

BioSolveIT

Biology Problems **S**olved using **I**nformation **T**echnology

A Combinatorial Approach for Handling of Protonation and Tautomer Ambiguities in Docking Experiments

Ingo Dramburg

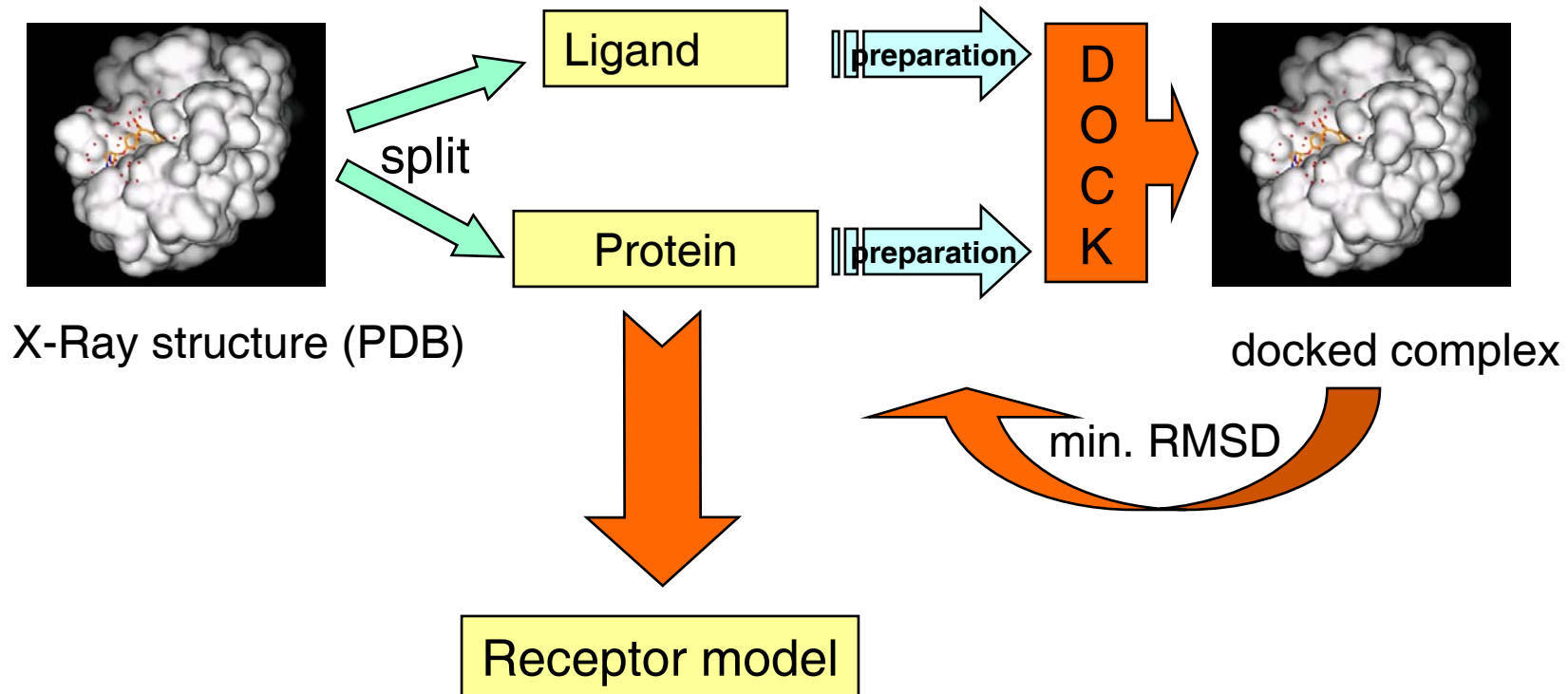
BioSolve IT GmbH • An der Ziegelei 75 • 53757 Sankt Augustin • Germany
www.biosolveit.de

Overview

- Motivation, current state
- Automatic protonation during docking
- Benchmarking the method
- Outlook: Automatic screening of close analogues
- Conclusions

