PharmShapeCC: 3D pharmacophore searching against ten trillion combinatorially accessible compounds

Qiang Zhang and Ingo Muegge

9 th International Conference on Chemical Structures
June 5-9, 2011, Noordwijkerhout, The Netherlands

Boehringer Ingelheim

125 Jahre mehr Gesundheit

Extremely Large Virtual Compound Collection

Boehringer Ingelheim
125 Jahre mehr Gesundheit

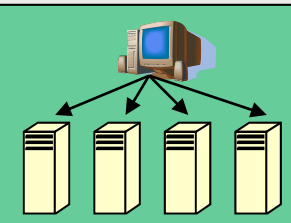
BI CLAIM

10+ trillion virtual library compounds of BI chemistry

Colibri

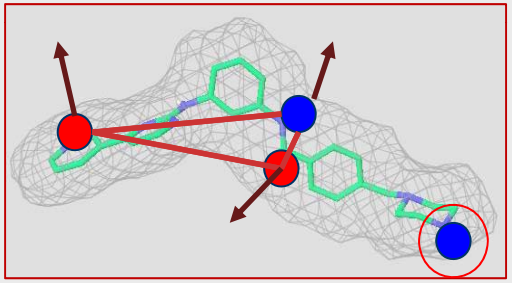
BioSolveIT

FeatureTree-FS



local clusters
GRID computing

PharmShapeCC



2D Similarity

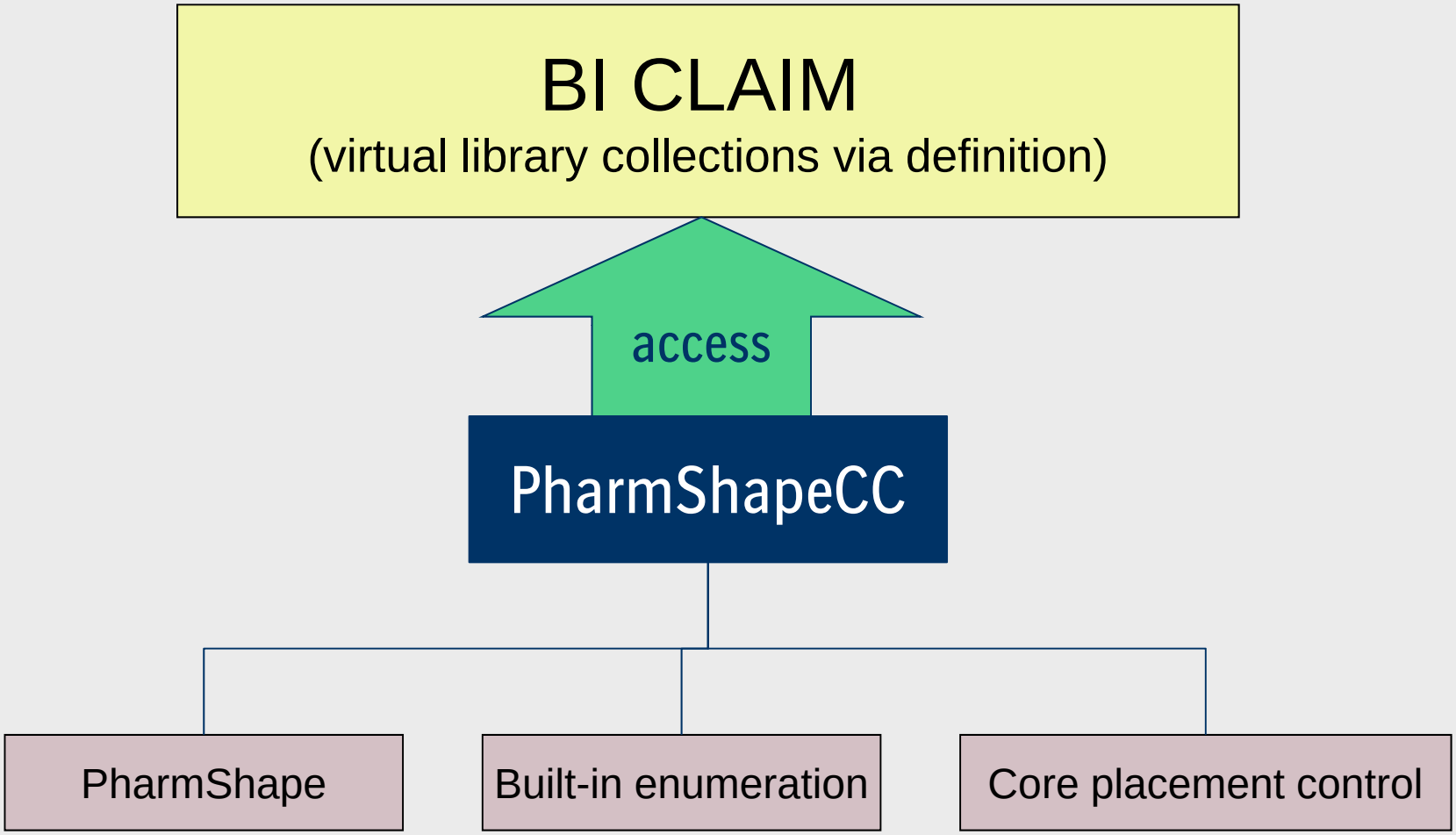
dimensionality reduction

3D Features

Traditional VS methods (10^3 - 10^4 compounds)

Synthesize and test hit libraries (~100 compounds)

