Boehringer Ingelheim

Daylight User Group Meeting
Cambridge, 4.-5.11.04

Application of Daylight Fingerprints to Virtual Screening

Uta Lessel
Boehringer Ingelheim Pharma GmbH & Co. KG
Department of Lead Discovery

Ligand Based Virtual Screening

Boehringer Ingelheim

Goal:

- Selection of subsets with increased hit rates from a data set

Procedure:

- Looking for compounds similar to known actives
- Ranking of data sets with actives and inactives according to decreasing similarities

Evaluation:

- E.g. determination of enrichment curves

Study

Boehringer Ingelheim

Aim:
Comparison of different methods for the search for similar compounds

Methods analyzed:

- Tanimoto coefficients on the basis of Daylight Fingerprints
- Euklidean distances in a 5-dimensional BCUT property space (R.S. Pearlman, K.M. Smith, Perspectives in Drug Discovery and Design, 9/10/11, 339-353, 1998)
- Feature Trees
  (M. Rarey, J.S. Dixon, J. of Computer-Aided Molecular Design, 12, 471-490, 1998)

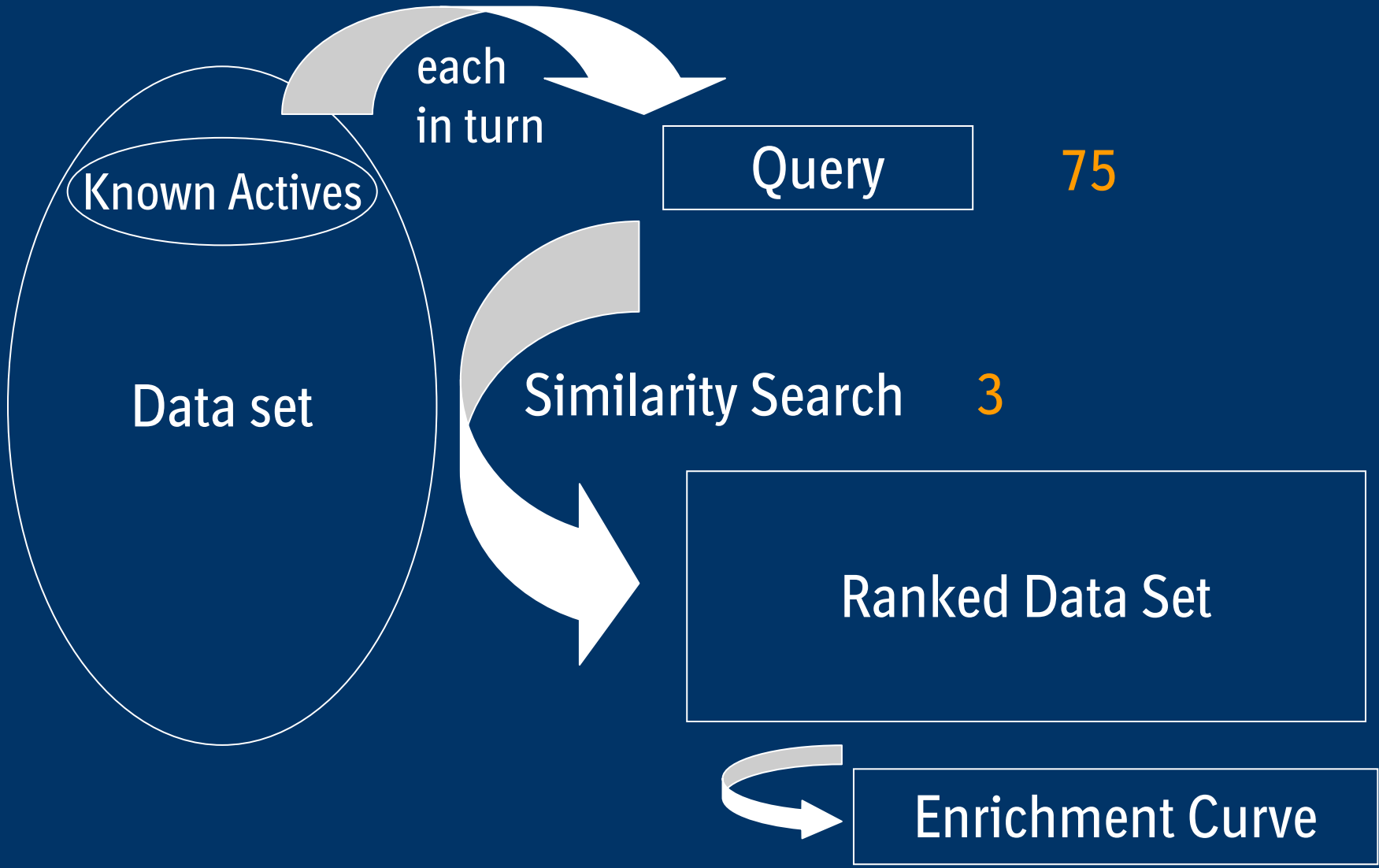
Data Set

Boehringer Ingelheim

75 *$5HT_{1A}$ agonists*
75 *H2 antagonists*
75 *MAO_A inhibitors*
75 *Thrombin inhibitors*
\+ ~ 15.000 compounds chosen randomly from MDDR data base

Examples shown for the $5HT_{1A}$ agonists

First Step


each
in turn
Known Actives
Data set
Query
75
Similarity Search
3
Ranked Data Set
Enrichment Curve

