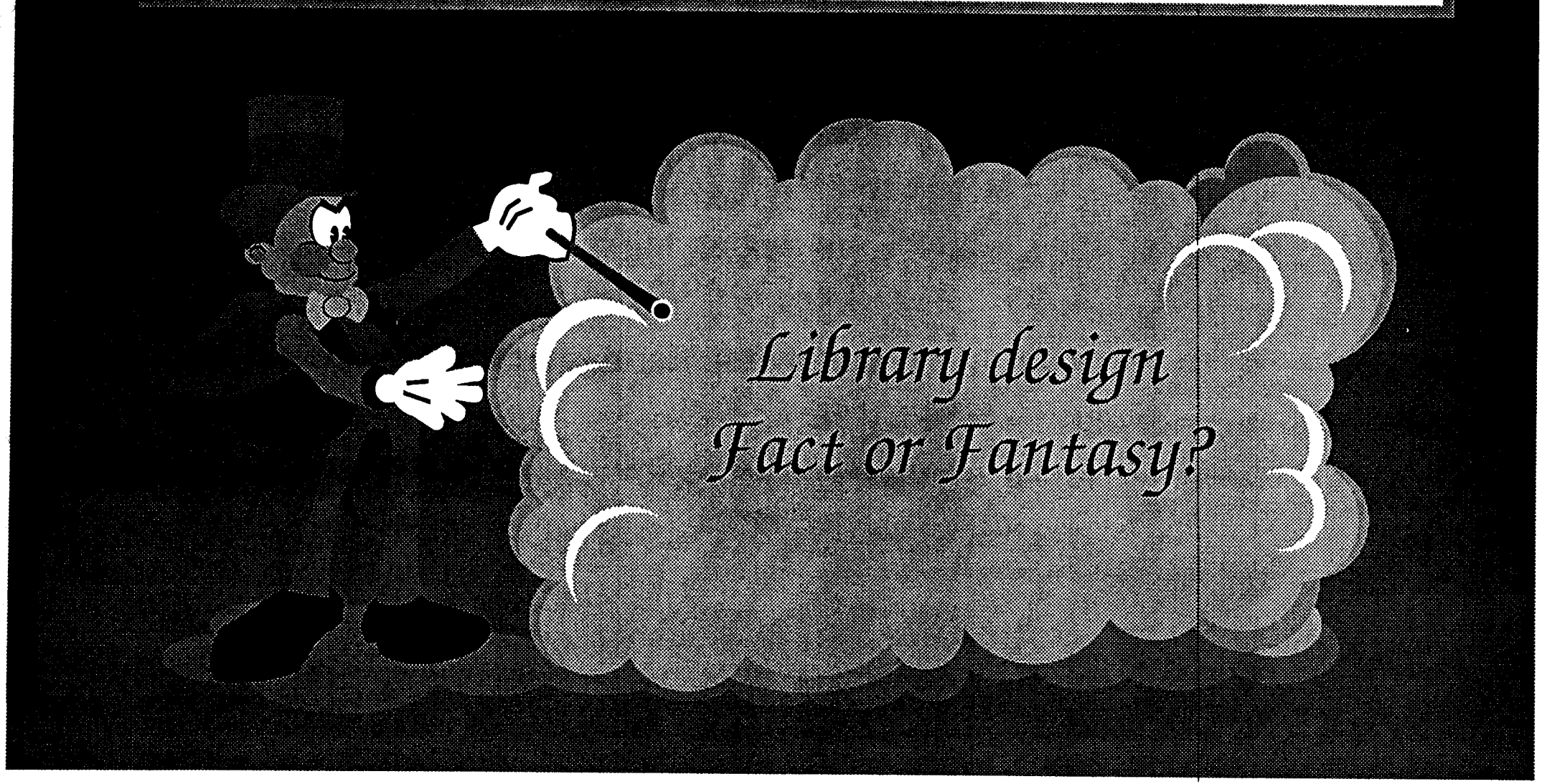
TYPES OF LIBRARIES



HIT GENERATION	FROM HIT TO LEAD	LEAD PROGRESSION
primary libraries	semi-focused	focused libraries

- knowledge based (thematic, biased) libraries
- pharmacophore based
- structure based

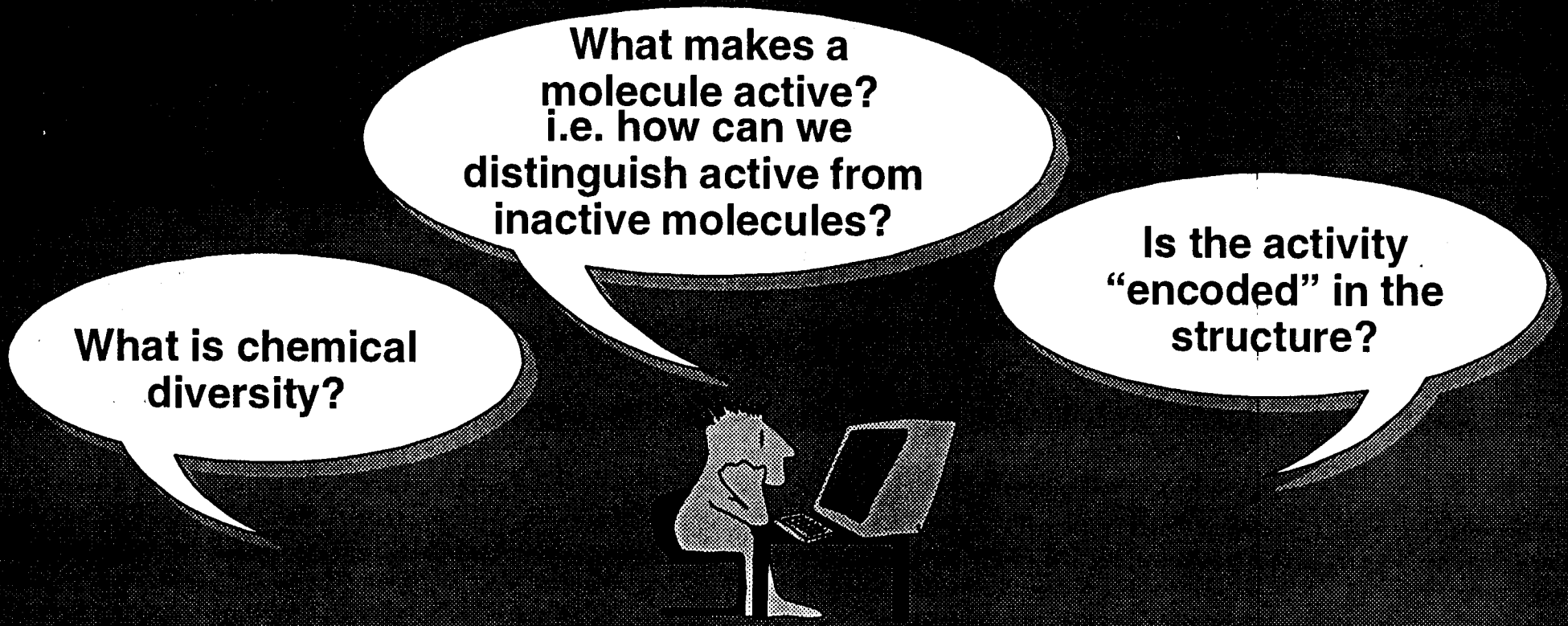
Are computational chemistry tools of any help in **Library Design ?**

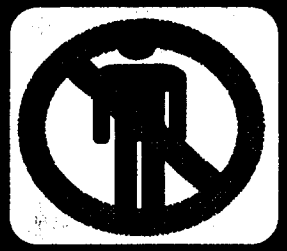


WHY LIBRARY DESIGN ?

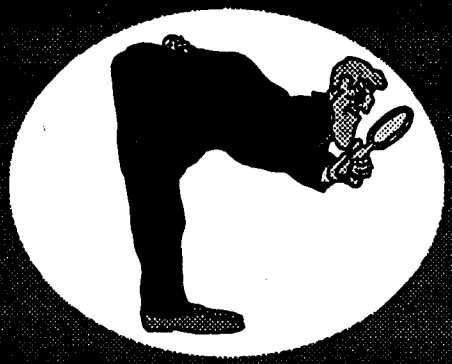
- To reduce the number of theoretically possible molecules to an appropriate set without decreasing the "Diversity" of the library
- To collect the maximum information with the minimum number of structures

More questions than answers





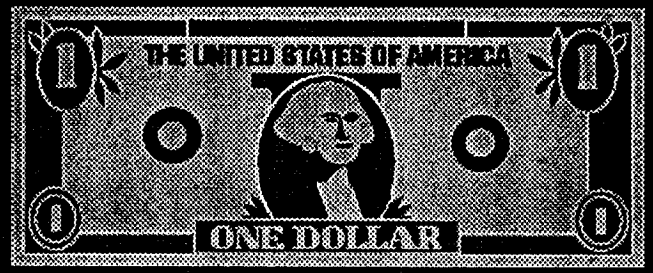
Only for people thinking that DIVERSITY is not necessarily coincident with QUANTITY



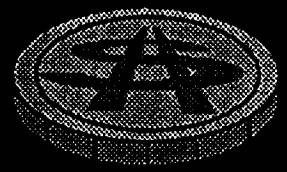
- Wrong question !
- How do we describe the objects ?