Components of PharmShapeCC



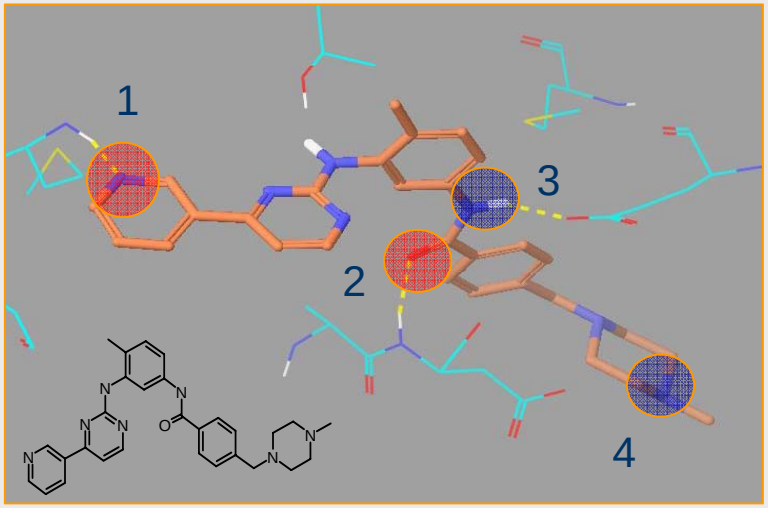
Most important components for PharmShapeCC to achieve screening 10+ trillion compounds quickly (in about one day with 1000 CPUs)

PharmShape – A 3D Pharmacophore Searching Program

Gleevec as an example

Pharmacophore Requirements:

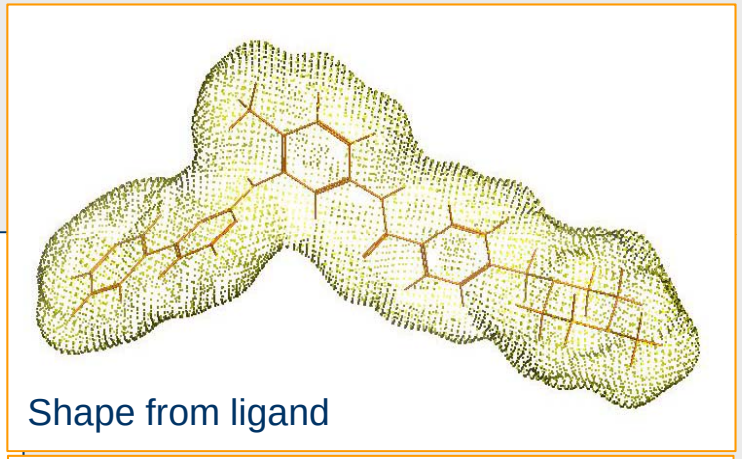
1 = HA; 2 = HA; 3 = HD; 4 = Basic amine



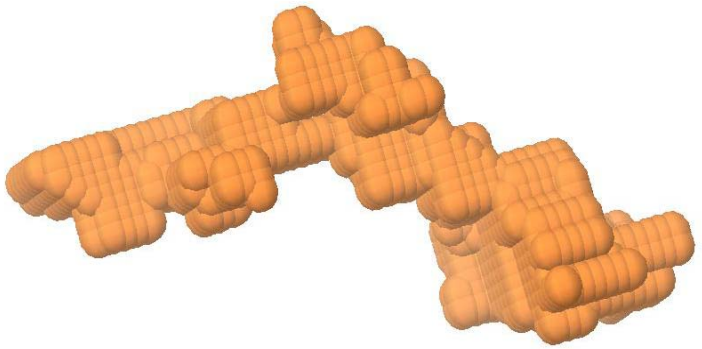
+

Shape Requirements:

Composite shape of the aligned ligands and/or the shape of binding pocket



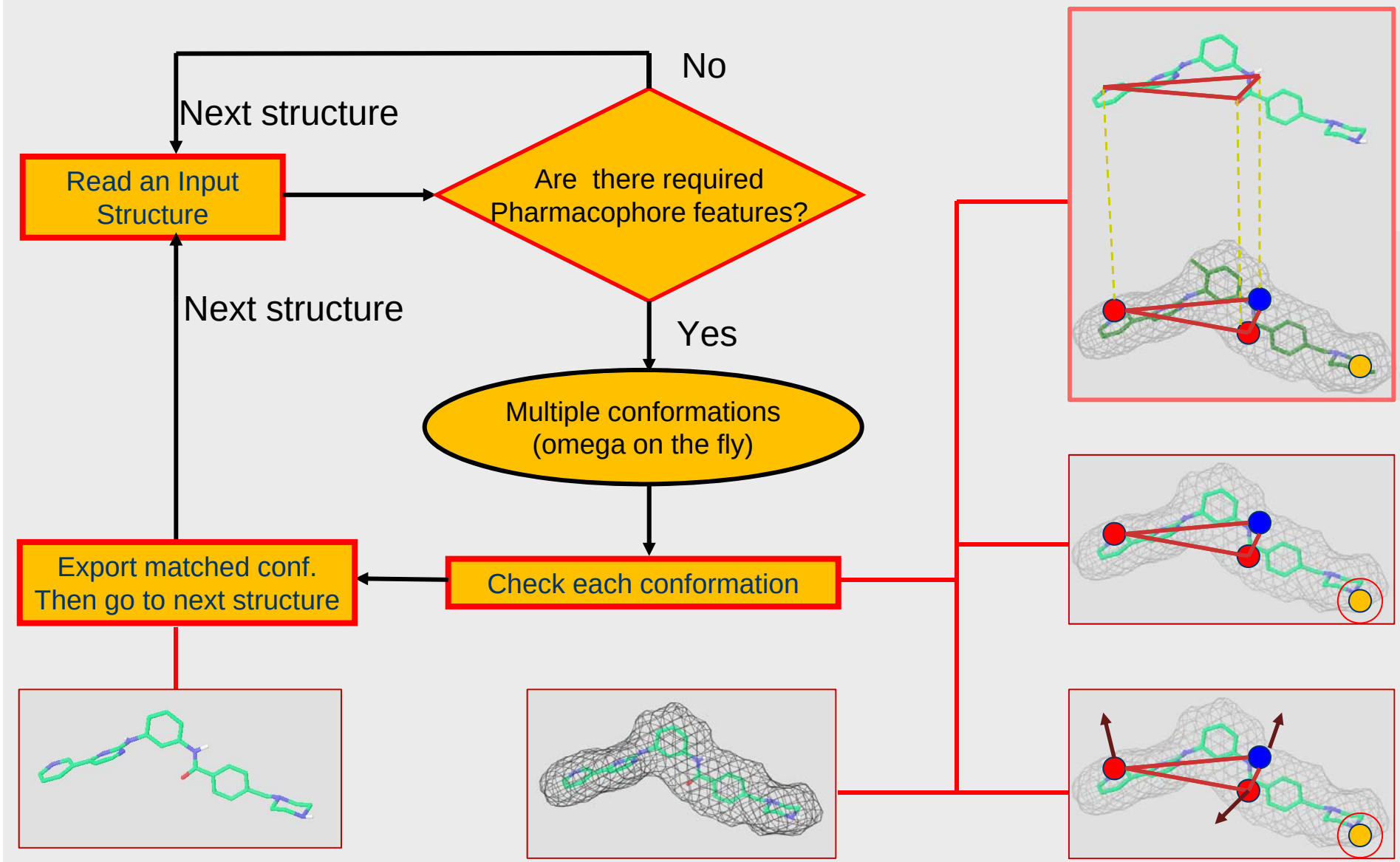
Shape from ligand



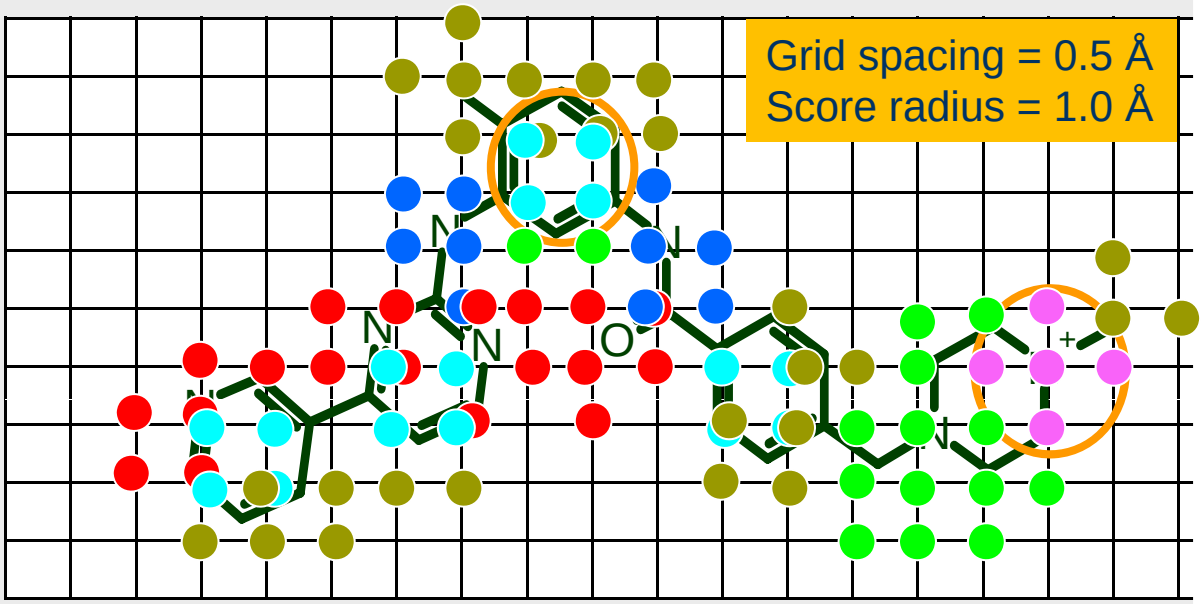
Shape from binding pocket

- User defined features allows the accurate search of compounds with desired features.
- Searching result has no bias towards template, favoring scaffold hopping.
- Any 3D model can be used as template.

Procedures for Hit Identification
(PharmShape)



