MOLECULAR PRINCIPLES OF DRUG DESIGN
MGR. ANNA HUDCOVÁ
DEPARTMENT OF CHEMICAL DRUGS
UVPS BRNO

CHEMOGENOMICS AND THEORY OF PRIVILEGED STRUCTURES

CHEMOGENOMICS

aims towards
the systematic identification
of small molecules
that interact with the products of genome and
modulate their biological function

CHEMOGENOMICS

=

the discovery and description of

all possible drugs

for

all possible drug targets

Some definitions

- GENOMICS = an interdisciplinary field of science focusing on the structure, function, evolution, mapping, and editing of genome
- Genome = a genome is an organism's complete set of DNA, including all of its genes
  - Human Genome Project (1990-2003): identification of 20000-25000 genes in human DNA
- Genetics = study of individual genes and their role in inheritance

- PROTEOMICS = refers to the large-scale experimental analysis of protein
- Proteome = the entire set of proteins that are produced or modified by an organism or systhem

More definitions

- Chemical genomics / CHEMOGENOMICS
  - the systematic identification of small molecules that interact via a specific molecular recognition mode with target proteins encoded by the genome
  - the term chemogenomics is applied more specifically to target family approaches in drug discovery
- Chemical genetics
  - identify chemical compounds which induce or revert specific biological phenotypes by using cell-based or microorganism-based screening of compound
- Chemical biology
  - the functional and mechanistic investigation of biological systems using chemical compounds and constitutes a more general discipline

The chemogenomic terminology

- 1996
- Glaco Wellcome
- Systematization of drug discovery within target families based on the analysis of gene families

Traditional approach

- Based on therapeutic areas
- Genomically unrelated targets are addressed together
- Example: therapy of neurodegenerative diseases
  - Acetylcholine and their derivarives
  - Antagonists of muscarinic and nicotinic receptors (target = **receptor)**
  - Inhibitors of acetylcholinesterase (target = **enzyme**)
  - Inhibitors of β-amyloid agregation (target = **protein**)

Chemogenomic approach

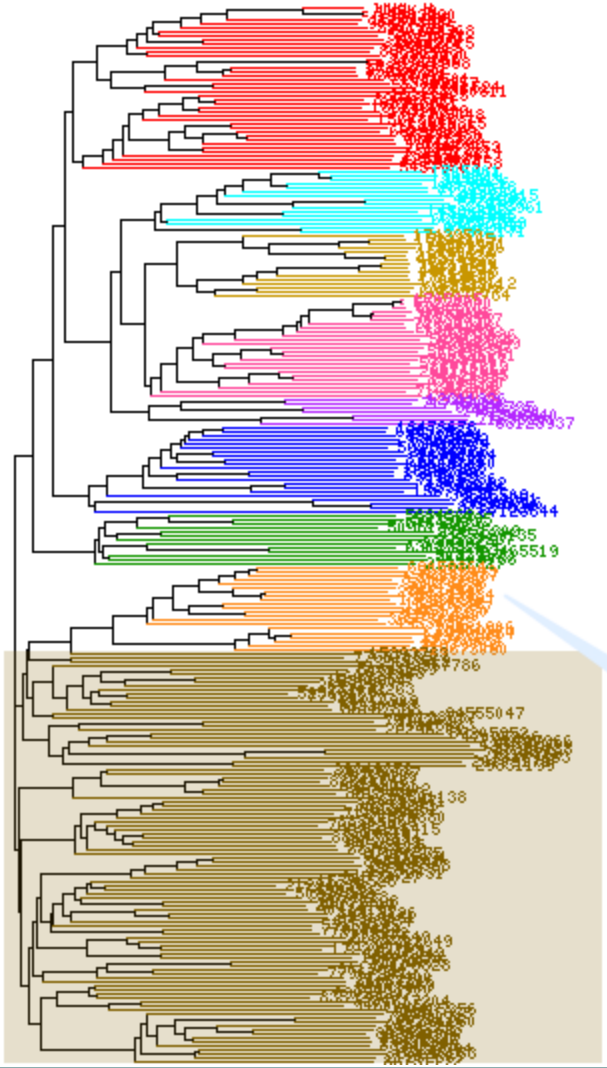
- Analysis does not depend on knowledge of biological function
- Members of the same protein family can share important practical aspects
- Similar ligands should bind to similar target → knowledge previously obtained → transferable to new related projects

How?