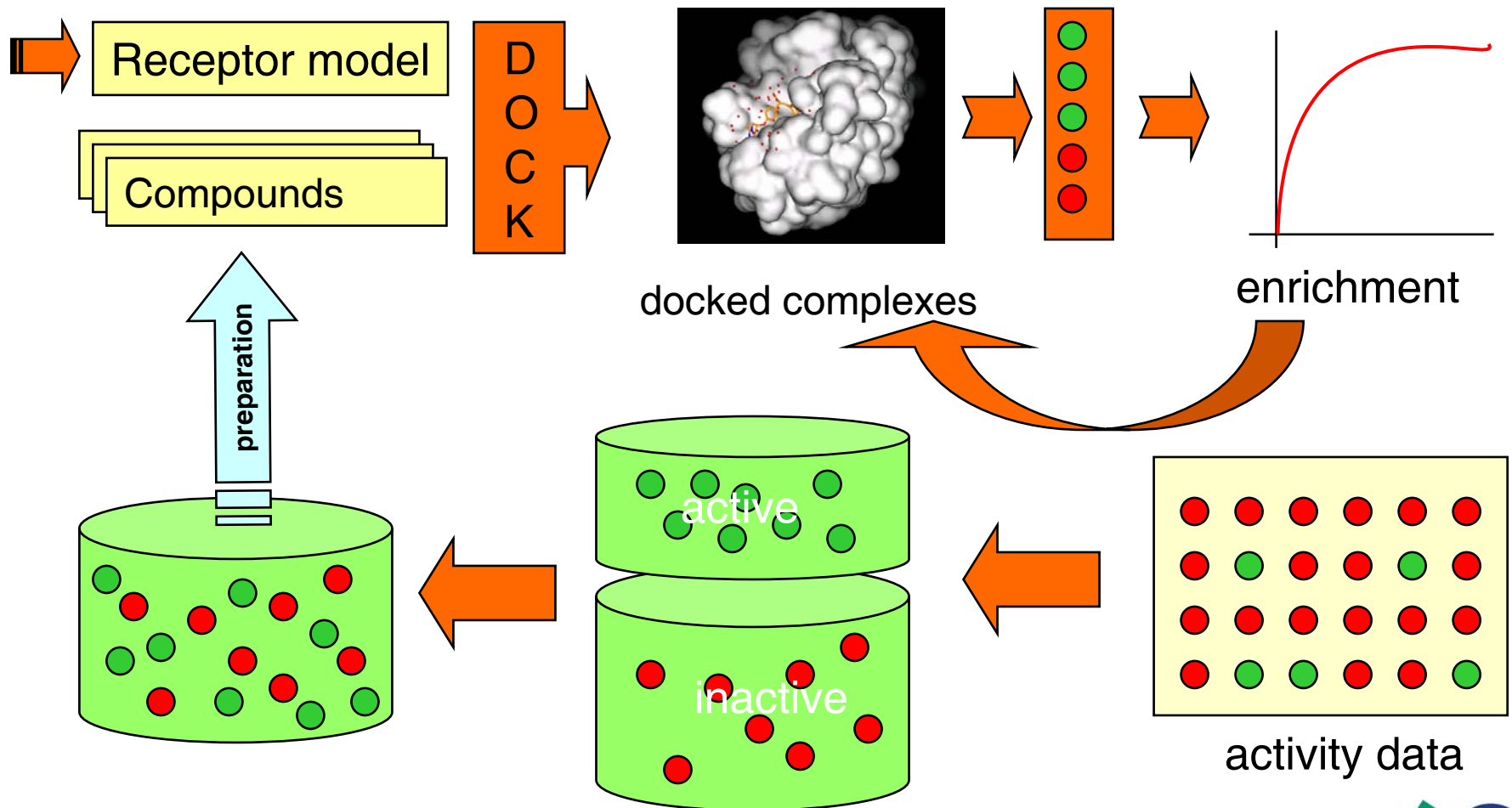
A drug design workflow

Phase I: Building a receptor model



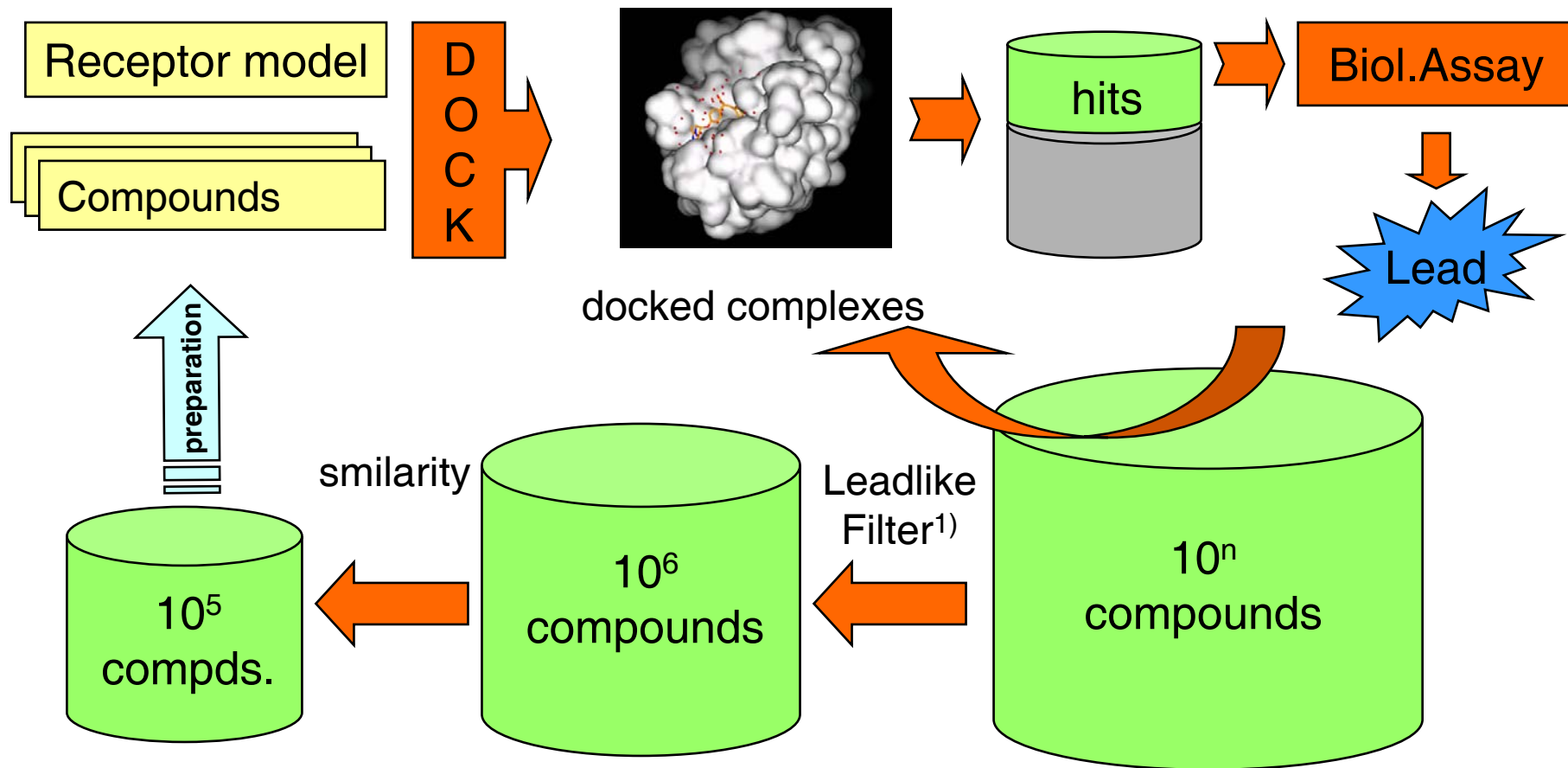
A drug design workflow

Phase II: The screening model



A drug design workflow

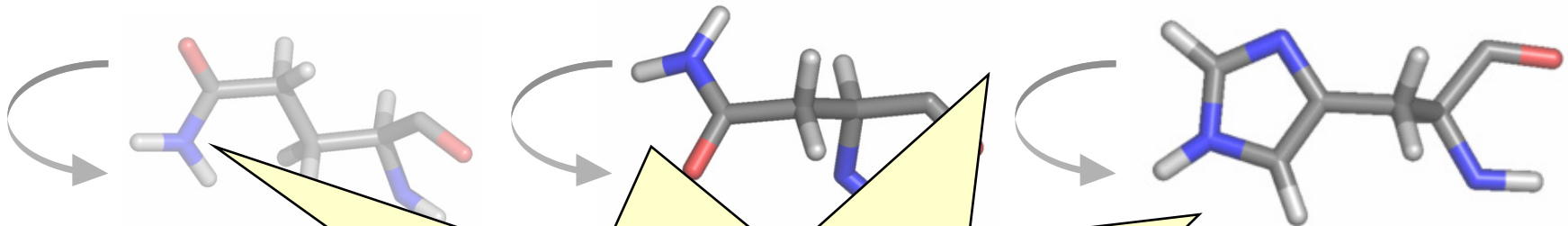
Phase III: Virtual screening



¹⁾C.A. Lipinski et. Al., *Adv Drug Delivery Revs* (1997)23:3-25

Ambiguities in X-ray structures

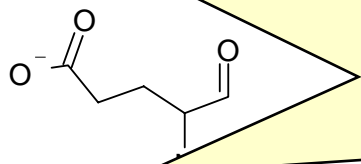
- Gln,Asn,His are „flippable“



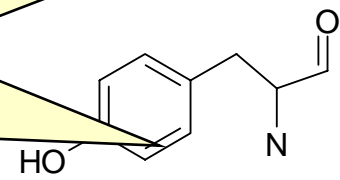
- Glu-,Asp-,Lys

„titrateable“

