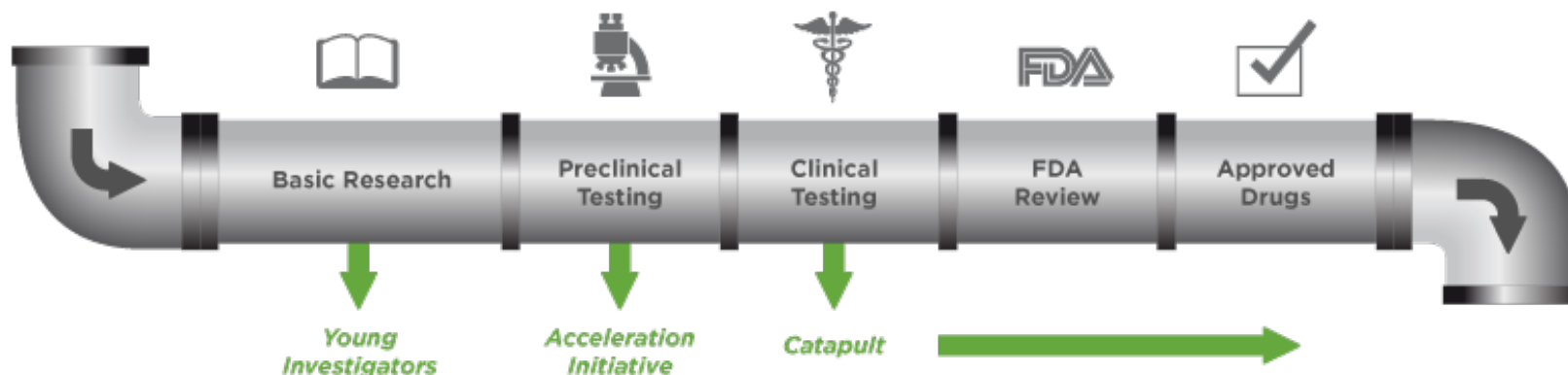


Modern Methods in Drug Discovery

Aims of this course:

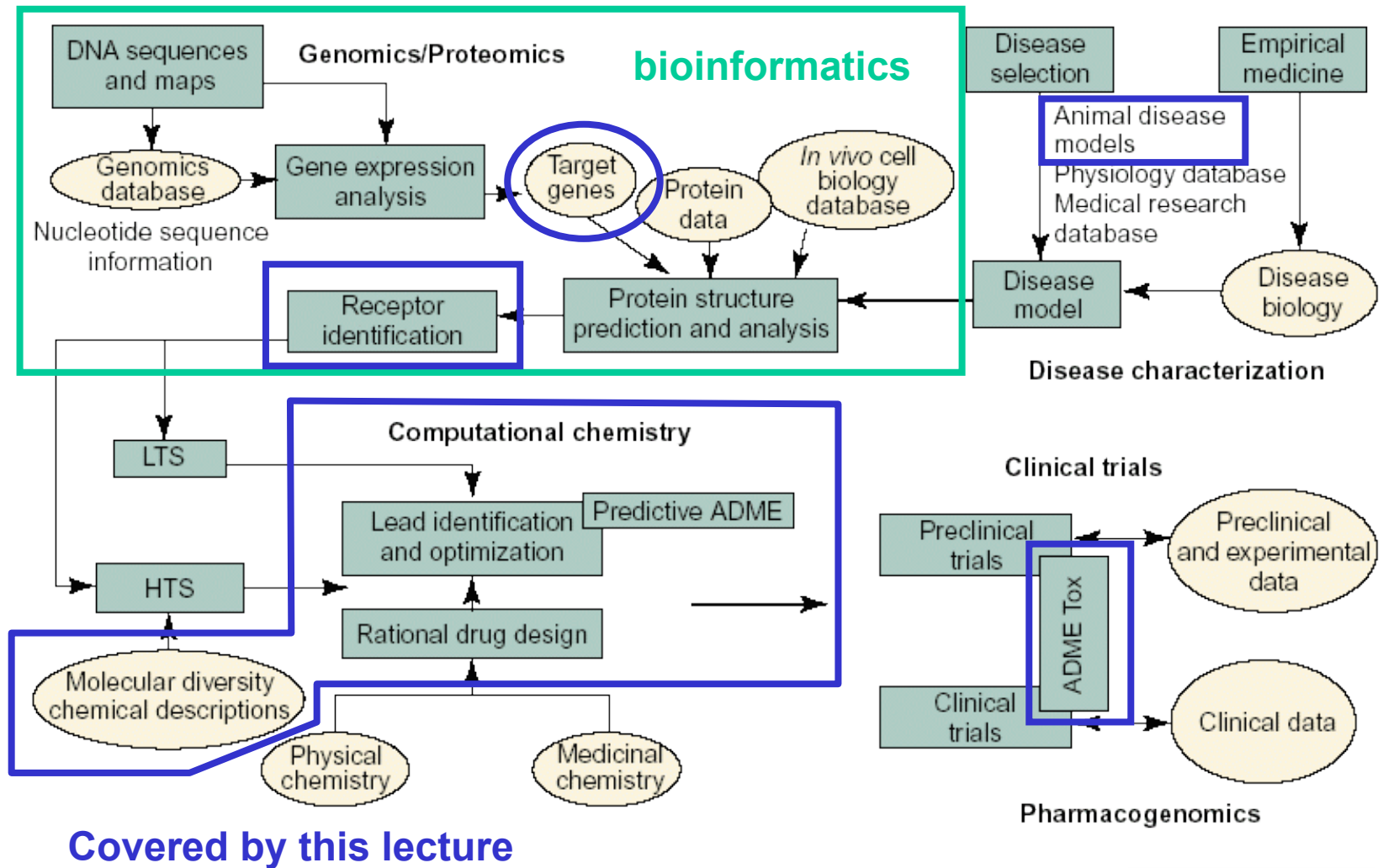
- comprehensive knowledge about all processes in the *drug discovery pipeline*
- in particular *in silico* methods of *drug design*
- performing substance queries in databases



Picture source:

<https://curesearch.org/Impact-Report-Winter-2016/images/research-pipeline.png>

Flow of information in a *drug discovery pipeline*



Drug Discovery Today

Related topics not covered by this lecture

medicinal chemistry

organic synthesis

biopharmaceutical aspects

(tissue models, non-oral administration)

immune system and monoclonal antibodies

clinical aspects

molecular modelling theory

homology modelling theory

docking basics and applications

computational chemistry

genome, proteome, metabolome, targetome,...

bioethics, legal aspects, regulatory issues, and patent law



Required knowledge

Use of tools for sequence analysis, e.g. BLAST, CLUSTALO

Use of visualizing tools, e.g. BALL, Chimera, Pymol, VMD, SPDBV

recommended prior courses:

Softwarewerkzeuge der Bioinformatik	}	Bachelor program
Computational Chemistry		
Bioinformatics I & II		
Structural Bioinformatics	}	Master program

Actual applications during the exercises and homework:

multiple sequence alignment, homology in sequences
analyzing protein-ligand interactions

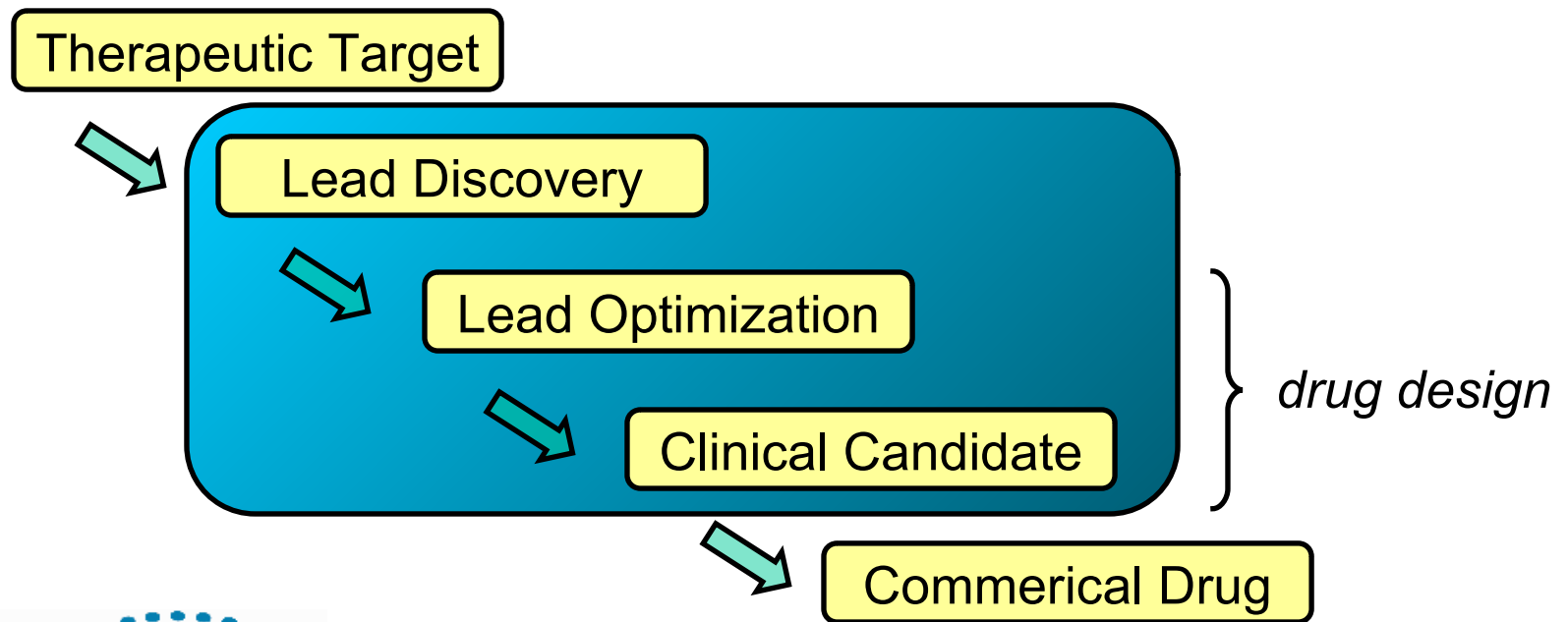
SMILES and SMARTS notation of chemical structures
using SMARTS with Open Babel

Database queries (PubChem, ChEMBL, DrugBank,
ZINC, UniProt,...)

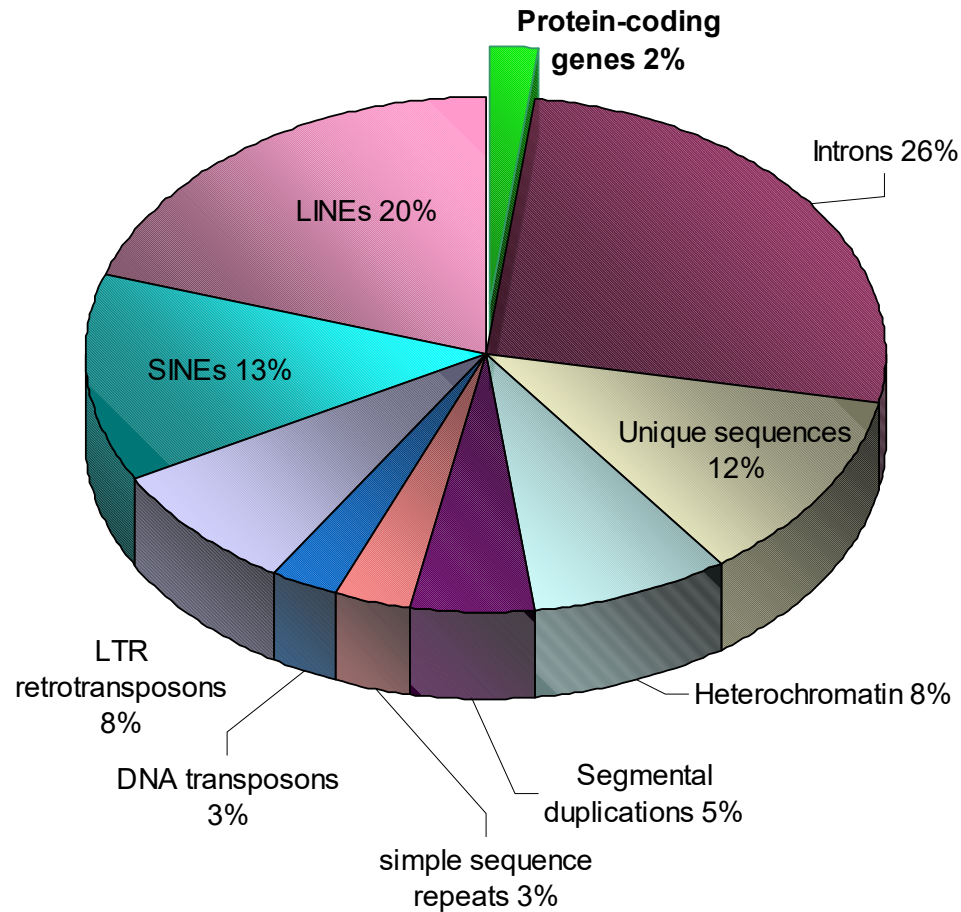


What is *drug discovery* ?