I. identification of all members of a gene family
II. classification into subfamilies
III. revelation of common elements and patterns in the sequence and tertiary structures

NCBI curated domain hierarchy for voltage-gated chloride channel

cd00400 Sequence Cluster



parent node, in this case, encompasses 3 kingdoms of life: archaea, eubacteria, eukaryota. Domain has unique double-barreled architecture and voltage-dependent gating mechanism.

Sub-family Hierarchy

- cd00400 Voltage_gated_ClC
  - cd01031 EriC
  - cd01033 ClC_like
  - cd01034 EriC_like
  - cd01036 ClC_euk
    - cd03683 ClC_1_like
    - cd03684 ClC_3_like
    - cd03685 ClC_6_like
  - cd03682 ClC_sycA_like

child node: domain model specifically found in archaea and eubacteria and associated with extreme acid resistance

putative domain models split out as separate child nodes due to the natural phylogenetic clustering of their member protein sequences, as shown in the sequence cluster tree

intermediate parent: domain model found only in eukaryotes

found in human CLCN1 gene (myotonia), CLCN2 gene (epilepsy), and CLCNKA and CLCNKB (Bartter syndrome)

found in human CLCN3, CLCN4, and CLCN5 (Dent disease, nephrolithiasis, proteinuria, and hypophosphatemic rickets)

found in human CLCN6, CLCN7 (osteoporosis)

found in bacteria; facilitates acid resistance in acidic soil

colors in sequence cluster and subfamily hierarchy correspond to each other

The goal of the NCBI conserved domain curation project is to provide insights into how patterns of residue conservation and divergence in a family relate to functional properties, and to provide useful links to more detailed information that may help to understand those

Classification of proteins

- Conventional phylogenetic

  WHOLE STRUCTURE

- Chemogenomic

  SMALL-MOLECULE BINDING SITES

  (only part of enzyme structure, most commonly in active site of the enzyme)

Example

CHEMOGENOMIC APPROACH IN CARDIOVASCULAR DISEASES

1. identification of target structures associated with cardiovascular diseases (literature research)
2. organisation of cardiovascular targets in protein families
   - target families: GPCR (G-related couple proteins), enzyme, kinase, proteinase, nuclear receptor, catalytic receptor, ion channel, transporter, other protein

Example

3. the establishment of knowledge base of the cardiovascular target space
   - determining the level of applicability of structure-based methods for *in silico* target profiling
   - is the structure of targets known?
   - if it is – is there any relationship, similarity?

4. using only known structures
   - which belong to *Homo sapiens*?