Scoring Mechanism in PharmShape

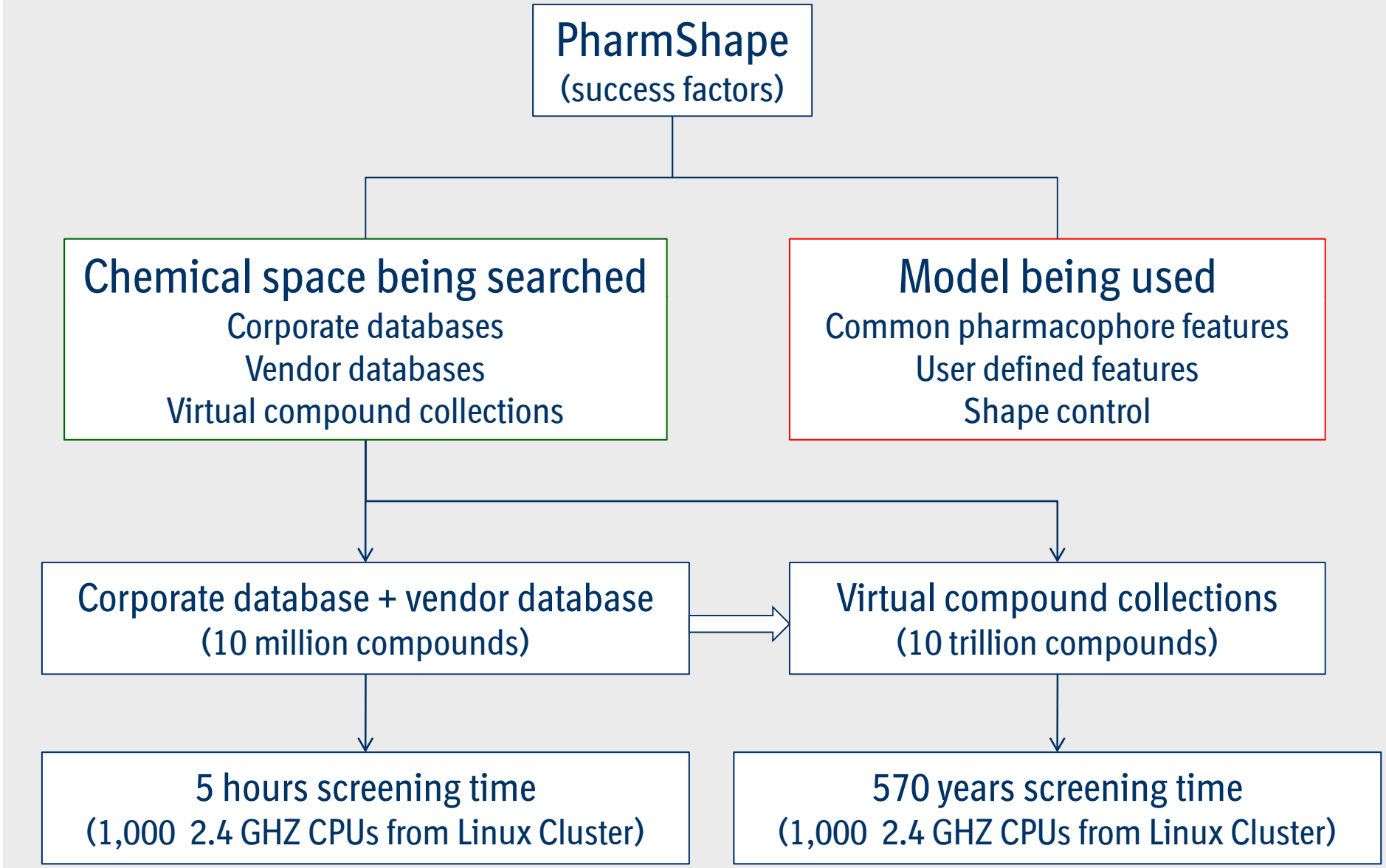


- A: Acceptor
- D: Donor
- R: Aromatic
- H: Hydrophobic
- P: Positive
- N: Negative
- X: Any

- Grid points are annotated first based on template ligands in 3D space
- Subsequent annotation is done for each searched compound after overlay

  PharmShape _score = num_matched / num_occupied

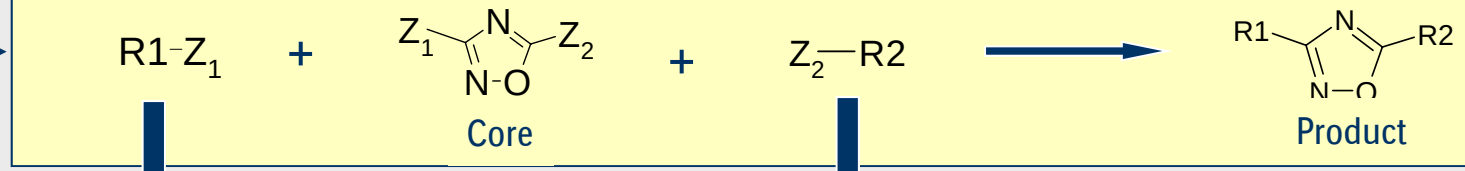
PharmShape Summary

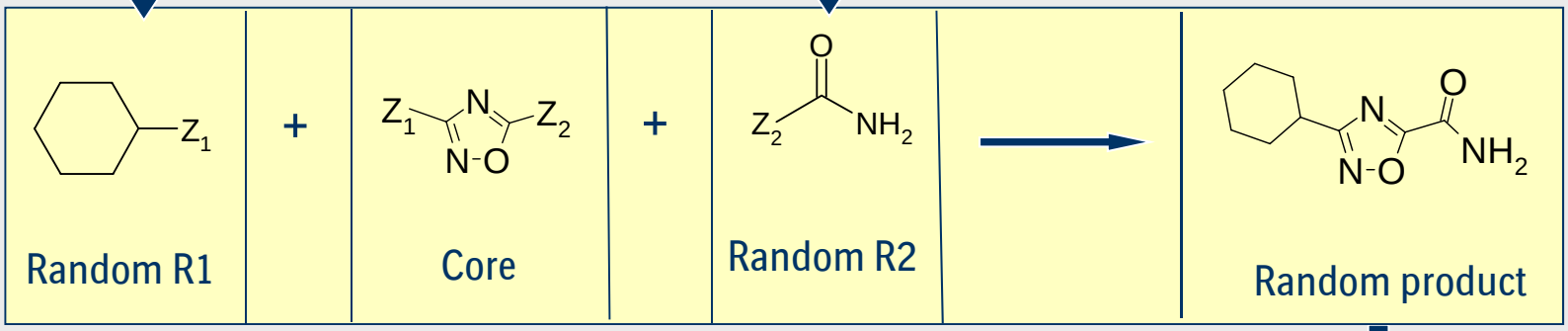


Step 1: Determine the Core Placements
(PharmShapeCC)

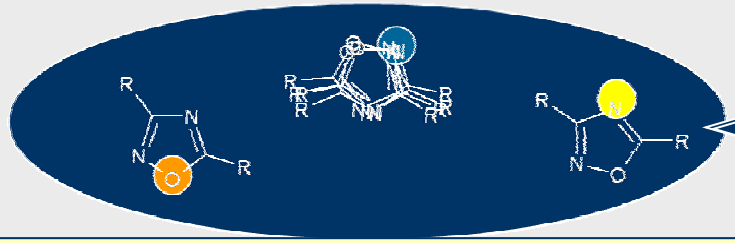
Boehringer Ingelheim
125 Jahre mehr Gesundheit