Diversity is a relative concept depending on how objects are defined

Diversity is a relative concept depending on how objects are defined



1$ note



1$ coin

Description	Diversity
Value	similar (equal)
Size, shape, weight, colour, ...	dissimilar

.....there exists a determinable relationship between the structure and biological activity...

•Molecular descriptors

what we want to describe :

- •the whole molecule
- •only relavant parts (pharmacophore)

descriptors must be :

- •meaningful
- •easily calculated

Two main categories according to the nature of space they define

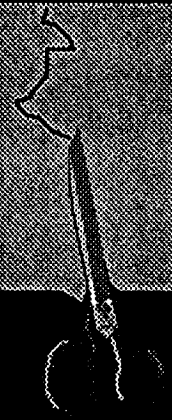
CONTINUOUS
(or metric)

parametric space

NON CONTINUOUS
(or non-metric)

non parametric space

•Molecular descriptors



CONTINUOUS (or metric)

•Lipophilicity : CLOGP,HINT, LOGKOW ...

•electronic :
- total E state sum
- total dipole
...

•steric :
- elipsoidal volume
- Molar refractivity: CMR
- VdW volume
- IX (smallest moment of inertia)
....

NON CONTINUOUS (or non-metric)

•Topological descriptors (based on the connection table)

-"Chi" and "K" indices encoding size, shape, flexibility, branching, arrangements of cycles *Molconn-X*

- MW

-count of things about each molecule (profiles) i.e.

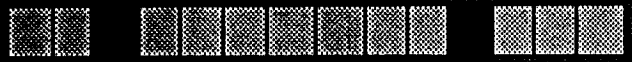
> functional groups
> substructures
> features
> distances
>....

•Molecular descriptors

NON CONTINUOUS
(or non-metric)

•2D screens : expressed as a bit string containing structural information about the presence of 2D features in the structure .

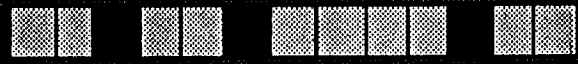
Structural keys
Maccs II



O C A A A
C A A A
O O A A

fragments

hashed fingerprints
Daylight, Unity



C=CN

0-bond path
1-bond path
2-bond path

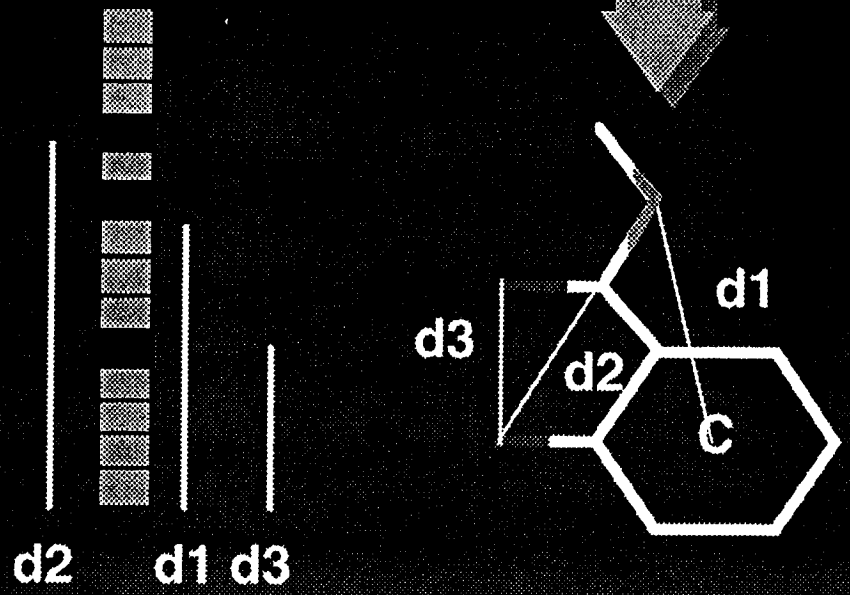
C N
C=C CN
C=CN

patterns

•Molecular descriptors

NON CONTINUOUS
(or non-metric)

•3D screens : expressed as a bit string encoding the spatial relationships (distances, angles etc.) between atoms, rings, centroids and planes.



3D-Unity

ChemDBS-3D

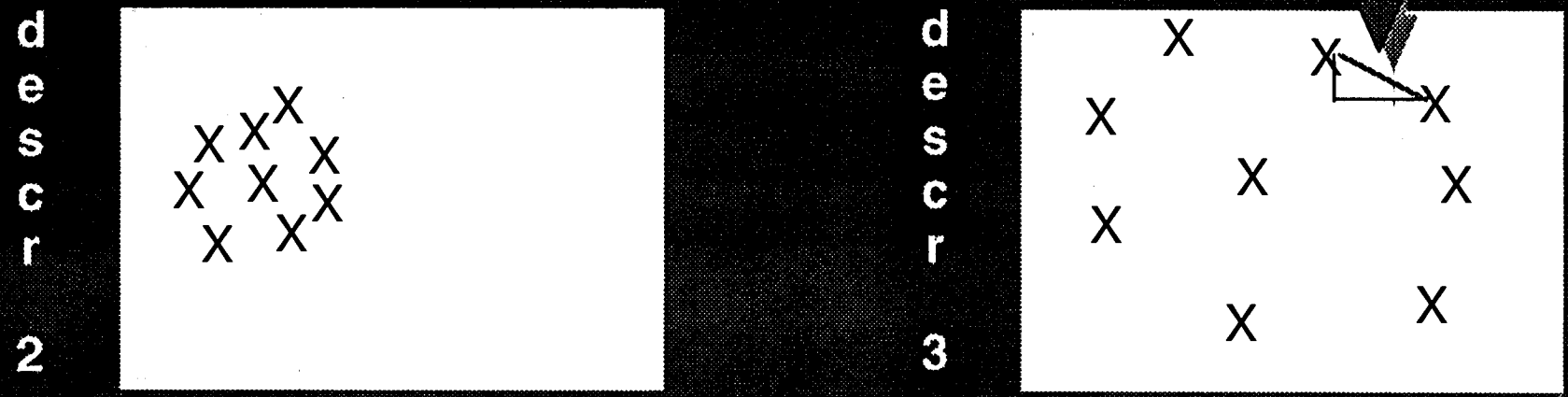
3DSEARCH

Diversity is a relative concept depending on how objects are defined

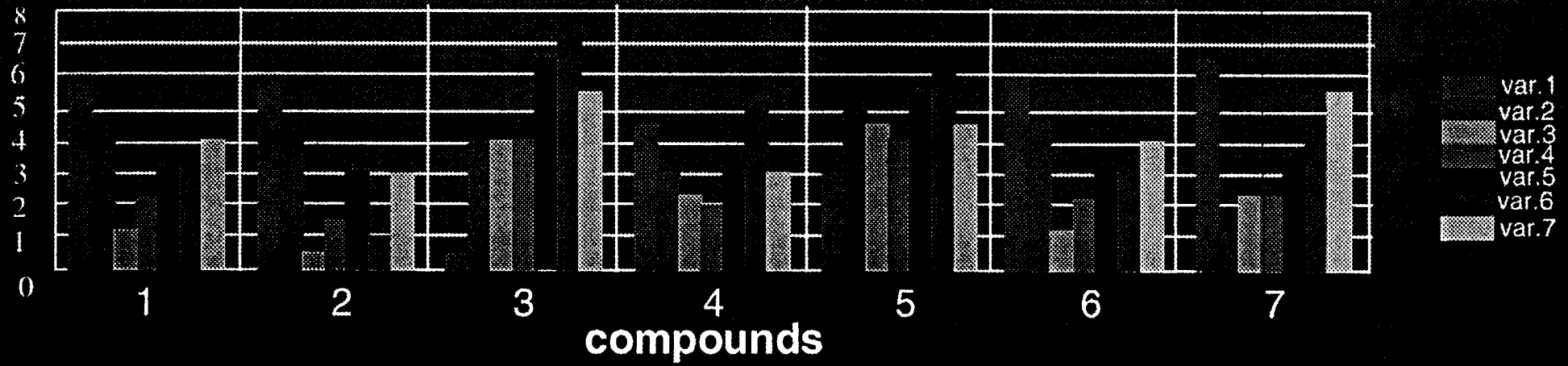
Assessing the proximity of objects

Euclidean distance

- parametric space (continuous descriptors)



- feature counts profile (topological descriptors)



Diversity is a relative concept depending on how objects are defined

Assessing the proximity of objects

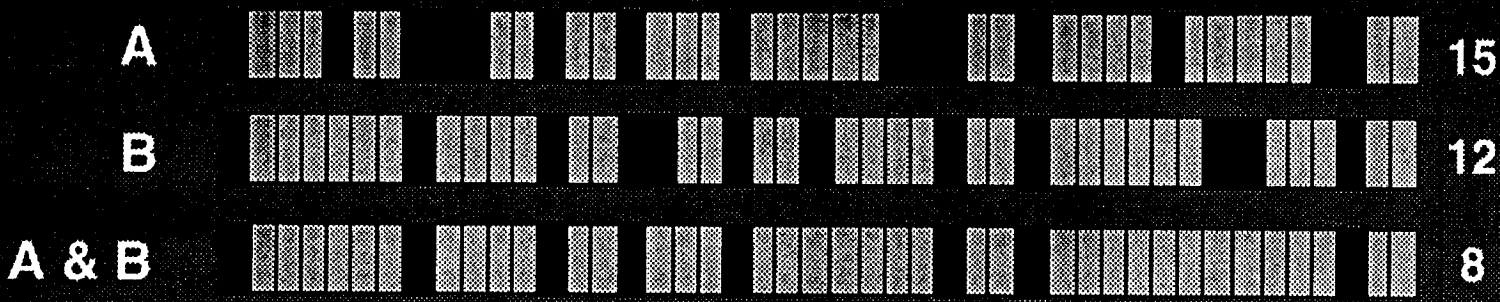
- non parametric space (fingerprints)

