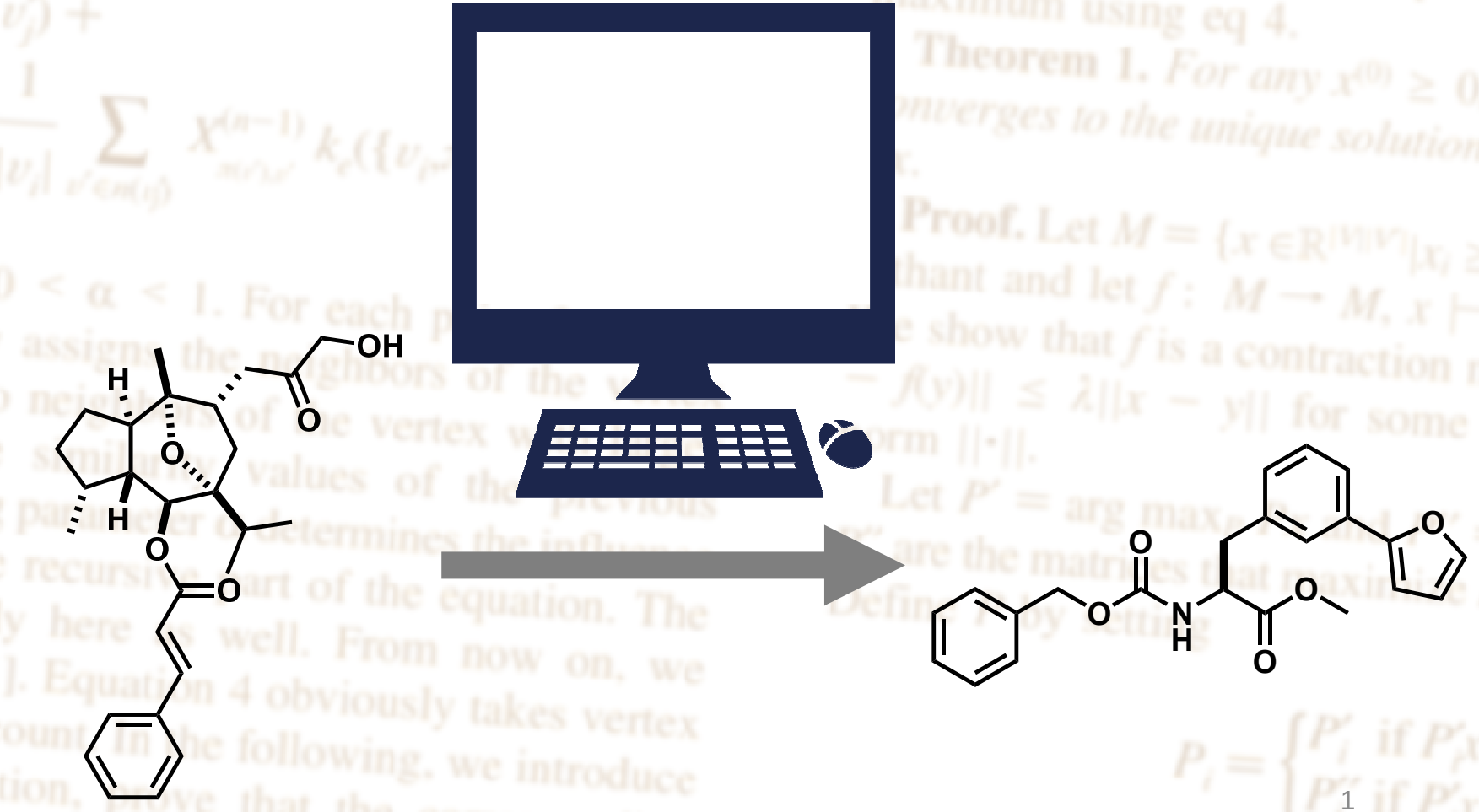
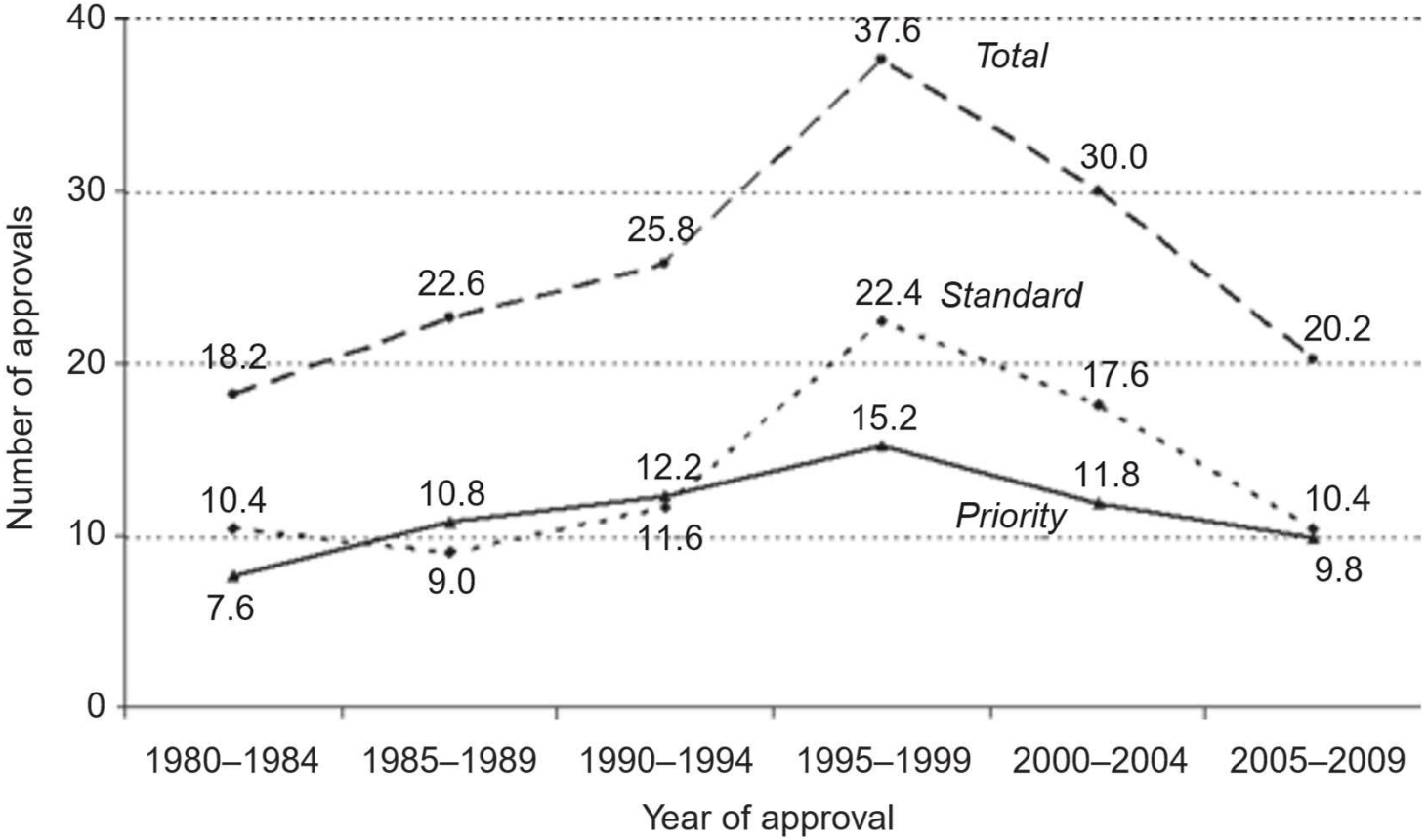
Ligand-Based *De Novo* Drug Design

2017/7/22 Kotaro Tokumoto



Contents

1-1. Motivation for *De Novo* Design



Average annual number of approved priority, standard, and total new molecular entities and significant biologicals (shown in figure) is decreasing.

New methods to exploit whole chemical space are saught.

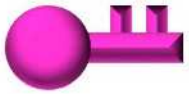
(a) Kaitin, K. I.; DiMasi, J. A. *Clin. Pharmacol. Ther.* **2011**, *89*, 183. (b) Honma, T. *Med. Res. Rev.* **2003**, *23*, 606.

1-2. *De Novo* Drug Design

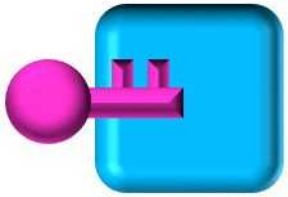
De Novo
≃ *Ab Initio*, From the Beginning

chemical synthesis

drug
(ligand)



target
(mainly protein)