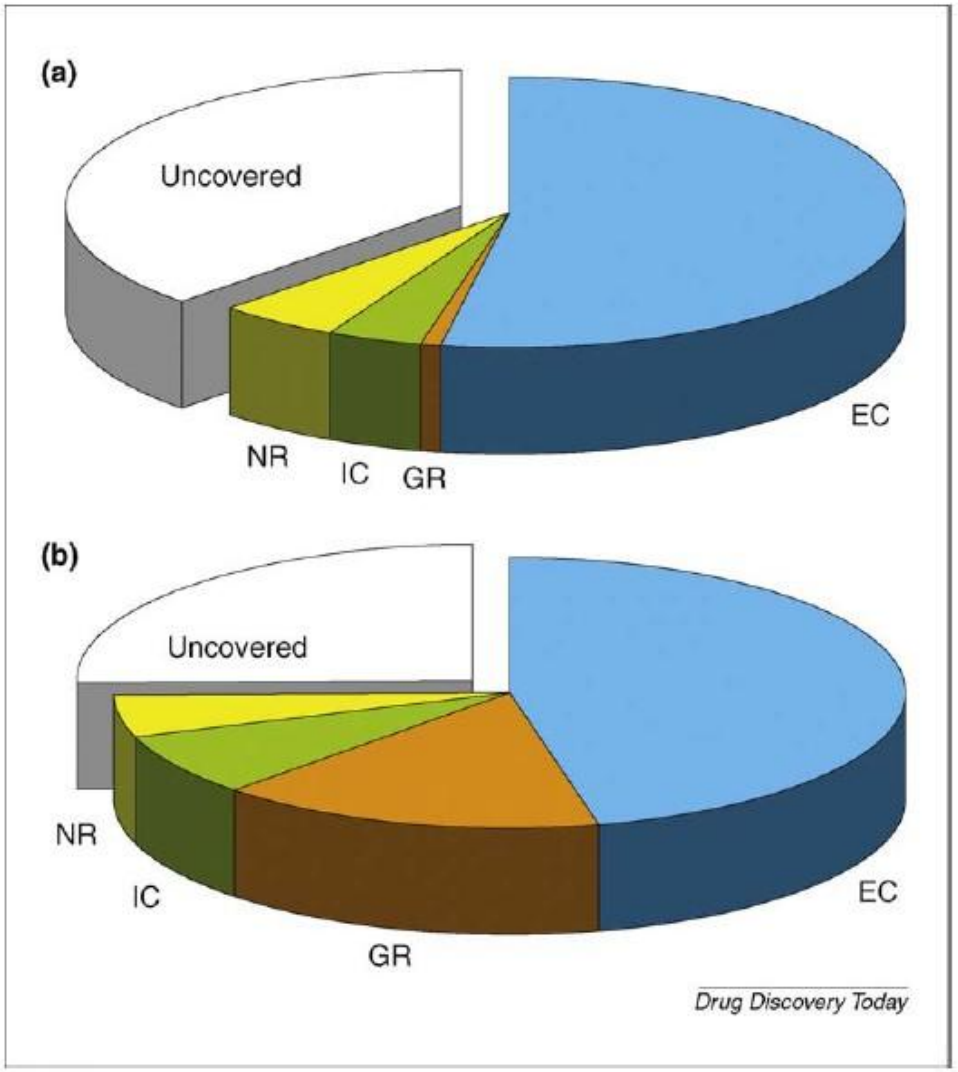
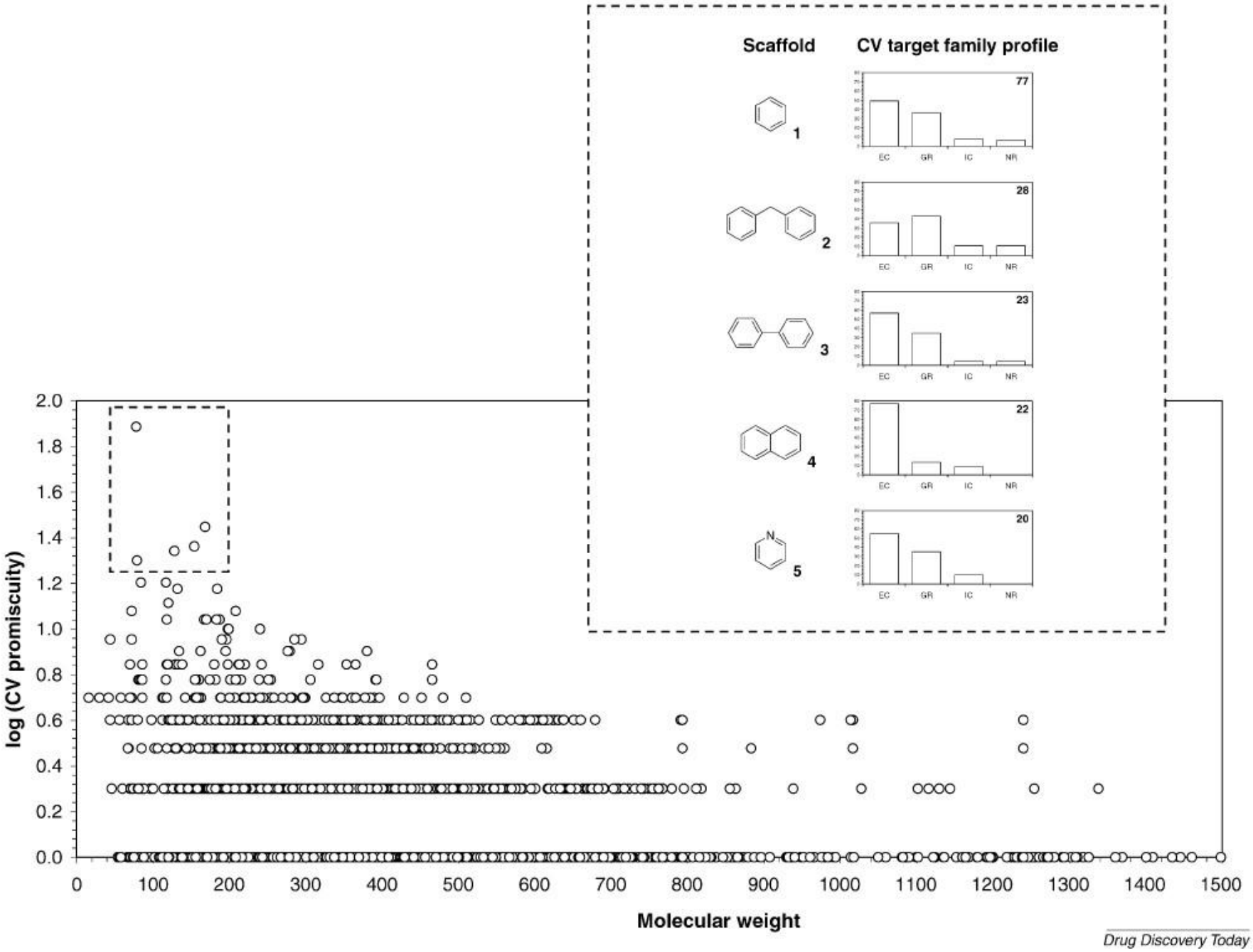