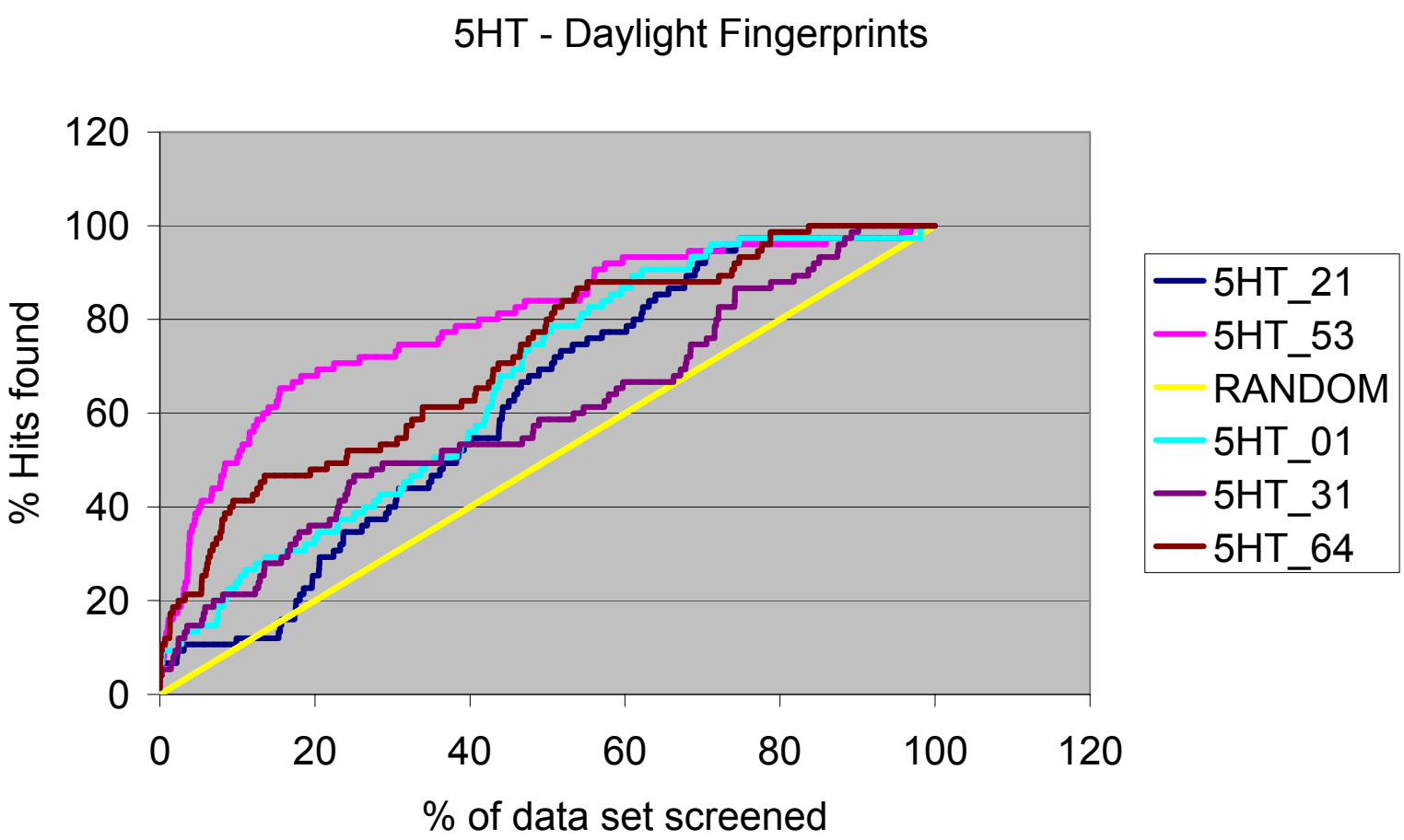
Results from First Step

Boehringer Ingelheim

1. Shapes of individual enrichment curves depend on the query, shown for Daylight Fingerprints

Individual Enrichment Curves - Daylight Fingerprints

Boehringer Ingelheim



Results from First Step

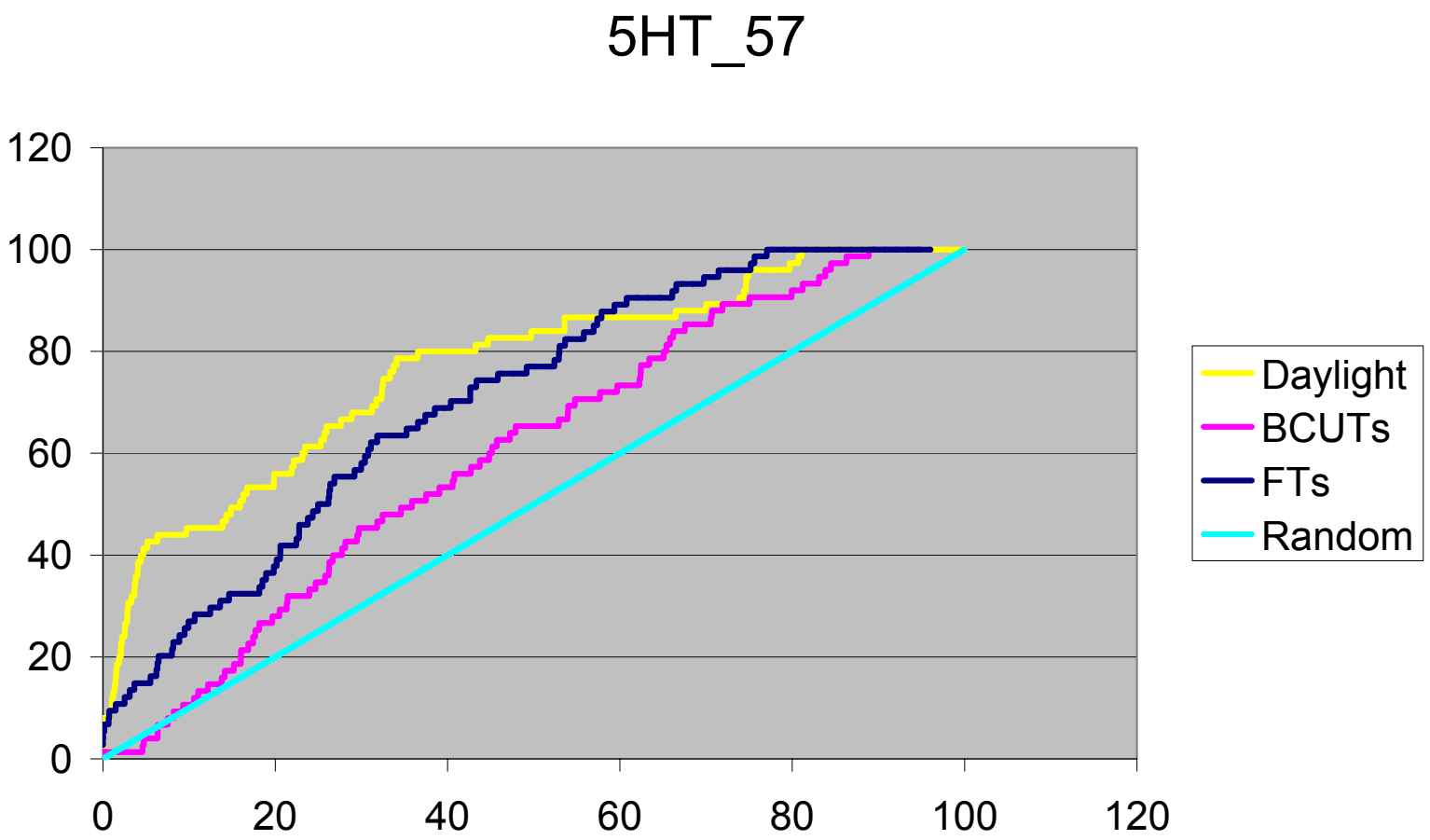
Boehringer Ingelheim

1. Shapes of individual enrichment curves depend on the query
   Valid for all three methods

2. Shapes of individual enrichment curves depend on the method used for similarity searches,
   shown for 5HT_57

Corresponding Results Achieved with Daylight Fingerprints, BCUTs, and FTs

Boehringer Ingelheim



Results from First Step

Boehringer Ingelheim

1. Shapes of individual enrichment curves depend on the query
   Valid for all three methods

2. Shapes of individual enrichment curves depend on the method used for similarity searches,
   shown for 5HT_59

3. Ranking of the 3 methods depends on the queries