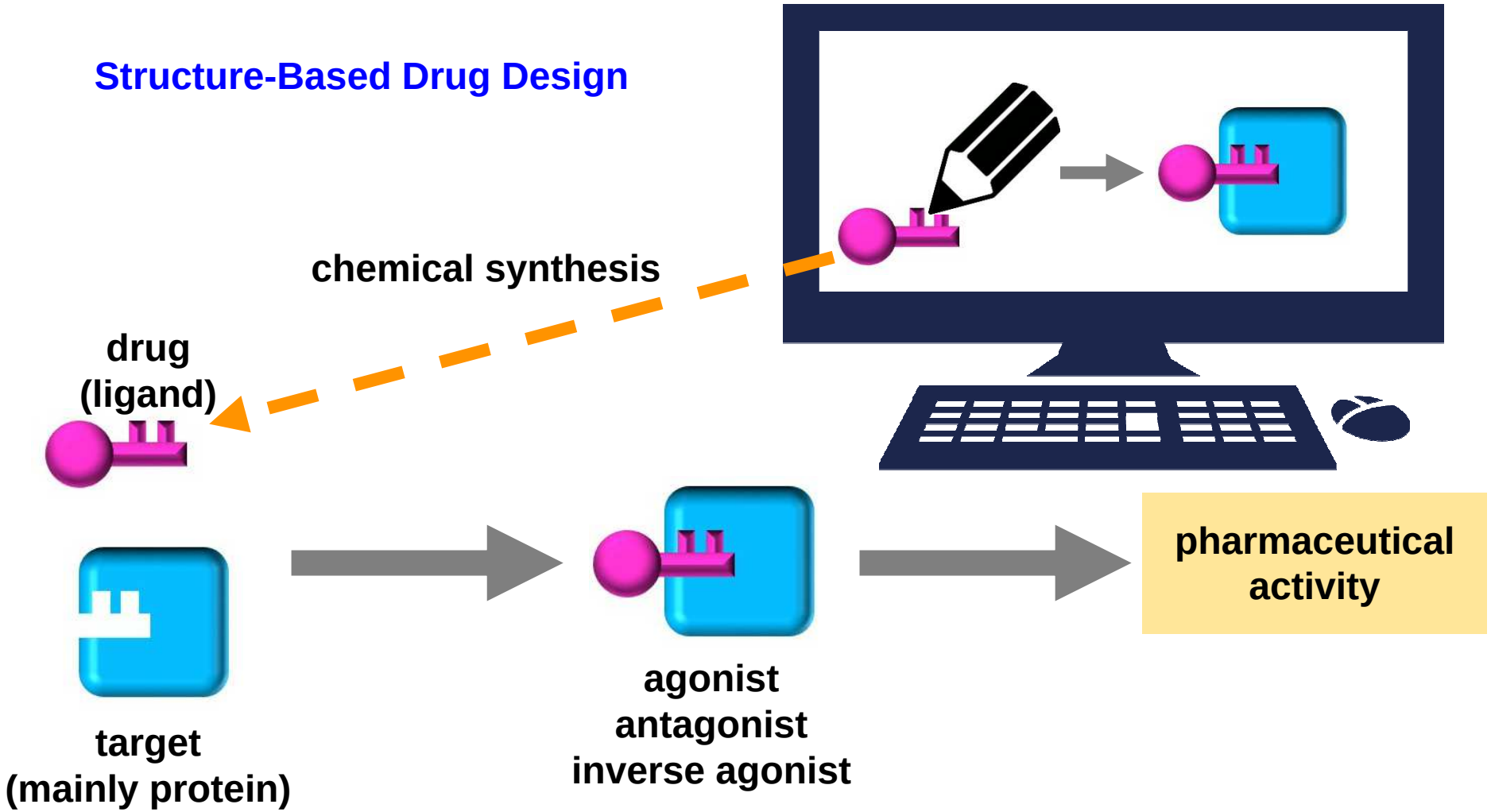
agonist
antagonist
inverse agonist

pharmaceutical activity

- **structure- (receptor-) based approach**
- **ligand-based approach**

Aparoy, P.; Reddy, K. K.; Reddanna, P. *Curr. Med. Chem.* **2012**, *19*, 3763

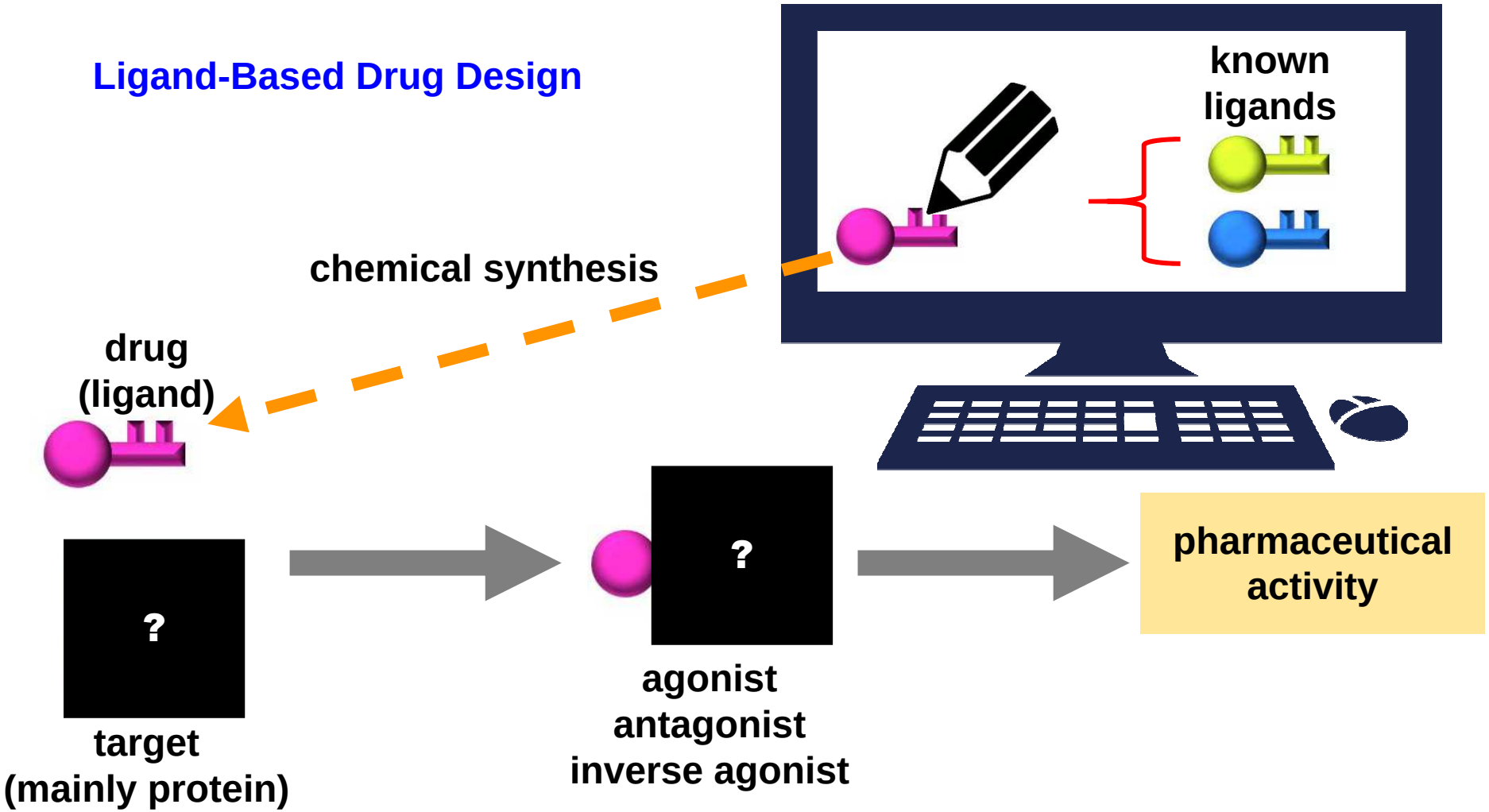
1-3. Structure-Based Drug Design



- designing a molecule so that it binds to the target protein
- popular approach (cf. docking study)
- requirement for sufficient structural information of the target protein

Aparoy, P.; Reddy, K. K.; Reddanna, P. *Curr. Med. Chem.* **2012**, *19*, 3763

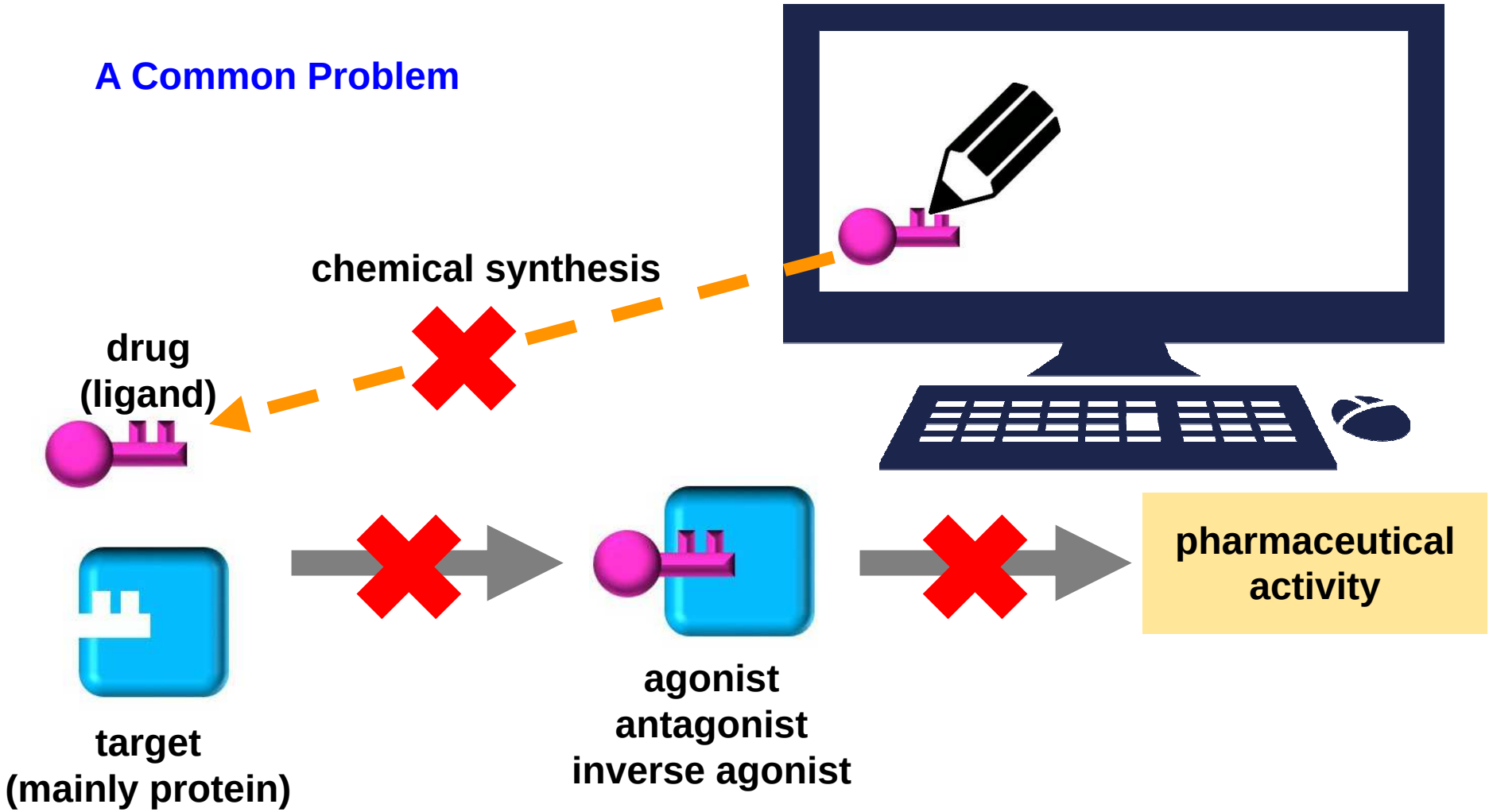
1-4. Ligand-Based Drug Design



- designing a molecule taking the structures of the known ligands into consideration.
- exploiting pharmacophore, machine learning, etc.
- available for target proteins without sufficient structural information.

Aparoy, P.; Reddy, K. K.; Reddanna, P. *Curr. Med. Chem.* **2012**, *19*, 3763

1-5. A Common Problem in Drug Design



- **Virtually constructed molecules often fail to be synthesized in reasonable time frames and with acceptable effort.**

Hartenfeller, M.; Zettl, H.; Walter, M.; Rupp, M.; Reisen, F.; Proschak, E.; Weggen, S.; Stark, H.; Schneider, G. *PLoS Comp. Biol.* **2012**, *8*, e1002380.

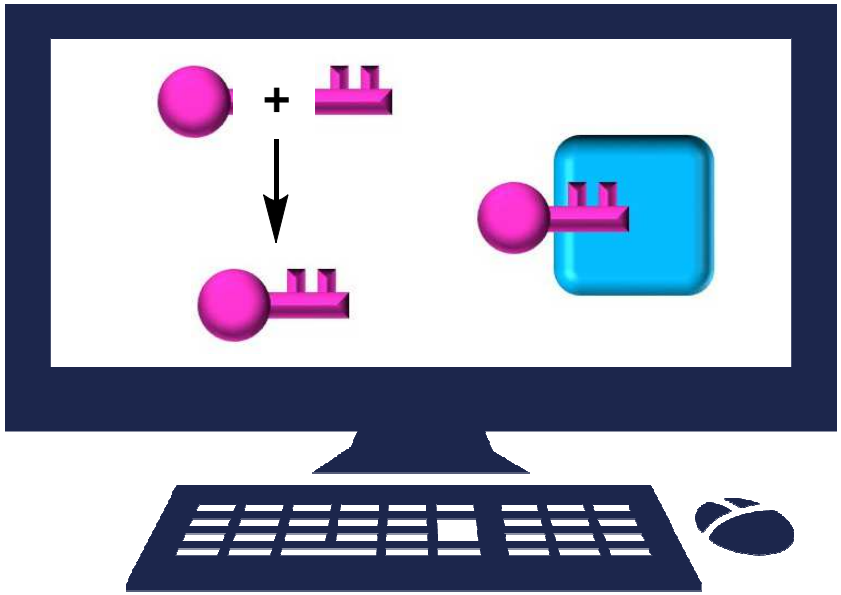
1-6. Attempts to Resolve Synthetic Problem

Solution to the Synthesizability Problem:

- employing rules to guide the assembly process
- computing the synthetic process during the virtual optimization of the structure

For Structure-Based Drug Design:

- RECAP (Retrosynthetic Combinatorial Analysis Procedure)[1)]
- SYNOPSIS (SYNthesize and OPtimize System *In Silico*)[2)]



1) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511. 2) Vinkers, H. M.; de Jonge, M. R.; Daeyaert, F. F. D.; Heeres, J.; Koymans, L. M. H.; van Lenthe, Lewi, P. J.; Timmerman, H.; Aken, K. V.; Janssen, P. A. J. *J. Med. Chem.* **2003**, *46*, 2765.