FIGURE 2

Structural and chemical coverage in cardiovascular target space. (a) Distribution of targets for which at least one representative structure exists in the Protein Data Bank and (b) distribution of targets for which at least one bioactive ligand is present in the annotated chemical libraries considered. EC: enzymes; GR: GPCRs; IC: ion channels; NR: nuclear receptors.

Example

5. data from chemical libraries
   - ligands with pharmacological potency at (at least one of) cardiovascular target structure
6. identification of atomic frameworks or scaffolds
7. synthesis
8. testing on cells



Drug Discovery Today

PRIVILEGED STRUCTURES

Privileged structure is
a single molecular framework
able to provide ligands
for diverse receptors.

Evans and co-workers, 1988

Why are so interesting?

- New drug targets
- Bring a new drug to market
- Modeling of potential “small molecules”
  - number of possible small molecules is 10E200
  - drug-like properties may have 10E60

Characteristics

- constitute a significant portion of the total mass of the molecule
- represent the core element of the molecule
- physico-chemical characteristics for promiscuous binding
- smaller molecule has higher capacity of binding to multiple receptors
- simple ligand surface

Characteristics 2

Cyclic structures are ideal scaffolds for drug development

- molecular rigidity
- better bioavailability
- bicyclic and tricyclic molecules are ideal size for library synthesis

Characteristics 3

Promiscuous ligands:

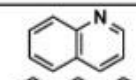
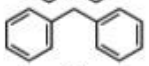
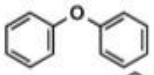
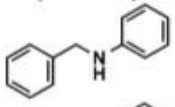
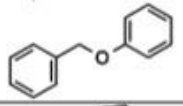
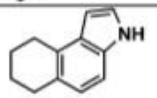
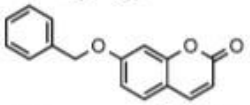
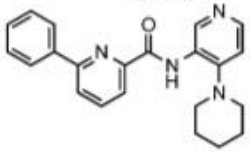
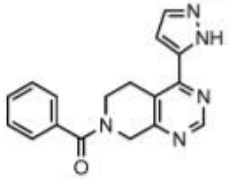
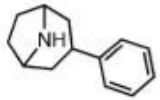
- potently
- specifically
- reversibly
- but not selectively

bind to members of different macromolecular target families

Limites?