Tanimoto index :

$$\frac{N_{A\,\&\,B}}{N_A + N_B - N_{A\,\&\,B}}$$

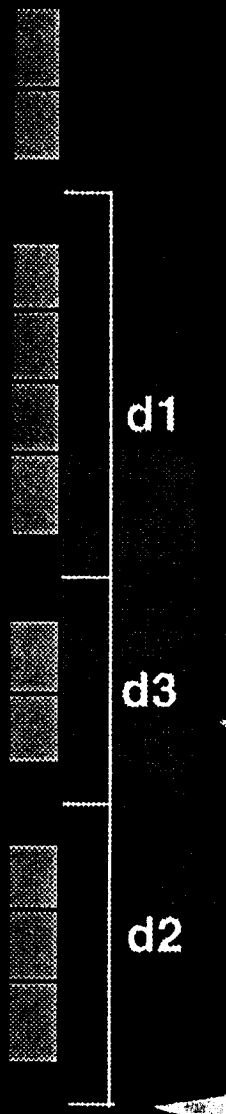
Is a measure of the number of common substructures shared by two molecules

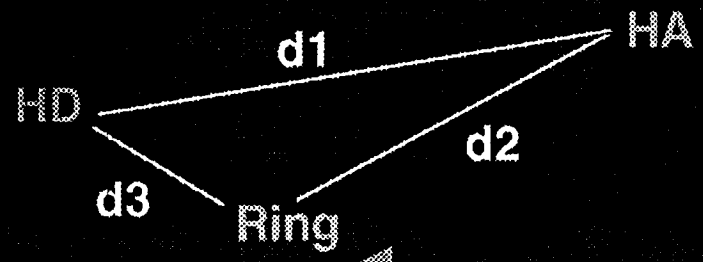
no similarity $0 < T_I < 1$ perfect similarity



pharmacophore diversity

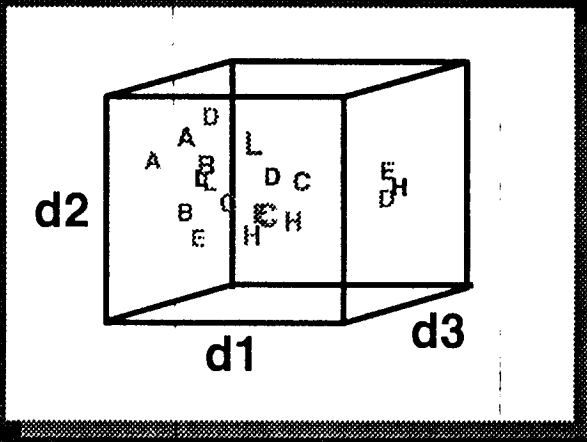
Diversity : the types and 3D geometries of pharmacophores exibited by the molecules in the library.





- 3 points pharmacophore definition

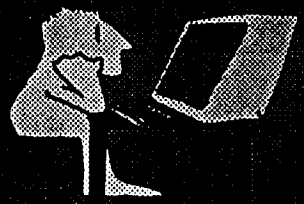
- **bits string encoding** *pharmacophore types and arrangments in low energy conformers for any molecule in the library*



pharmacophore key

Keith Davies and Caterine White, Chemical Design Ltd

What makes a molecule active? i.e. how can we distinguish active from inactive molecules?

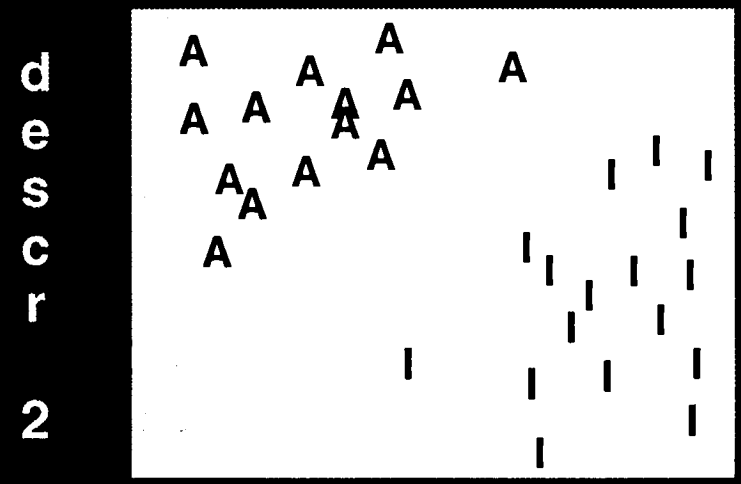


- Is structural diversity related to biological diversity ?

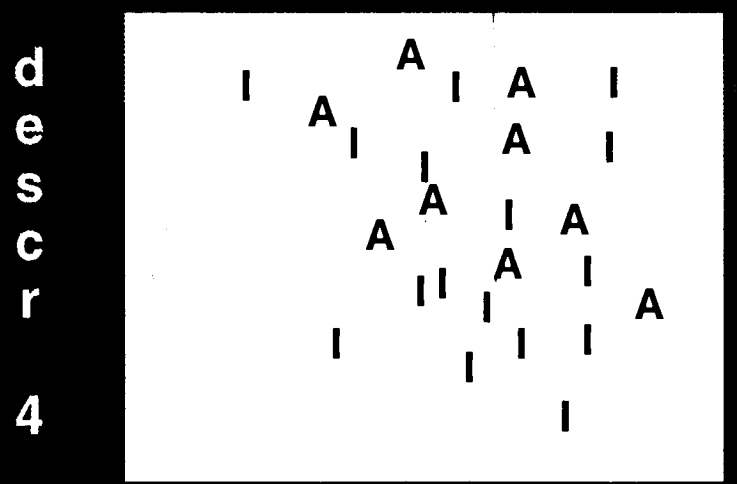
- *we are interested in what molecules do, rather than what they are*

- when using a set of suitable descriptors clustering the activity

the right space

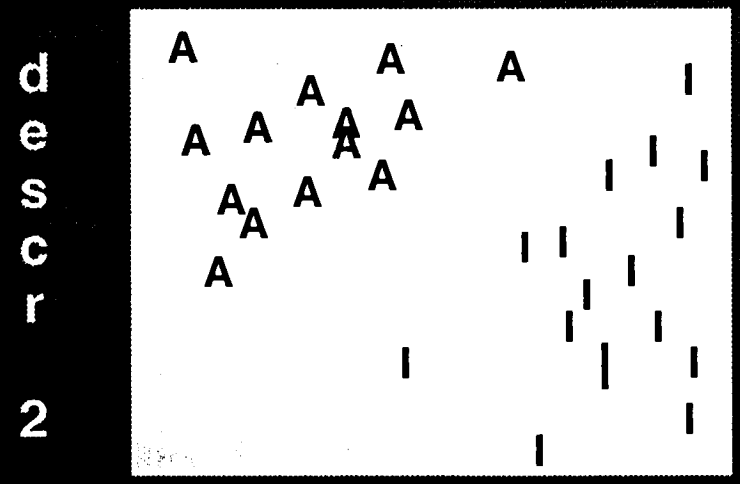


the wrong space

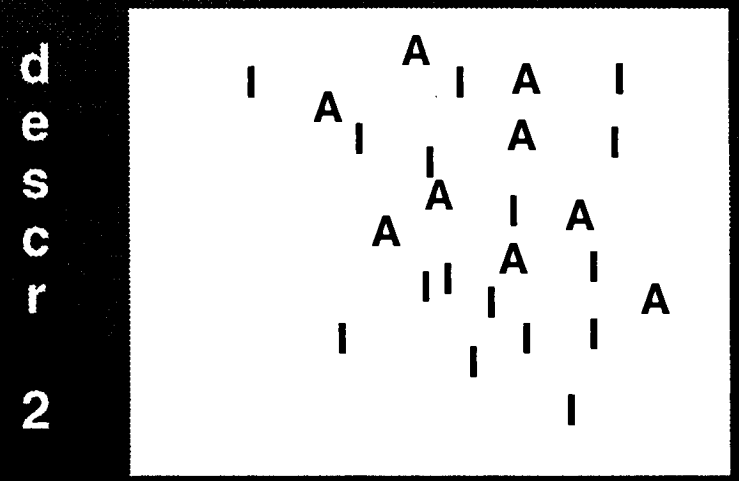


•Biology also exhibits classes !