Contents

2-1. Prof. Gisbert Schneider (1)

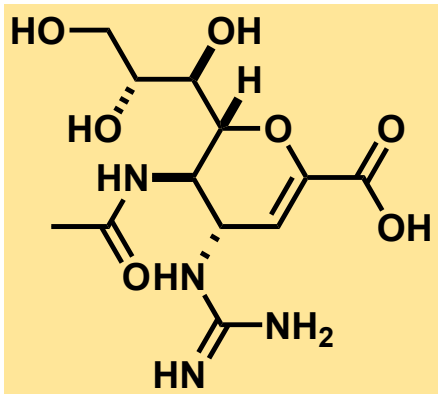
Prof. Gisbert Schneider

1965	Born in Fulda, Germany
1991	B.S. biochemistry at the Free University of Berlin, Germany (Prof. Paul Wrede)
1994	Ph.D. biochemistry at the Free University of Berlin, Germany (Prof. Paul Wrede)
1994-1997	Postdoctoral fellow at the Benjamin Franklin University Clinic, Berlin, the Massachusetts Institute of Technology in Cambridge, MA, the University of Stockholm, Sweden, and the Max-Planck-Institute of Biophysics in Frankfurt, Germany
1997-2002	Scientific specialist in industrial research at Hoffmann – La Roche Ltd.
2002-2009	Professor at Goethe University in Frankfurt, Germany (Beilstein Endowed Chair for Chem- and Bioinformatics)
2010-	Professor at ETH Zürich (Computer-Assisted Drug Design at the Institute of Pharmaceutical Sciences in the Department of Chemistry and Applied Biosciences)

2-2. Prof. Gisbert Schneider (2)

Research Project

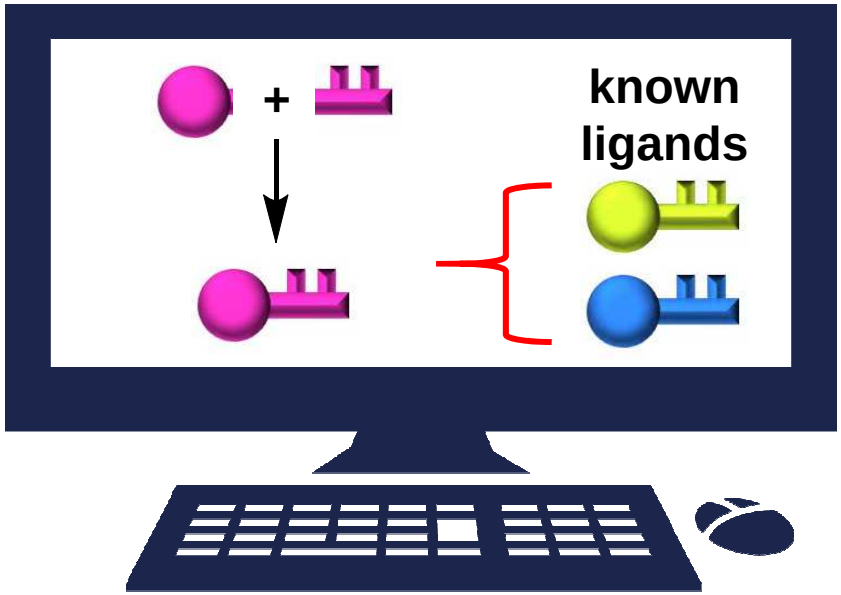
development of methods for virtual screening, molecular *de novo* design and adaptive autonomous systems in drug research



2-3. DOGS (Design of Genuine Structures) (1)

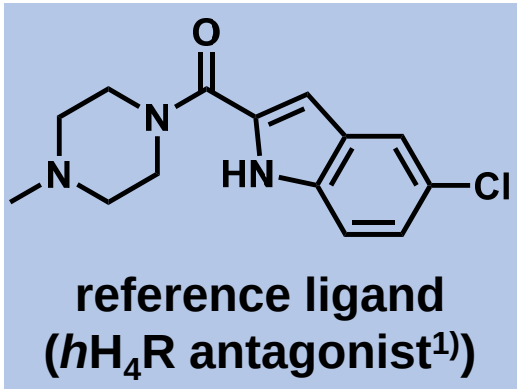
What is DOGS?

- a ligand-based software tool for '*in silico*' assembly of potentially novel bioactive compounds
- a computational method for ligand-based *de novo* design through chemical reaction.
- comprising a set of 25,114 commercially available building blocks for performing 'virtual syntheses' of compounds with 83 reactions



2-4. DOGS (Design of Genuine Structures) (2)

An Example of Output of DOGS

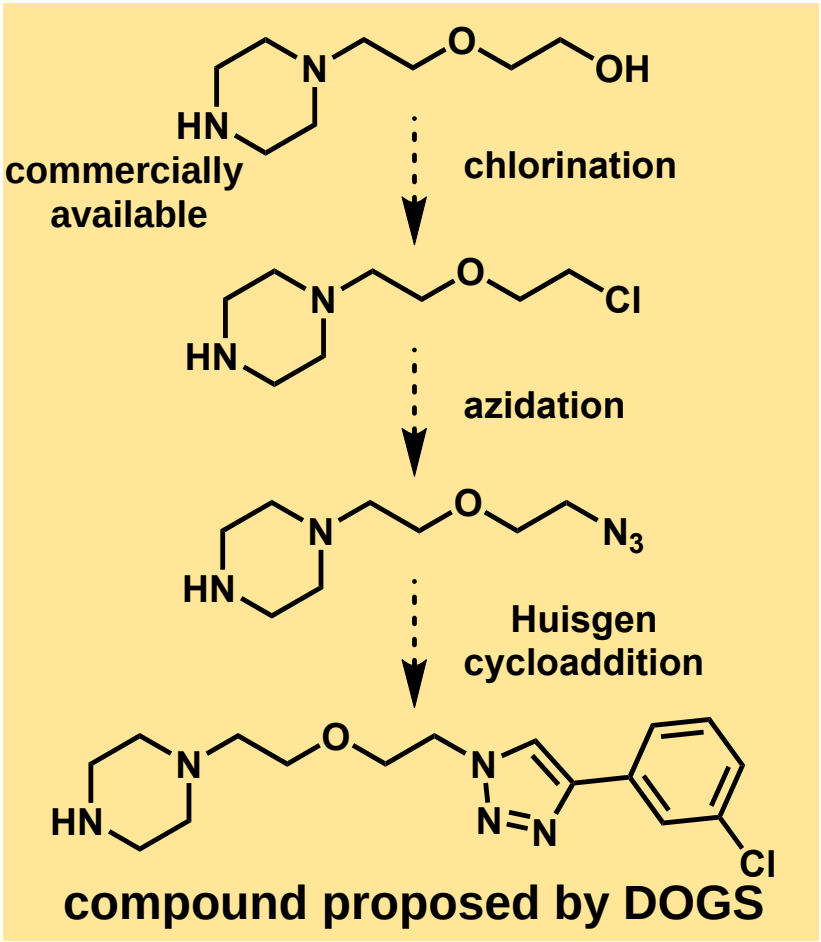


reference ligand
(*h*H_4R antagonist[1])

input

DOGS

"*scaffold-hopping*[2]"

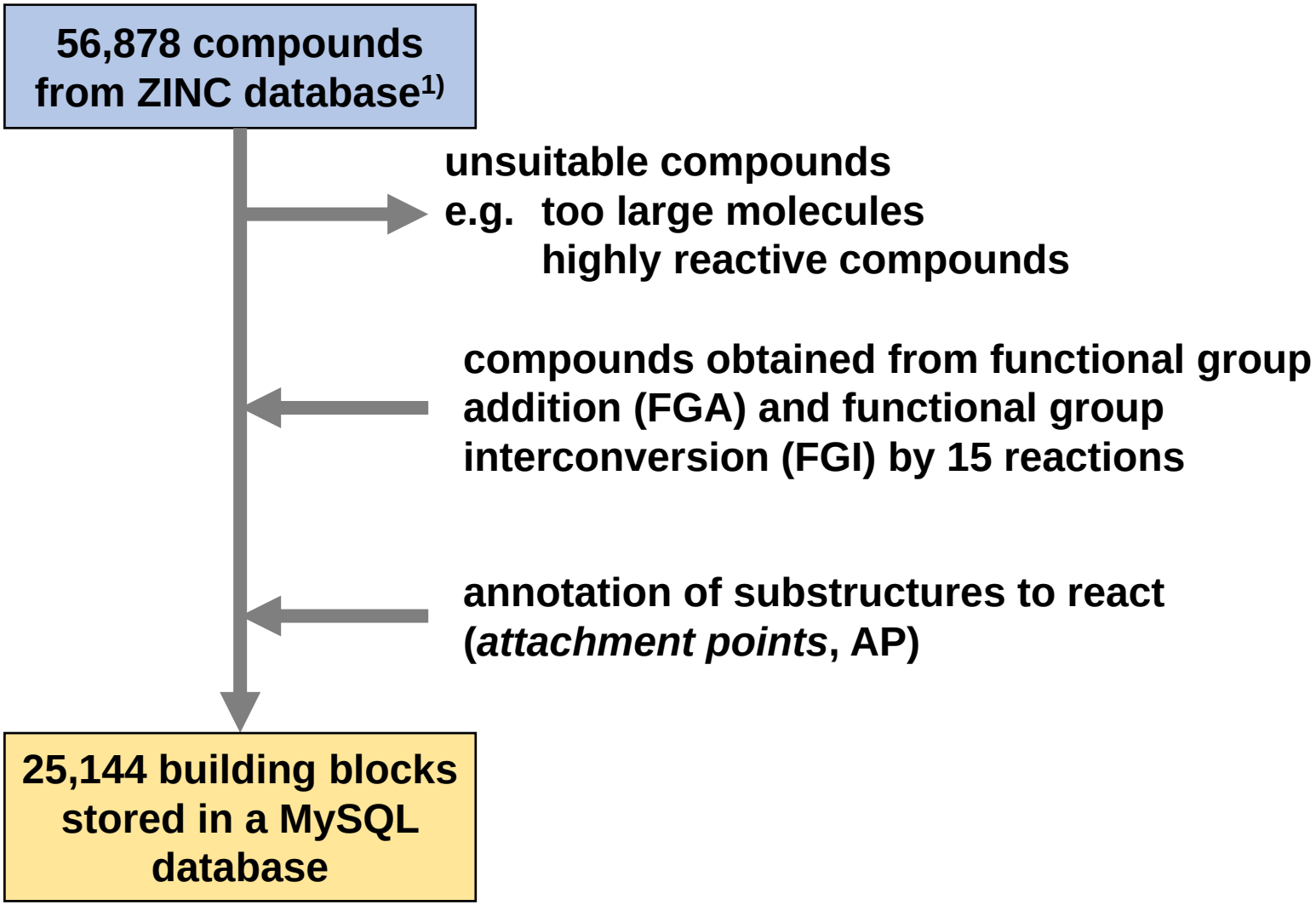


compound proposed by DOGS

output

1) Thurmond, R. L.; Desai, P. J.; Dunford, P. J.; Fung-Leung, W.-P.; Hofstra, C. L.; Jiang, W.; Nguyen, S.; Riley, J. P.; Sun, S.; Williams, K. N.; Edwards, J. P.; Karlsson, L. *J. Pharm. Exp. Ther.* **2004**, *309*, 404. 2) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894.

2-5. Building Block Library



1) Irwin, J. J.; Shoichet, B. K. *J. Chem. Inf. Model.* **2005**, *45*, 177.