   Complementarity?

Consequences from First Step

Boehringer Ingelheim

Global conclusions on this basis questionable!

⇒ Try to reduce variance and / or dependence on the queries

⇒ Analyze complementarity of the methods

Strategy to Reduce Variance

Boehringer Ingelheim

Combination of Queries:

75 x random selection of 3 actives

for each combination:

- determine distances to all 3 actives for the whole data set
- for each compound:
  select the distance to the nearest of the 3 actives
- rank all compounds according to those distances

perform this procedure for all 3 methods

Results for Combinations of 3 Queries

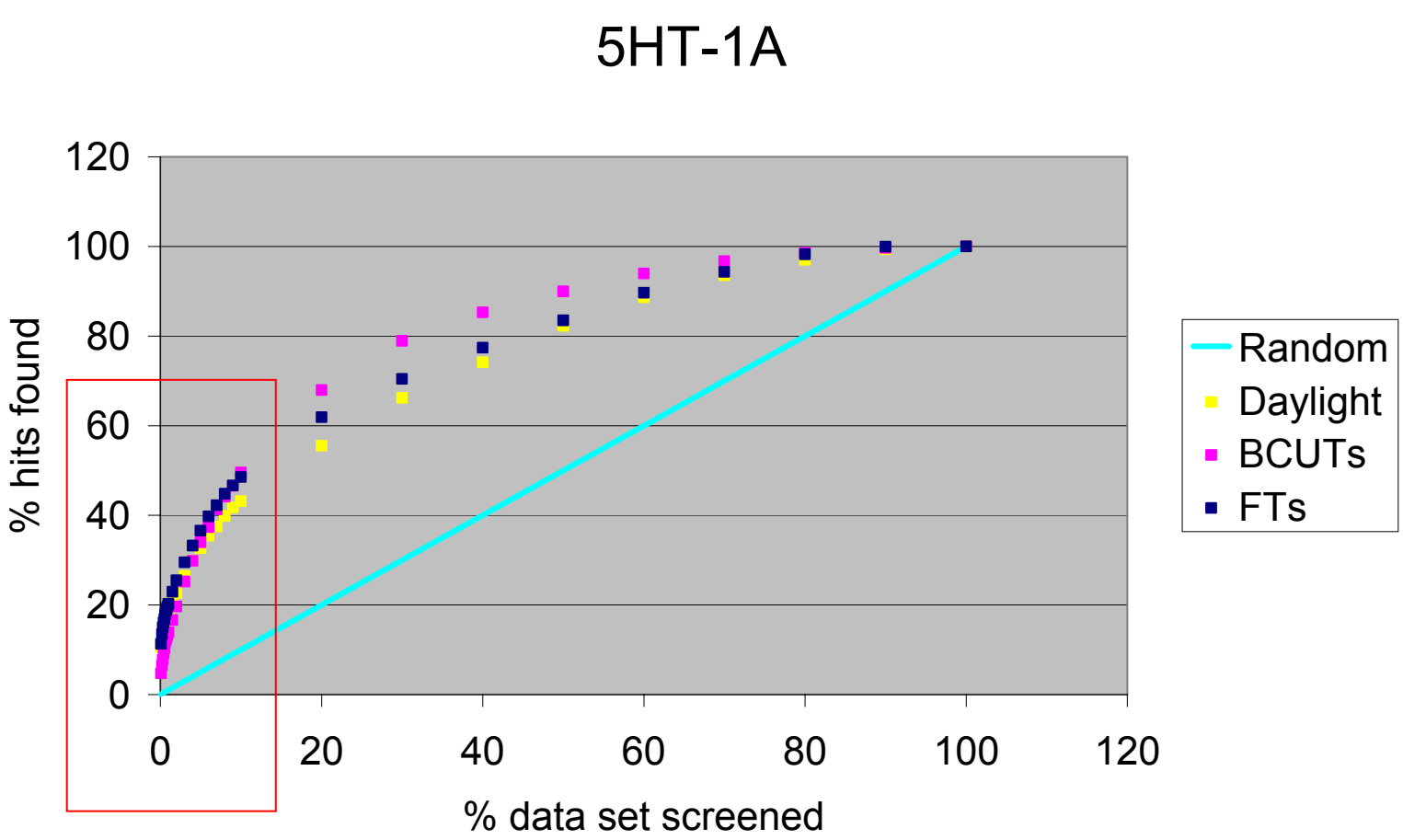
Boehringer Ingelheim

# comp.	method	Single queries		combinations	
		Average # hits	SD	Average # hits	SD
75	Daylight	5.5	2.2	11.1	3.0
	BCUTs	4.2	3.3	7.4	2.9
	FTs	6.4	3.0	12.1	3.5
1500	Daylight	22.2	8.3	30.9	7.0
	BCUTs	29.1	12.1	35.2	6.6
	FTs	26.4	9.3	34.7	8.2