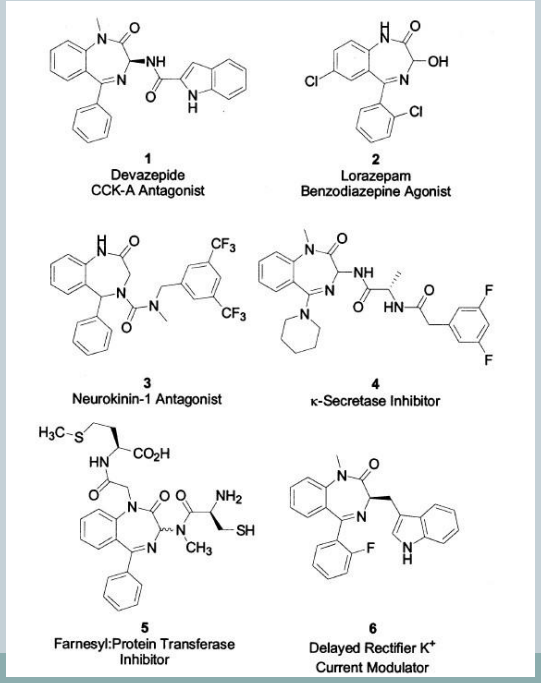
- Have they limited utility due to their promiscuous nature?
- Ex. Biphenyl framework has been described as a preferred substructure for protein binding – appears in 4,3% of all known drugs
- This may indicate that although privileged substructures have the capacity to bind nonspecifically to a number of receptors, THE SUBSTITUENTS attached to the scaffold may be responsible for its receptor specificity, while the scaffold itself provides a number of features conductive to binding

Table 1: Top-ranking scaffolds according to their promiscuity expressed as Shannon entropy *H* (**1a–5a**), and according to maximal information content *I* (**1b–5b**). *N* = number of potent (p*Activity* ≥ 6) compounds containing only the respective atom scaffold and no other ring system.

ID	Scaffold	*N*	Welsch list[8]	False positives (Rishton)[3]	False positives (Hann et al.)[4]	H (bit)	*I* (bit)	GPCR	Enzyme	Kinase	Proteinase	Nuclear receptor	Catalytic receptor	Ion channel	Transporter	Protein	Undefined target
1a		285	yes	7%	0%	2.75	0.57	15%	25%	15%	19%	9%	0%	8%	1%	7%	0%
2a		545		23%	4%	2.54	0.78	34%	22%	3%	2%	1%	3%	11%	20%	3%	1%
3a		1184	yes	5%	3%	2.46	0.86	21%	19%	0%	17%	0%	1%	18%	21%	2%	0%
4a		245		1%	18%	2.39	0.93	10%	23%	0%	1%	39%	6%	1%	16%	4%	1%
5a		828	yes	17%	6%	2.38	0.94	23%	35%	0%	11%	0%	0%	19%	9%	2%	1%
1b		280	yes	0%	4%	0	3.32	100%	0%	0%	0%	0%	0%	0%	0%	0%	0%
2b		282	yes	3%	0%	0	3.32	0%	100%	0%	0%	0%	0%	0%	0%	0%	0%
3b		103		0%	3%	0	3.32	0%	0%	100%	0%	0%	0%	0%	0%	0%	0%
4b		190		0%	0%	0	3.32	0%	0%	0%	0%	0%	0%	100%	0%	0%	0%
5b		870		2%	0%	0	3.32	0%	0%	0%	0%	0%	0%	0%	100%	0%	0%

1. Library based upon one core scaffold

- Screen against a variety of different receptors
- Example on 1,4-benzodiazepin-2-ones
  - Small library of 192 molecules => cholecystokinin A receptor => several active compounds => larger library of 1680 compounds
  - Possible explanation?
  - Targets
    - Cholecystokinin (devazepine)
    - Gastrin
    - Central benzodiazepine receptors (lorazepam)
    - Neurokinin-1 antagonistS
    - K-secretase inhibitors
    - Farnesyl:protein transferase inhibitors
    - Delayed rectifier K+ current modulator



2. DOS

- **Diversity Oriented Synthesis**
- Novel drug-likecompounds with a high degree of molecular diversity
- Resulting DOS-derived libraries contain complex and diverse structures with a high fraction of sp3-hybridized carbon atoms and more stereogenic centers