2-6. Library of Chemical Reactions

Selection of Reactions

- **a single product from one or two reactants**

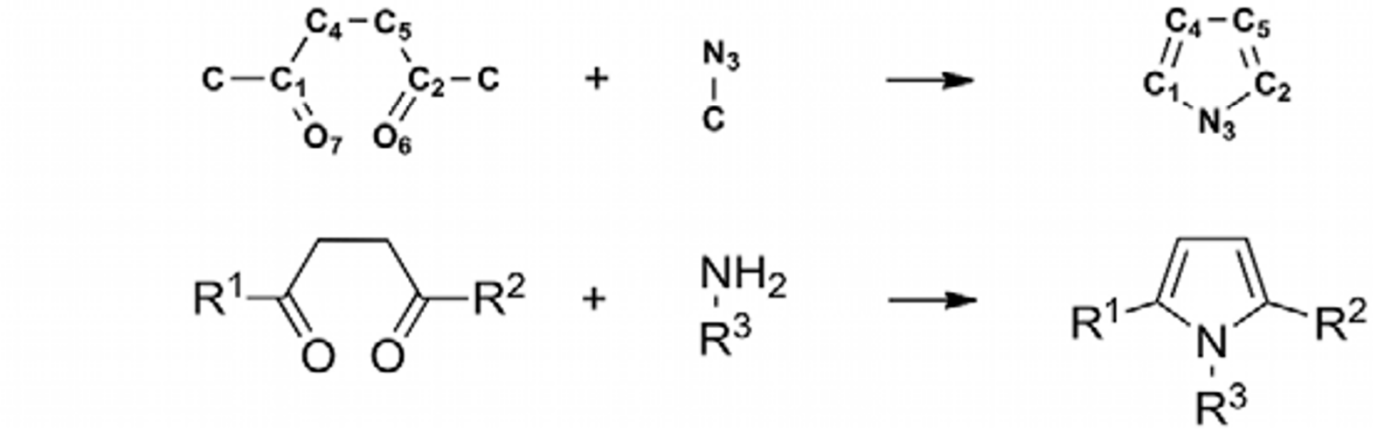
  A → B or A + B → C

Encoding of Reactions (83 reactions)

- **Reaction-MQL[1)]**

  an example

C-C1[!ring](=O7)-C4[!aromatic & bound(-H)]-C5[!aromatic & bound(-H)]-C2(=O6)-C ++ C[!bound(=O)]-N3[allHydrogens=2 & charge=0] >> Paal-Knorr pyrrole >> C1$1-N3-C2=C5-C4$=1



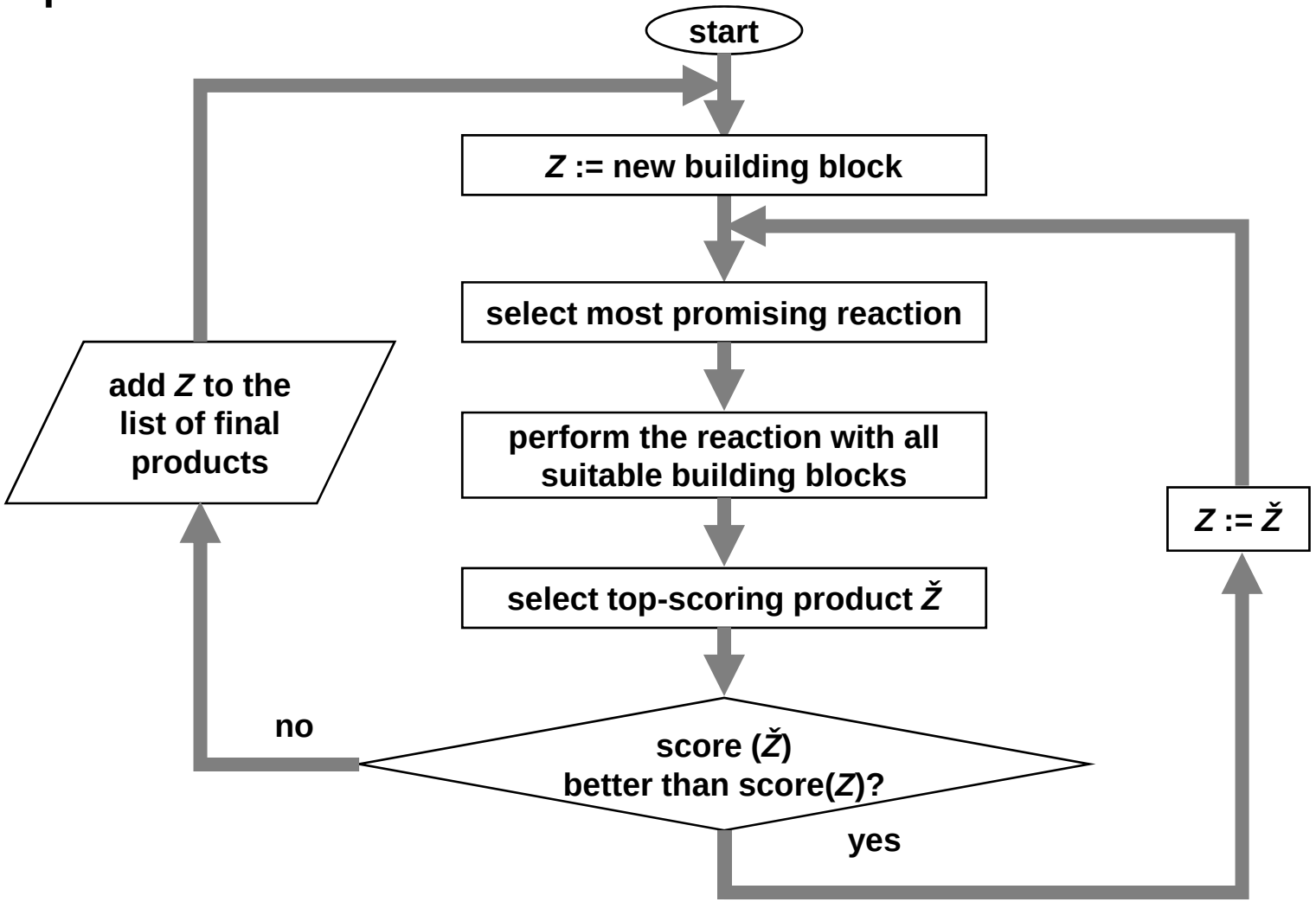
1) Reisen, F. H.; Schneider, G.; Proschak, E. *J. Chem. Inf. Model.* **2009**, *49*, 6.

2-7. Construction Algorithm

Flowchart of the Construction Algorithm

(Some processes were omitted for clarification.)

Score function returns the similarity of the compound to the reference compound. The details will follow later.



2-8. Scoring Function (1)

ISOAK (Iterative Similarity Optimal Assignment Kernel)[1)]

Formation of Molecular Graph[2)]

- **A molecular graph is consisted of nodes (atoms except for hydrogen) and edges (bonds).**
- **Each node is labeled with one of the following "colors":**

A: hydrogen bond acceptor	**D: hydrogen bond donor**
E: hydrogen bond donor & acceptor	**P: positive charge**
N: negative charge	**R: aromatic**
L: lipophilic	**0: no type**

- **Each edge is labeled with one of the following properties: single, double or triple.**

1) Rupp, M; Proschak, E.; Schneider, G. *J. Chem. Inf. Model.* **2007**, *47*, 2280. 2) Cayley, A. *Philos. Mag.* **1874**, *47*, 444.

2-9. Scoring Function (2)

Evaluation of Topological Similarity

molecular graph

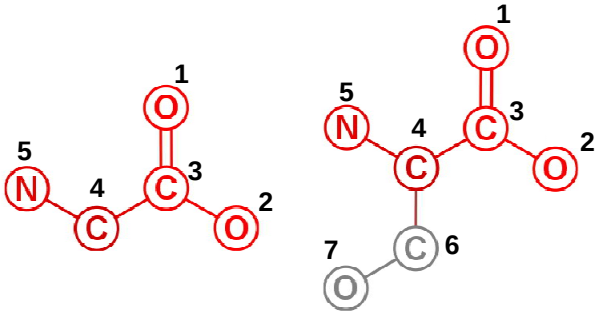
iterative computation

(i) calculation of atom pair similarities ("Two atoms are similar if their neighbors are similar")

(ii) optimal assignment
Each atom of the smaller graph is assigned to one atom of the larger graph, so that the sum of atom pair similarities is maximized.

	1	2	3	4	5	6	7
1	0.98	0.50	0.00	0.00	0.00	0.00	0.50
2	0.50	0.98	0.11	0.16	0.34	0.17	0.89
3	0.00	0.11	0.96	0.68	0.14	0.78	0.13
4	0.00	0.24	0.67	0.81	0.17	0.77	0.20
5	0.00	0.33	0.14	0.13	0.91	0.20	0.38

(e) Pairwise vertex similarities. Boxes indicate vertex assignments.



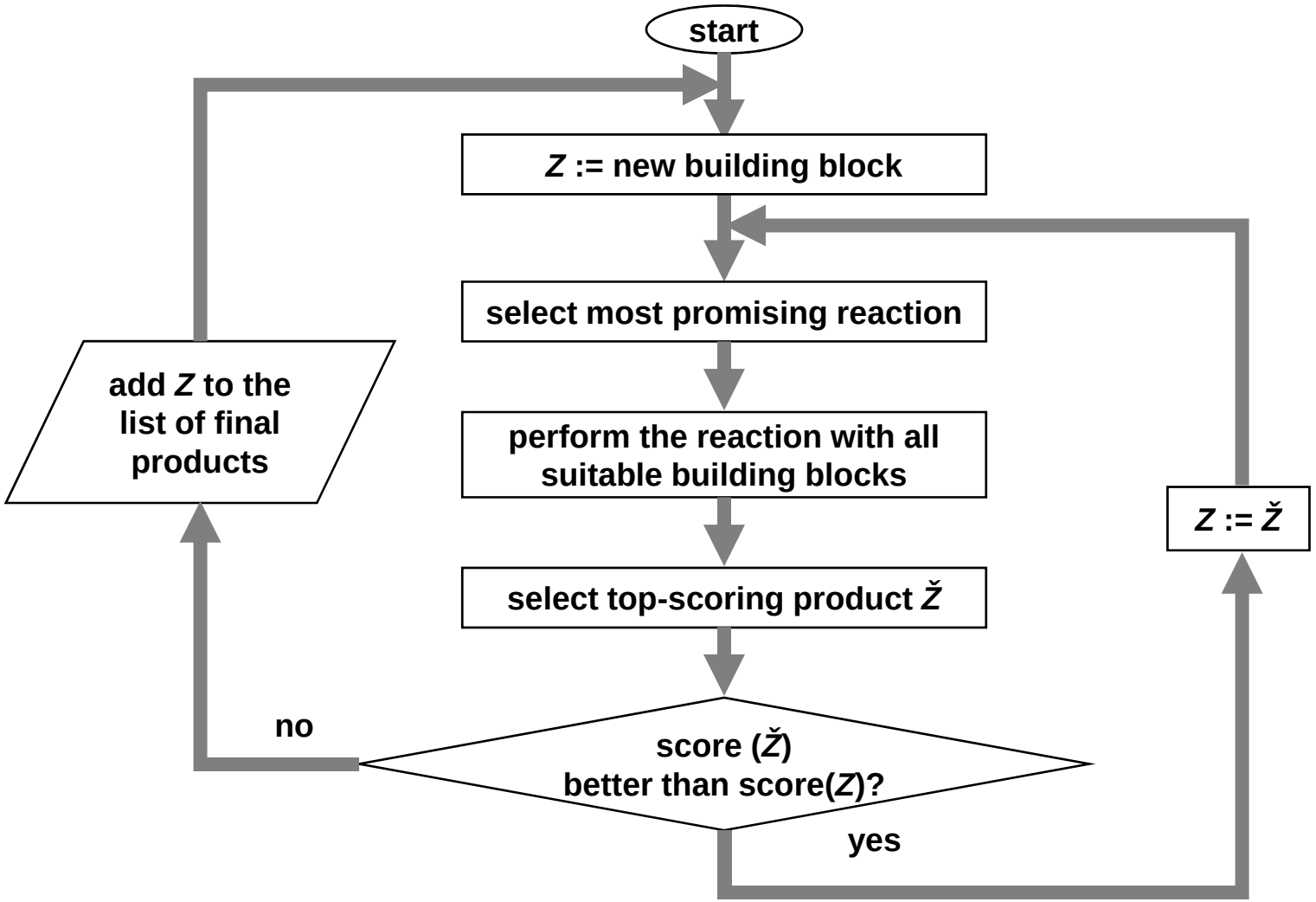
(f) Vertex assignments, color-coded from red (highest similarity) to black (lowest similarity); gray vertices were not assigned.

(a) Rupp, M; Proschak, E.; Schneider, G. *J. Chem. Inf. Model.* **2007**, *47*, 2280. (b) Rupp, M. *Doctor thesis, Goethe University Frankfurt*, 2009.