1. Average number of hits found increased
2. Relative SD decreased
3. Trends instead of global conclusions

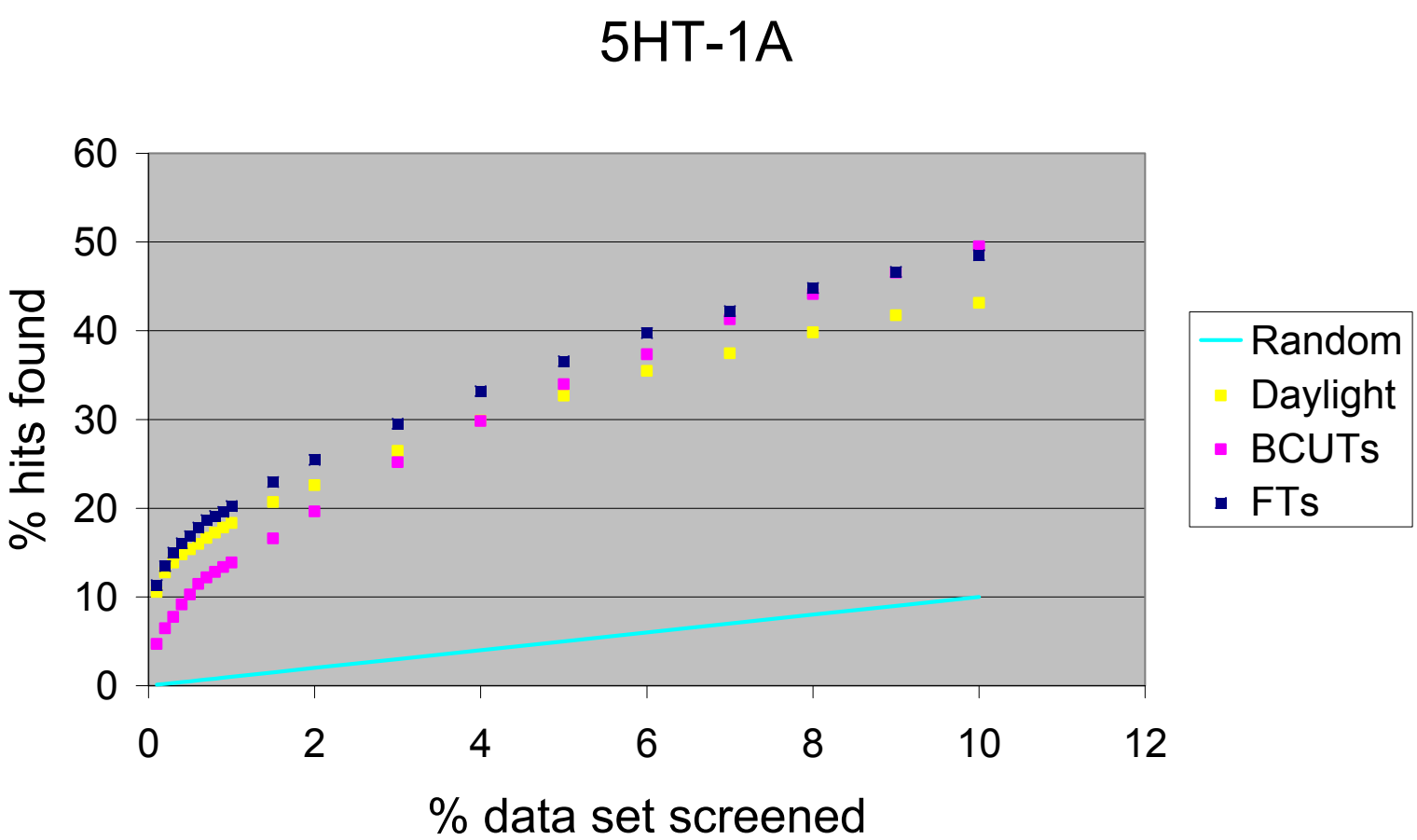
Average Enrichment Curves for 75 Combinations of 3 Queries