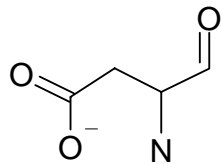
Single binding-site representation
can be inappropriate !



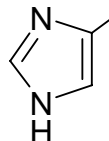
Glu⁻



Tyr

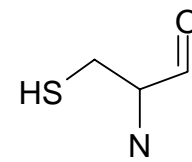


Asp⁻



His

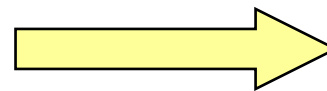
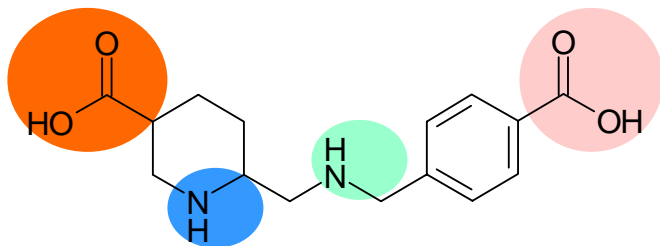
Arg⁺



Cys

Ambiguities in compound libraries

- Virtual compound libraries differ
 - Sources
 - File-formats (MOL2, SDF, SMILES, PDB,...)
 - Conformations (stereo isomers,...)
 - Protonation, tautomers (neutral, charged)
- Which protonation state is the right one ?