rational and targeted search for new drugs



Content of the Human Genome

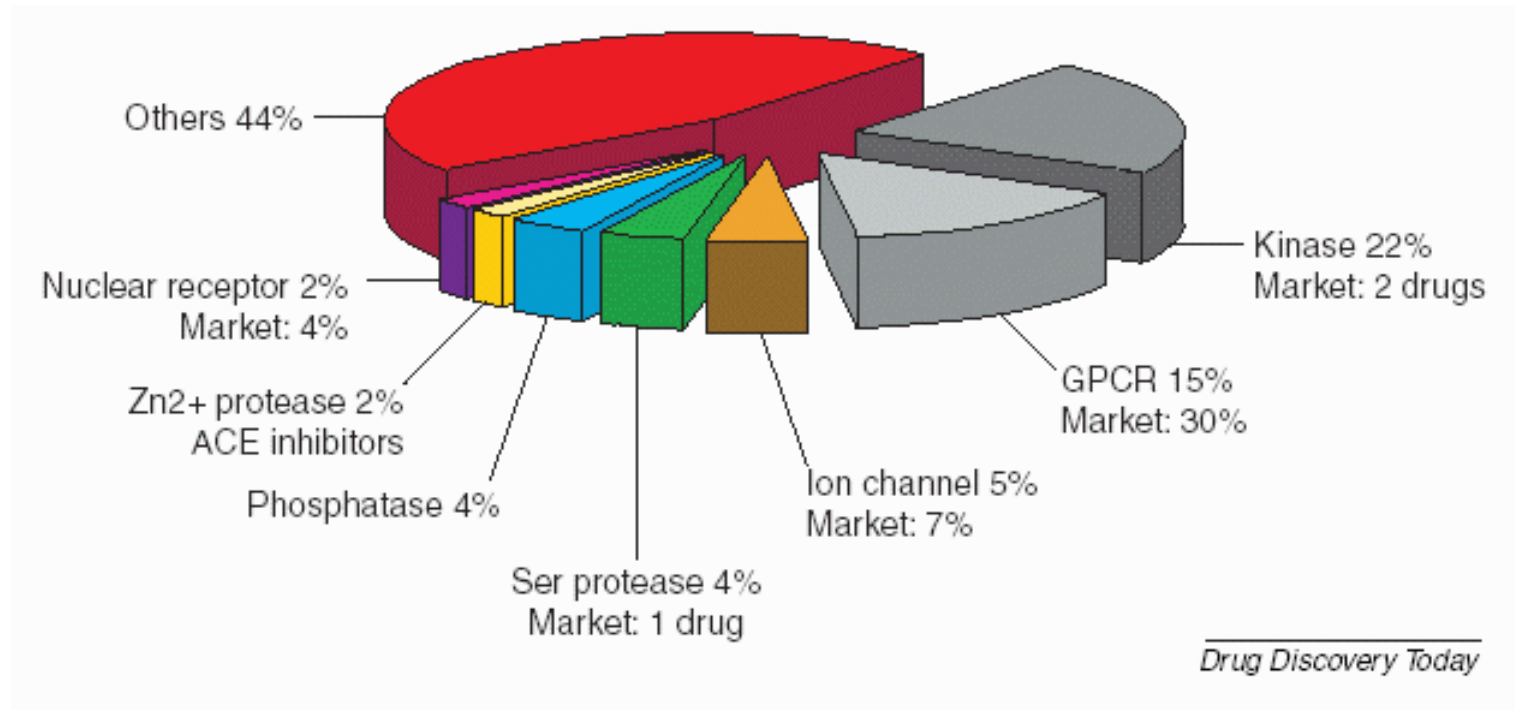


Fraction of protein-coding sequences

Data from the NHS National Genetics and Genomic Education Centre estimated in 2014.



Typical targets (I)



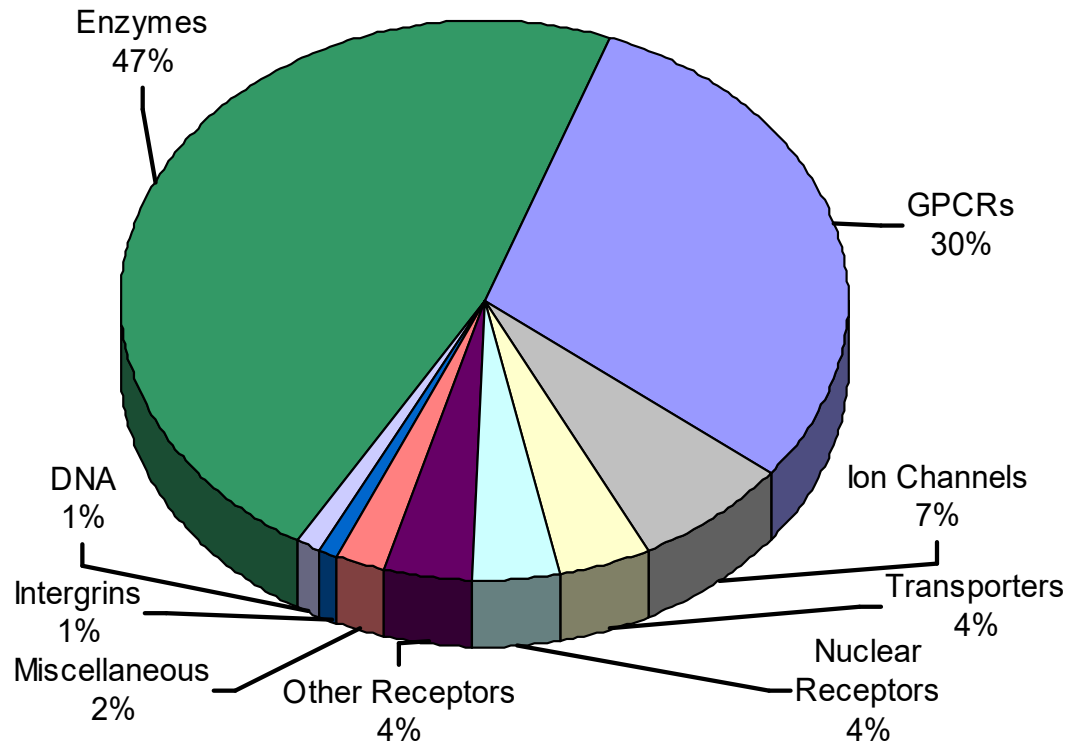
Contribution to the human genome and marketed drugs
Estimated drug targets: >3.000

Source: G.Müller *Drug Discov. Today* **8** (2002) 681



Typical targets (II)

drug targets by biochemical class



Fractional content of marketed drugs according to their biochemical targets

data: Hopkins & Groom, *Nat.Rev.Drug.Disc.* 1 (2002) 727



Preliminary schedule (lectures/topics)

1. Introduction, overview, recap of chemical structures
2. typical diseases
3. properties of drugs and their mode of action
4. Substance databases and bioisosteric compounds
5. QSAR, statistics and descriptors
6. More about QSAR and statistics
7. ADME models
8. metabolism and toxicology
9. target identification, animal models
10. cytochrome P450, polymorphisms, transporters
11. more complex diseases
malaria, obesity
12. current trends



Preliminary schedule (exercises/tutorials)

Biweekly online via MS-Teams meeting

- you will be added to the corresponding MS-Teams after registration by email (conditions and restrictions apply, see our web-site)
- discussion of the assignments
 - Hands-on tutorials about:
 - chemical structures of drugs: SMILES and SMARTS
 - substance databases and SMARTS queries
 - enzyme-ligand interactions, analysis of .pdb files
 - using PubChem and ChEMBL databases
 - ortholog targets in model organisms



Requirements to obtain the certificate and the credit points

1. You will be enrolled to the course in the moodle system.
2. Passing the two online tests (will be available in the moodle system) covering the topics of the previous assignments.
→ You don't have to hand in the assignments!
3. 50% of all accomplishable points from the final exam taking place at the end of the lecture period. If necessary, repeated written exam or oral exam (subject to the corresponding study regulation).

Applies only to students enrolled in Bioinformatics:
Please register for the final exam in the LSF timely.

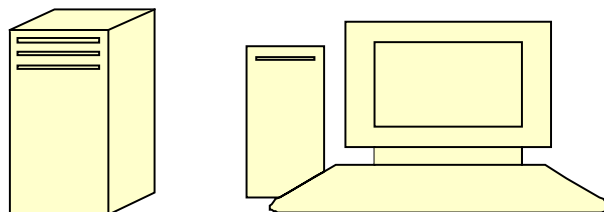


Compound Databases

Size of typical substance libraries (2026)

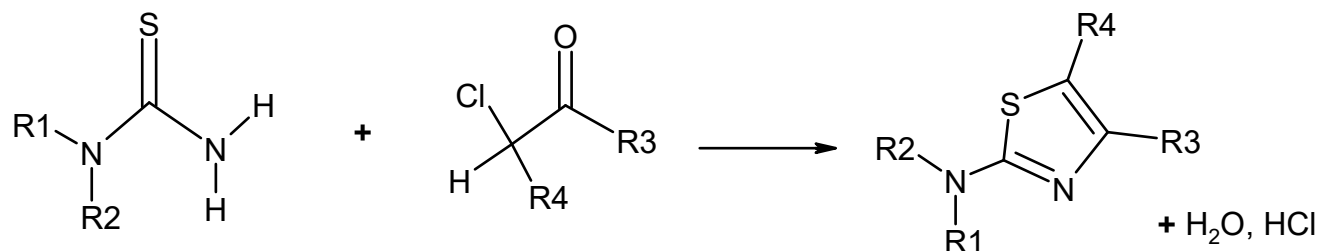
ACD	12,000,000 chemicals	} commercial
World Drug Index	80,000 compounds	
USAN	<10,000 in clinical trials	
typical virtual library	≈100,000 compounds	company, in house

PubChem	> 123,000,000 compounds	NCBI	} academic
ChEMBL	> 2,900,000 compounds	EMBL	
DrugBank	> 19,000 drugs	Uni. Alberta	
ZINC20	>750,000,000 compounds	UCSF	

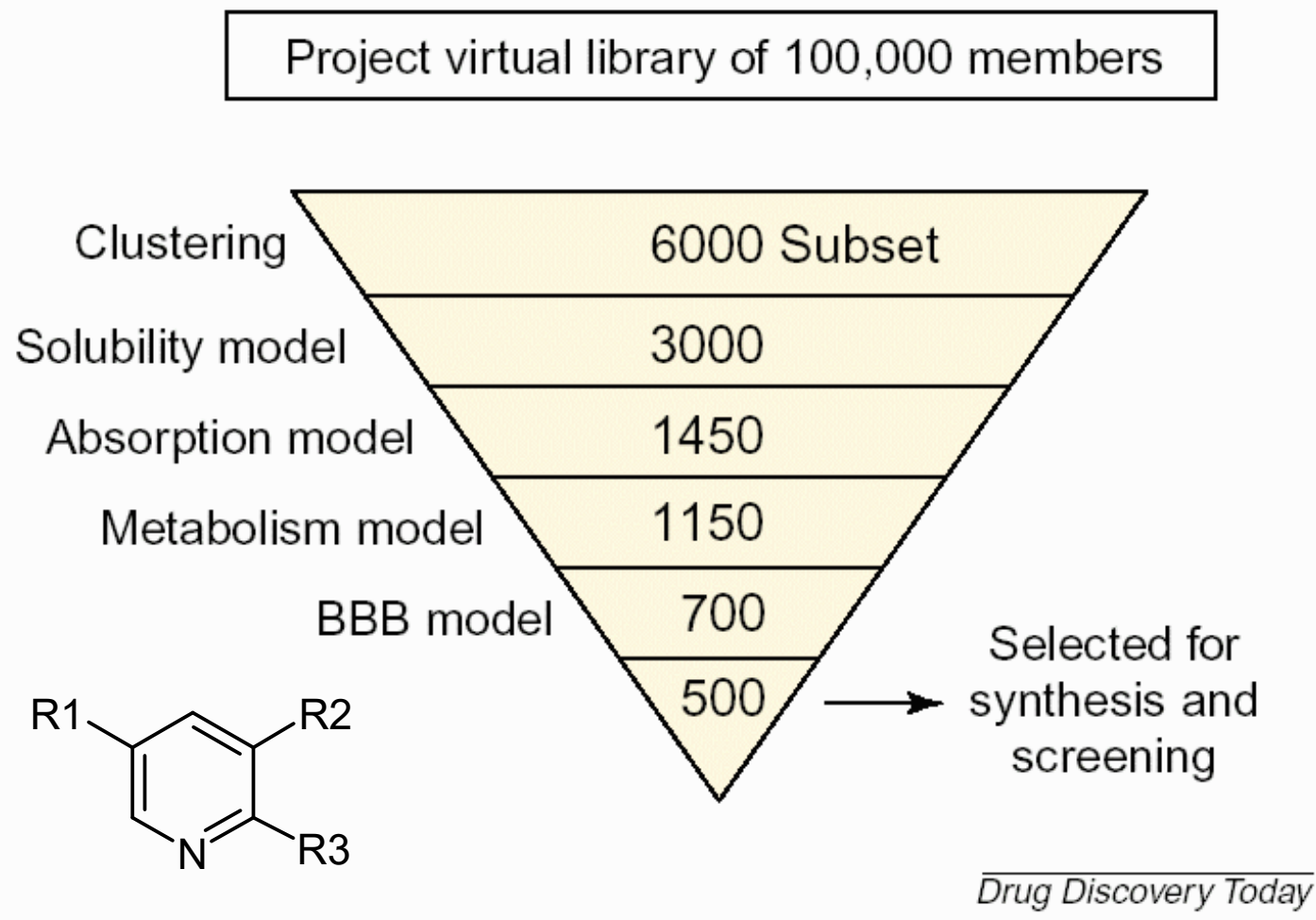


Investment per new chemical entity: >1,000,000 \$
New chemical entities per year: ca. 20 (strongly fluctuating)