2-10. Construction Algorithm

Flowchart of the Construction Algorithm

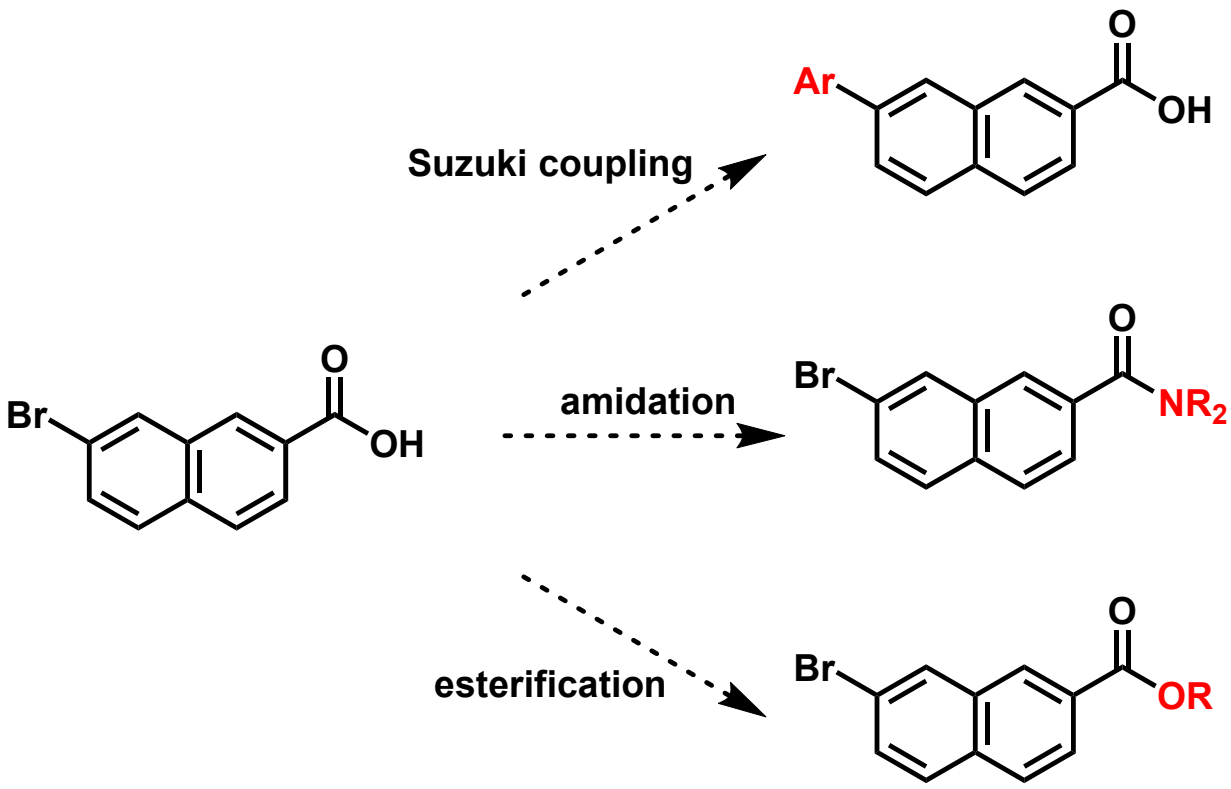
(Some processes were omitted for clarification.)



2-11. Selection of Reaction (1)

Selection of the Applied Reaction

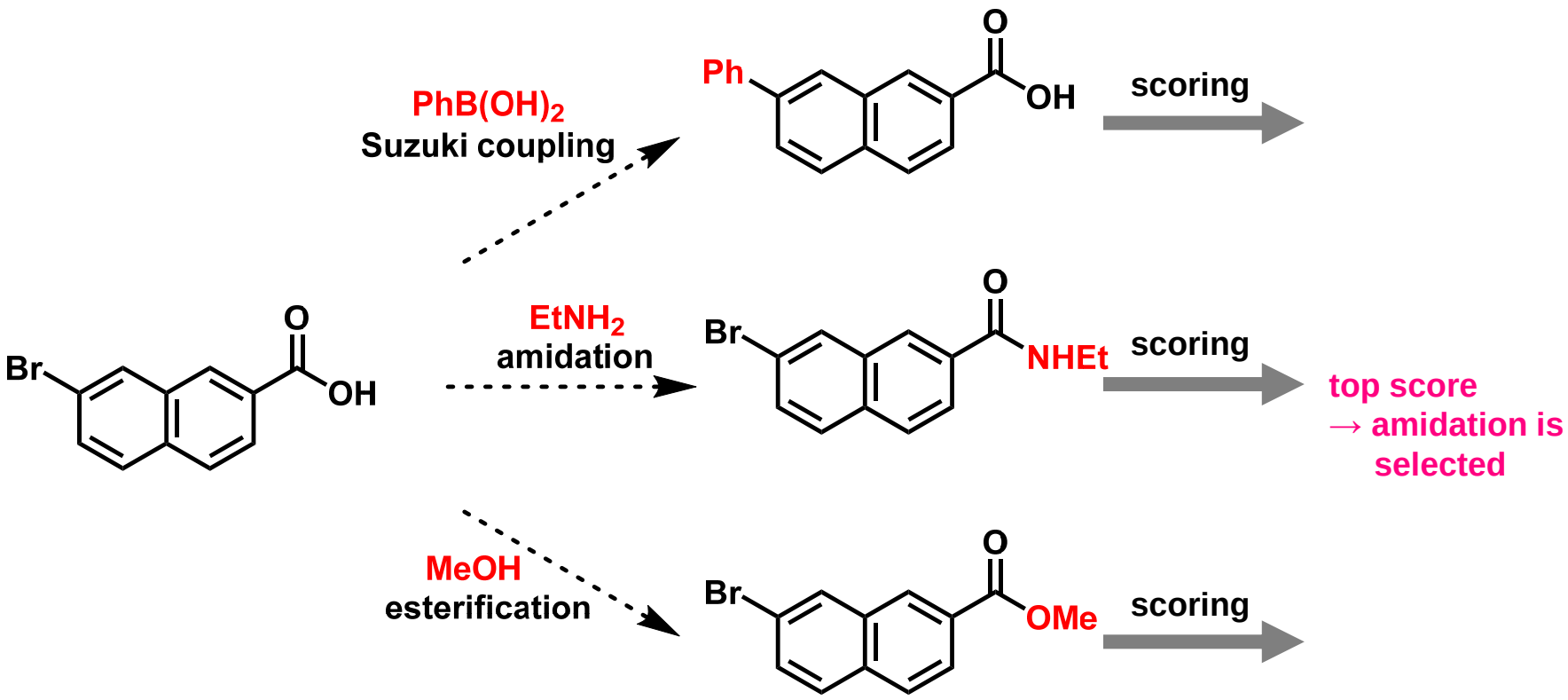
- Which reaction is chosen?
- Which building block is chosen?



2-12. Selection of Reaction (2)

"Minimal dummy fragments"

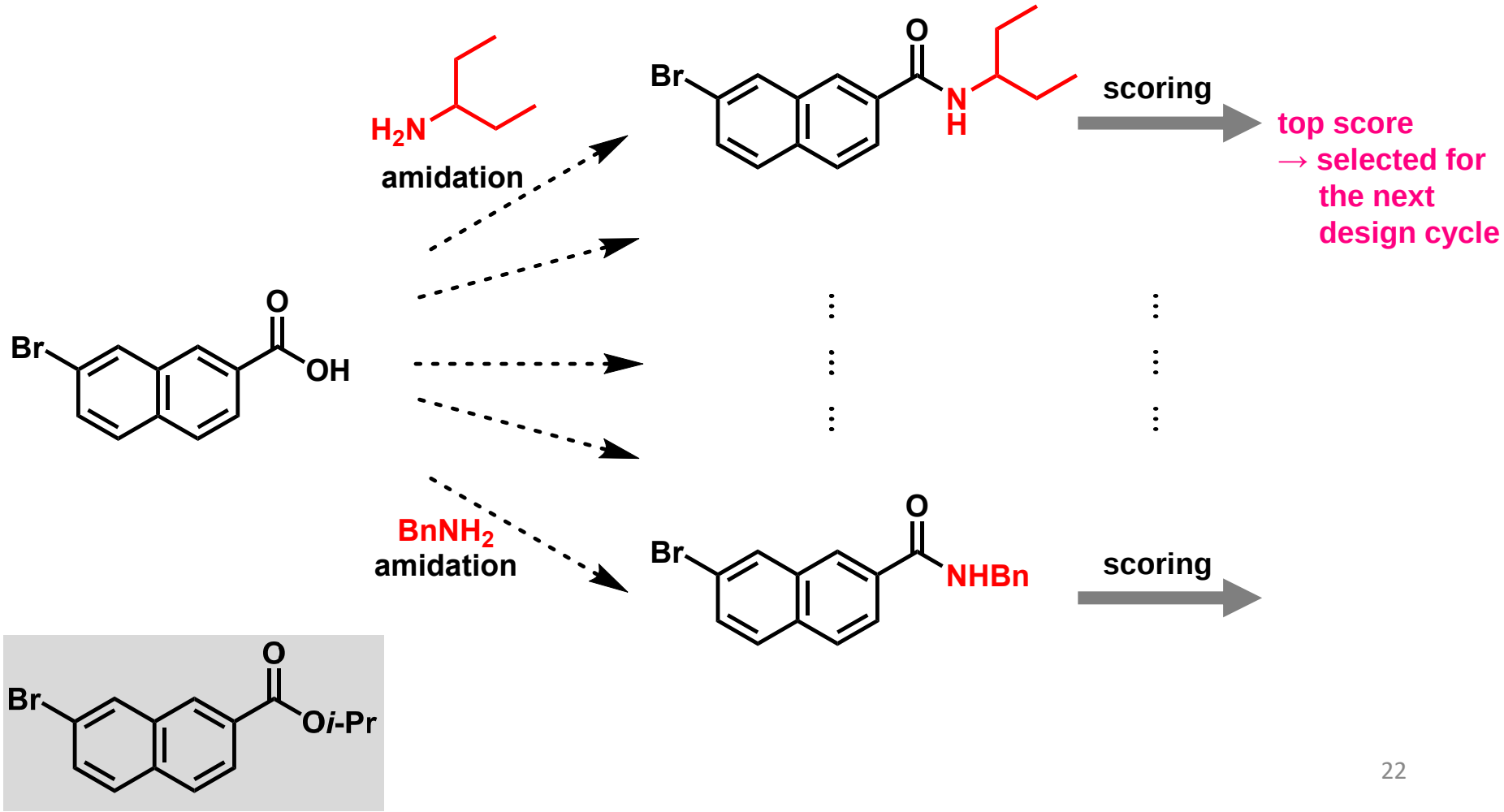
- **a virtual molecule that satisfies the minimal structural demand to participate in a specific reaction**



2-13. Selection of Reactant

Selection of Reactant for the Determined Reaction

- All building blocks which can participate in the reaction is tested.