thrombin inhibitors



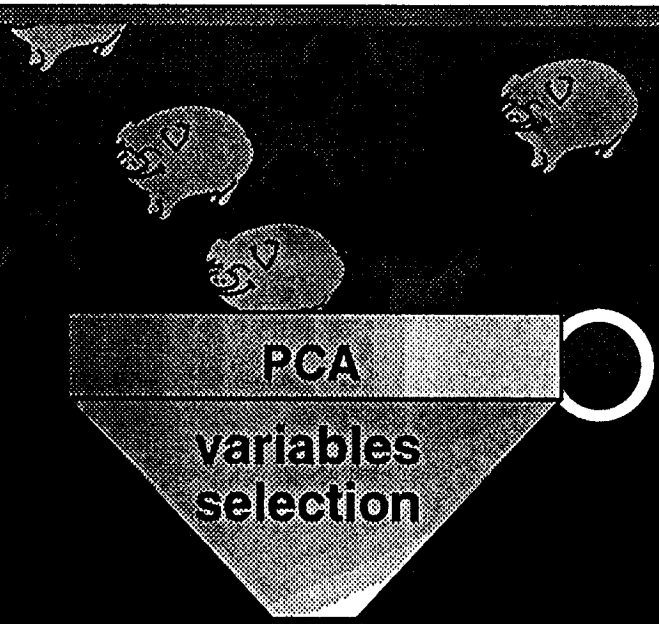
histamine antagonists



•We are looking to classify things in chemistry in a way that is consistent with the particular biology of interest

- *We should know the biological results before we can decide on the appropriate space in which to represent our compounds*

- *The largest number of descriptors should be explored in creating the "diversity" of primary libraries but...*

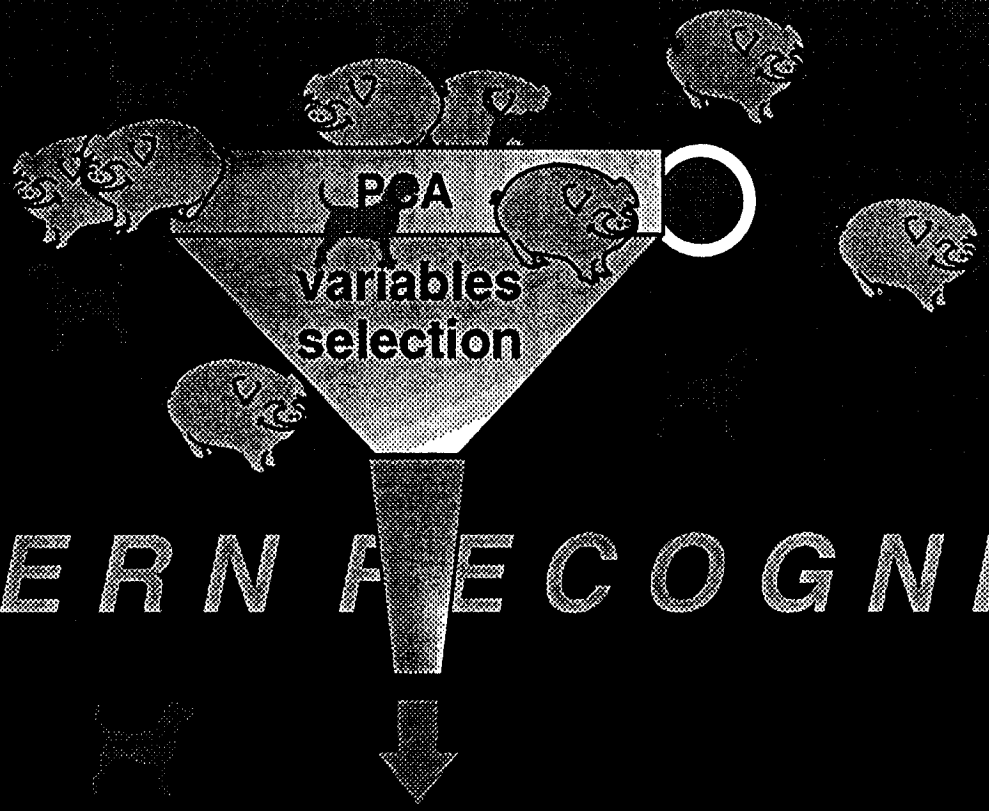


PATTERN RECOGNITION

Right space

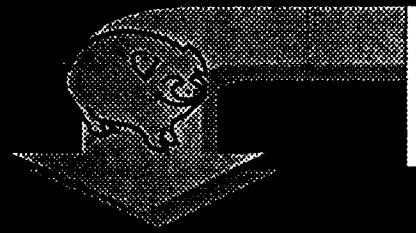
...for pattern recognition problems as the number of variables increased, the classification performance initially increases but then deteriorates, thus...

D.J. Hand ,Discrimination and Classification, Wiley 1981,122

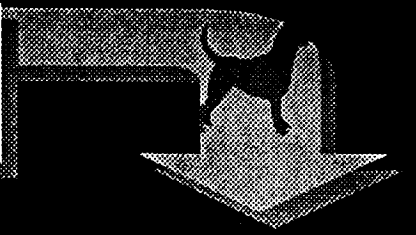


PATTERN RECOGNITION

It is unrealistic to expect a set of compounds chosen by some relevant criteria to be appropriate for testing in several biological screens (universal libraries).



Two main categories according to the nature of space they define

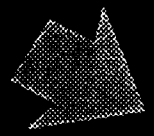


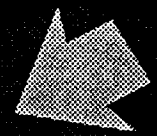
CONTINUOUS (or metric)

- *Lipophilicity*
- *CMR*
- *VdW volume*
- *Topological descriptors*

NON CONTINUOUS (or non-metric)

- *fingerprints*



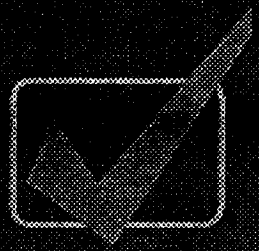


parametric space

EUCLIDEAN SPACE

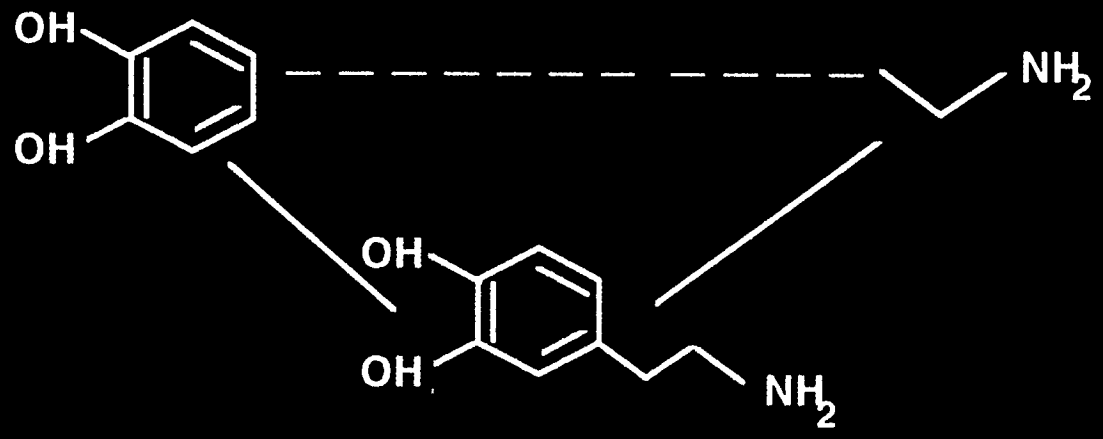
non parametric space

CHEMICAL SPACE

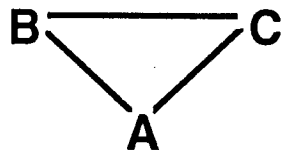


CHEMICAL SPACE

Without a real understanding of the nature of chemical space measures of the relationships between chemicals may be meaningless