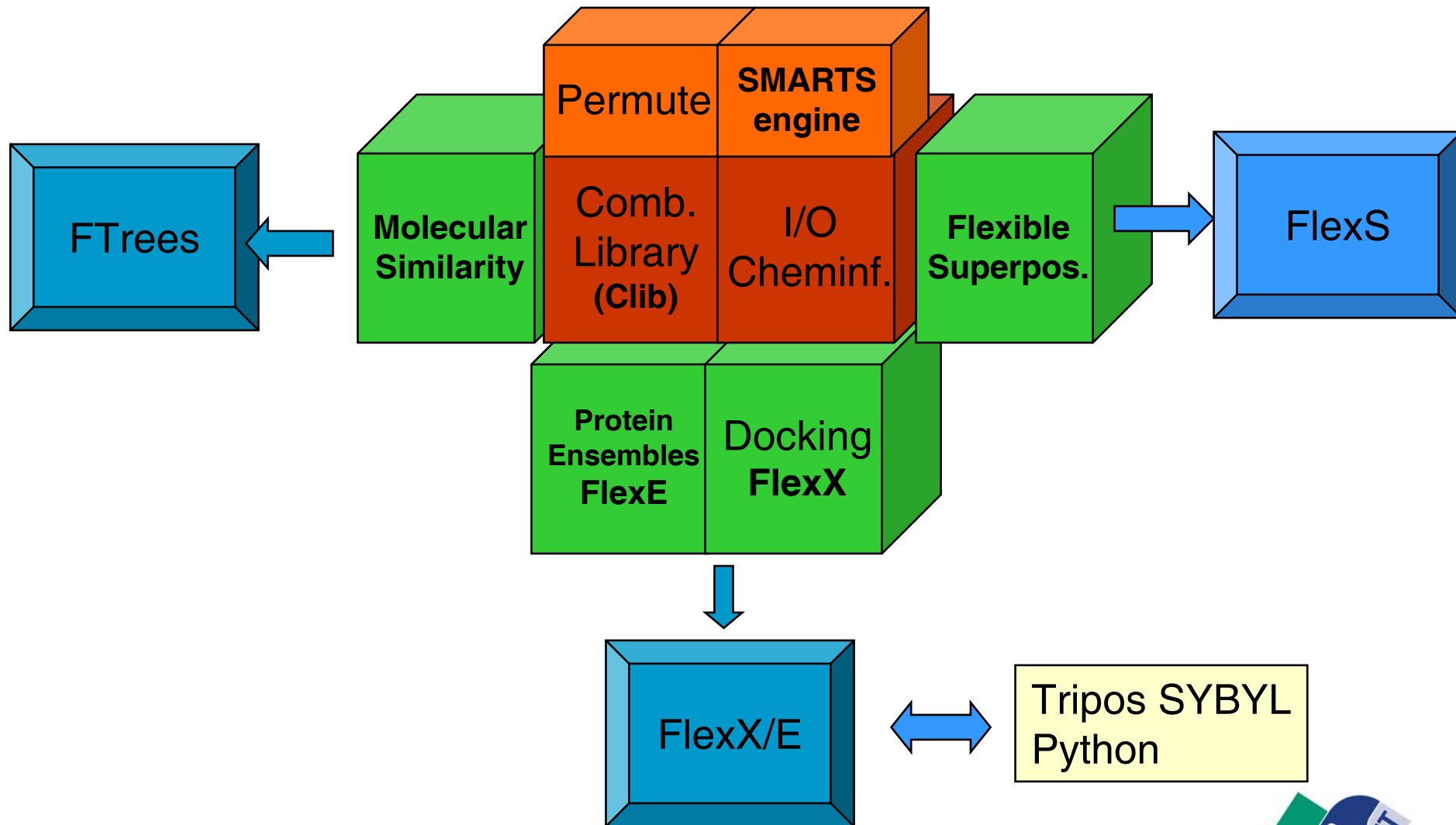


			-
		+	
		+	-
	+		
	+		-
	+		-
	+	+	-
-			
-			-
-		+	
-		+	-
-	+		
-	+		-
-	+	+	
-	+	+	-

Typical data preparation steps

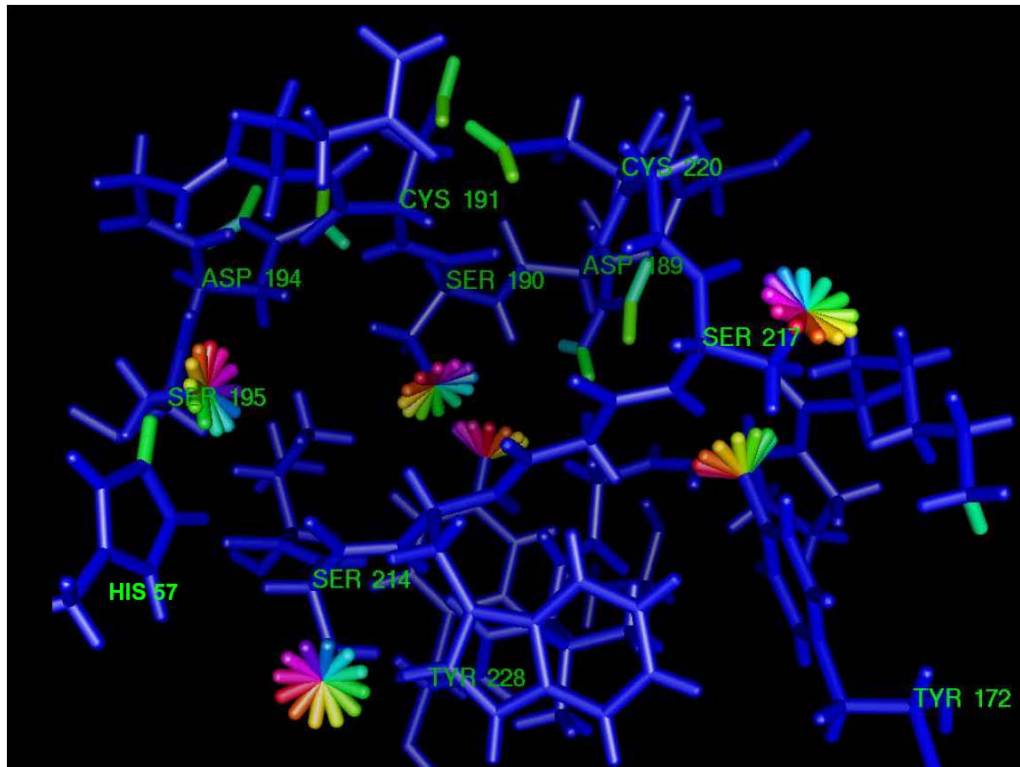
1. Complex-structure: separate protein and ligand
2. Protein preparation
 - structure validity
 - ions, water, cofactor handling
 - [protonation states for titratable sites](#)
 - atomtypes (interaction properties)
3. Ligand preparation
 - structure validity
 - [protonation state](#)
 - atomtypes (interaction properties)