Methods of Combinatorial Synthesis for High Throughput Screening (HTS)



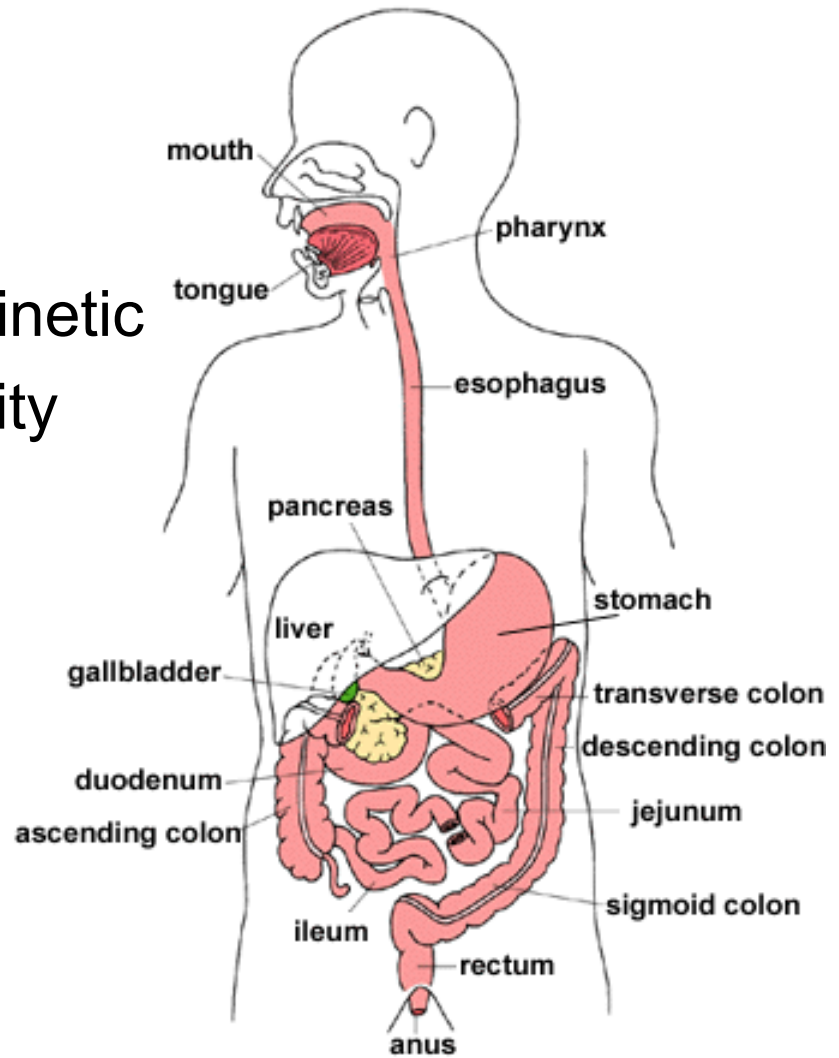
Selection of compounds for High Throughput Screening (HTS)



Predictive ADME

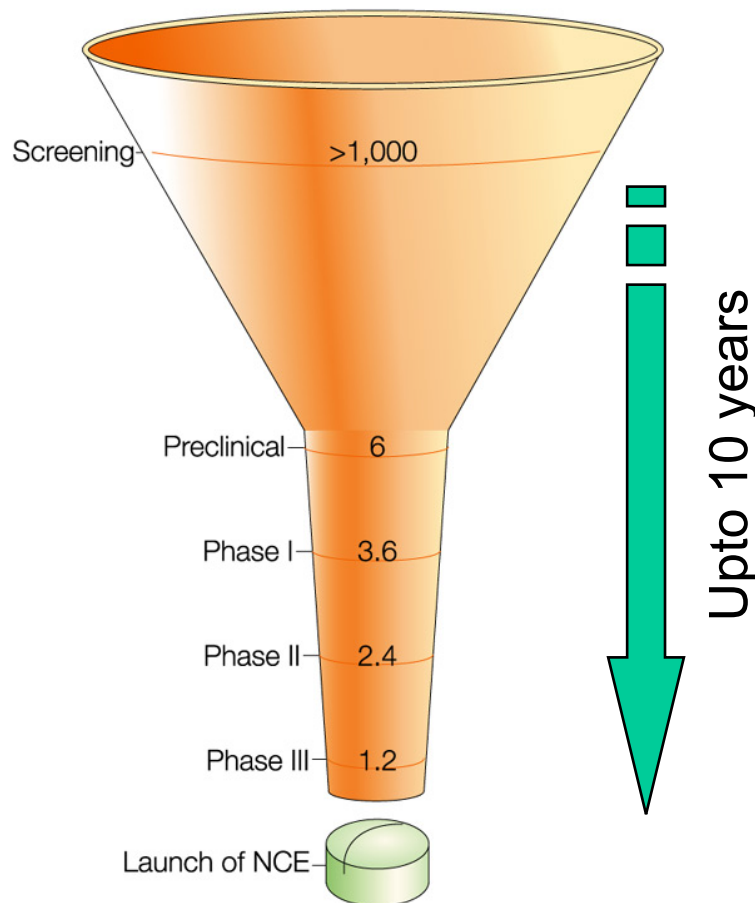
Absorption
Distribution
Metabolism
Elimination

Pharmacokinetic
Bioavailability



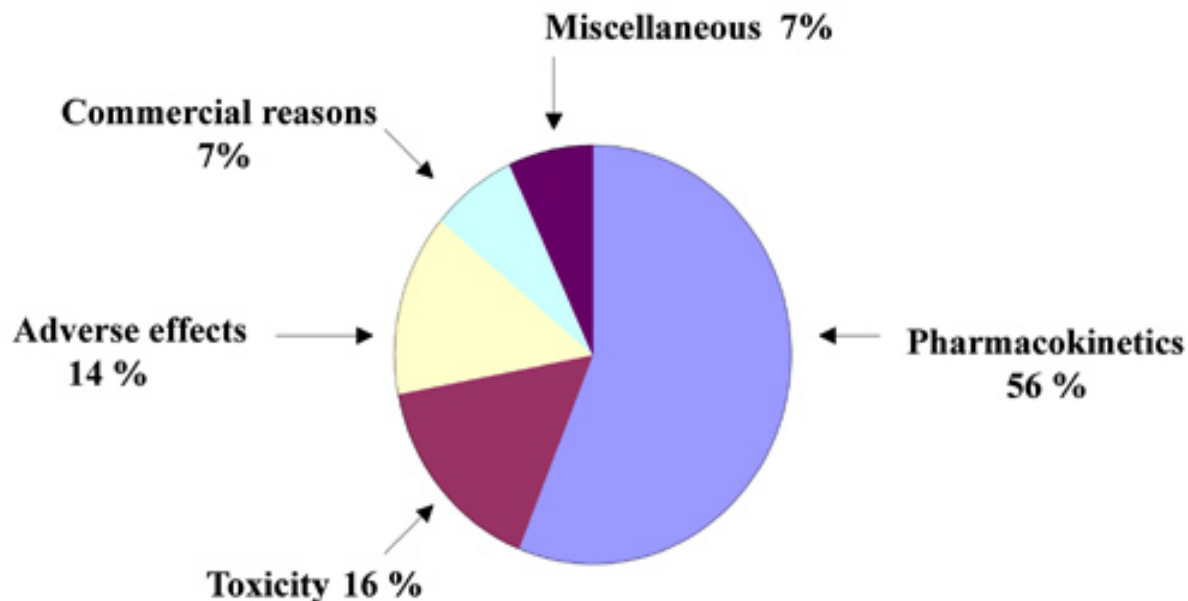
From the pipeline until the commercial launch

For each actual marketed drug (*new chemical entity, NCE*) there have been more than 1000 substances that underwent screened *in vitro*. Without the use of available computer-based ADMET filters, this number would be even larger.



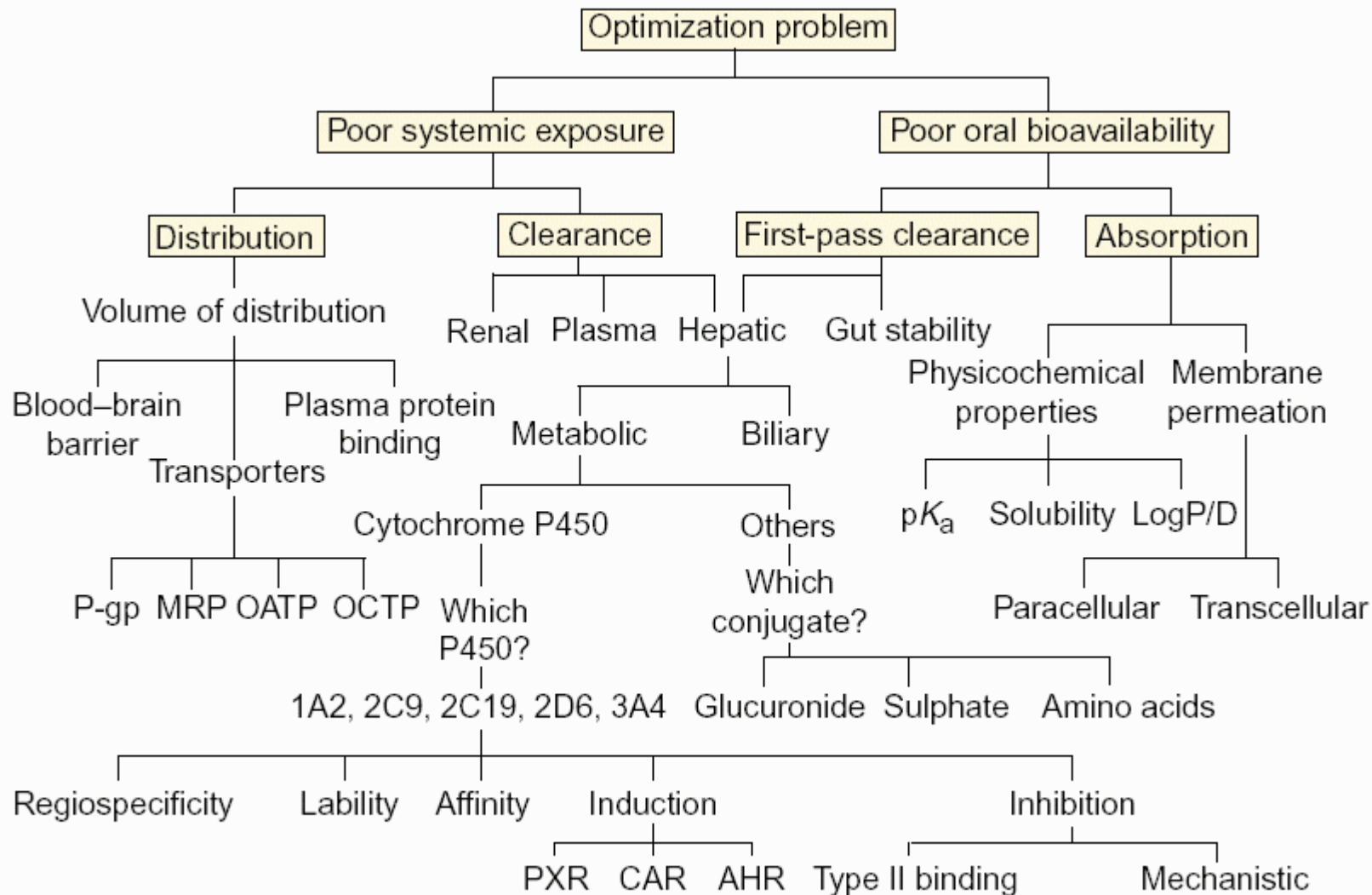
Nature Reviews | Drug Discovery

Why is the prediction of ADME parameters that important ?