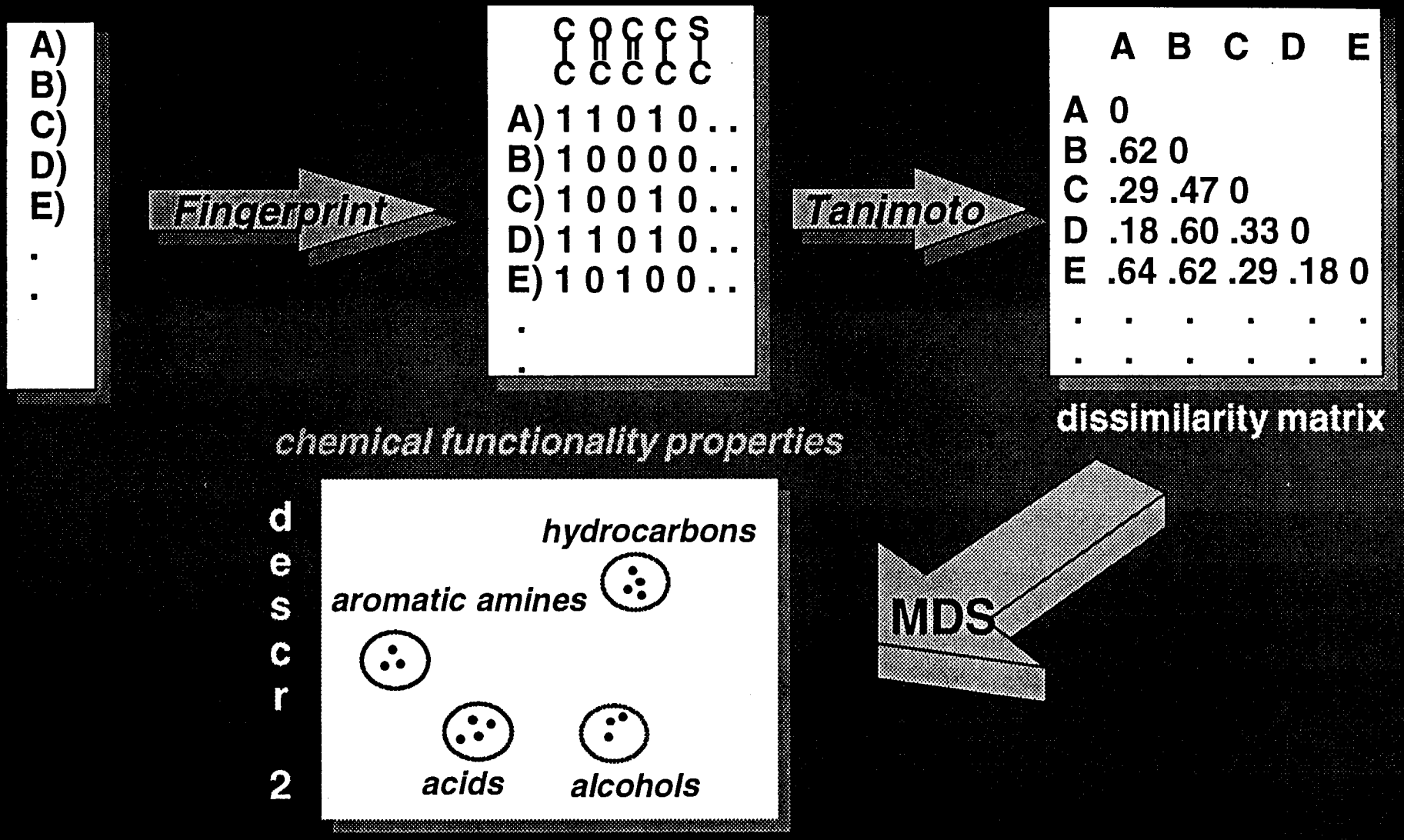


triangle inequality rule



BA + AC ≥ BC
is not true

to design a diverse set in which dissimilarity is important an appropiate distance metric must be used...

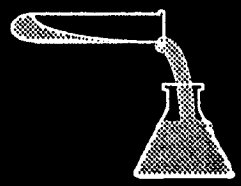


SAS PROC MDS SAS institute Inc

Martin E.J. et al., J.Med.Chem. 1995, 38, 1431-1436

Compound selection

Space filling



•Coverage

A "coverage " design seeks a set of points such that any point not in the design is not very far away from a design point ○.

> cluster analysis

> uniform shell

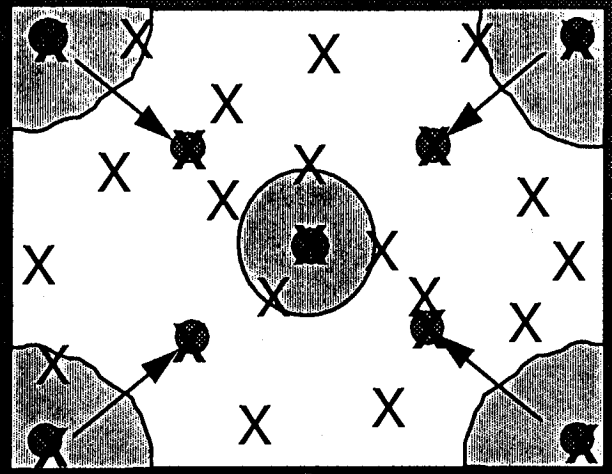
•Spread

A "spread "design seeks a set of design points that are mutually as far from each other as possible ●

> experimental design

D-optimal design

Factorial design

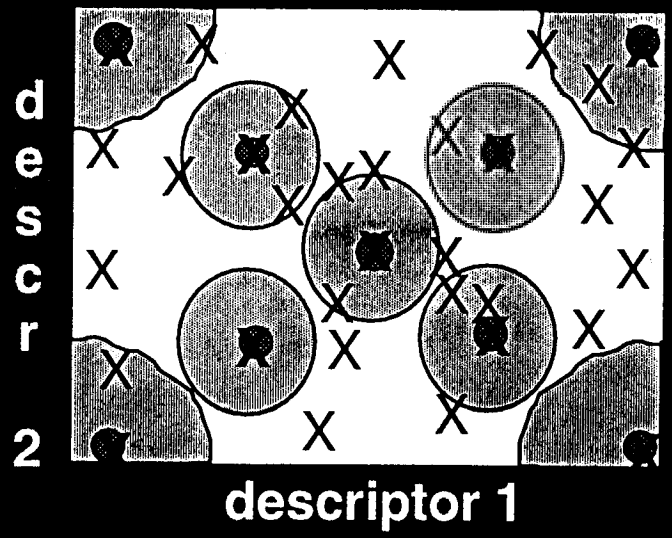


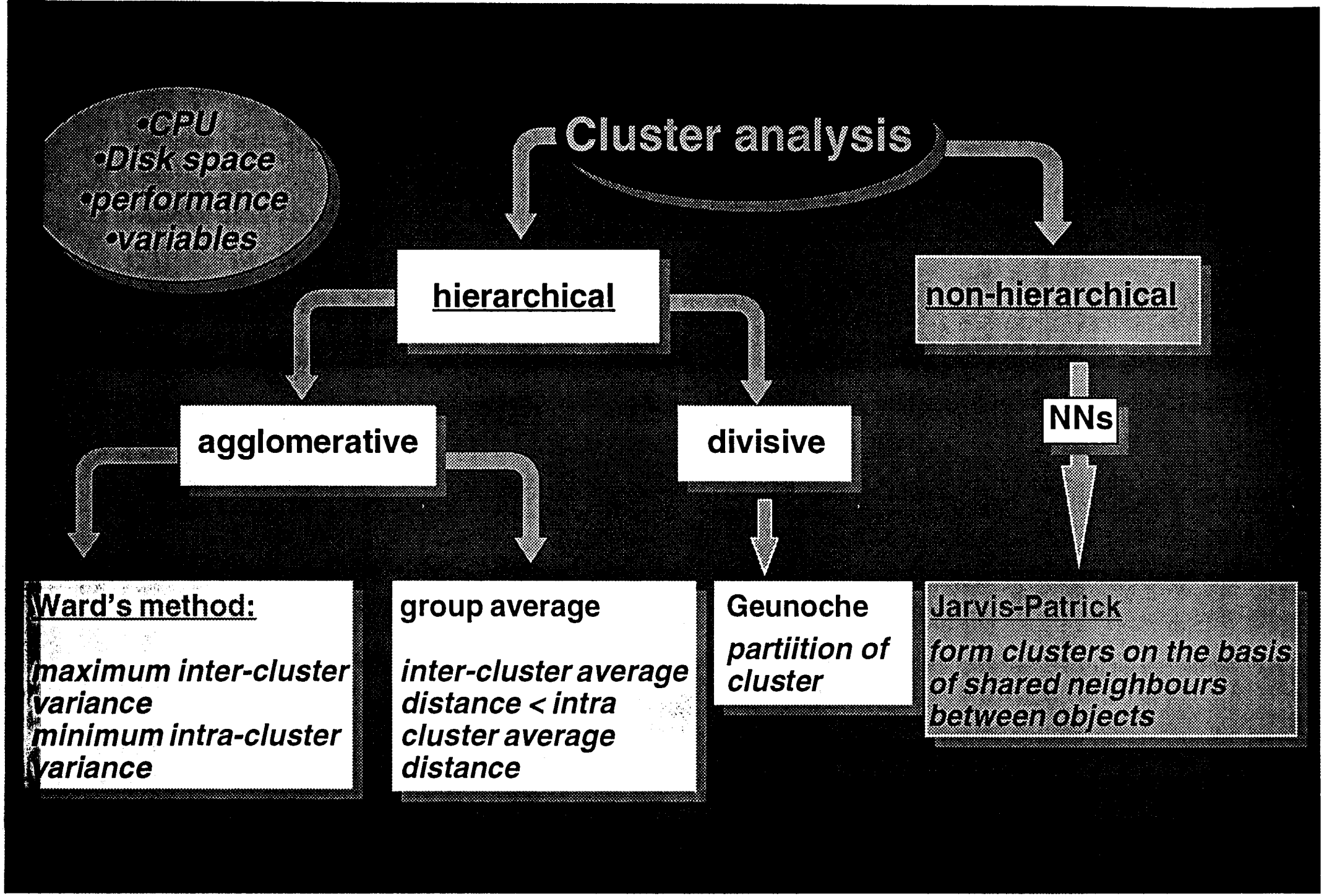
Cluster analysis

- Similar Property Principle : similar molecules behave similarly

Structures which cluster togheter may be assumed to have sufficiently similar activities that only one or few may be selected as representative of the cluster as a whole for testing in biological assays