Boehringer Ingelheim



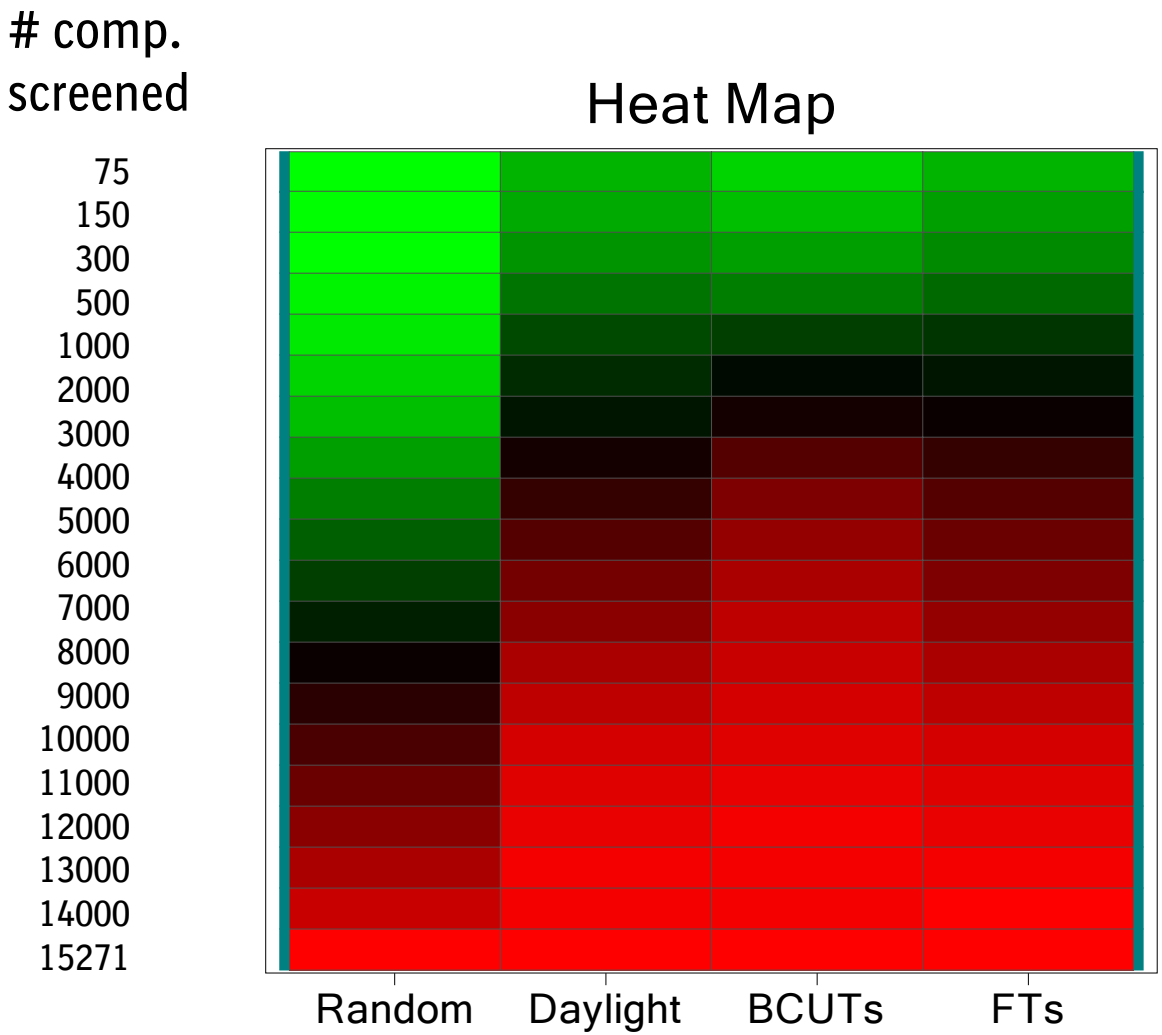
Average Enrichment Curves for 75 Combinations of 3 Queries - Detail