Reasons that lead to failure or withdrawal of a potential drug by the mid 1990's

Pharmacokinetics and Bioavailability



Drug Discovery Today

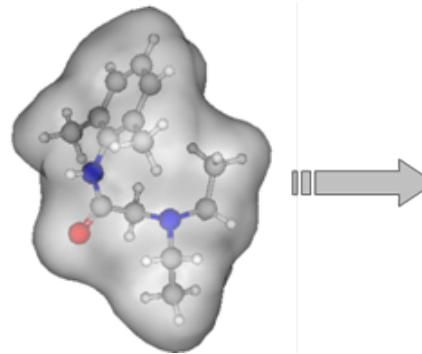
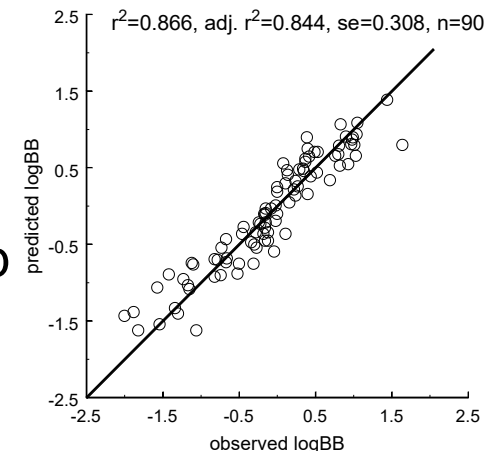
(Some) descriptors based on molecular properties used to predict ADME properties

logP water/octanol partitioning coefficient

topological indices

count of hydrogen-bonds

polar surface area

[illegible]

Multiple linear regression analysis

QSAR quantitative structure activity relationship

QSPR quantitative structure property rel.

Metabolism

(bio-)chemical reactions of xenobiotics in the body

First pass effect:

Extensive metabolization of mainly lipophilic molecules, such with $MW > 500$, or those that have a specific affinity to certain transporters, during the first passage through the liver

Phase I:

Oxidation, reduction and hydrolysis →
esp. cytochrome P450 enzymes

Phase II:

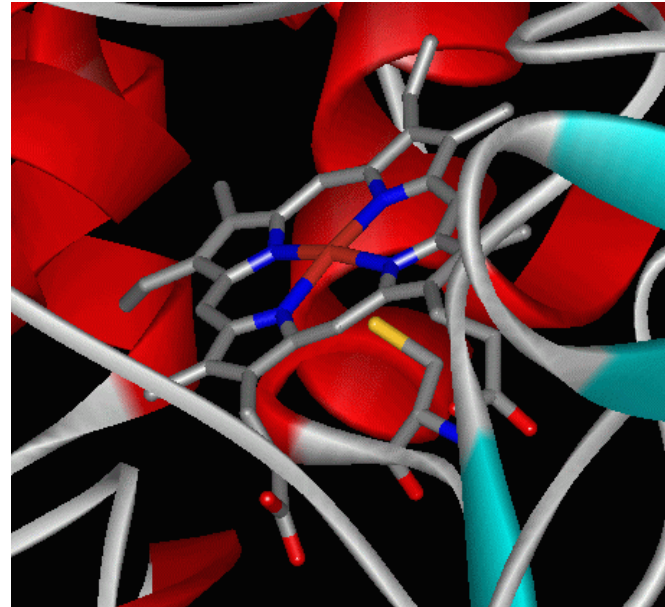
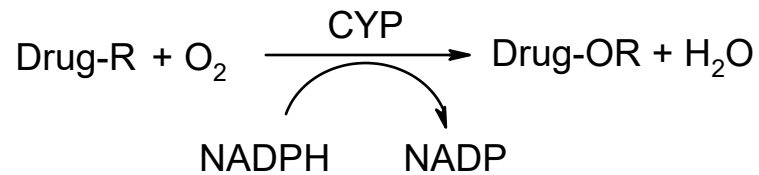
Conjugation with small molecules (e.g. glutamine)
mediated by transferases

Phase III:

elimination by transporters



Cytochrome P450 Enzymes (I)



flavin monooxygenase isoenzyme (FMO)
monoamine dehydrogenase (MAO)
aldo-keto reductase (AKR)
alcohol dehydrogenase
aldehyde oxidase

} Further phase I enzymes

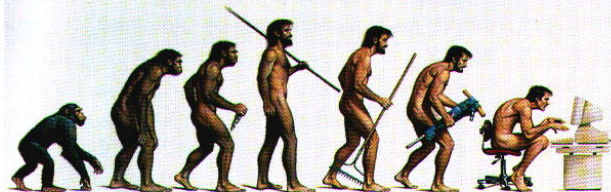
Cytochrome P450 gene families



Plants 22



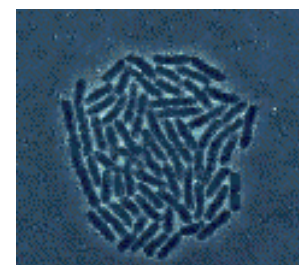
Insects 3



Human 17+



Molluscs 1



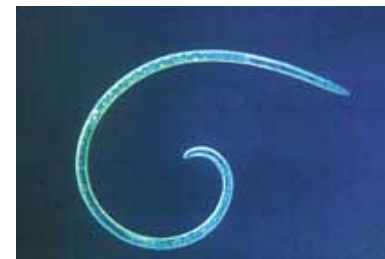
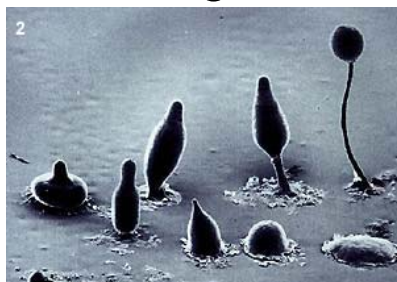
Bacteria 18

CYP450

Fungi 11

Yeasts 2

Nematodes 3



Cytochrome P450 polymorphism

„Every human is (more or less) different“

Determination of the phenotype by the actual activity or the amount of the expressed enzyme.

In contrast, the genotype is determined by the individual DNA sequence.

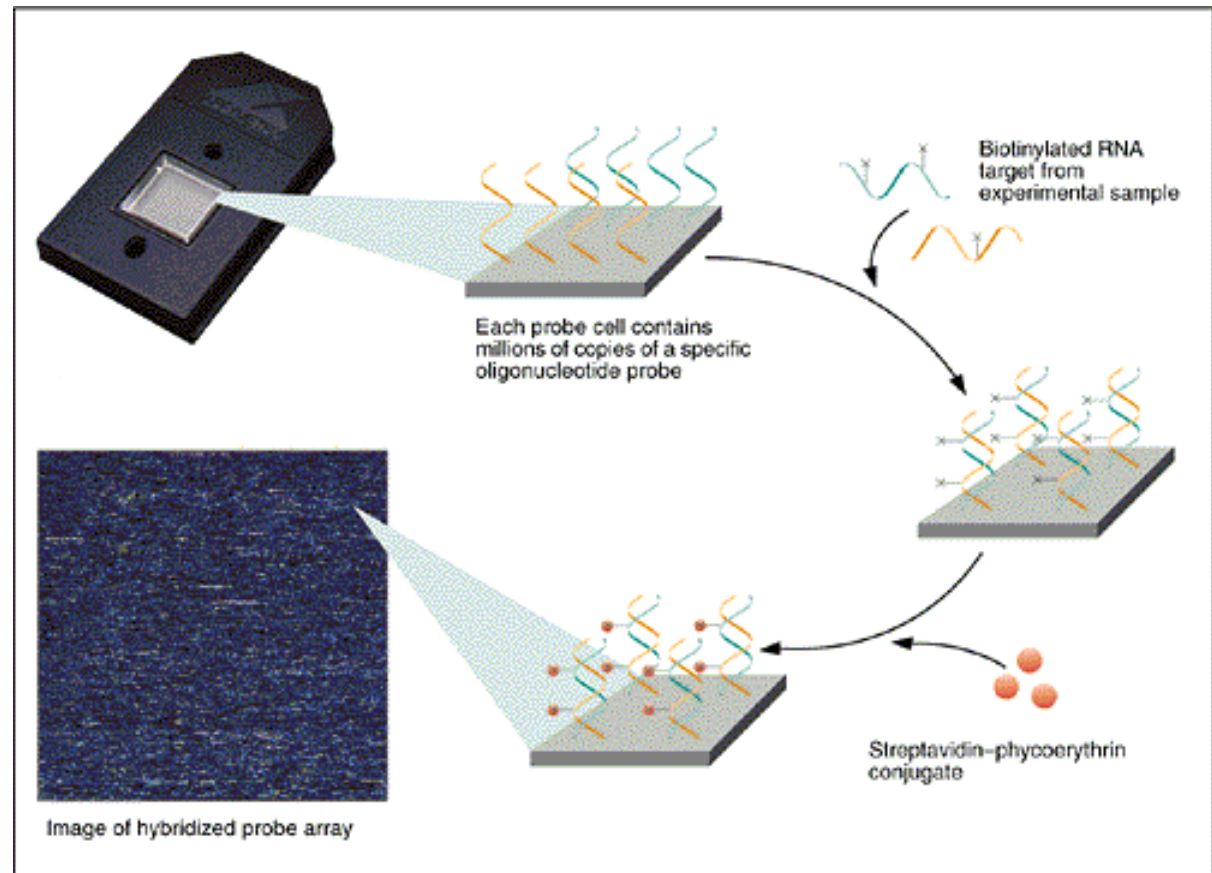
Thus, the same genotype enables several different phenotypes

According to their metabolic activity of CYP there is a classification into normal (*extensive metabolizer*), weak (*poor metabolizer*), and accelerated (*ultra-rapid metabolizer*) metabolism.

Lit: K. Nagata et al. *Drug Metabol. Pharmacokin* **3** (2002) 167

Genotyping of CYP P450 alleles

By using immobilized, synthetic copies of P450 nucleotides, the Affymetrix company (USA) has developed microarrays (gene chips) that allow the identification of all clinically relevant alleles. E.g. the AmpliChip CYP450 contains 32 variant haplotypes of CYP2D6, and 2 of CYP2C19.



Prediction of molecular properties (I)

The keynote of *rational drug design*

The general question is:

What is the connection between the biological space (activity) and the chemical space (structure)?

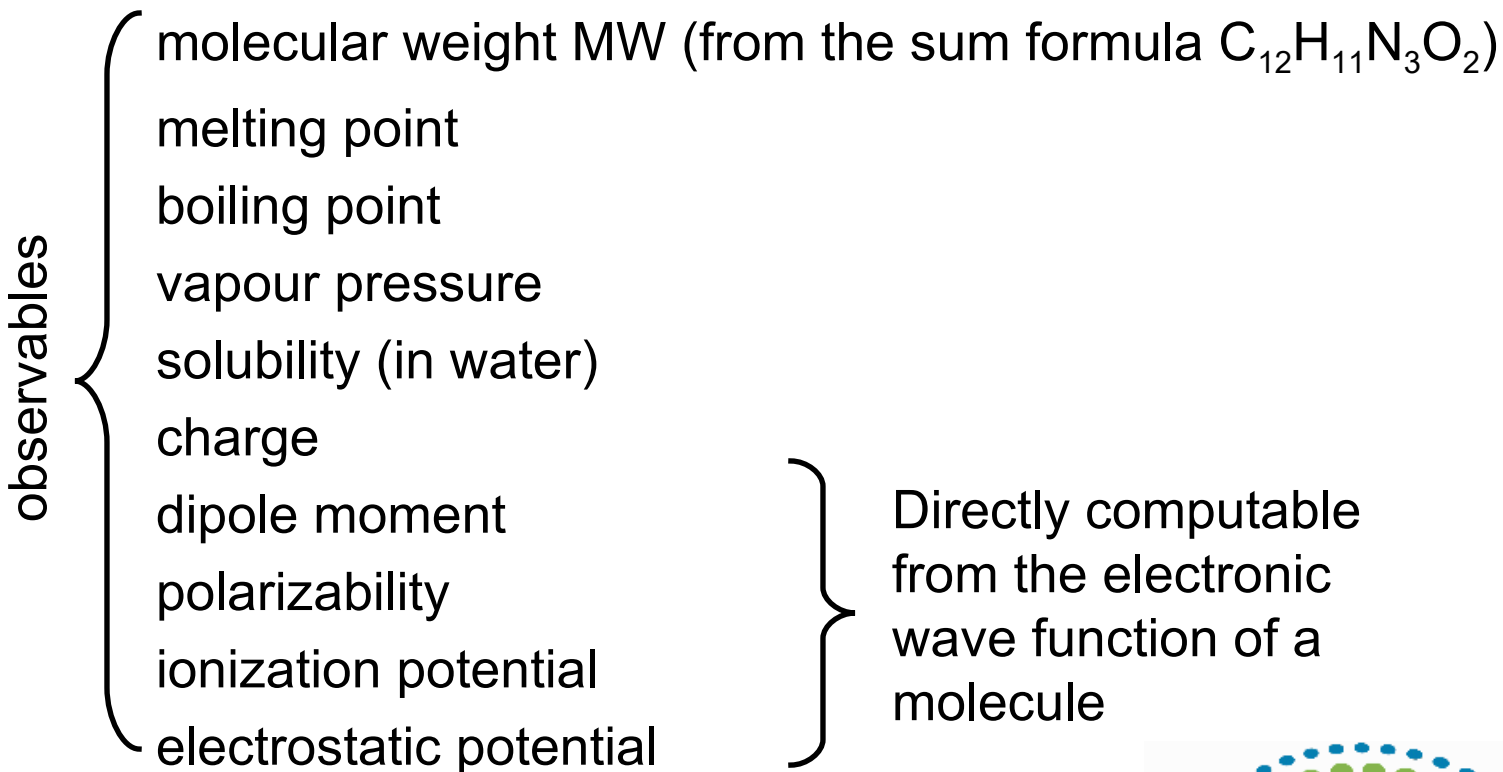
How are we able to make structure-based prediction?