From library definition table

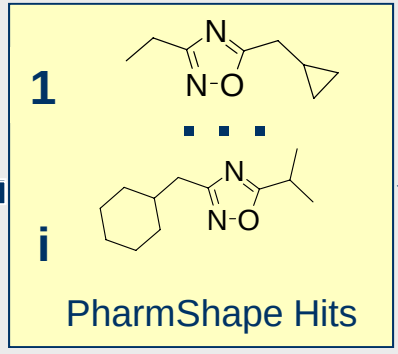




Repeat until pre-set number of compounds tested

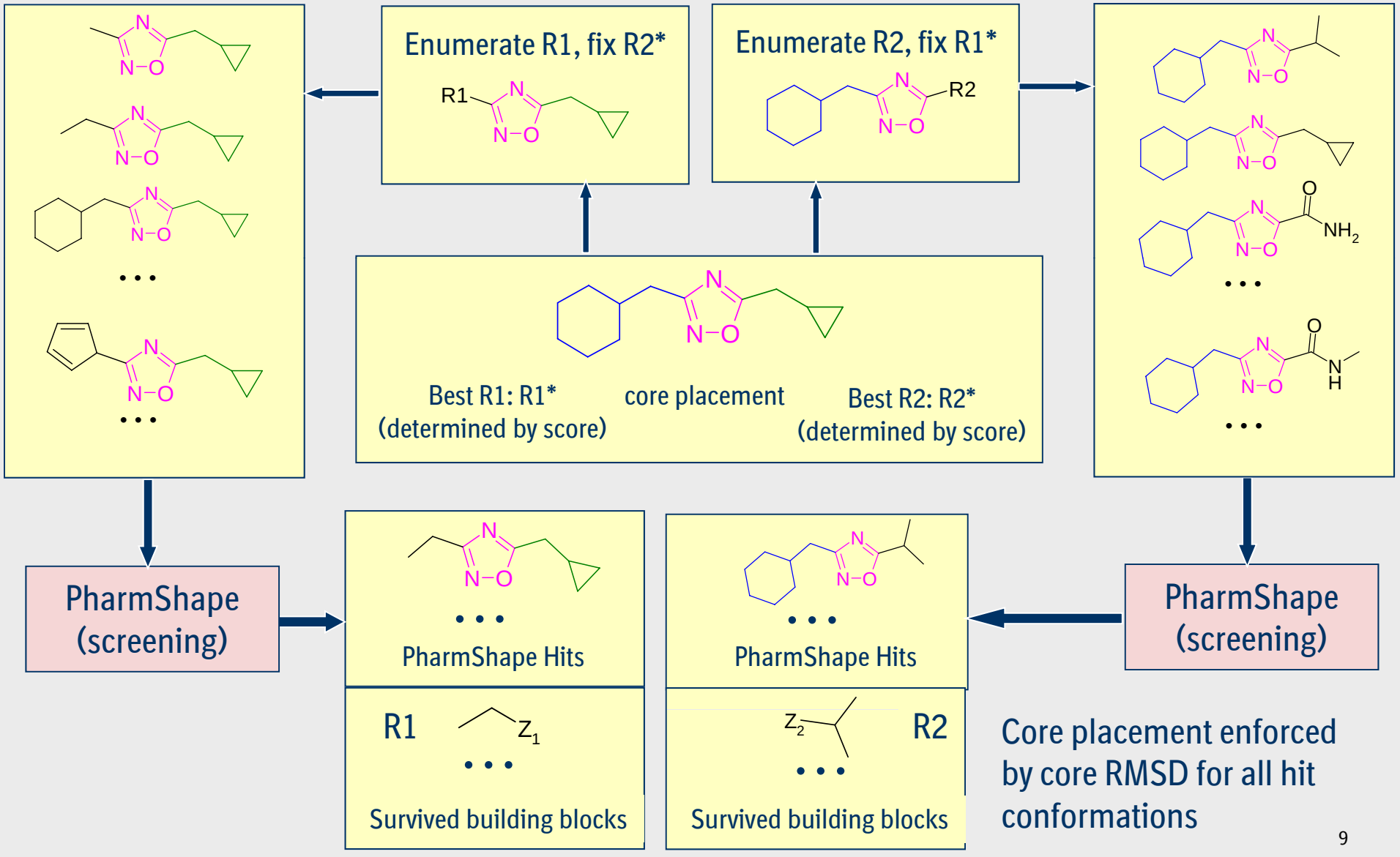


Use core consensus to determine the core placement

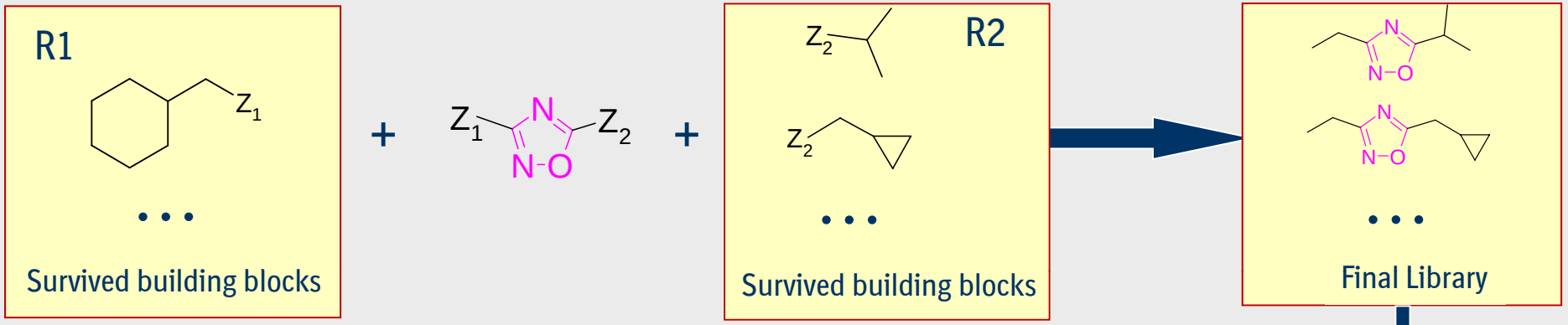


PharmShape (screening)

Step2: Screen Building Blocks Using PharmShape (PharmShapeCC)



Step3: Screen Final Library with Survived Building Blocks (PharmShapeCC)



Report hit libraries:

- Number of hits involved for each building block.
- Average score for each building block.