Boehringer Ingelheim



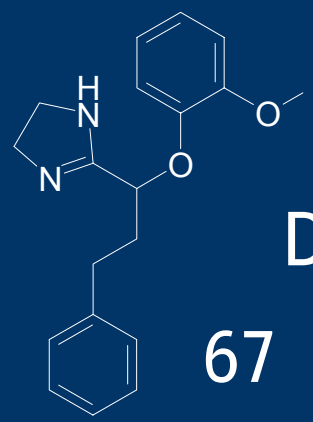
Average Number of Hits Found

Boehringer Ingelheim



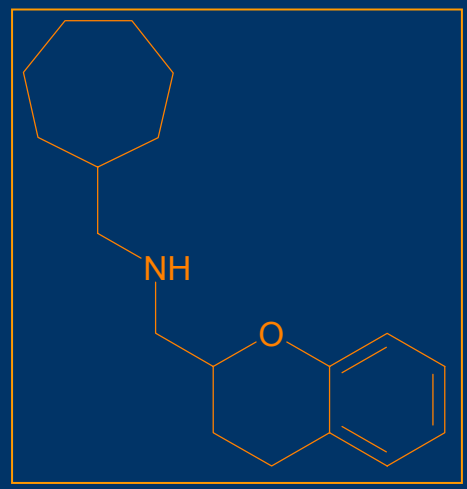
Nearest Neighbors (Actives) to 5HT_59

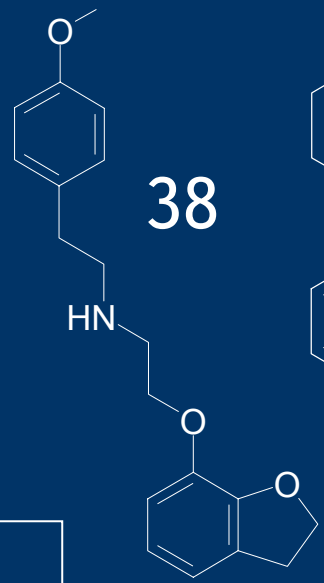
Boehringer Ingelheim



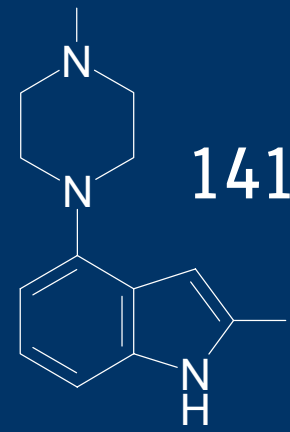
Daylight

67

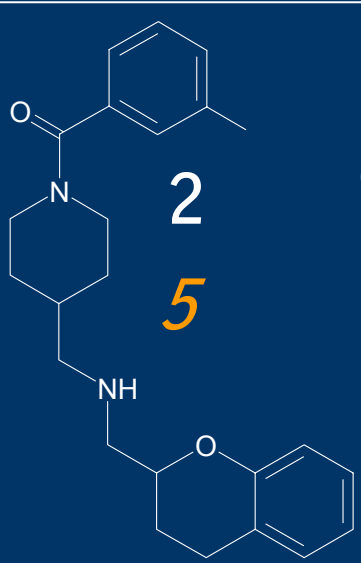




38

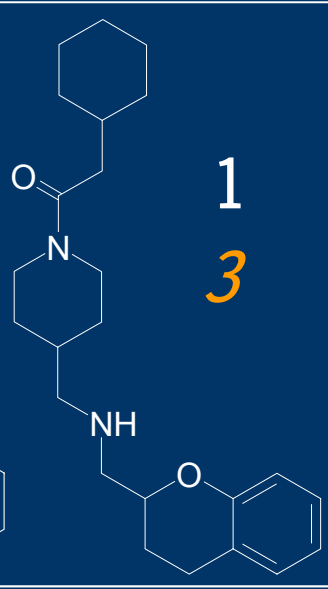


141



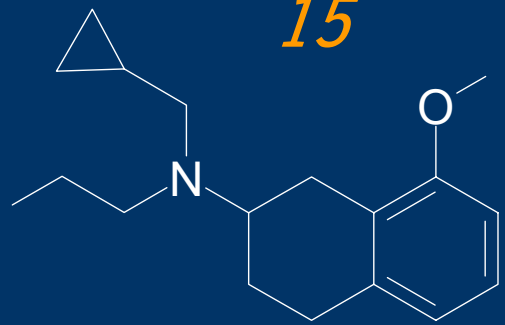
2

5



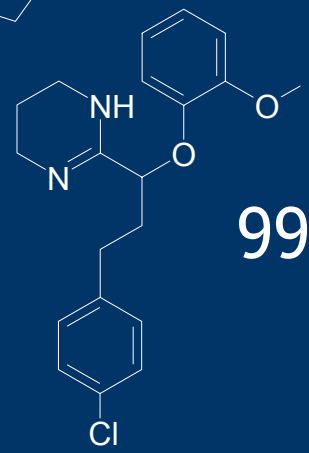
1

3



15

Feature Trees



99

BCUTs

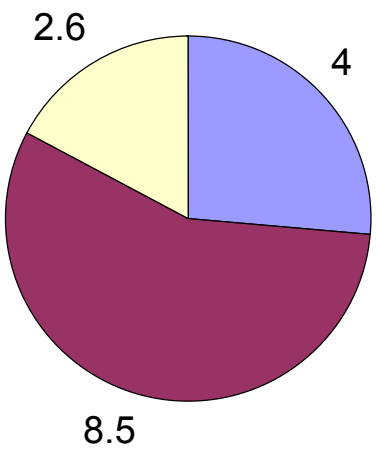
Overlap Daylight – Feature Trees

Boehringer Ingelheim

Average # hits detected by screening x% of the data set

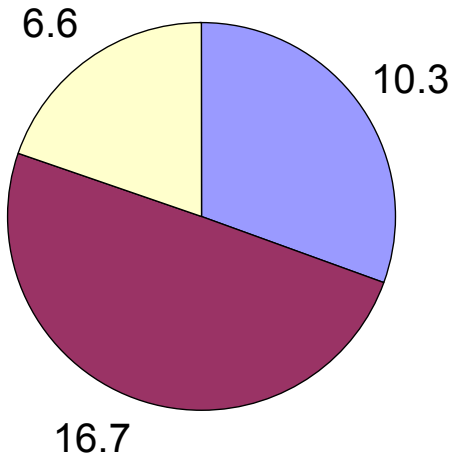
x = 0. 5

15.1 hits found:



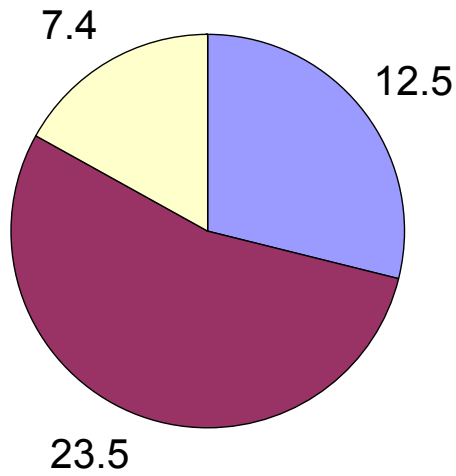
x = 5

33.6 hits found:



x = 10

43.4 hits found:



only Feature Trees only Daylight both

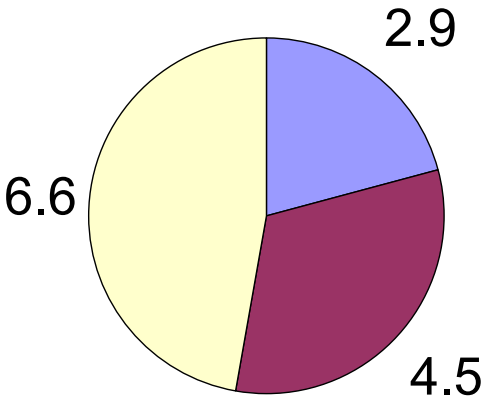
Overlap Daylight - BCUTs

Boehringer Ingelheim

Average # hits detected by screening x% of the data set

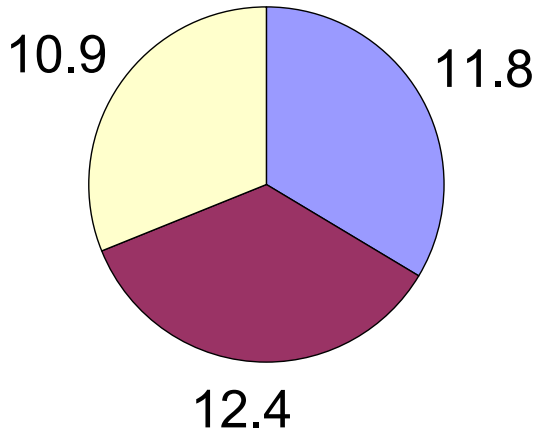
x = 0. 5

14 hits found:



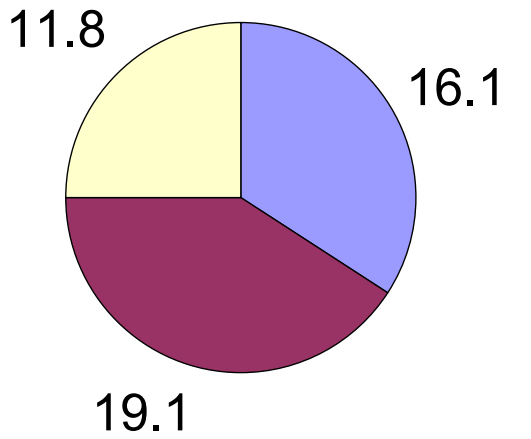
x = 5

35.1 hits found:



x = 10

47 hits found:



only BCUTs only Daylight both

Combination of methods - but how?