- QSAR and QSRP, regression analysis
- other statistical methods, e.g. partial least square
- decision trees, machine learning algorithms, e.g. support vector machines, boosting, neural networks, deep learning, AI, ...



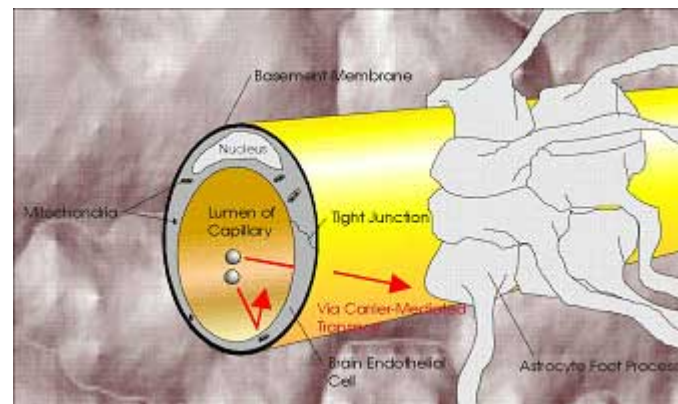
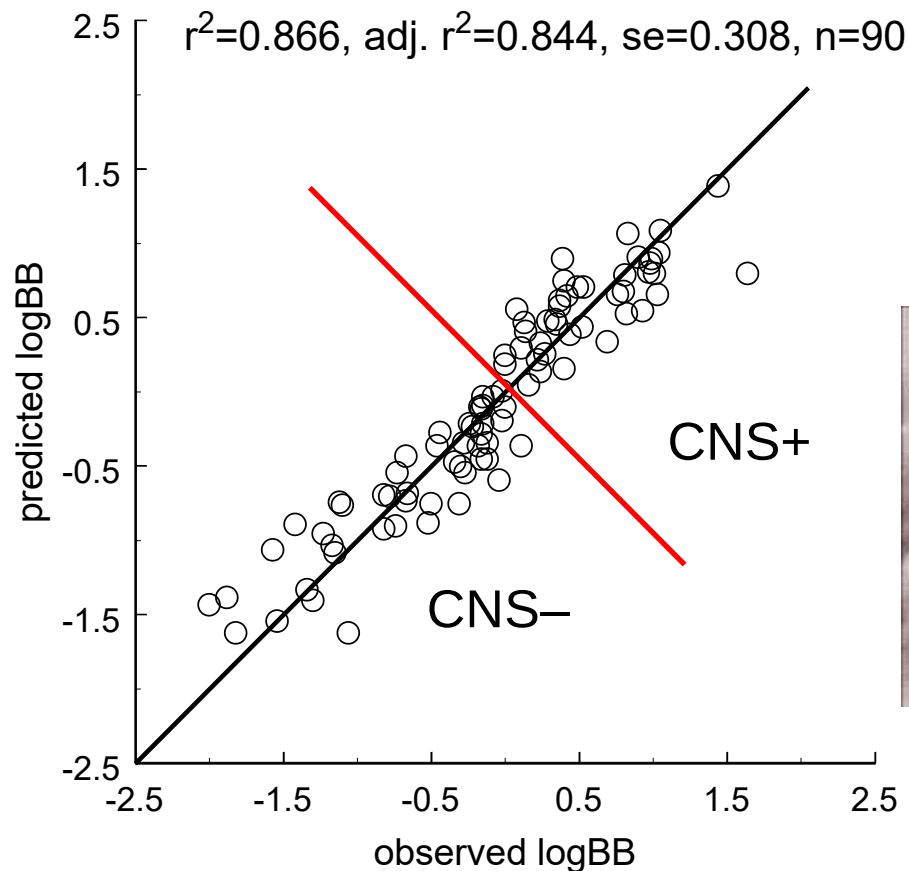
Prediction of molecular properties (II)

What are (physical) molecular properties?



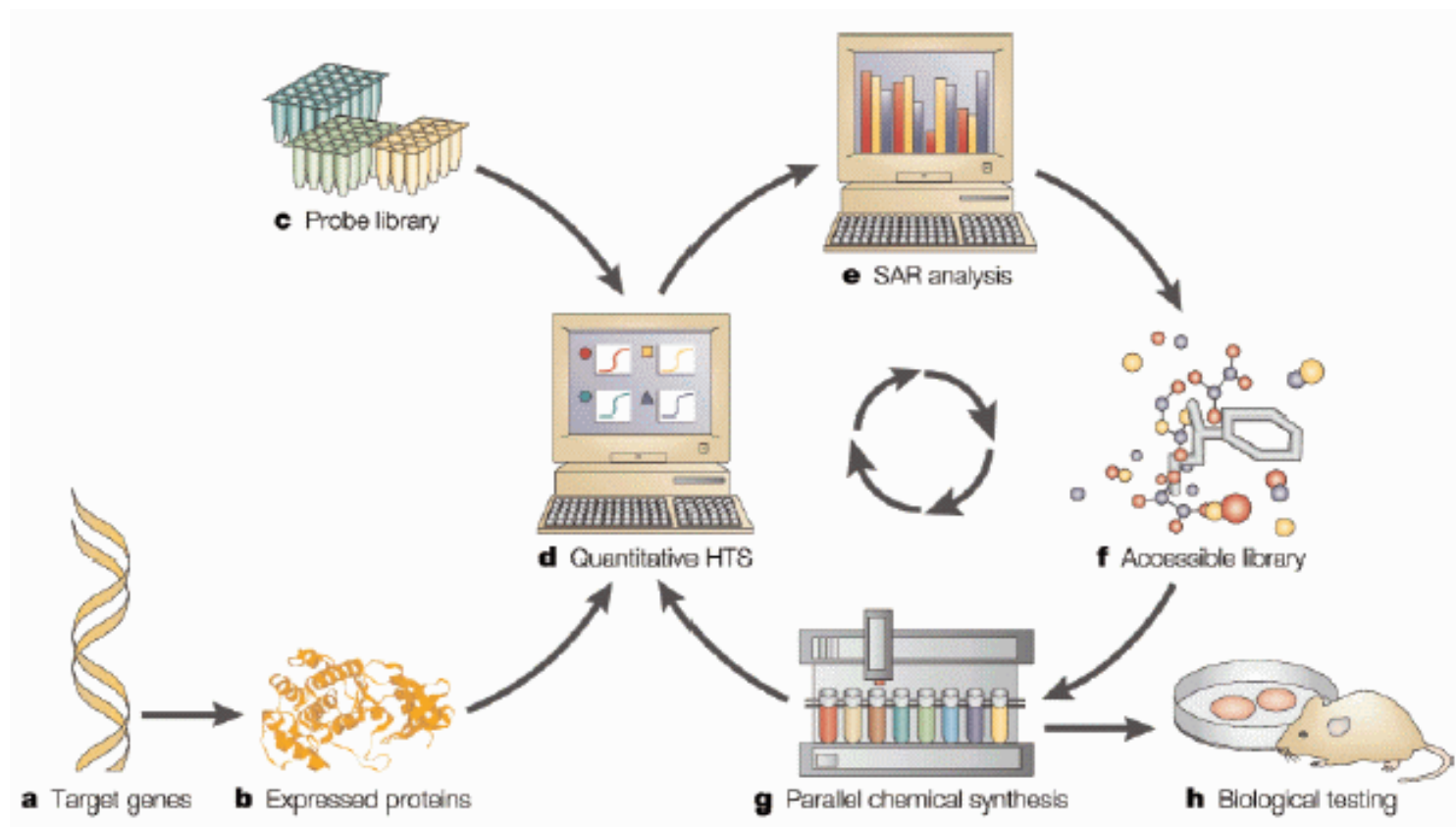
BBB-model with 12 descriptors

Descriptors mainly from QM calculations: electrostatic surface, principal components of molecular geometry, H-bond properties



Lit: M. Hutter *J.Comput.-Aided.Mol.Des.* **17** (2003) 415.

Cycle of optimization in the drug discovery pipeline

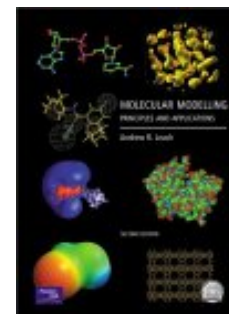


Source: D.K. Agrafiotis et al. *Nature.Rev.Drug.Discov.* **1** (2002) 337.

Accompanying books and further reading (I)

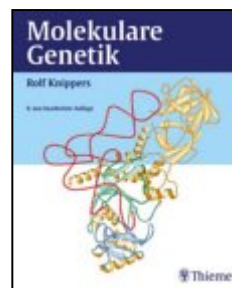
Andrew R. Leach*

Molecular Modelling. Principles and Applications
2nd edition, Prentice Hall, 2001



Rolf Knippers*

Molekulare Genetik
8. Auflage, Thieme, 2001



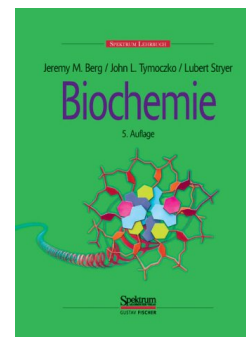
The Merck Index*

13th edition, Merck & CO., Inc., 2001



J.M. Berg, L. Stryer*

Biochemie, Spektrum Verlag
Biochemistry, W.H. Freeman & Co Ltd.



*Available in the „Semesterapparat“



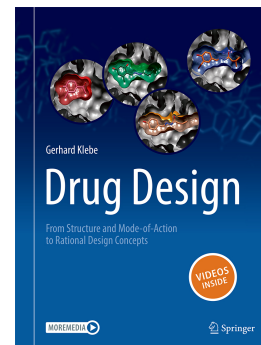
Accompanying books and further reading (II)

Gerhard Klebe

Drug Design, Springer, 2025

„Wirkstoffdesign“ 3. Auflage 2023

online verfügbar über SULB

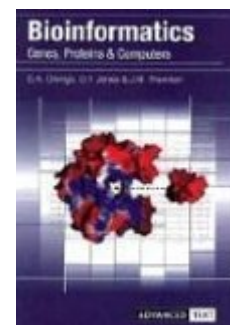


C.A. Orengo, D.T. Jones, J.M. Thornton*

Bioinformatics

Genes, Proteins & Computers

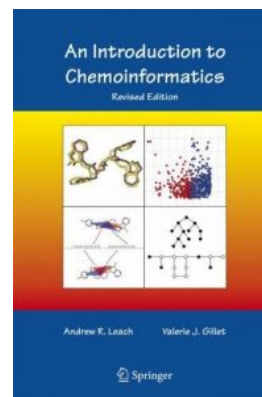
1st ed., Bios Scientific Publishers, 2003



A.R. Leach, V. Gillet*

An Introduction to Chemoinformatics

revised ed., Springer, 2007

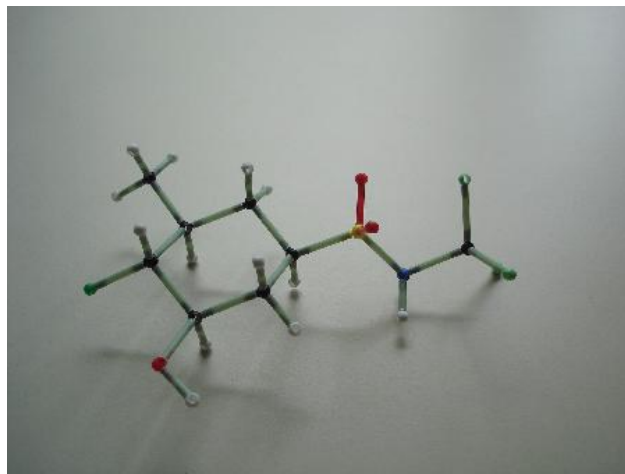
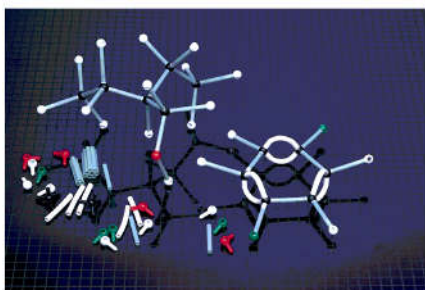