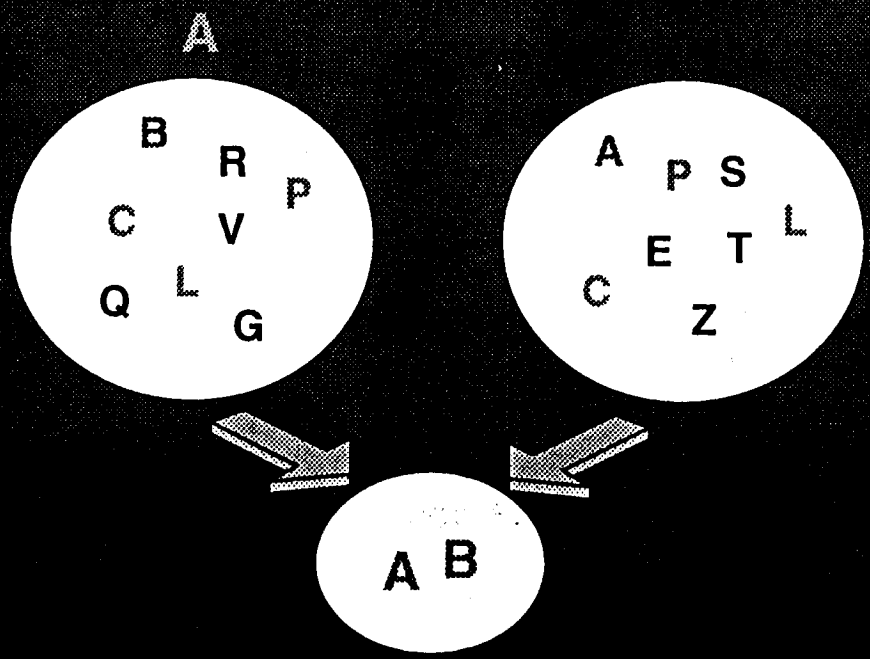
region of predictability around a tested compound




•CPU
•Disk space
•performance
•variables
Cluster analysis
hierarchical
non-hierarchical
agglomerative
divisive
NNs
Ward's method:
maximum inter-cluster variance
minimum intra-cluster variance
group average
inter-cluster average distance < intra cluster average distance
Geunoche
partiition of cluster
Jarvis-Patrick
form clusters on the basis of shared neighbours between objects

Jarvis-Patrick

- It's a non-parametric method
- For each item, find its nearest neighbors. This requires order (N^2) CPU time, but needs to be done only once
- The clustering step is vary fast
- Two structures cluster together if:
  - *They are in each other's list of J jearest neighbors*
  - *K of their J nearest neighbors are in common*



STRUCTURAL DESCRIPTORS & CLUSTERING ALGORITHMS : WHICH PERFORM BETTER ?

- proprietary enzyme assay , dataset of about 1000 structures

- Descriptors : - Daylight2D, Maccs2D, Unity2D fingerprints
  - Unity 3D rigid and flex fingerprints

- Clustering methods : - hierarchical Ward's (*at various levels of clustering*)
  - non - hierarchical Jarvis-Patrick (*varying need/near ratio and similarity treshold*)

performance : the ability to clusters active structures separately from inactives

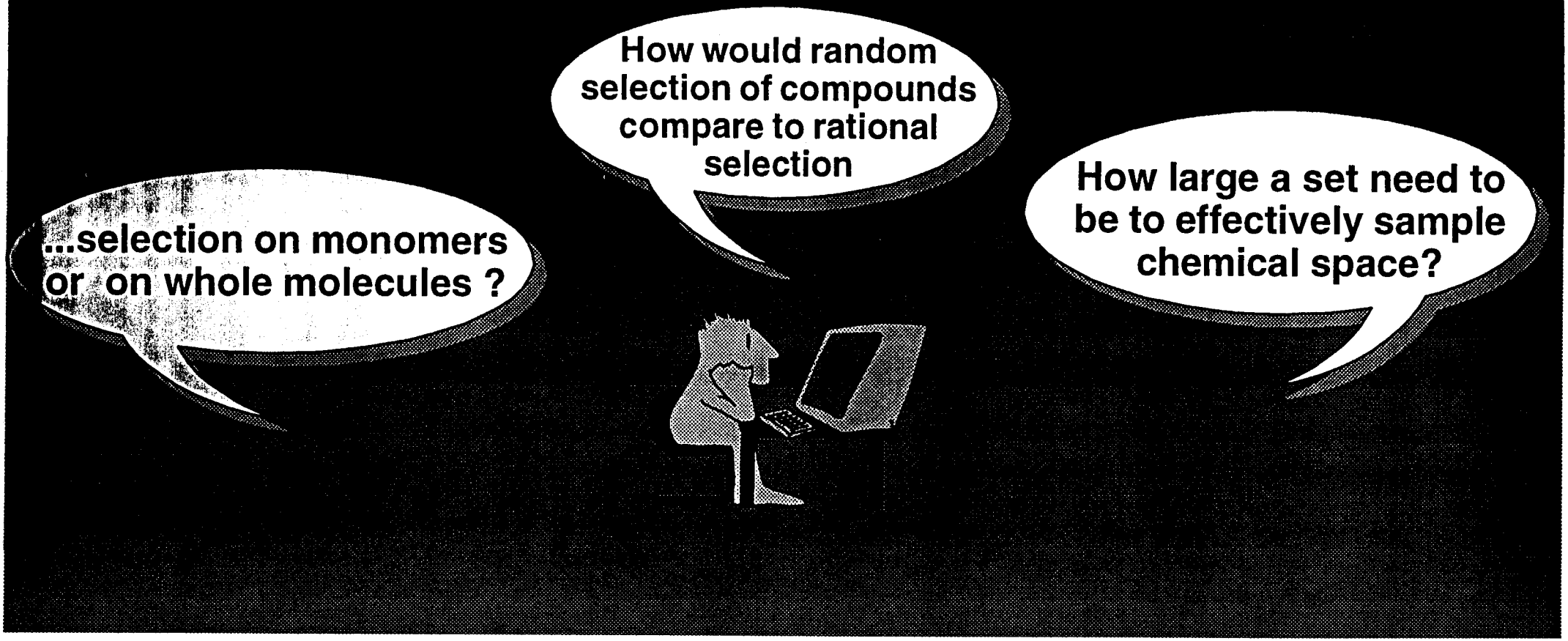
Results

- **Both flex and rigid 3D fingerprints perform worse than any of the 2D ones (3Dflex performing particularly badly !)**
- **Ward's agglomerative clustering gives better results than J-P with good descriptors, similar in the other cases**

R. D.Brown et al. in Proc. First Electron.Comput. Chem. Conf.(1995)

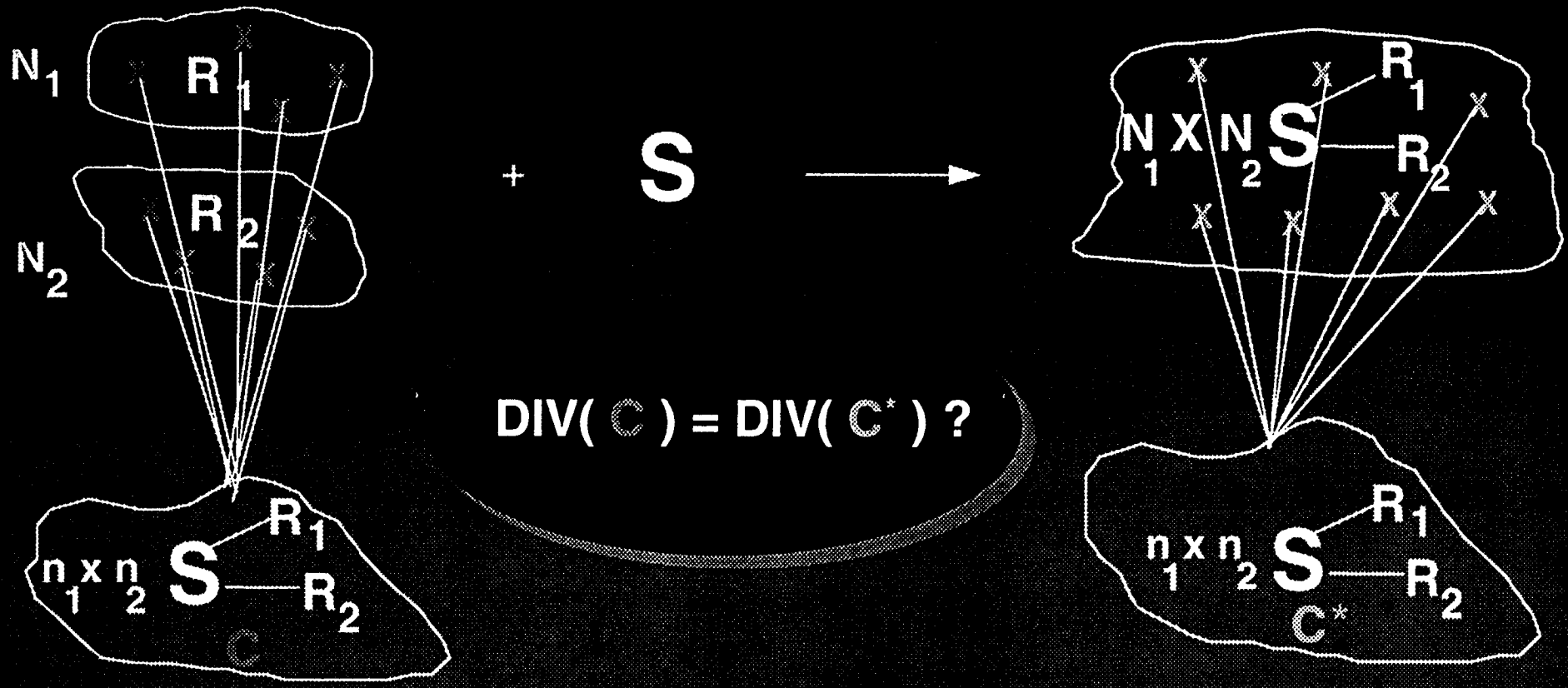
WHEN TO DESIGN A LIBRARY ?