Characteristics of Methods

Boehringer Ingelheim

BCUTs:

- Allow scaffold hopping
- Higher percentages of the data set have to be screened to make full use of the method‘s potential

Daylight Fingerprints:

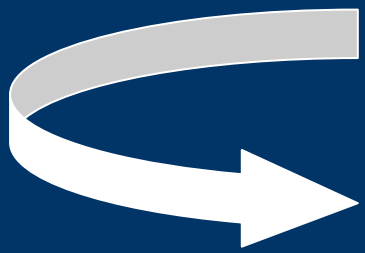
- Especially useful for the detection of actives from the same structural class
- Extremely high enrichments among the very nearest neighbors
- High hit rates among nearest neighbors within a Tanimoto threshold

Similarity Search with Daylight Fingerprints Using a Tanimoto Threshold - Procedure

Boehringer Ingelheim

Combined query:

Act1
Act2
Act3



Rank data set using Daylight Fingerprints

A	0.95
B	0.83
C	0.79
D	0.72
E	0.69
F	0.68
...	

1. Number of combined queries with any nearest neighbors within Tanimoto threshold
2. Average hit rate of subsets from queries with any nearest neighbors within Tanimoto threshold
3. Sum of hits and sum of non-hits within all subsets from all queries

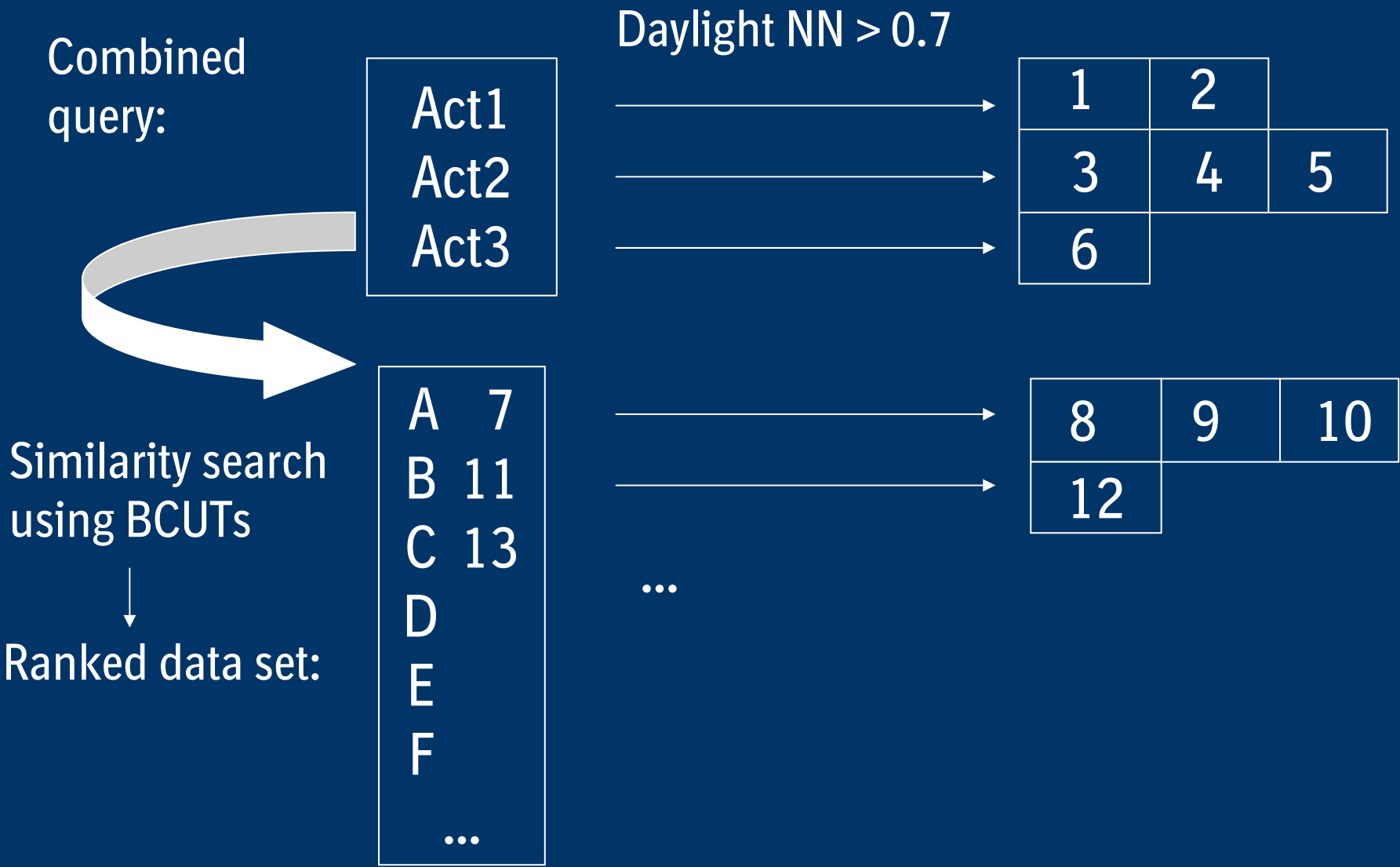
Similarity Search with Daylight Fingerprints Using a Tanimoto Threshold - Results

Boehringer Ingelheim

Tanimoto Threshold	# Queries with NNs	Average hit rate	# hits	# non-hits
0.8	73	94.1 %	233	8
0.7	75	88.0 %	387	60
0.6	75	55.6 %	549	602

Procedure

Boehringer Ingelheim



Average Number of Hits Found

Boehringer Ingelheim

# comp. screened	Daylight	BCUTs	Daylight + BCUTs	Random
75	11.1	7.4	9.9	0.4
500	19.9	19.0	21.9	2.4
1500	30.9	35.2	39.6	7.1

1. Combination better than BCUTs for screening 75 compounds
2. Combination better than both methods for all other cases
3. Single methods as well as combination clearly superior to random selection