*Available in the „Semesterapparat“
located at the library in E2.3



Further hands-on tools

Molecular model sets / Molekülbaukasten



Commerically available at various price ranges

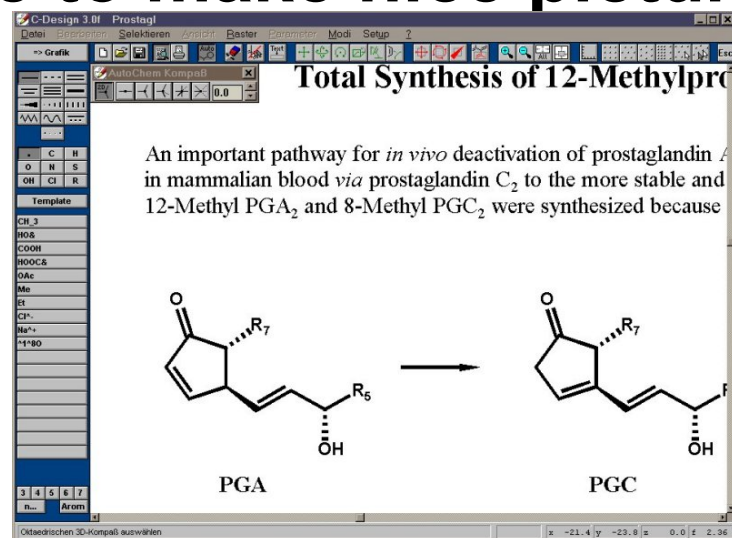


Other useful software to make nice pictures

Chemical structures
and other objects:

C-Design 3.0f

Windows-Platform

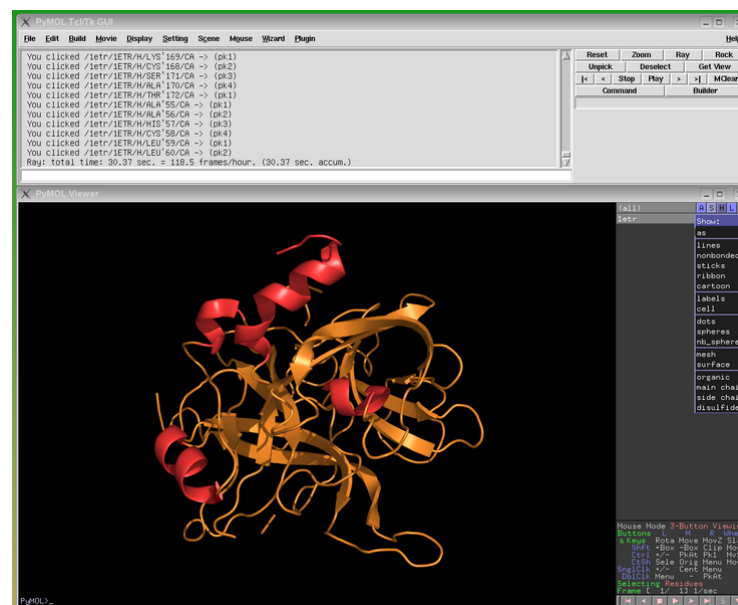


Protein structures:

PyMOL www.pymol.org

Chimera

Linux, Mac OS X, Windows



1st assignment (I)

Refer to a prescription medicine of your own choice

Write down the active ingredient

Try to find out its molecular structure:

<https://pubchem.ncbi.nlm.nih.gov/>

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

PubChem About Blog Submit Contact

PubChem presents at the American Chemical Society National Meeting in San Diego (August 25-29, 2019) [Read More >](#)

Explore Chemistry

Quickly find chemical information from authoritative sources

capoten

Try aspirin EGFR C9H8O4 57-27-2 C1=CC=C(C=C1)C=O InChI=1S/C3H6O/c1-3(2)4/h1-2H3

☐ Use Entrez ☒ Compounds ☐ Substances ☐ BioAssays

1st assignment (II)

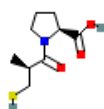
SEARCH FOR

capoten



Treating this as a text search.

COMPOUND BEST MATCH



Captopril; 62571-86-2; L-Captopril; Capoten; Lopirin; Captopryl; Cesplon; Tensoprel; ...

Compound CID: 44093

MF: $C_9H_{15}NO_3S$ MW: 217.29g/mol

InChIKey: FAKRSMQSSFJEIM-RQJHMYQMSA-N

IUPAC Name: (2S)-1-[(2S)-2-methyl-3-sulfanylpropanoyl]pyrrolidine-2-carboxylic acid

Create Date: 2005-06-24

[Summary](#)

[Similar Structures Search](#)

[Related Records](#)

[PubMed \(MeSH Keyword\)](#)

Compounds
(2)

Substances
(19)

Literature
(80)

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that annotations text from compound summary pages is not searched. [Read More...](#)

1st assignment (III)

PubChem Captopril (Compound)

7 Drug and Medication Information

7.1 Drug Indication

For the treatment of essential or renovascular hypertension (usually administered with other drugs, particularly thiazide diuretics). May be used to treat congestive heart failure in combination with other drugs (e.g. cardiac glycosides, diuretics, β-adrenergic blockers). May improve survival in patients with left ventricular dysfunction following myocardial infarction. May be used to treat nephropathy, including diabetic nephropathy.

▶ from DrugBank

Treatment of heart failure

▶ from European Medicines Agency (EMA)

7.2 LiverTox Summary

Captopril is an angiotensin-converting enzyme (ACE) inhibitor used in the therapy of hypertension and heart failure. Captopril is associated with a low rate of transient serum aminotransferase elevations and has been linked to rare instances of acute liver injury.

▶ from LiverTox

7.3 Drug Classes

Angiotensin-Converting Enzyme Inhibitors

Share Tweet Email

Cite Download

CONTENTS

- 1 Spectral Information
- 5 Related Records
- 6 Chemical Vendors
- 7 Drug and Medication Information
- 8 Pharmacology and Biochemistry
- 9 Use and Manufacturing
- 10 Identification
- 11 Safety and Hazards
- 12 Toxicity
- 13 Literature
- 14 Patents
- 15 Biomolecular Interactions and Pathways
- 16 Biological Test Results
- 17 Classification
- 18 Information Sources

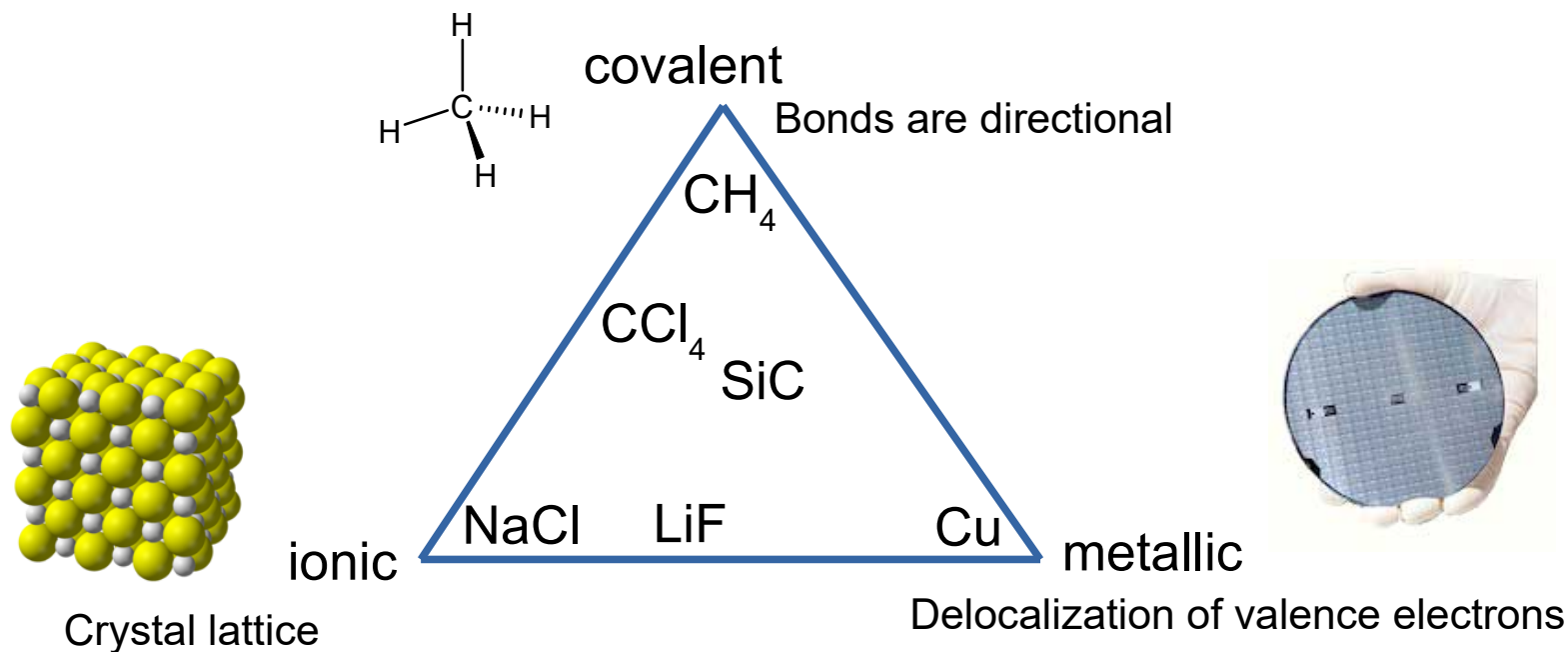
Explain why the medicine has a completely different name compared to the actual substance.

Try to find out some information about its actual *molecular target* (here: Angiotensin-Coverting Enzyme) e.g. using Wikipedia

Recapitulation: The Chemical Bond

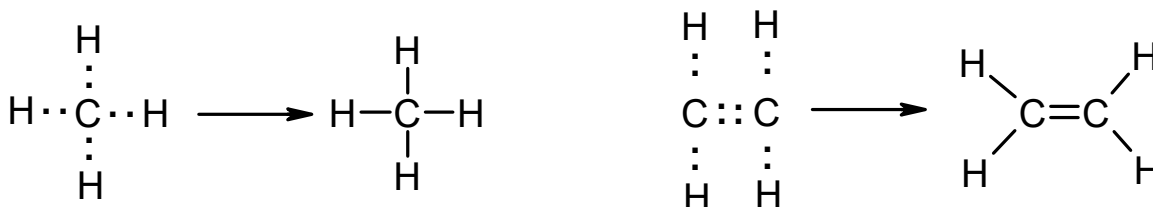
Entirely covalent, ionic, or metallic bonds between atoms are the extrem cases. More often the actual kind of bonding is somewhere in between and is subject to the difference in electronegativity of the atoms/elements involved.

→ Uneven distribution of electrons between the atoms



Representation of chemical structures (I)

The valence electrons of the atoms are pairwise grouped together



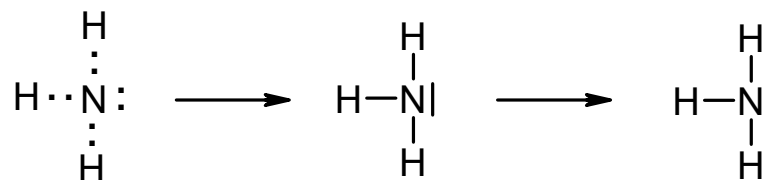
Such Lewis structures reflect covalent bonds between atoms in a molecule.

Therefore any molecule can be regarded as graph with the atoms being the nodes and the bonds as vertices.