The BioSolveIT Tool Family



Receptor preparation: A united protein model

- FlexE: Handle multiple conformations at once



Claussen et al., J.Mol.Biol.(2001)308; 377-395

Ingo Dramburg © BioSolve IT GmbH

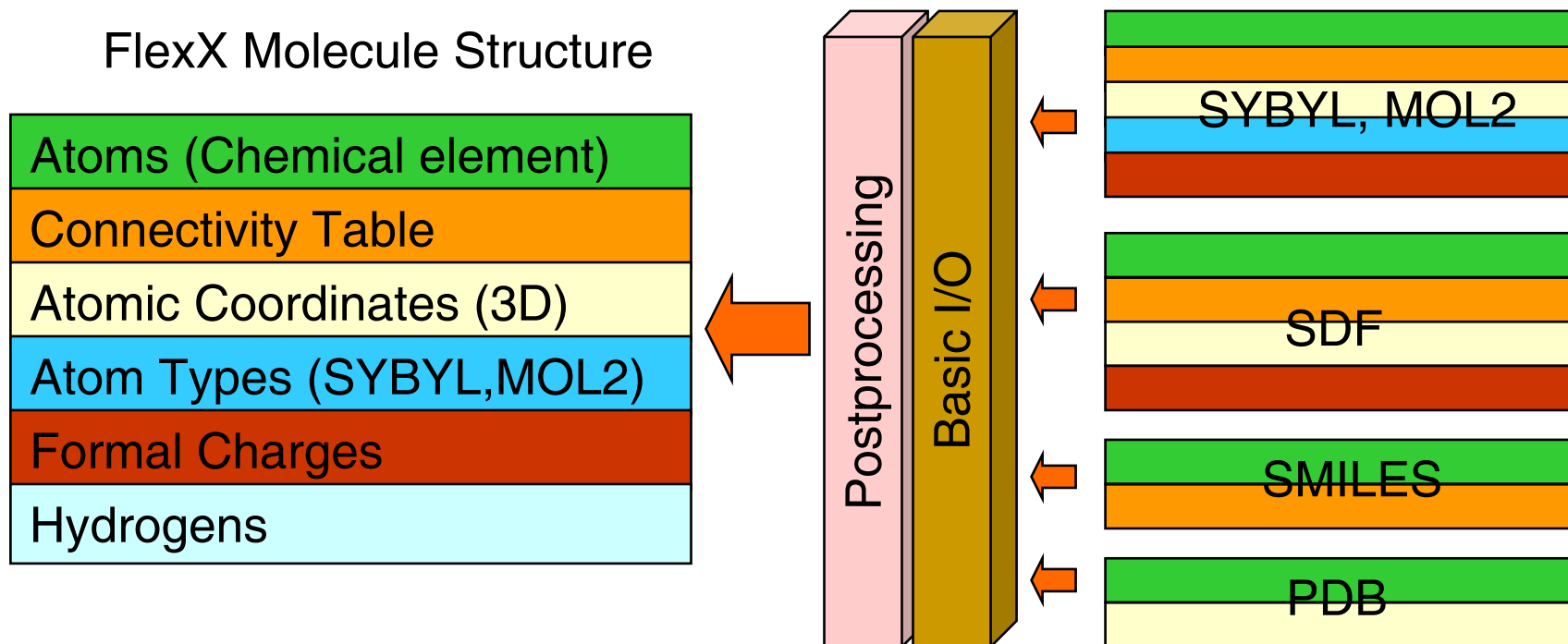
www.biosolveit.de

14-Apr-04

10

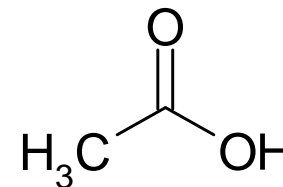


Ligand preparation

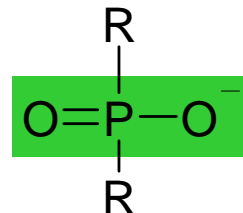
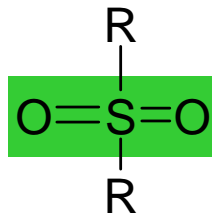
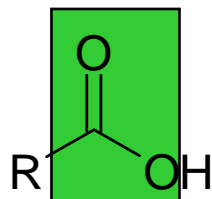


SMILES and SMARTS

- Line notation for molecule and subgraph description
- Initially developed by Daylight Inc.
 - SMILES → Molecule (Unique description of molecules)
 - SMILES „CC(=O)O“ (acetic acid)



- SMARTS → Subgraph (complex description of substructures)
 - SMARTS „[P,S,C](=[OD1])[OD1]“ (,acidic' groups)



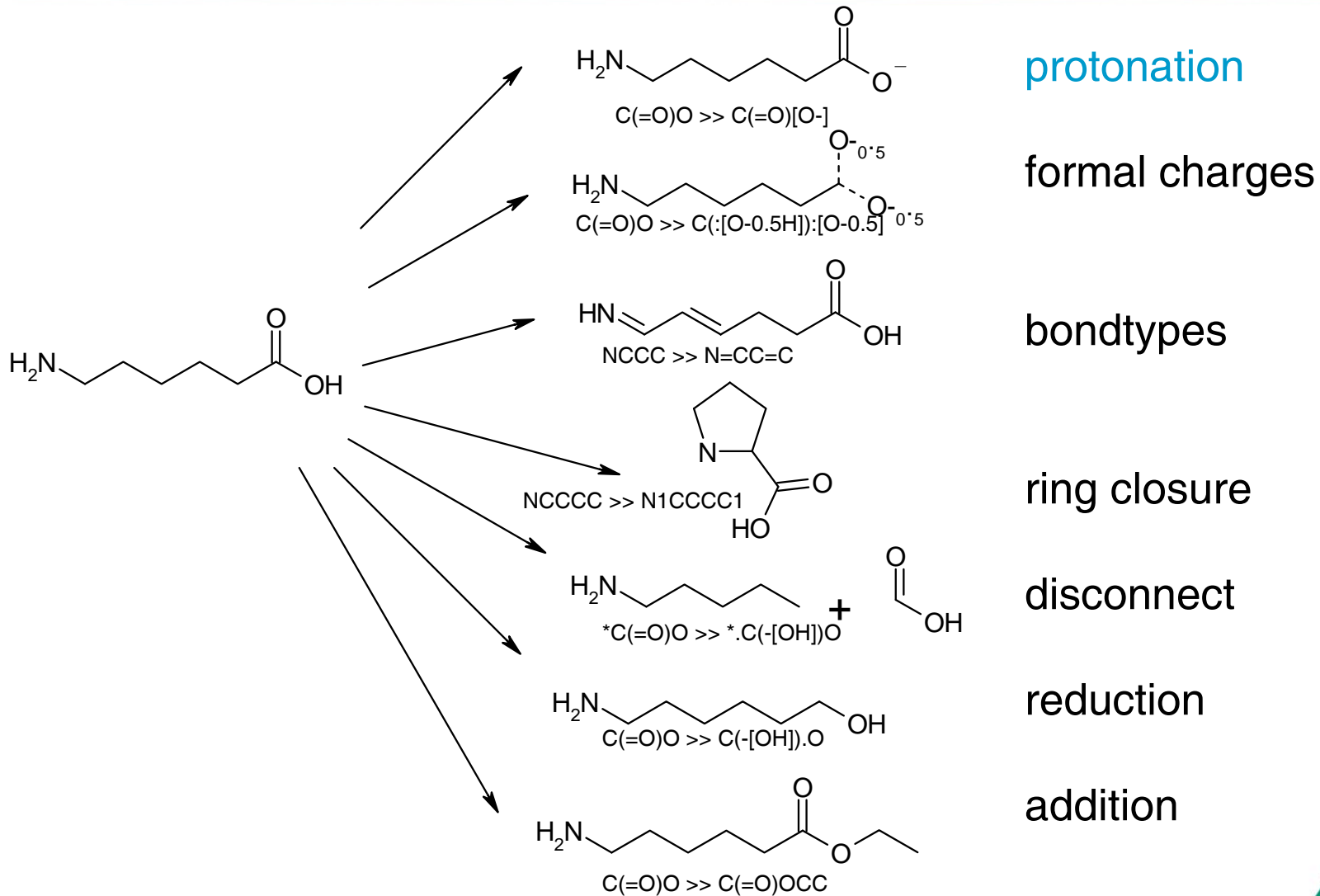
FlexX` SMART_S Engine

- Subgraph Matching
 - Property assignment (aromaticity, atomtypes, charges,....)
 - Descriptor assignment (interaction-types, torsion angles, atomtypes)
 - Structure checking
- Static structure manipulation (similar to SMIRKS)
 - Structure correction
 - Assignment of bond-patterns (mesomerism, tautomerism,...)
 - Chemical modifications, reactions
- Combinatorial structure manipulation: Permutation
 - Permutation of protonation states
 - Generation of close analoges