More questions than answers



Does "Diversity" in monomers imply "Diversity" in libraries?

•we test the final compounds, not the monomers



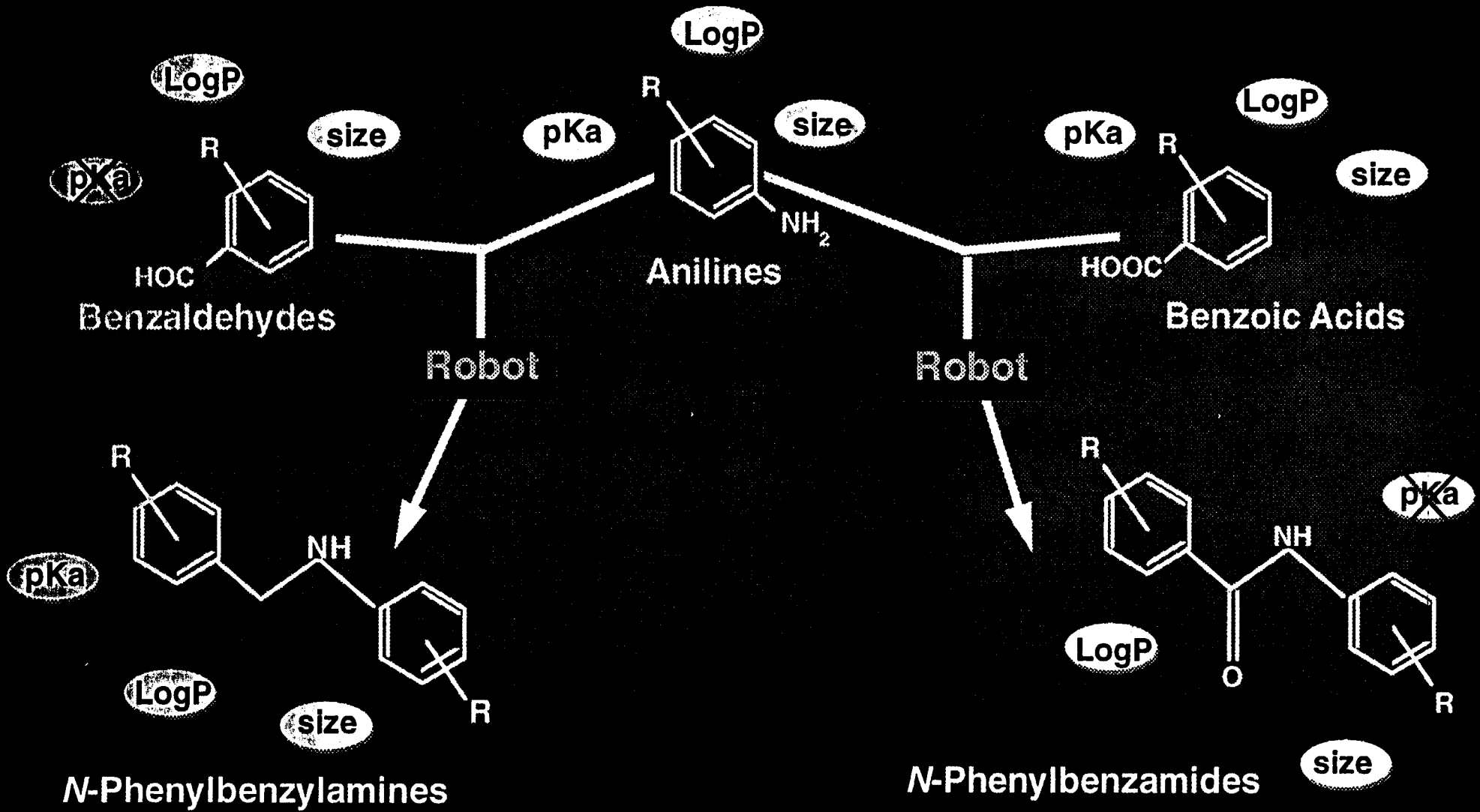
extrimely efficient for structural space exploration

$O(n_1N_1) + (n_2N_2)$

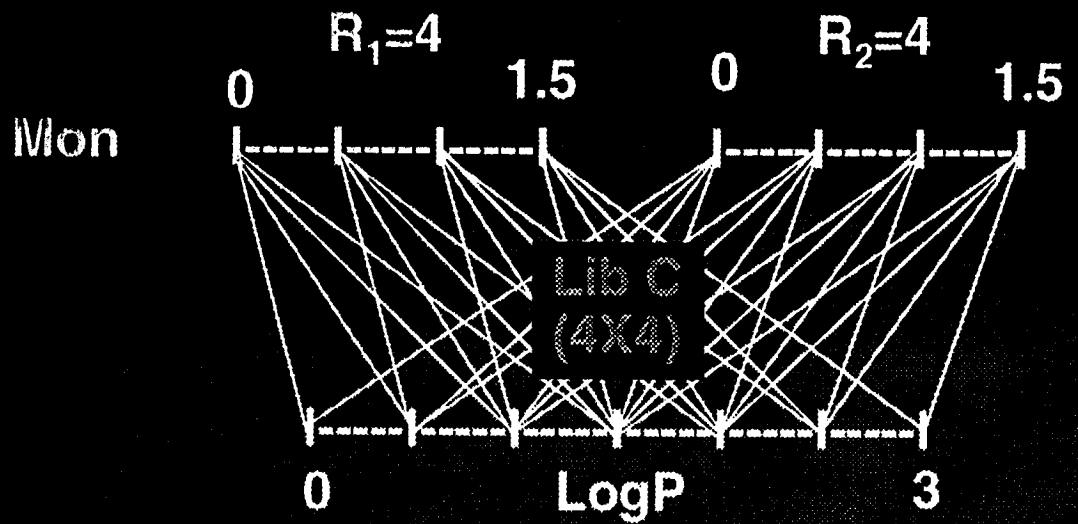
computationally expensive

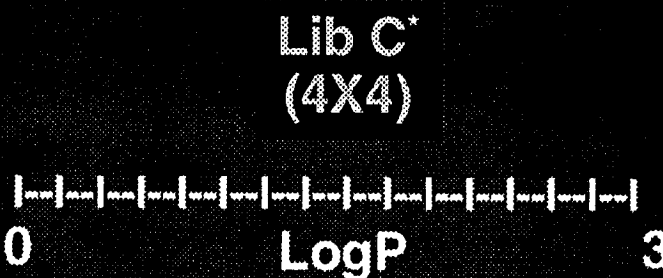
$O(n_1n_2N_1N_2)$

Does "Diversity" in monomers imply "Diversity" in libraries?



Does "Diversity" in monomers imply "Diversity" in [illegible]





- Example
- R1=400 from 1000 primary amines • R2=400 from 1000 carbox. acids · 160,000 products

DIV (C*) = 0.651
DIV (C) = 0.594
DIV (R) = 0.512

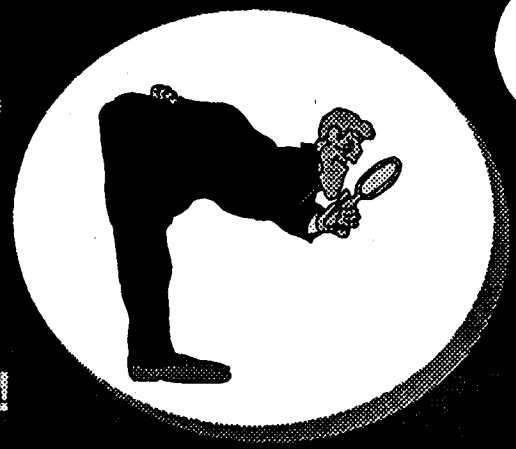
...the diversities of the subset C are greater than those of the random subset, R, but less than those of the maximally-diverse subsets, C *, chosen from the full library.

Selection procedures should operate in product space but....
combichem needs to reduce the nunber of monomers not reactions

P. Willet oral communication

Only for people thinking that rational methods are not always better than random methods

How large a set need to be to effectively sample chemical space?



How would random selection of compounds compare to rational selection?

more questions than answers again !