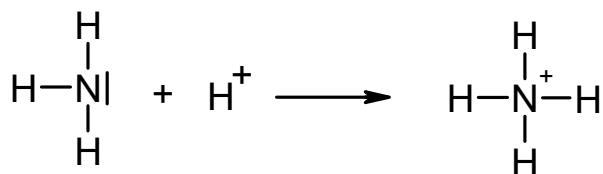


Representation of chemical structures (II)

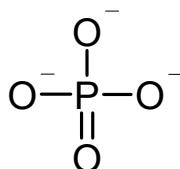
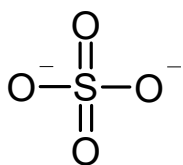
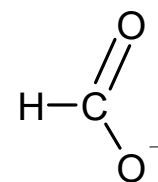
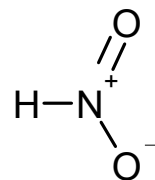
(electron) lone pairs are often not shown for visual clarity



octet rule and hypervalent atoms

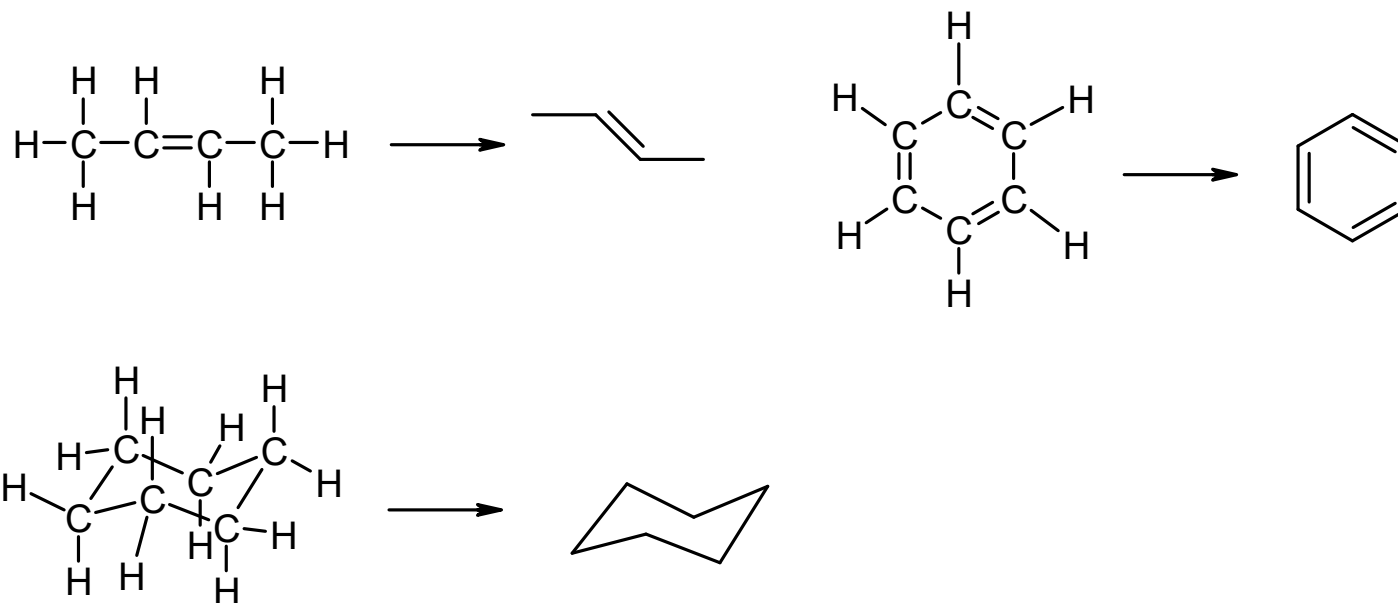


Equal bond lengths!



Representation of chemical structures (III)

Also carbon atoms are often omitted

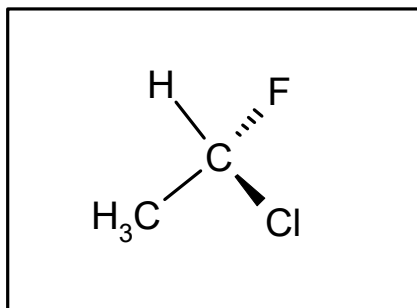


Corners and end of lines denote carbon atoms saturated with the appropriate number of hydrogen atoms



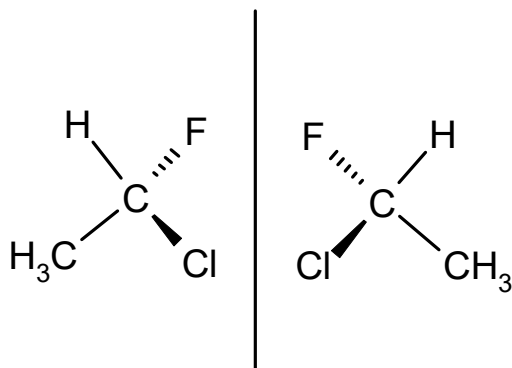
Representation of chemical structures (IV)

Stereochemistry and chirality



Solid wedges denote atoms in front of the plane, dashed wedges denote atoms behind

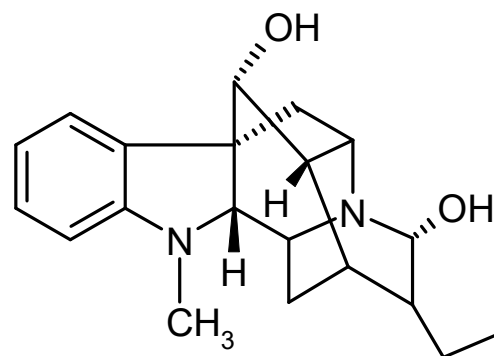
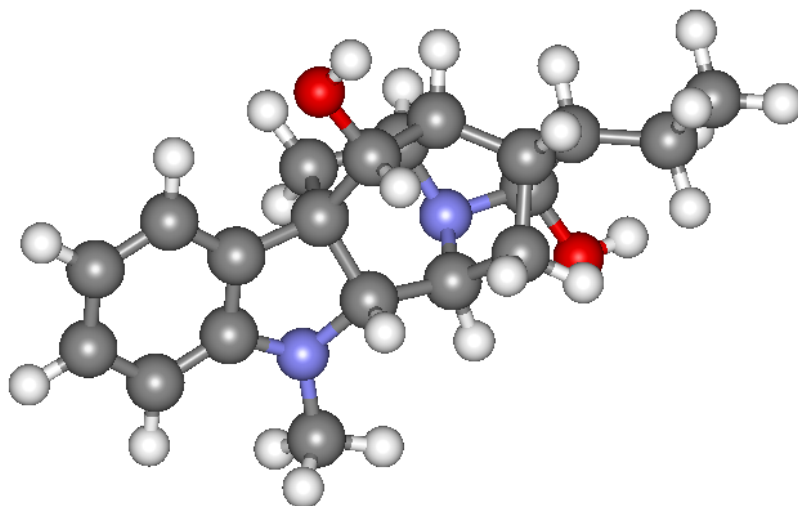
Four different substituents at a carbon atom cause chirality. Similar: silicon, sulfur, phosphorus,...



Examples: Most amino acids, sugars, many natural compounds,...

Representation of chemical structures (V)

Particular for more complex molecules, these 2D structural drawings provide more clarity than a picture of an actual 3D representation does.



Exercise: Construct this molecule using a molecular model set. Specify the chiral carbon atoms.



Bond distances and bond dissociation energies (I)

bond distance [Å] D_0 [kJ/mol] (homolytic cleavage)

H–H 0.742 432 e.g. $\text{H}\cdot \cdot \text{H}$

C–H 1.09 ± 0.01 411 ± 7

C–C 1.54 345

C=C $1.34 - 1.40^*$ 602 ± 21 *aromatic bond

$\text{C}\equiv\text{C}$ 1.20 835

C–N 1.47 305

C=N 1.35 615

$\text{C}\equiv\text{N}$ 1.16 887

C–O 1.43 358

C=O 1.20 526

C–Si 1.85 318

C–P 1.84 264

C–S 1.82 272

C=S 1.60 577 ± 21

longer
←

H							He
Li	Be	B	C	N	O	F	Ne
Na	Mg	Al	Si	P	S	Cl	Ar
K	Ca	Ga	Ge	As	Se	Br	Kr
Rb	Sr	In	Sn	Sb	Te	I	Xn
Cs	Ba	Tl	Pb	Bi	Po	At	Rn

longer,
weaker
↓

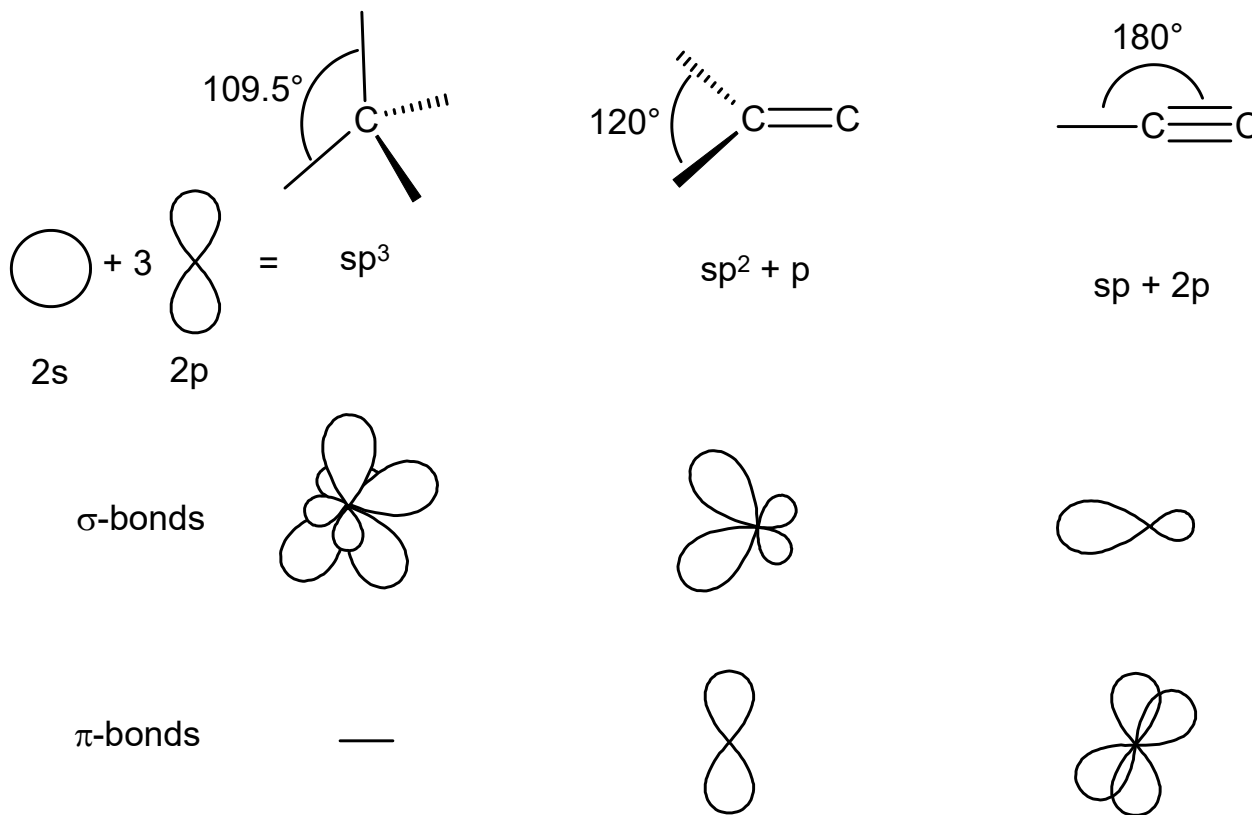
Adapted from: J.E.Huheey
Inorganic Chemistry, Wiley.

Bond distances and bond dissociation energies (II)

bond	distance [Å]	D ₀ [kJ/mol]	
C–F	1.35	485	
C–Cl	1.77	327	
C–Br	1.94	285	
C–I	2.14	213	
C–H	1.09	411	non-polar hydrogen
O–H	0.96	459	} polar hydrogens, exchangable in polar solvents, e.g. water
N–H	1.01	386 ± 8	
S–H	1.34	363 ± 5	
N–N	1.45	247 ± 13	reason:
N=N	1.25	418	N, O, and S are more electronegative than C;
N–O	1.40	201	
N=O	1.21	607	heterolytic cleavage that leads to ions e.g. O [−] H ⁺
P–O	1.63	≈335	
P=O	≈1.50	≈544	

Bond angles (I)

Strongly depend on the hybridization



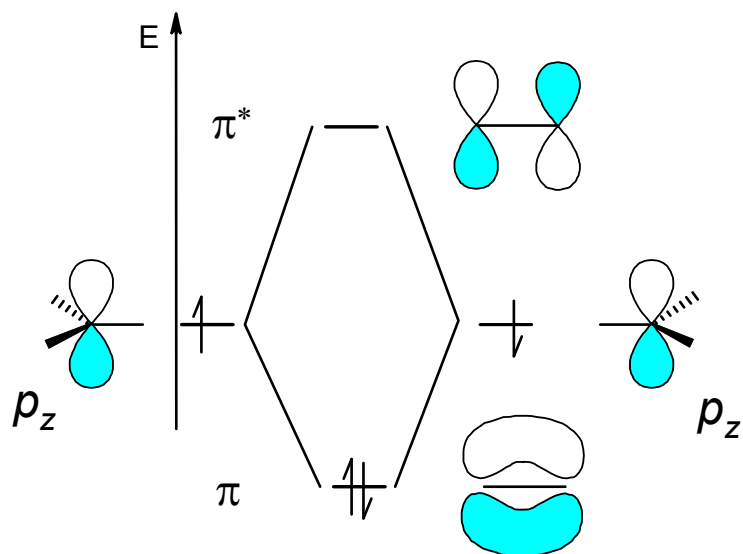
The C–C σ -bond is formed by overlap of the 1s orbitals

These are hybridized atomic orbitals. Do not confuse with molecular orbitals (=linear combination of atomic orbitals)

Molecular Orbitals

MO = linear combination of atomic orbitals (LCAO)