Retrieve library ranking using post processing software:

- Average score for entire library.
- Size of the library (R1, R2, …)

Select libraries for synthesis based on multiple factors:

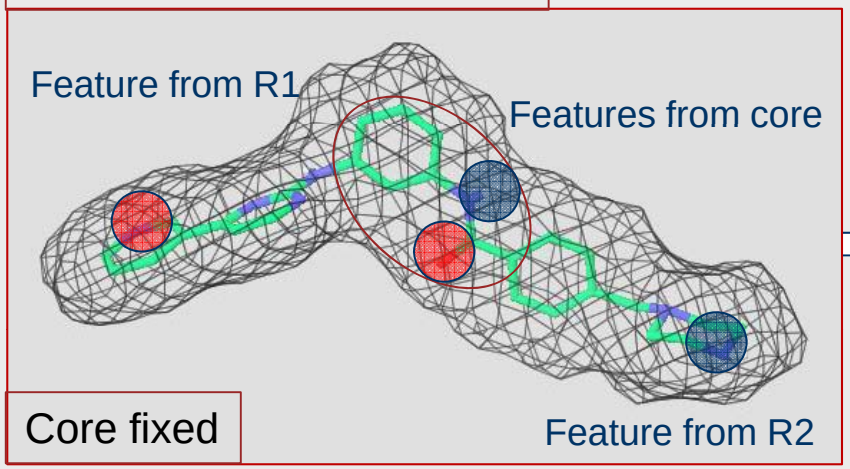
- PharmShape score, size, and other conventional measures

Can PharmShapeCC identify all the PharmShape hits?

(Yes if core placement predicted correctly)

Boehringer Ingelheim
125 Jahre mehr Gesundheit

1. Theoretical explanation:



Fixing core placement allows the identification of the majority hits by screening building blocks

2. Validation:

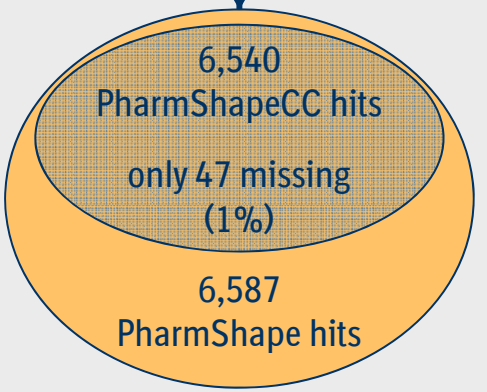
A BIClaim library
(~10 million compounds)

Enumeration

PharmShape
(~10 million compounds)

PharmShapeCC
(~6,500 compounds)

comparison

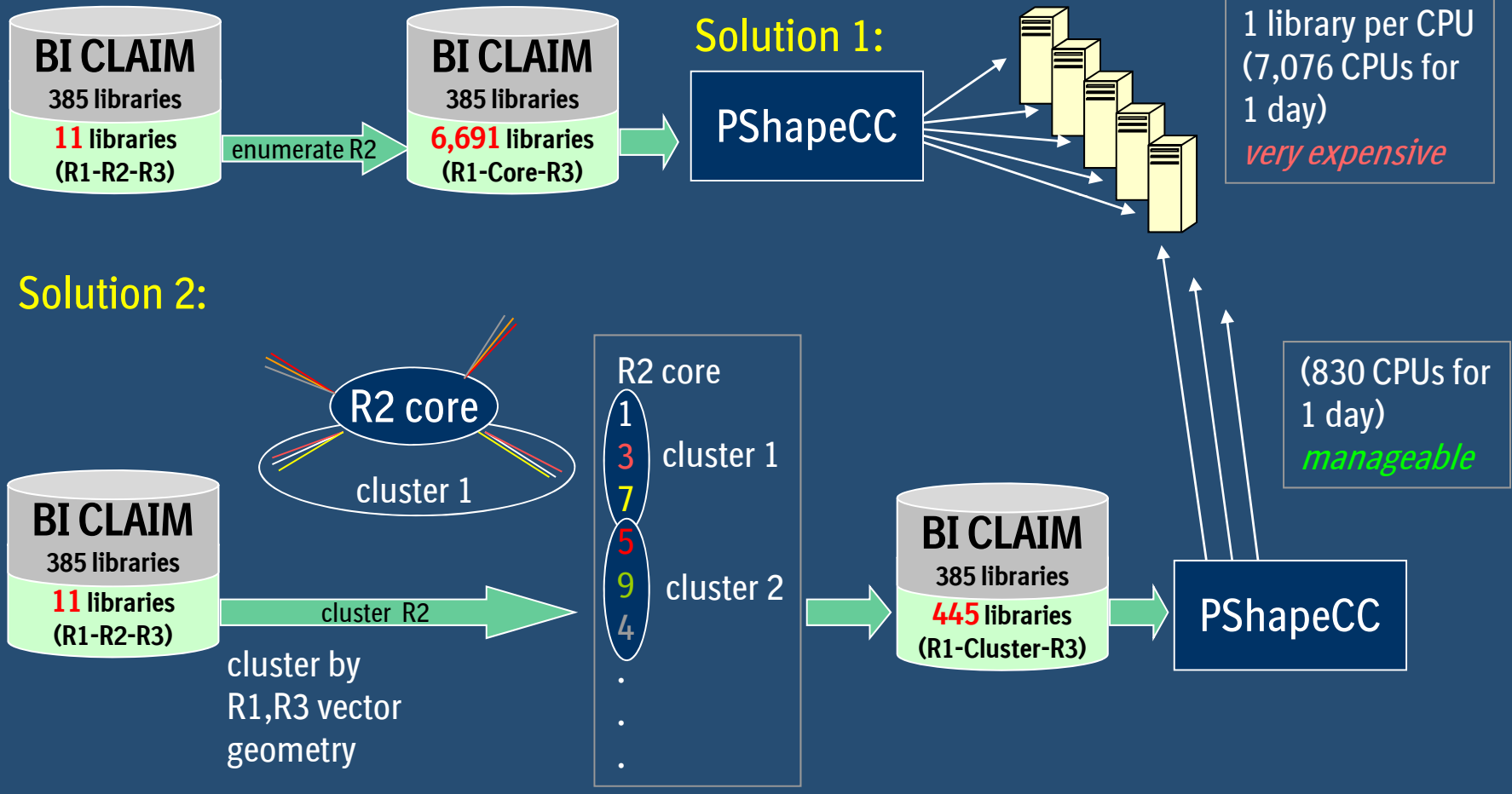


- Single core placement used for PharmShapeCC. The 47 missing hits represent significant shift of core placement (RMSD > 2.0 Å).
- Similar results obtained with two additional validations.