- What is the inherent dimensionality of chemical space ?
  (How many features need to be correct for a compound to be active?)

spatial arrangement of pharmacophoric features	6 - 10 +
correct size and alignment	6 +
roughly correct physical properties (lipophilicity,electrostatic potential)	2 =
10 cps to span each dimension	14 - 18

Without information (primary universal libraries)

10^{14} - 10^{18} cps to fully span the space

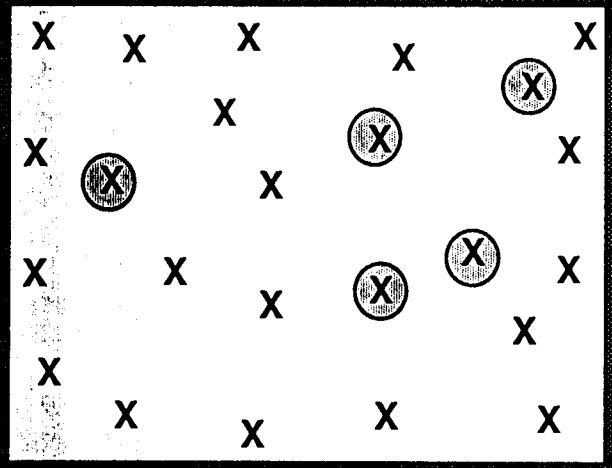
" The curse of high dimensions "

How would random selection of compounds compare to rational selection?

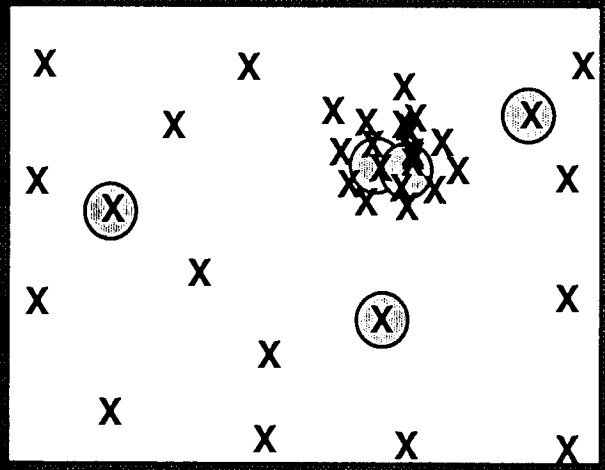
region of predictability around a selected compound

•chemical space >> coverage region, relatively few cps

EVEN DISTRIBUTION



UNEVEN DISTRIBUTION



randomly selected cps will cover as much space as carefully selected cps

as we expect little overlap among randomly selected cps

very dense areas should be removed and treated in a special way

Cover the volume of these areas with the same density as in other regions

S.S. Young et al. Net. Science (1996)

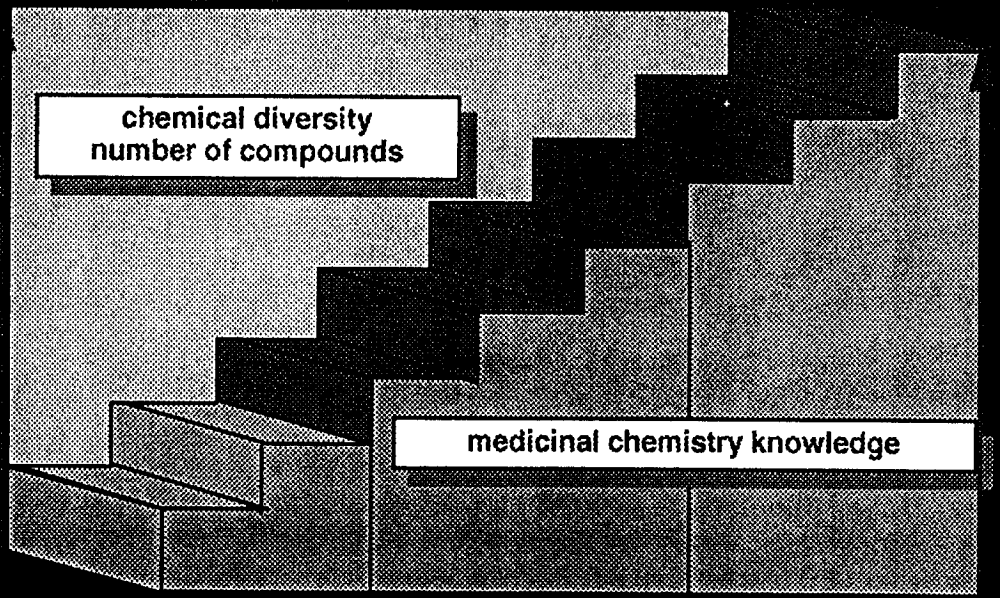
" The curse of high dimensions "

It would be very useful to have a small number of variables that are pertinent to biological activity

The design of the chemical libraries for a particular target should use accumulated biological knowledge for variable selection when it is available to reduce the chemical space (knowledge based libraries)

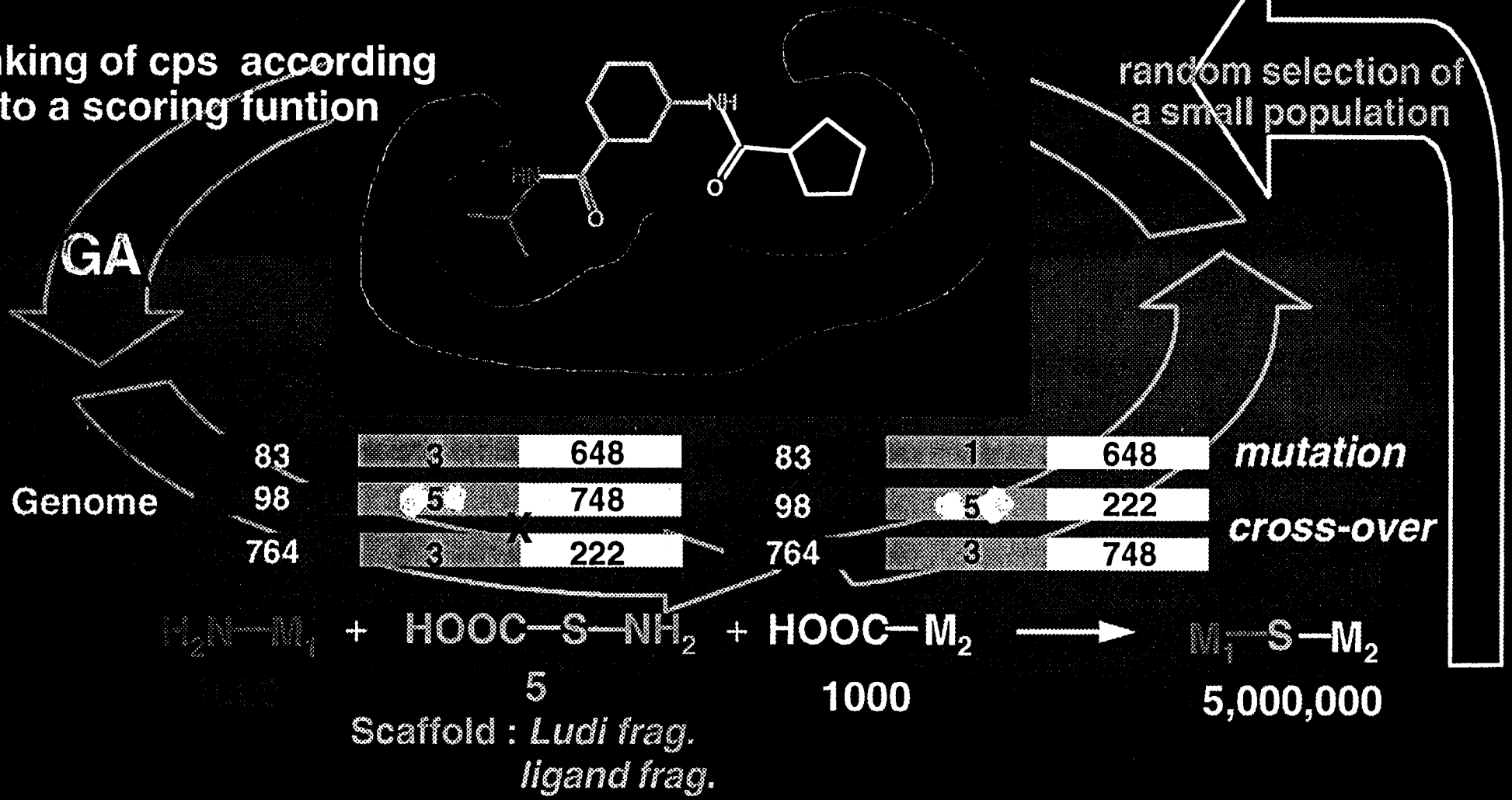
Chemical space reducion for primary libraries:

- *comparing "diversity" with other libraries (i.e. DDR, 7TM etc)*
- *pharmacokinetics*



STRUCTURE BASED LIBRARY DESIGN

Quick docking or pharmacophore fitting procedure (Dock, Catalyst, Chem-X)



R.P. Sheridan and S.K.Kearsley *J.Chem.Inf.Comput.Sci.*(1995),**35** 310

Conclusions

- Chemical space is very large . Unless a very large number of cps are used to fill space, randomly selected compounds will cover as much space as carefully selected compounds

- If diverse cps are desired, realizing that coverage will not be improved by rational methods, then fast spread methodologies can be used to ensure that selection will be as far apart as possible while eliminating possibility of overlapping selection (especially for uneven didstributions)

- If the important dimensions for a particular problem are identified, and if a focused set of cps is desired, then rational selection should be more effective than random design.

S.S. Young et al. Net. Science (1996)