Conclusions

Boehringer Ingelheim

- Reasonable enrichments of actives can be achieved using each of the three methods to measure similarity
- Results of the three methods are complementary to each other
- Daylight Fingerprints show
  - extremely high enrichments among the very nearest neighbors (actives from the same structural class)
  - High hit rates among nearest neighbors within a Tanimoto threshold (e.g. 0.8 / 0.7)
- BCUT distances allow scaffold hopping, but higher percentages of the data set have to be screened to make full use of the method‘s potential
- Feature Trees allow scaffold hopping, but they are also useful for the detection of actives from the same structural class
- Improvement of results by combining methods

Acknowledgements

Michael Bieler

Bernd Wellenzohn

Herbert Köppen

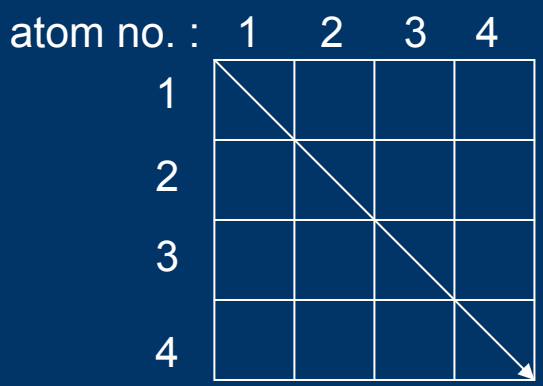
BACKUP

Descriptors

Boehringer Ingelheim

Generally any kind of descriptors can be used!

Diverse Solutions provides **BCUT values**:



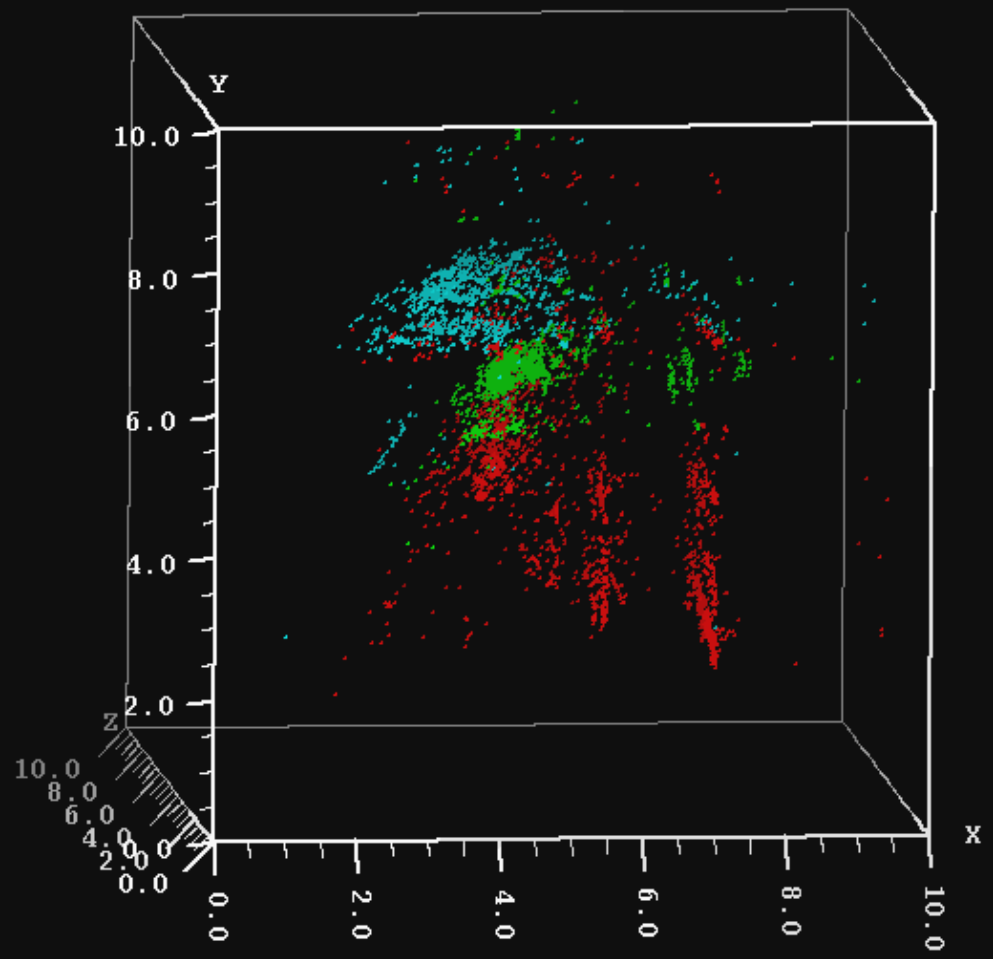
diagonal elements contain atomic properties:
- Gasteiger charges
- H-donor and H-acceptor abilities
- polarizabilities

off-diagonal elements reflect connectivity information: 2D, 3D, topological BCUTs

for each matrix different BCUT values:
- highest and lowest eigen values
- set of scaling factors

Clustering of Compounds from Different Activity Classes

GPCR ligands
Kinase inhibitors
Protease inhibitors

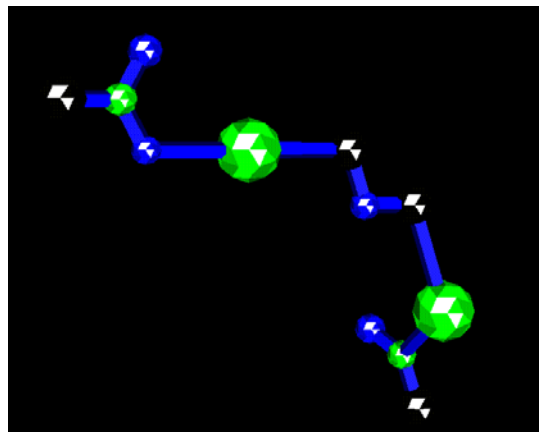


BCUT values useful for similarity searches / virtual screening?

Feature Trees

Boehringer Ingelheim

Instead of a linear representation of a molecule, the molecule is described by a tree structure representing its major chemical building blocks and the way they are connected.



Characteristics:

- conformation independent (2.5 D)
- fragment based
- can handle local similarity