2-14. Characteristics of DOGS

Merits

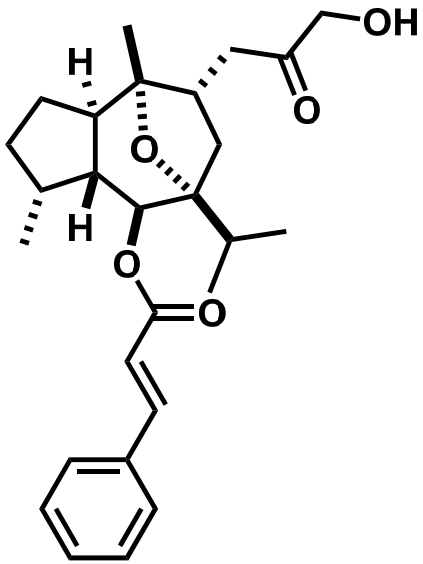
- high synthesizability
- proposal of synthetic route (basically, ≤ 10 steps)
- unnecessariness for 3D structure of the target protein
- elimination of inappropriate functional groups

Demerits

- disregard for the functional group tolerance
- disregard for stereochemistry and conformation of the molecule
- disregard for synthetic cost
- requirement for reference ligands
- lower hit rate compared to structure-based approach

Contents

3-1. (−)-Englerin A



(−)-Englerin A

Isolation and Structural Determination

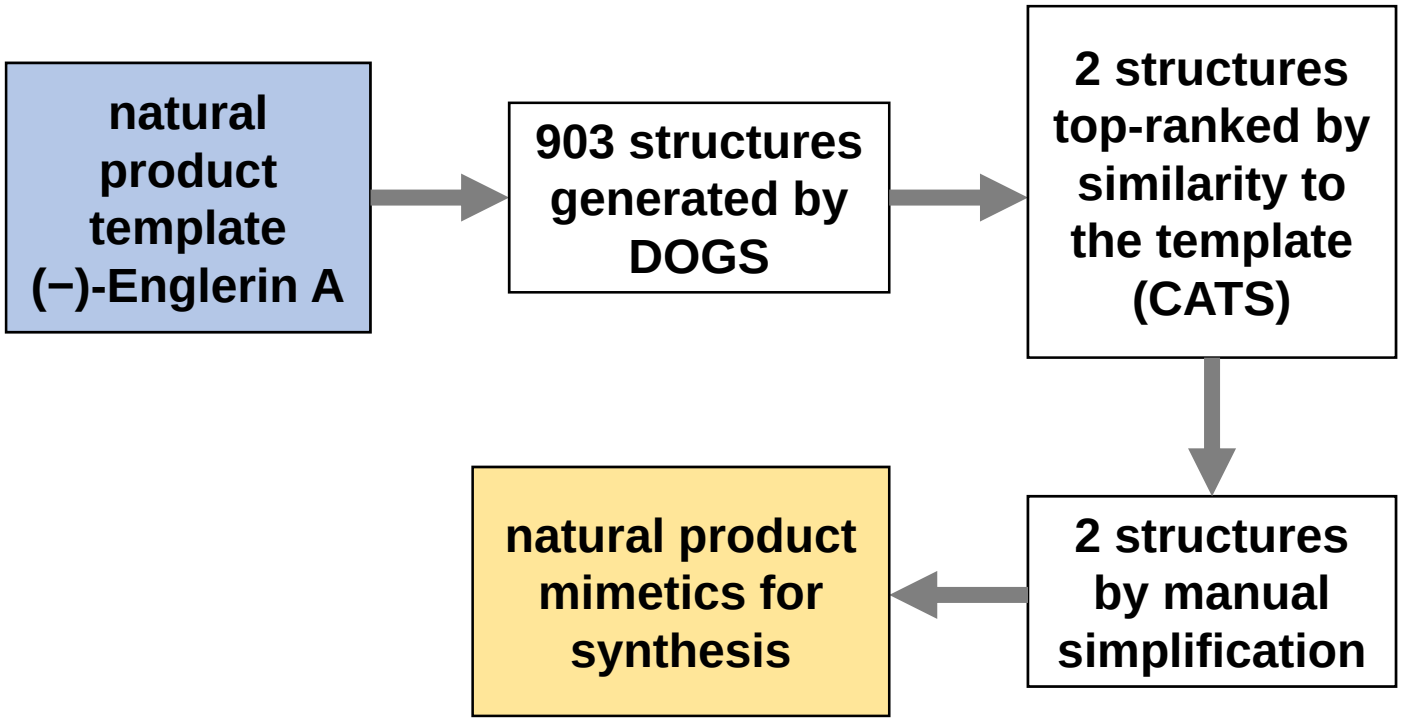
- by Beutler and co-workers, in 2008 from *Phyllanthus engleri*[1)]
- absolute configuration determined by Christmann's first total synthesis in 2009[2)]

Biological Activity[3)]

- agonist of the TRPC4/C5 Ca^{2+} channels → potent antiproliferative activity
- antagonist of the TRPM8 Ca^{2+} channel
- lethal in rodents

1) Ratnayake, R.; Covell, D.; Ransom, T. T.; Gustafson, K. R.; Beutler, J. A. *Org. Lett.* **2008**, *11*, 57. 2) Willot, M.; Radtke, L.; Könning, D.; Fröhlich, R.; Gessner, V. H.; Strohmann, C.; Christmann, M. *Angew. Chem., Int. Ed.* **2009**, *121*, 9369. 3) (a) Akbulut, Y.; Gaunt, H. J.; Muraki, K.; Ludlow, M. J.; Amer, M. S.; Bruns, A.; Vasudev, N. S.; Radtke, L.; Willot, M.; Hahn, S.; Seitz, T.; Ziegler, S.; Christmann, M.; Beech, D. J.; Waldmann, H. *Angew. Chem., Int. Ed.* **2015**, *54*, 3787. (b) Carson, C.; Raman, P.; Tullai, J.; Xu, L.; Henault, M.; Thomas, E.; Yeola, S.; Lao, J.; McPate, M.; Verkuyl, J. M.; Marsh, G.; Sarber, J.; Amaral, A.; Bailery, S.; Lubicka, D.; Pham, H.; Miranda, N.; Ding, J.; Tang, H.-M.; Ju, H.; Tranter, P.; Ji, N.; Krastel, P.; Jain, R. K.; Schumacher, A. M.; Loureiro, J. J.; George, E.; Berellini, G.; Ross, N. T.; Bushell, S. M.; Erdemli, G.; Solomon, J. M. *PLoS ONE* **2015**, *10*, e0127498.

3-2. Molecular Design Strategy

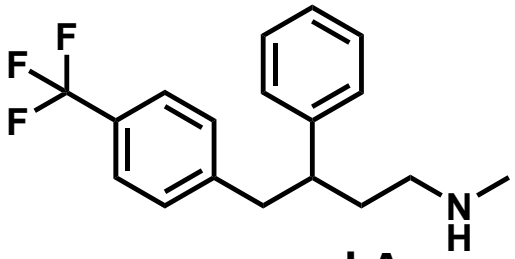


3-3. CATS (Chemically Advanced Template Search)

What is CATS?

- another method to compute topological pharmacophore similarity

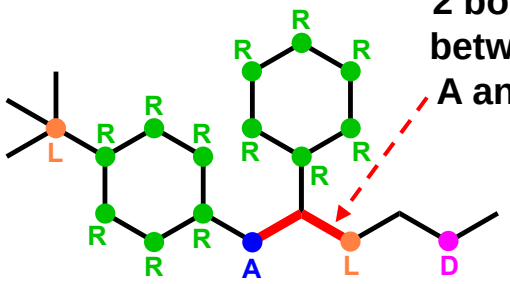
Calculation of CATS Distance



compound A

reducing to the molecular graph

assigning feature types



2 bonds between A and L

d / bonds	*N* / λ		
	D+R	A+L	...
1	0	0	...
2	0	0.0833	...
3	0	0	...
4	0.0833	0	...
5	0.250	0.0833	...
6	0.333	0	...
7	0.250	0	...
8	0.0833	0	...
9	0	0	...
10	0	0	...

scaling by λ

d / bonds	*N* = number of atom pairs		
	D+R	A+L	...
1	0	0	...
2	0	1	...
3	0	0	...
4	1	0	...
5	3	1	...
6	4	0	...
7	3	0	...
8	1	0	...
9	0	0	...
10	0	0	...
λ (sum of the column)	12	2	...

counting atom pairs

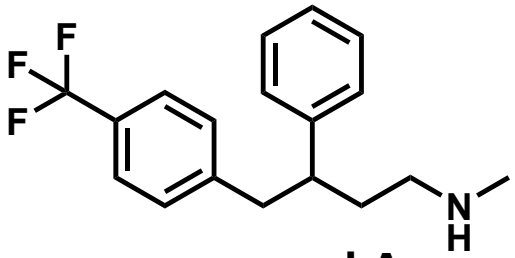
(a) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894. (b) Reutlinger, M.; Koch, C. P.; Reker, D.; Todoroff, N.; Schneider, P.; Rodrigues, T.; Schneider, G. *Mol. Inf.* **2013**, *32*, 133.

3-4. CATS (Chemically Advanced Template Search)

What is CATS?

- another method to compute topological pharmacophore similarity

Calculation of CATS Distance



compound A

d / bonds	N / λ		
	D+R	A+L	...
1	0	0	...
2	0	0.0833	...
3	0	0	...
4	0.0833	0	...
5	0.250	0.0833	...
6	0.333	0	...
7	0.250	0	...
8	0.0833	0	...
9	0	0	...
10	0	0	...

$$\overrightarrow{v^{A}} = \begin{matrix} 0 \\ 0 \\ 0 \\ 0.0833 \\ 0.250 \\ 0.333 \\ 0.250 \\ 0.0833 \\ 0 \\ 0 \end{matrix} \quad \begin{matrix} 0 \\ 0.0833 \\ 0 \\ 0 \\ 0.0833 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{matrix} \quad \begin{matrix} \vdots \\ \vdots \end{matrix}$$

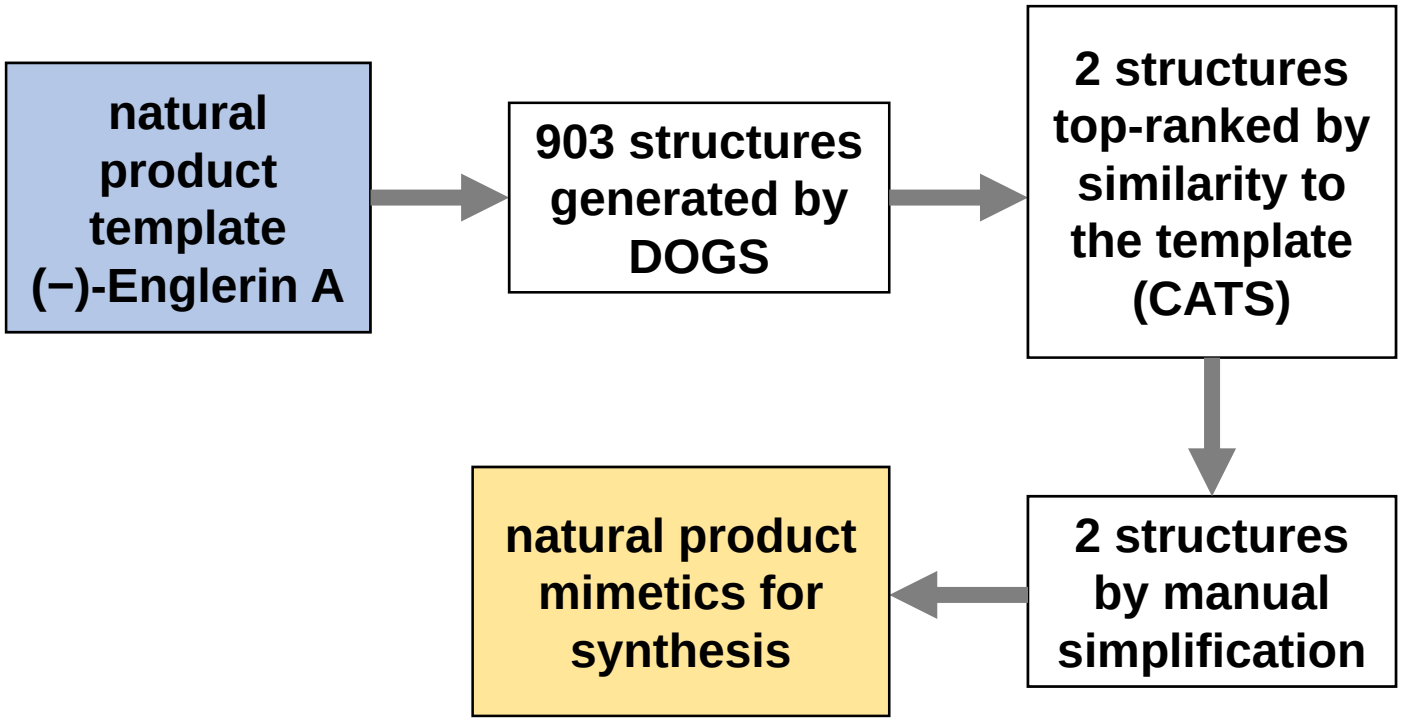
CATS distance (A, B)

$$= \left| \overrightarrow{v^{A}} - \overrightarrow{v^{B}} \right|$$

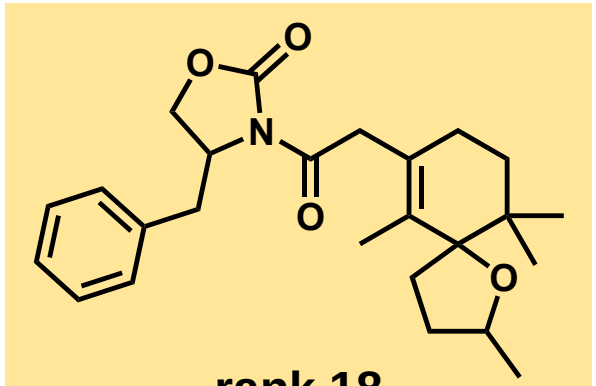
$$= \sqrt{\sum_{i=1} \left(v_i^{A} - v_i^{B} \right)^2}$$

(a) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. *Angew. Chem., Int. Ed.* **1999**, *38*, 2894. (b) Reutlinger, M.; Koch, C. P.; Reker, D.; Todoroff, N.; Schneider, P.; Rodrigues, T.; Schneider, G. *Mol. Inf.* **2013**, *32*, 133.