Reduce Resource Usage by Clustering of Cores

Boehringer Ingelheim
125 Jahre mehr Gesundheit

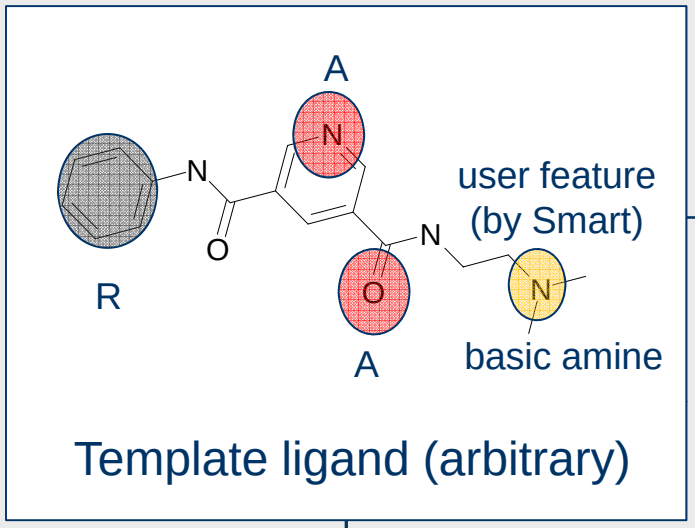
Problem: BI CLAIM library explosion due to R1-R2-R3 enumeration:



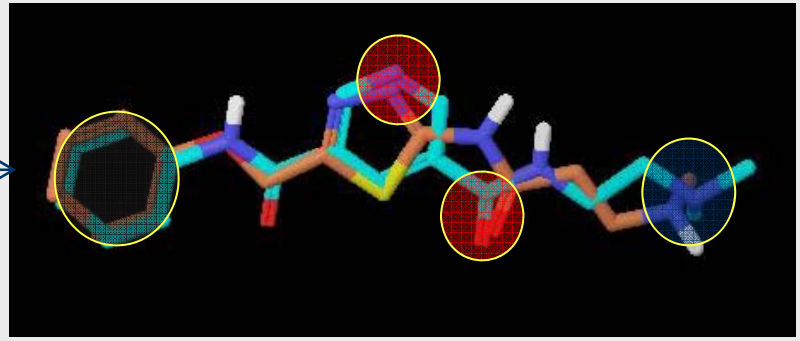
Virtual Hits Identified by PharmShape/CC

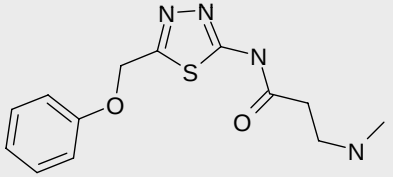
(examples from searching against 1.5 million vendor compounds)