3. pDOS

- **Privileged-substructure-based Diversity Oriented Synthesis**
- privileged substructures from natural products as chemical navigators
- polyheterocycles – example **benzopyranyl motif** was identified in biological evaluations:
  - a nonsteroidal androgen receptor antagonist
  - a small-molecule modulator with insulin-independent antidiabetic and antiobesity effects
  - a small-molecule anabolic activator of osteogenetic activity

4. Combinatorial chemistry

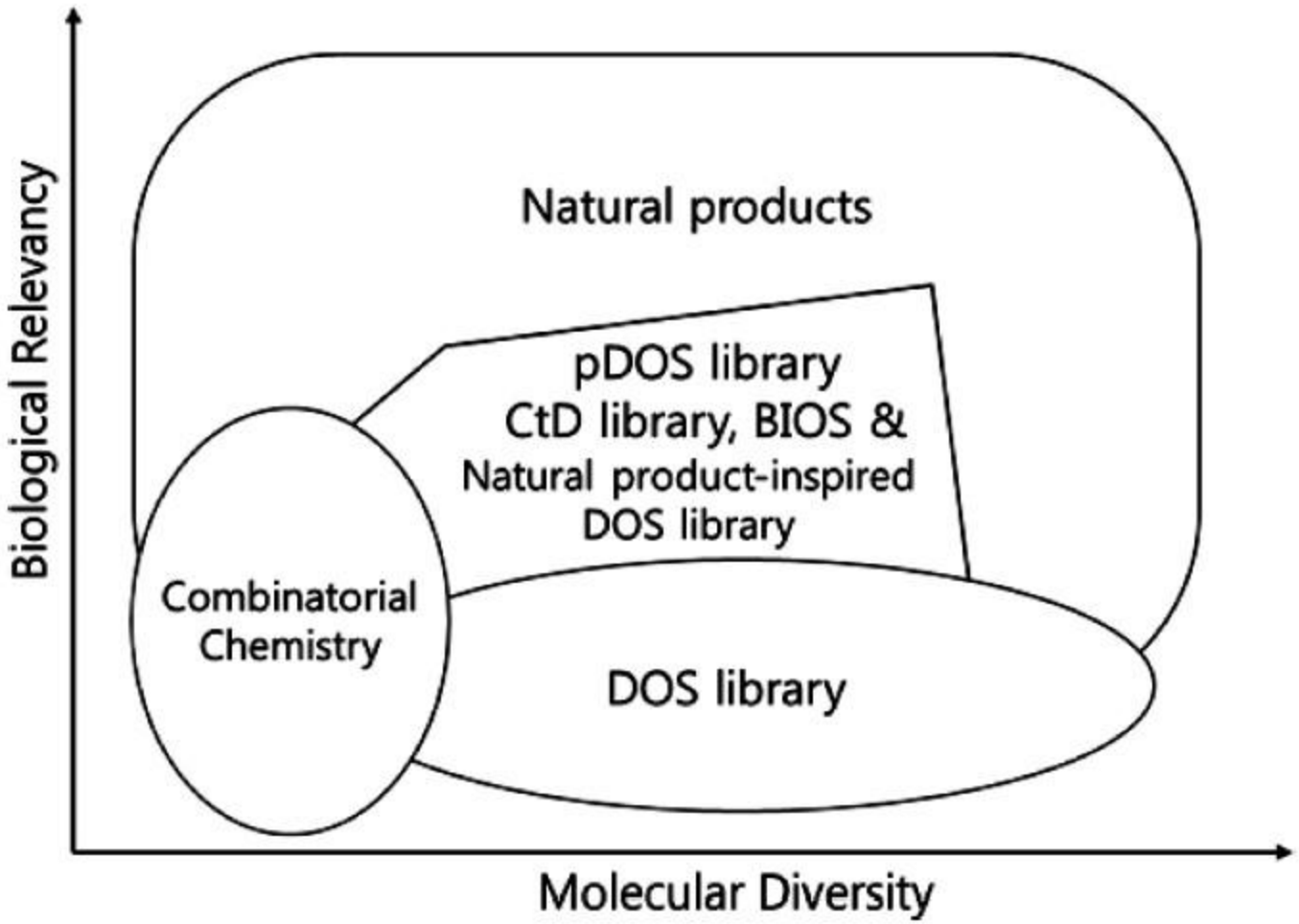
- Comprises chemical synthetic methods that make it possible to prepare a large number (tens to thousands or even millions) of compounds in a single process.
- **High-Throughput Screening** (HTS)
  - method for biological testing large chemical libraries
  - using robotics, data processing/control software, liquid handling devices, and sensitive detectors, high-throughput screening allows a researcher to quickly conduct millions of chemical, genetic, or pharmacological test
  - the microtiter plate: a small container, usually disposable and made of plastic, that features a grid of small, open divots called *wells* (multiples of 96)