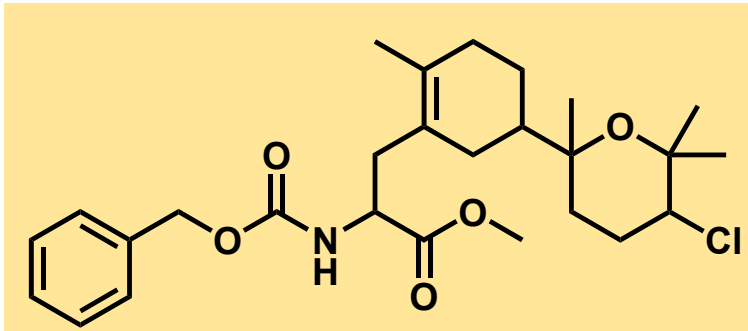
3-5. Molecular Design Strategy



3-6. Top-Scoring Compounds

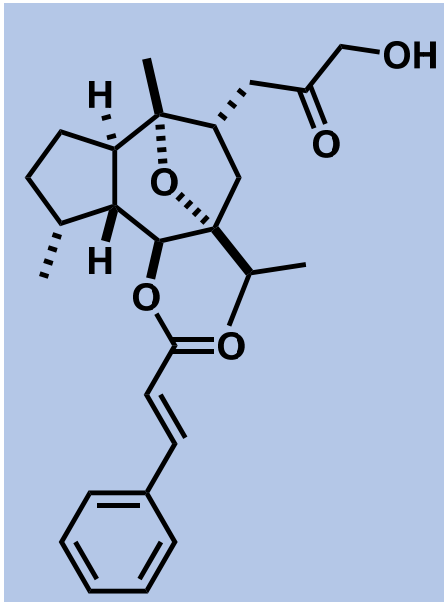


rank 18
(CATS distance = 1.490)



rank 4
(CATS distance = 1.377)

CATS

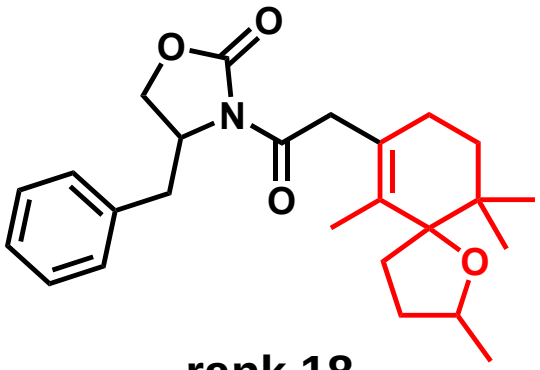


(−)-Englerin A

DOGS

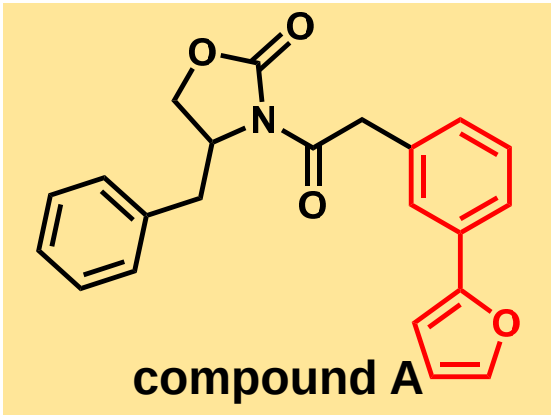
903 structures

3-7. Manual Simplification

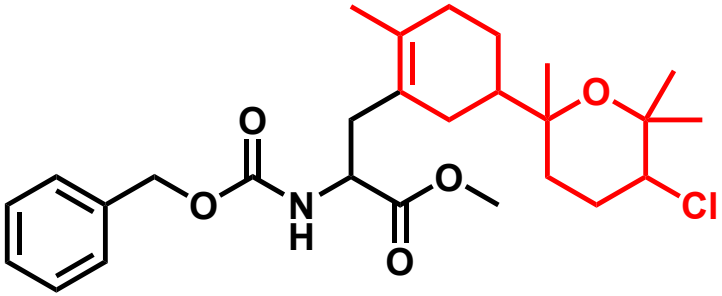


rank 18
(CATS distance = 1.490)

manual simplification

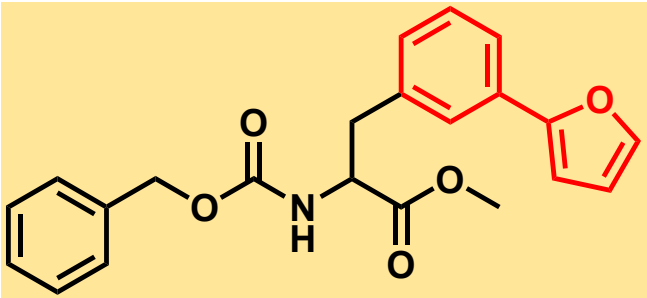


compound A



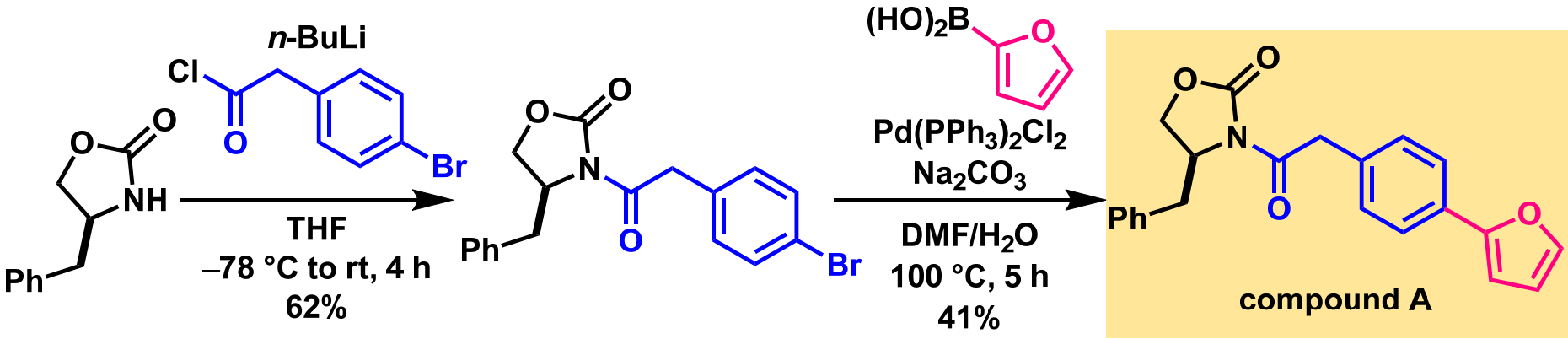
rank 4
(CATS distance = 1.377)