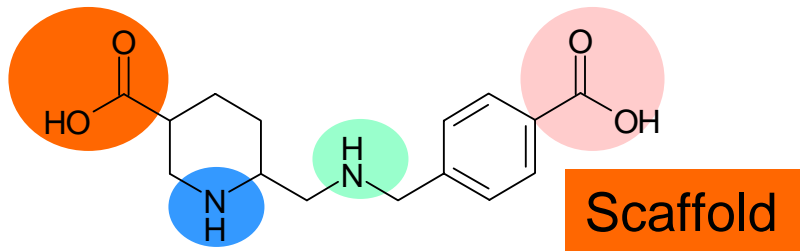
Static structure manipulation

old-substructure >> new-substructure

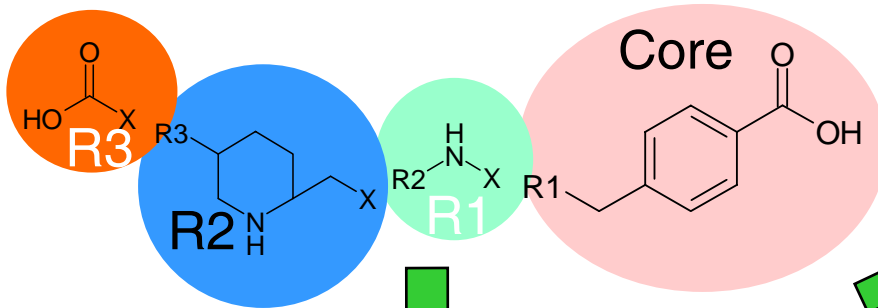


Combinatorial structure manipulation permutation of protonation

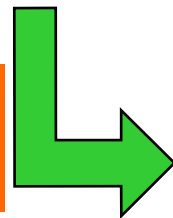
Group identification



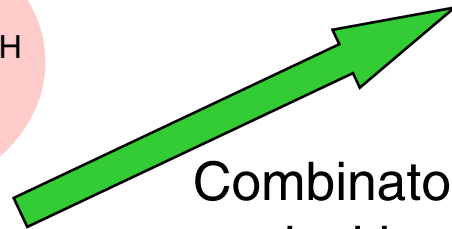
Decomposition



Combinatorial Library



Transformation ($X \gg Y$)
of groups

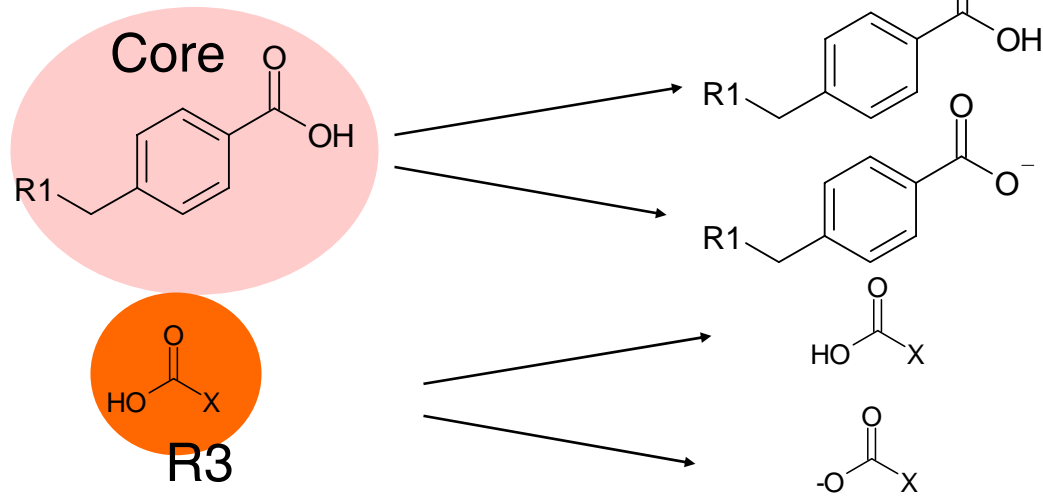


Combinatorial
docking

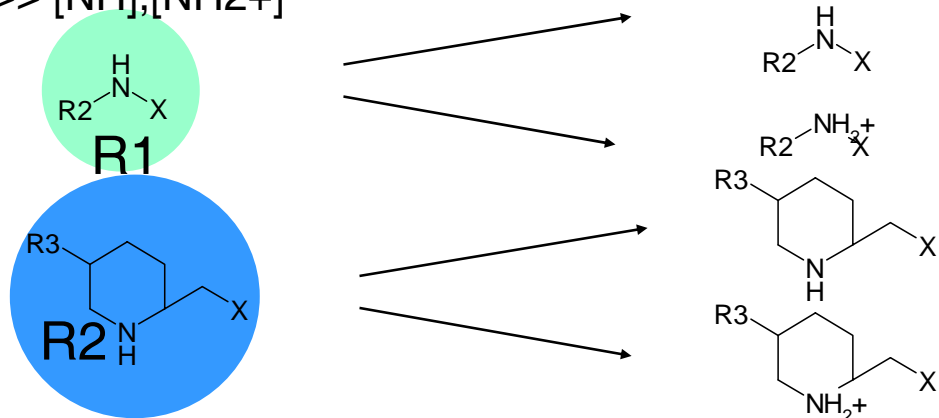
			-
		+	
		+	-
	+		
	+		-
		+	-
	+	+	-
-			
-			-
-		+	
-		+	-
-	+		
-	+		-
-	+	+	
-	+	+	-