Boehringer Ingelheim
125 Jahre mehr Gesundheit

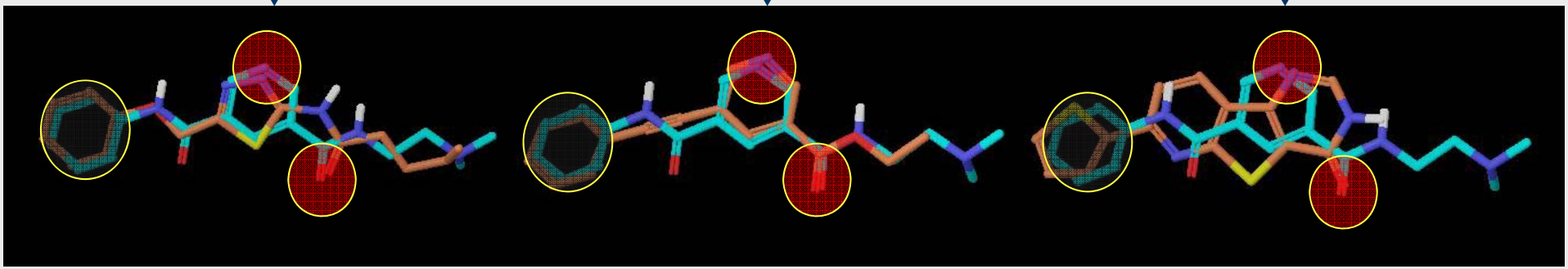


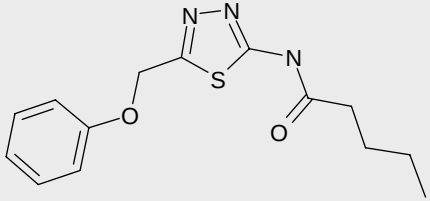
User feature turned on

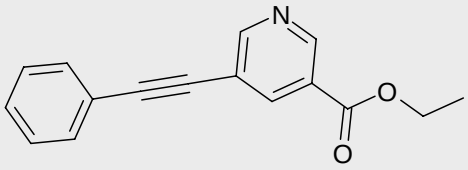


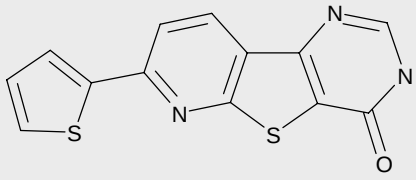


User feature turned off





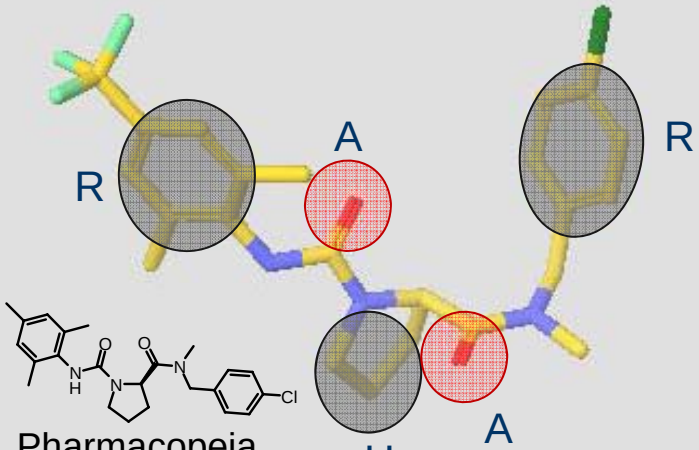




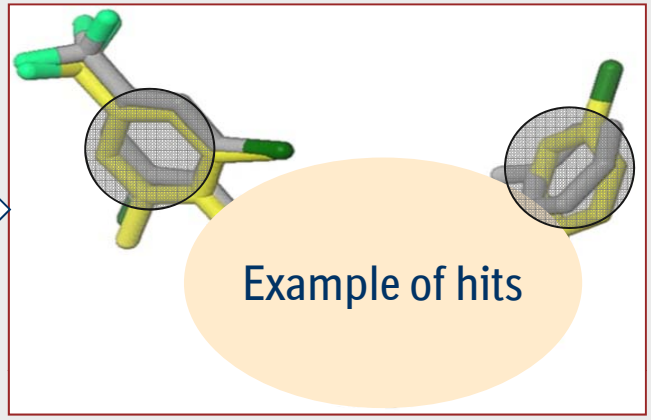
PharmShapeCC Results
(CCR1 project)

Boehringer Ingelheim
125 Jahre mehr Gesundheit

Conformation determined by consensus with a in-house ligand (not shown)



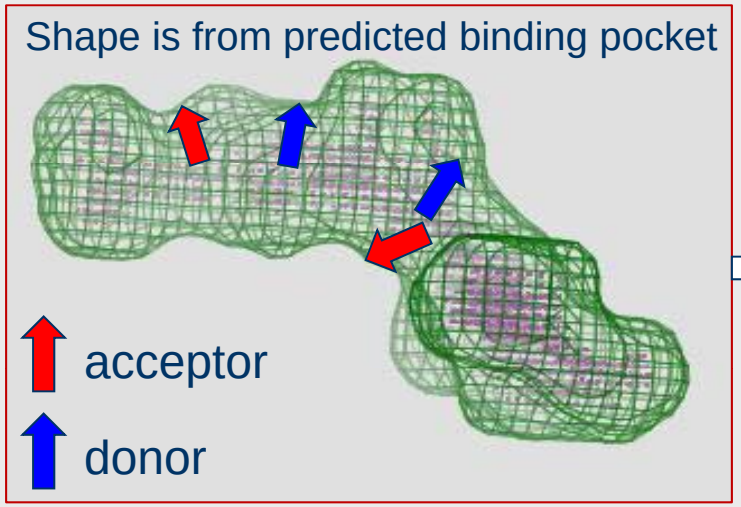
Pharmacopeia
IC50 ~ 10 nM

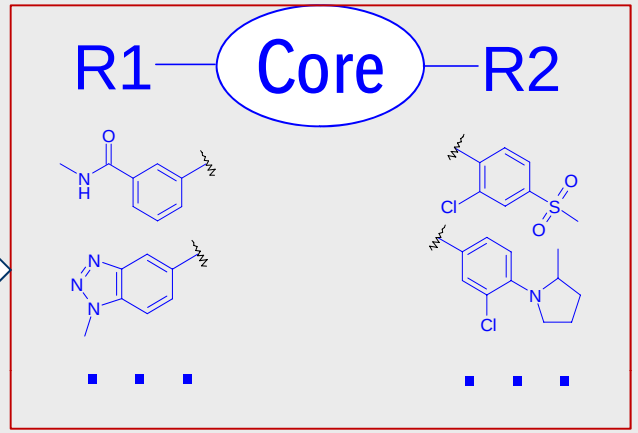


- 77 hit libraries identified from 1,916 libraries
- 20 libraries remain after filtering with score and size (score > 0.5; size > 2500)
- 4 libraries selected after visual inspection

Library No	Library 1	Library 2	Library 3	Library 4
Compounds synthesized	50	55	139	60
Hits identified (IC50 < 3000 nM)	5	7	3	0
IC50 of most potent hits (nM)	170	710	740	NA
Closest similarity to Pharmacopeia (Daylight 1024)	0.703	0.416	0.436	0.359

PharmShapeCC Results
(a kinase project)





Template ligand ~ 50 nM

- 15 hit libraries identified from 1,916 libraries
- 4 libraries remain after filtering with score and size (score > 0.5; size > 2500)
- 1 library selected after visual inspection

Library No	Compounds for testing ideas	Follow-up library
Compounds synthesized	6	108
Hits identified (IC50 < 1000 nM)	3	46
IC50 of most potent hits (nM)	< 1	1.3
Closest similarity to template (Daylight 1024)	0.492	0.583

Summary

- PharmShapeCC has been developed to do 3D pharmacophore search against extremely large combinatorial library pool.
- Increasing the size of compound collection significantly increases the chance of identifying potent compounds.
- PharmShape scoring mechanism and core clustering mechanism allows the better prediction of core placement, which is very important for success.
- Built-in enumeration routine avoids the need to do library enumeration before searching starts.
- Doubling the number of compounds in the library pool will not significantly increase the processing time

Acknowledgement

Structure Research:

- Minghong Hao

Parallel Synthesis:

- Younggi Choi
- Lana Smith
- Angela Berry