5. BIOS

- **Biology Oriented Synthesis**
- use of core structures derived from bioactive natural products as synthetic scaffolds
- structural motif and core skeletons from bioactive natural products can serve as chemical “navigators” for the synthesis of novel core skeletons

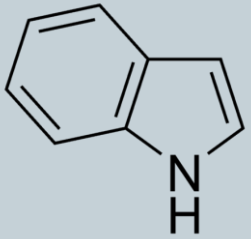
6. CtD

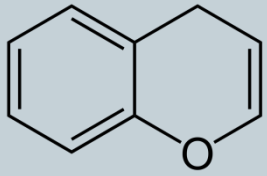
- **Complexity-to-Diversity strategy**
- construction of natural product-like small-molecule collections starting from commercially available natural products (abietic acid, adrenosterone, quinine)

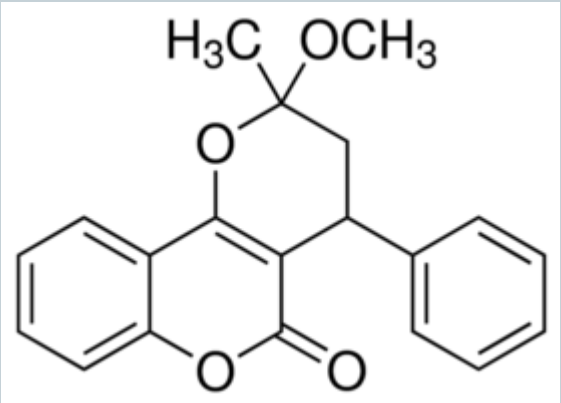

Biological Relevancy
Natural products
pDOS library
CtD library, BIOS &
Natural product-inspired
DOS library
Combinatorial
Chemistry
DOS library
Molecular Diversity

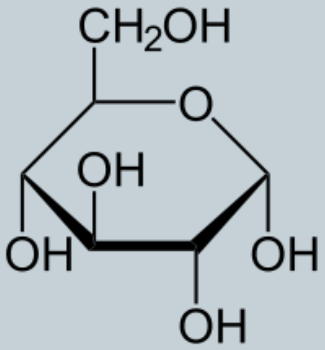
Organic scaffolds as privileged structures

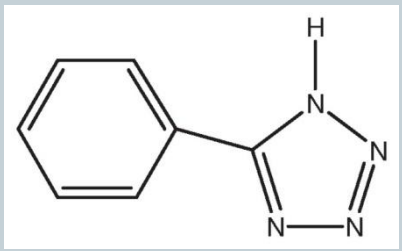
- 1,4-benzodiazepin-2-one
- biphenyl
- 1,4-dihydropyridine
- benzopyran
- pyranocoumarin
- 2,6-dichloro-9-thiabicyclo[3.3.1]nonane
- isoxazole
- 3,5-linked pyrrolin-4-ones
- β-glucose
- monosacharides in general
- benzazepinone
- diphenylmethane
- biphenyltetrazole
- spiropiperidine
- 4-substituted piperidine
- indole
- benzylpiperidine
- phenylpiperazine

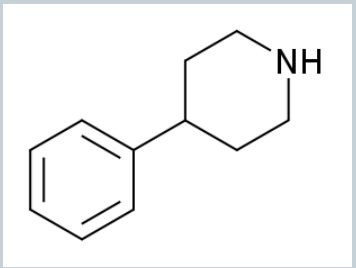












THANK YOU