Simple rules

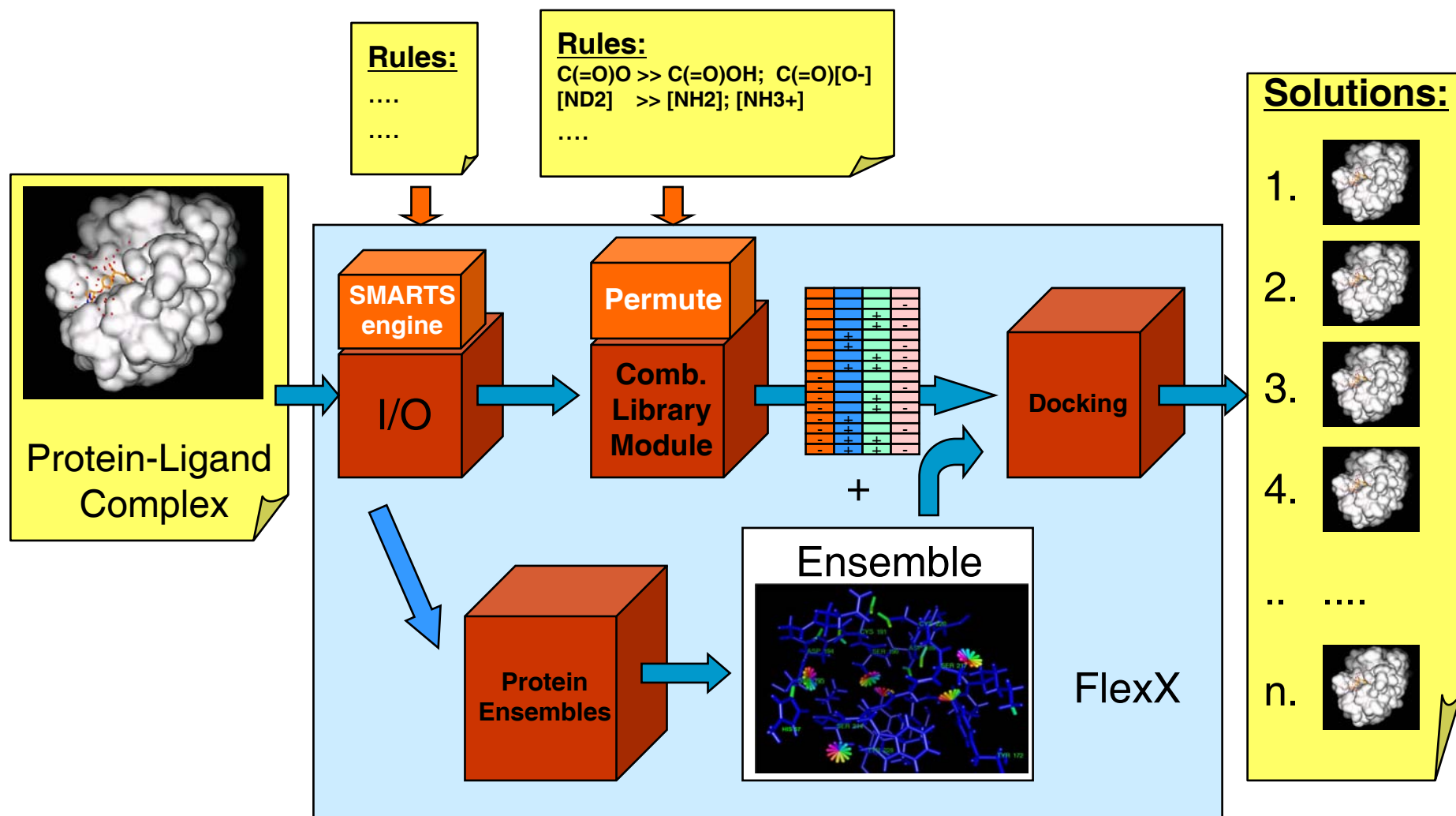
Rule 1: $C(=[OD1])[OD1] \gg C(=O)O; C(=O)[O-]$



Rule 2: $[ND2] \gg [NH]; [NH2+]$



The complete picture



Benchmarking: A Dataset XXL

Dataset from PDB: ~20000 complexes

- Protein constraints
 - X-Ray structures from PDB
 - Resolution < 3.5 Å
 - No DNA/RNA
- Ligand constraints
 - MW 78-750
 - Only C,H,N,P,S,F,Cl,Br,I
 - No pure CH compounds
 - Max. 10 rot. Bonds, max. ring size 9
 - Non-covalent
 - Non-cofactor (242 HET groups, freq. > 10)
- **2300 selected complexes**

Sadowski J., Buning C., Claussen H., (2004), in preparation

Ingo Dramburg © BioSolve IT GmbH

www.biosolveit.de

14-Apr-04

18



Benchmarking: current state

- Automatic complex setup
- Redocking with default settings
(6.5 Å around ligand as site, no water or cofactors included, chemscore)
- ‚hit‘ = rmsd of predicted complex on rank 1 < 2.0
- unified protonation model for receptor

	lig.setup	#	hitrate	any rank < 2.0
PDB2300	neutral	2300	510 22 %	1112 48 %
PDB2300	charged	2300	676 29 %	1324 57 %
FlexX200¹⁾	hand	200	93 46 %	138 69 %