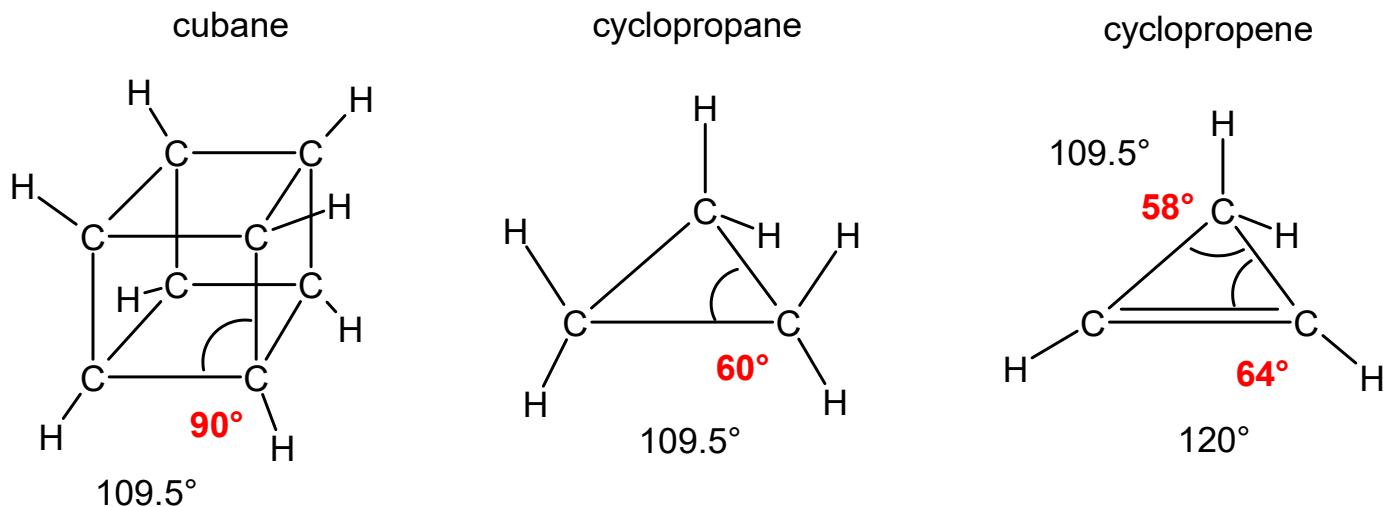
π -bond of ethylene $\text{H}_2\text{C}=\text{CH}_2$



The two combinations usually result in one bonding and one anti-bonding MO

Bond angles (II)

Extreme deviations from ideal bond angles

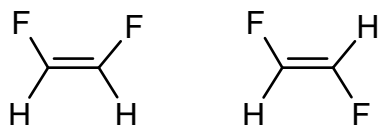


gives rise to strain energy in small rings

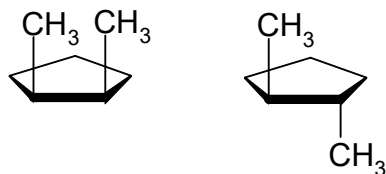
→ problems in force fields. More than one atom type for each hybridization state (e.g. Csp^3) necessary, or additional types for such small ring systems.

Stereoisomers

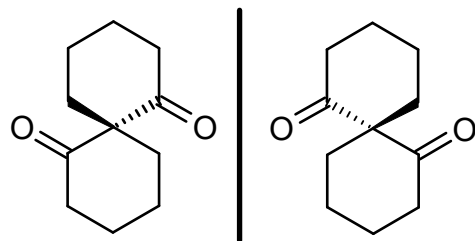
Gives rise to different orientation in 3D-space despite identical 2D-topology



cis/trans form at double bonds



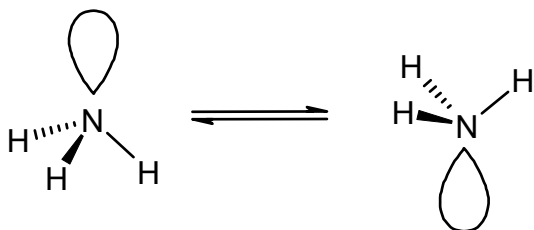
cis-trans isomers in ring systems



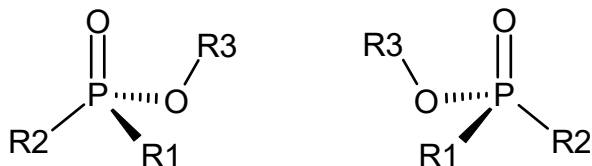
Chiral spiranes

Chiral atoms

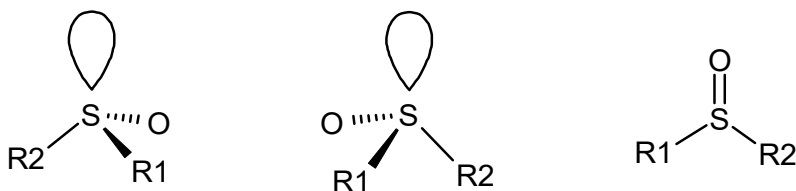
Further elements showing chirality/stereochemistry
(the lone electron pair behaves like a substituent)



Fast exchange at room temperature,
but slow at 77K



phosphorus inverts even slower

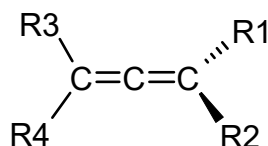
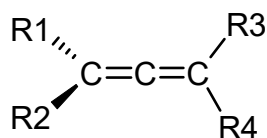


sulfoxides, sulfinic esters, etc
→ Different compounds at
room temperature

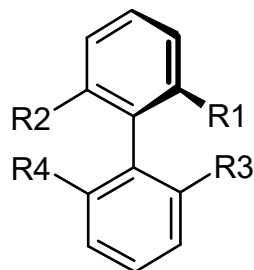
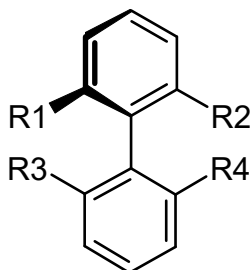
Furthermore: As, Si, ..., compounds with transition elements,
esp. octahedral and square planar metal complexes e.g. Pt

Axial Chirality

Even without a chiral center, stereochemistry can occur due to aggregated double bonds or steric hindrance caused by bulky groups.



Axial chirality in allenes

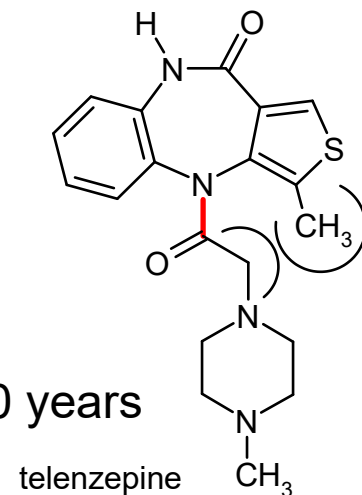


Atropisomers in ortho substituted biaryls.

Recent examples: daraxonrasib, navlimethostat, sosimerasib, zoldonrasib.

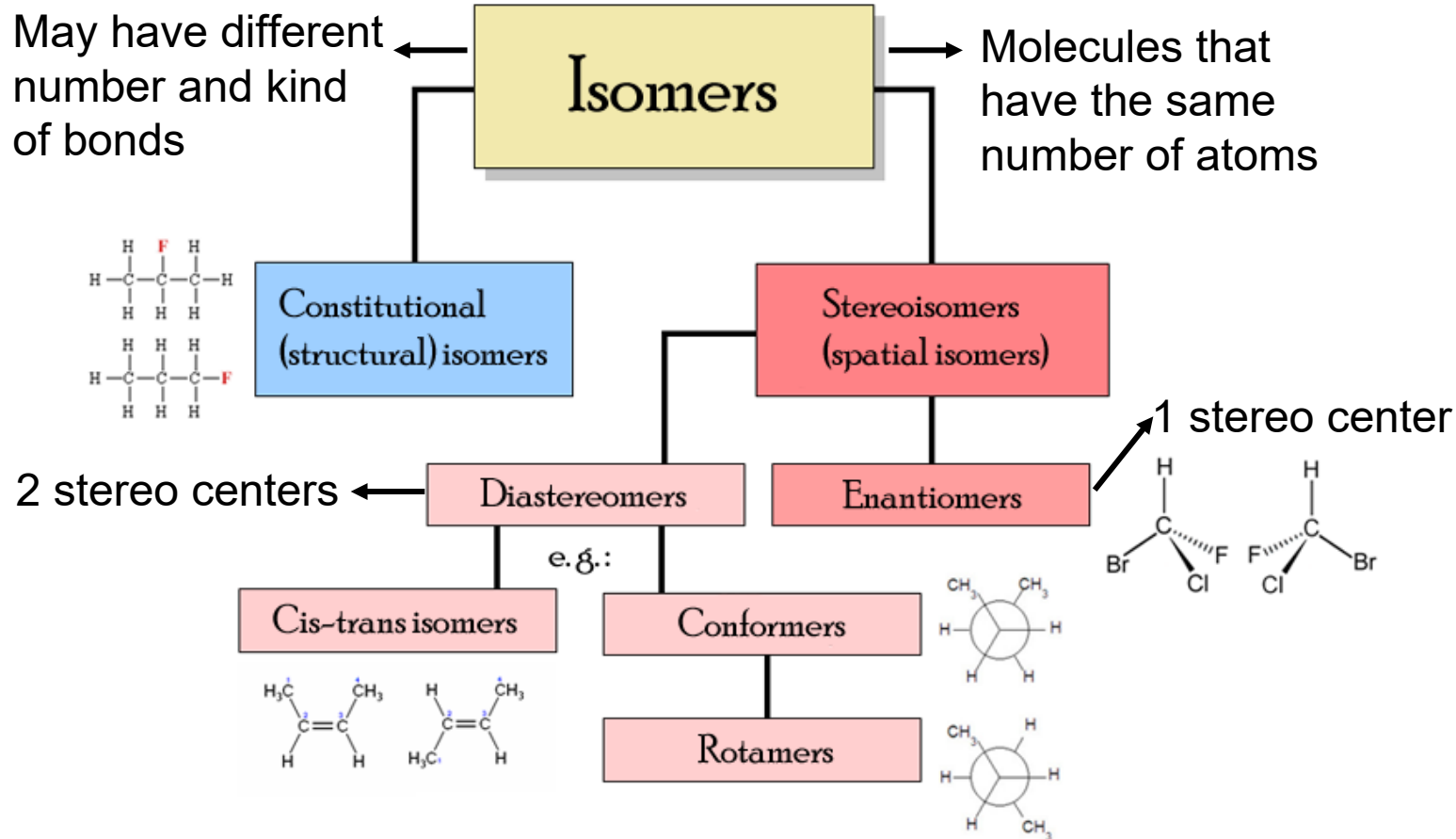
→ correct isomer needed for docking

Half-life of racemization > 1000 years



telenzepine

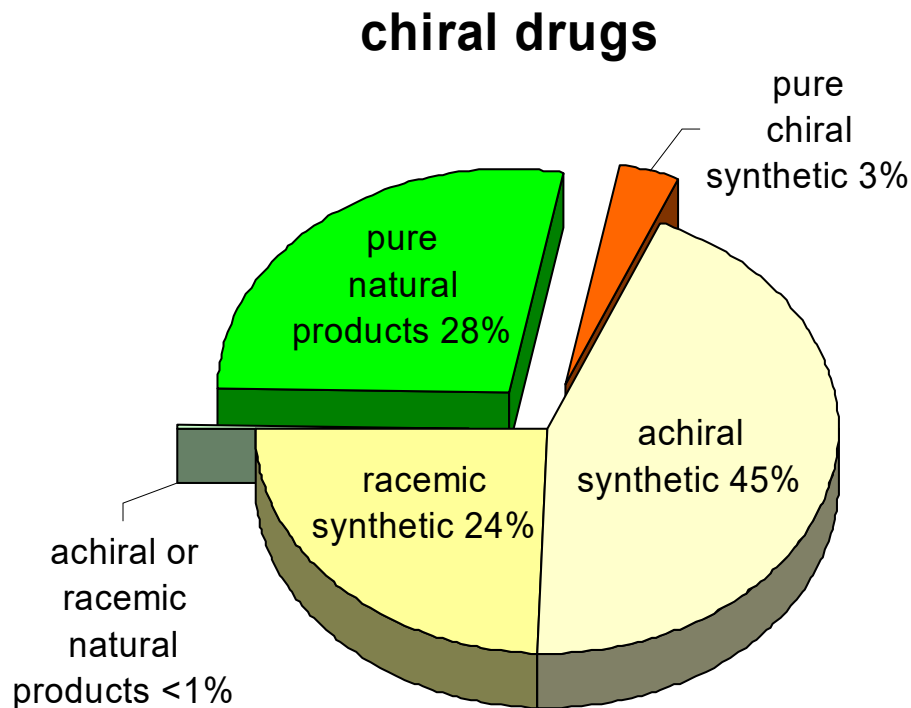
Isomers



Source: enhanced from wikipedia

Question: Which kind of computational method(s) allow(s) to calculate differences in energy between the respective isomers ?

Is stereochemistry important ?



Data from 1982: Böhm, Klebe & Kubinyi, *Wirkstoffdesign*