¹⁾Kramer et al., Proteins (1999) 37:228-241



Benchmarking: Permutation

- Ligands with titratable sites: 634

without permutation

lig.setup	#	hitrate	
PDB2300 neutral	634	144	22 %
PDB2300 charged	634	176	27 %

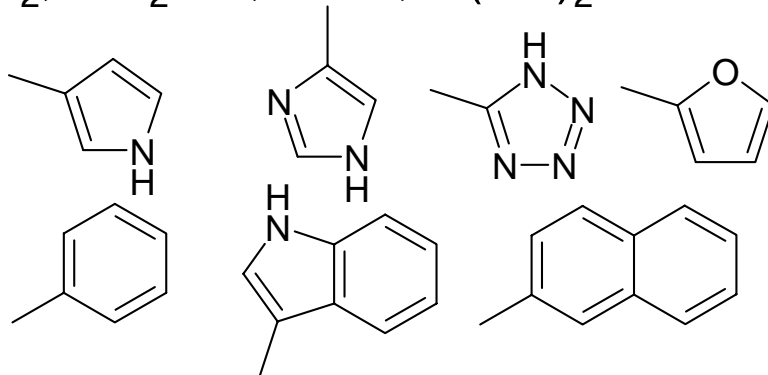
with permutation

lig.setup	#	hitrate	
PDB2300 neutral	634	179	28 % (+35/6 %)
PDB2300 charged	634	210	33 % (+34/6 %)

More complex applications: Isosteric replacement

■ Substitution ($[CD1, \$ (Hc)] \gg X$)

- F, Cl, Br, I, CF₃, NO₂,
- CH₃, C₂H₅, isoprop., t-butyl
- -OH, -SH, -NH₂, -CH₂OH, -OMe, N(Me)₂
- Rings:



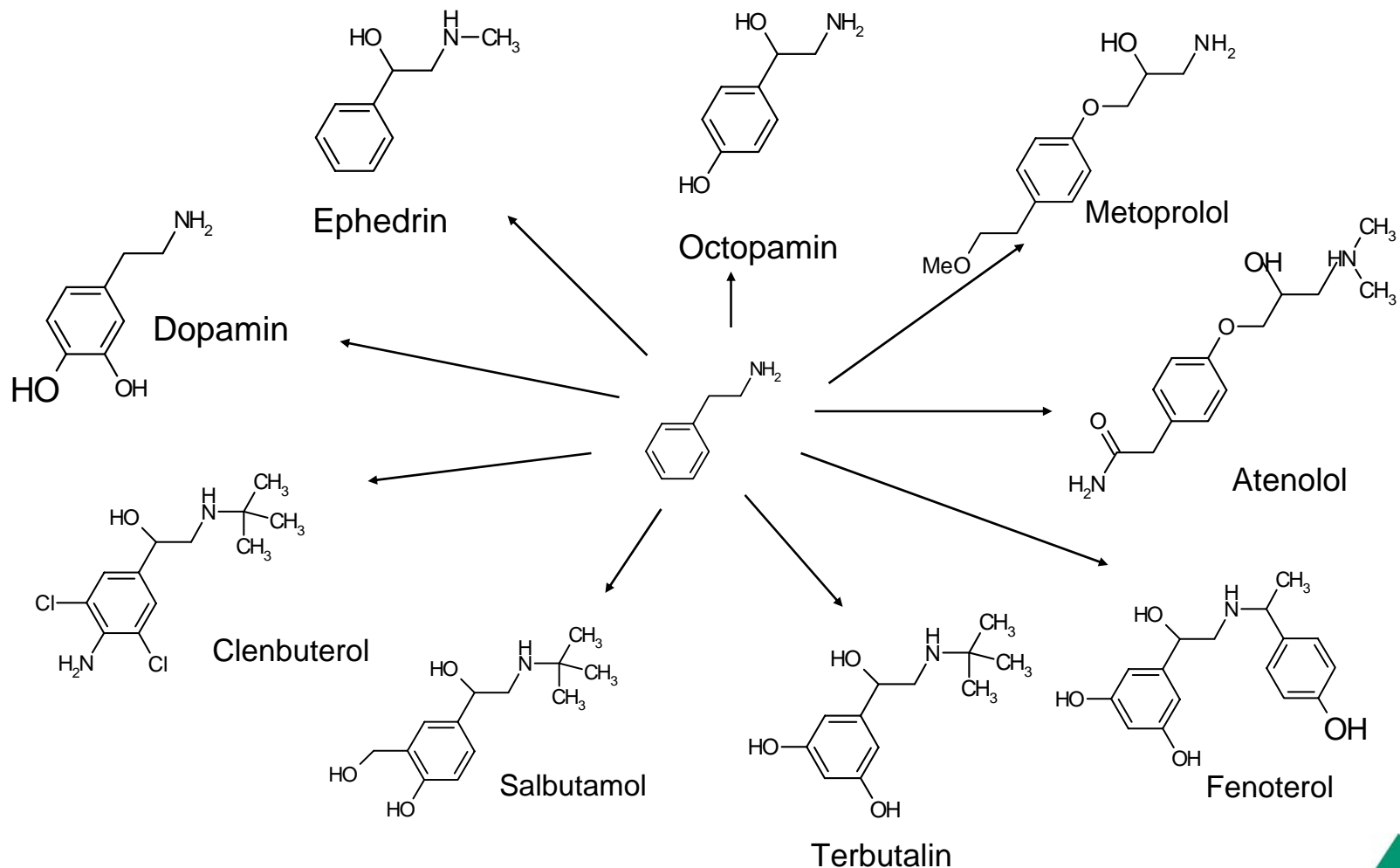
■ Insertions ($[CD2]-[CD2] \gg X$)

- -CH₂-, -NH-, -O-, -C₂H₄-, C₃H₆-, ...
- -COCH₂-, -CONH-, -COO-, ...
- >C=O, >C=S, >C=NH, >C=NOH, >C=NO-, ...

Outlook

Automatic generation of close analogues

Example: β -phenylethylamine family, explorable by simple rules



Conclusions

- FlexX can perform automatic docking run successfully
- Manual setup always beats automatism
 - Impossible for screening of large datasets !
- Significant increase of hitrate with permutation of isomers
- Only slight increase of time consumption for permutation
- Selection of best isomer by scoring function
- Method applicable to de novo design

Acknowledgements

AstraZeneca

J. Sadowski

MPI Saarbrücken

A. Kämper



BioSolve IT

H. Claussen
M. Gastreich
M. Lilienthal
C. Lemmen



ThanxX for your attention.