manual simplification

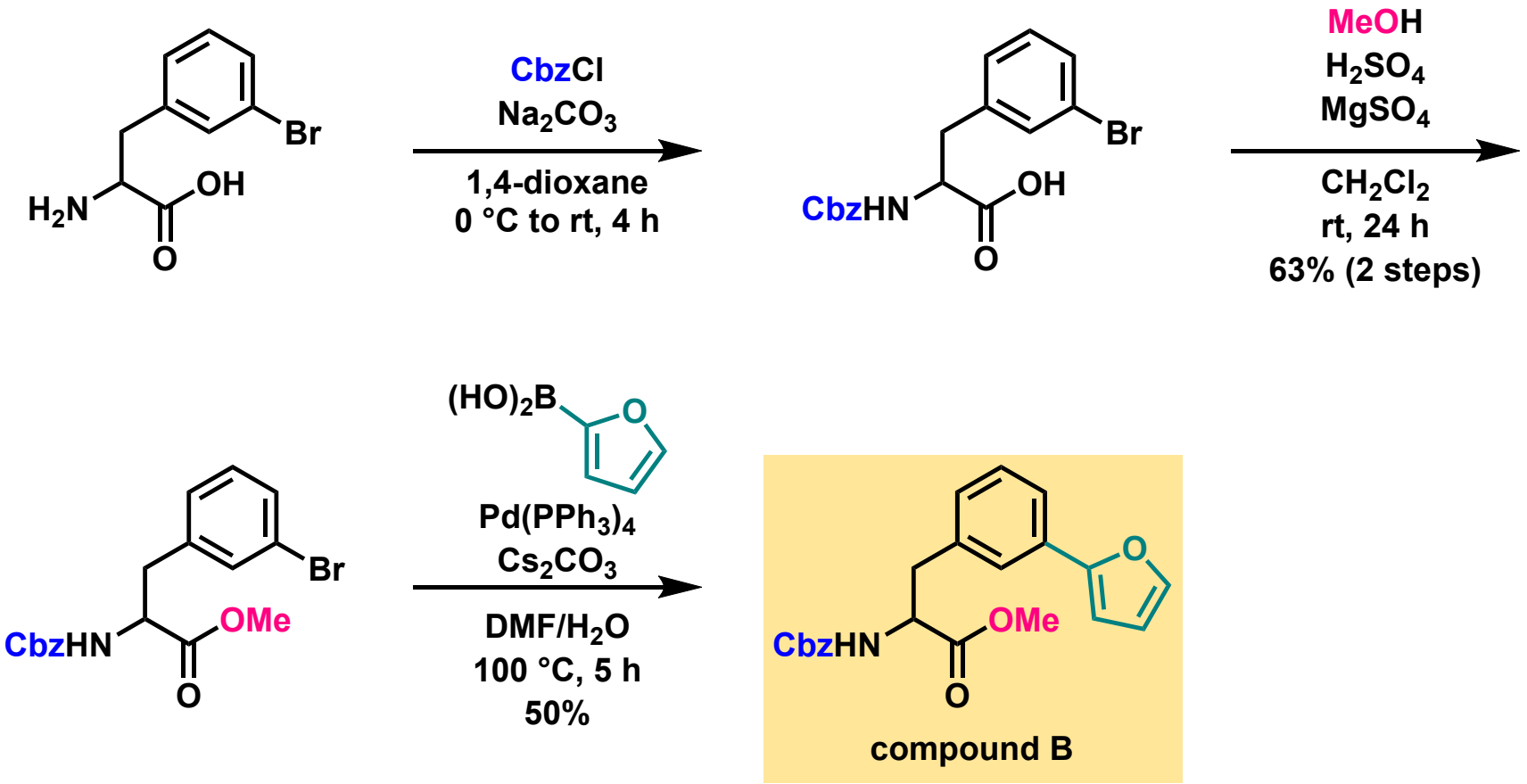


compound B

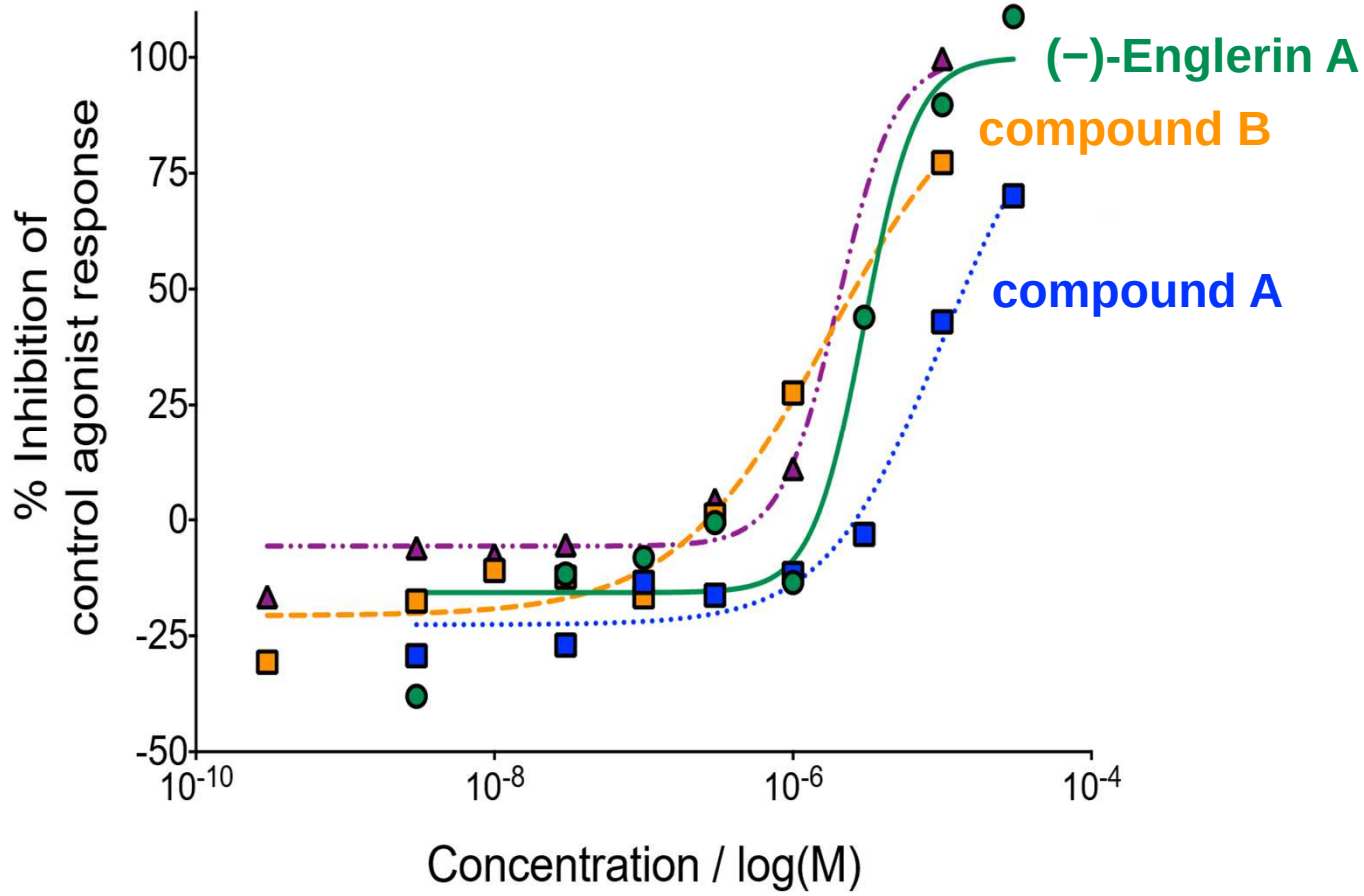
3-8. Synthesis of Compound A



3-9. Synthesis of Compound B

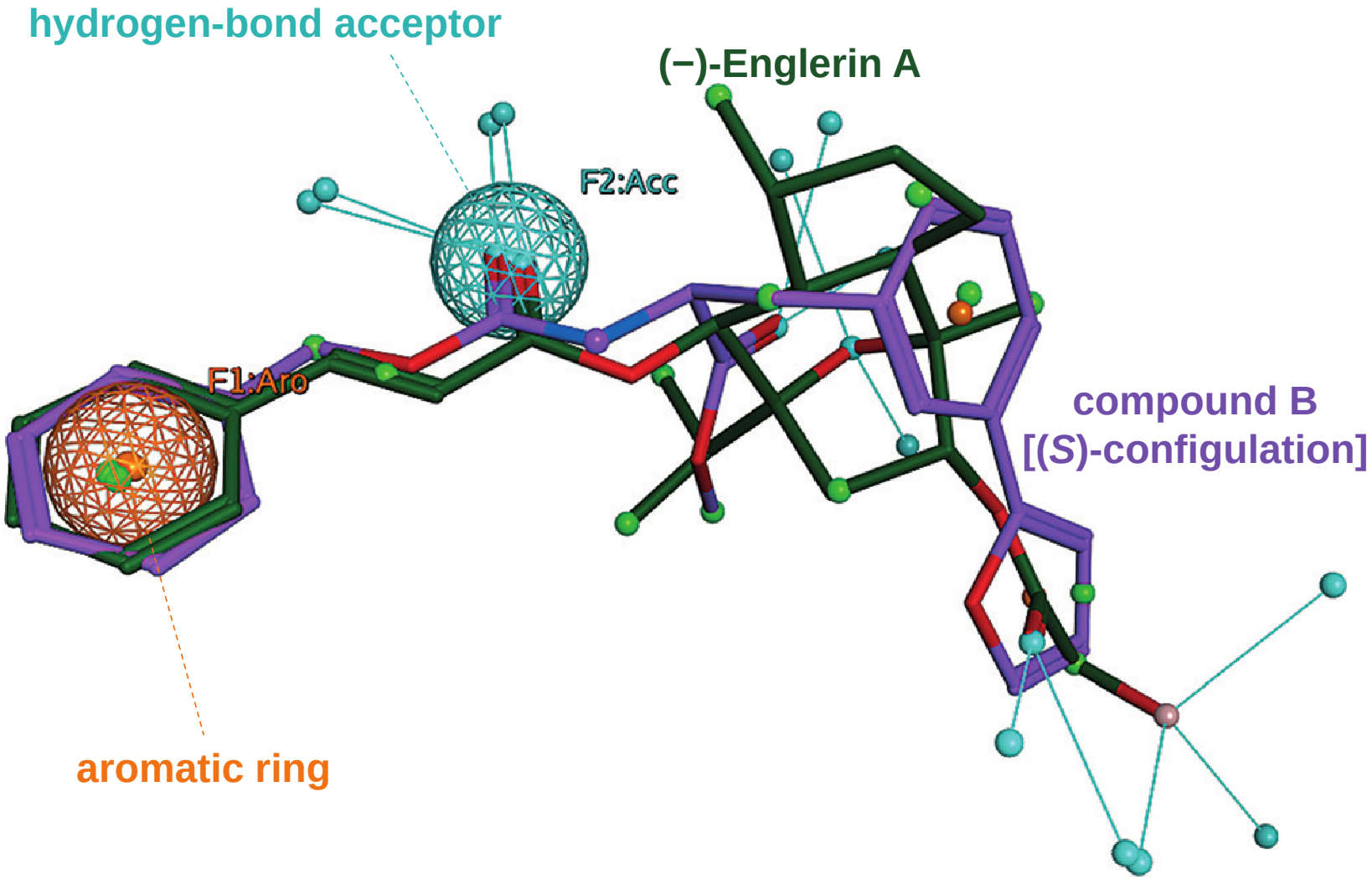


3-10. Competition Binding Experiment for TRPM8



compound	IC_{50} / μM	K_B / μM
(−)-Englerin A	3.0 ± 0.08 log units	0.4
compound A	10 ± 0.08 log units	1.4
compound B	1.7 ± 0.12 log units	0.2

3-11. Comparison of Conformations



Summary and Future Perspective

Summary

1. **DOGS is a novel method for ligand-based *de novo* drug design which focuses on synthesizability and molecular properties.**

2. **In reference to a complex natural product (−)-Englerin A, simple and synthetically accessible molecules were designed *in silico*. The obtained compounds showed comparable binding affinities to the target protein.**

Future Perspective

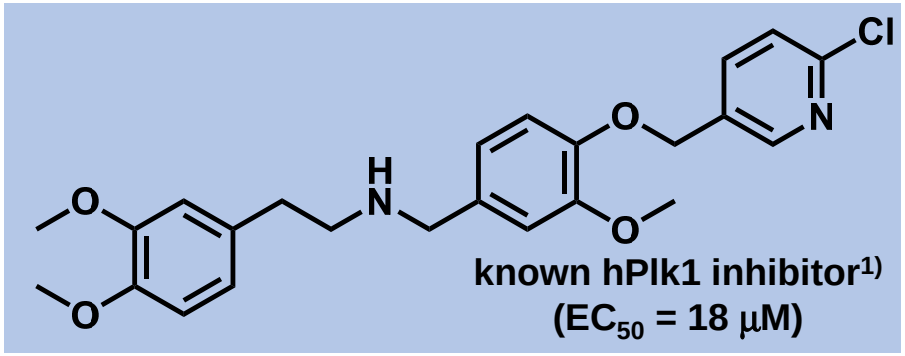
1. **The hit rate and structural diversity of designed molecules might be increased by referencing more than one known ligands.**

2. **If 3D structure of molecule is taken account of, the hit rate might be increase.**

3. **Similar techniques could be applied to antimicrobial peptides[1)].**

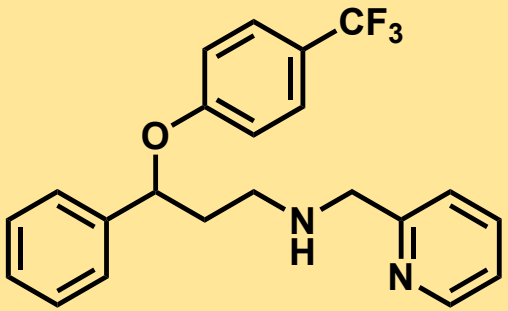
1) (a) Giguère, S.; Laviolette, F.; Marchand, M.; Tremblay, D.; Moineau, S.; Liang, X.; Biron, É.; Corbeil, J. *PLOS Comput. Biol.* **2015**, *11*, e1004074. (b) Bonaventura, I. D.; Jin, X.; Visine, R.; Probst, D.; Javor, S.; Gan, B.-H.; Michaud, G.; Natalello, A.; Doglia, S. M.; Köhler, T.; van Delden, C.; Stocker, A.; Darbre, T.; Reymond, J.-L. *Chem. Sci.* doi: 10.1039/C7SC01314K. (c) Schneider, P.; Müller, A. T.; Gabernet, G.; Button, A. L.; Posselt, G.; Wessler, S.; Hiss, J. A.; Schneider, G. *Mol. Inf.* doi: 10.1002/minf.201600011.

A-1. Prior Research Exploiting DOGS



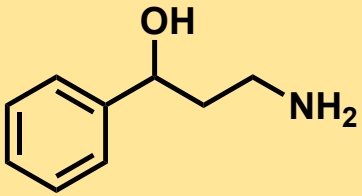
known hPlk1 inhibitor[1)]
(EC_{50} = 18 μM)

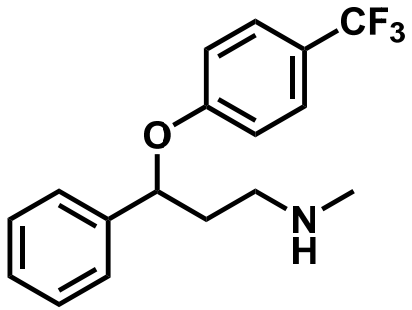
DOGS



new hPlk1 inhibitor[2)]
(EC_{50} = 4 μM)

2 steps





fluoxetine (Prozac)
(EC_{50} = 14 μM)

1) Keppner, S.; Proschak, E.; Schneider, G.; Spänkuch, B. *ChemMedChem* **2009**, *4*, 1806. 2) Spänkuch, B.; Keppner, S.; Lange, L. Rodrigues, T.; Zettl, H.; Koch, C. P.; Reutlinger, M.; Hartenfeller, M.; Schneider, P.; Schneider, G.; *Angew. Chem. Int. Ed.* **2013**, *52*, 4676.

A-2. Predicted Targets of Compounds by SPiDER

compound A

	Target	*Score*	*p Value*
1	Aryl Hydrocarbon Receptor	0.998876	0.001
2	Parathyroid Hormone Receptor	0.998812	0.001
3	Vasopressin Receptor	0.996618	0.002
4	Metabotropic Glutamate Receptor	0.995641	0.003
5	Nicotinic Acetylcholine Receptor	0.995511	0.003
6	Pregnane X Receptor	0.995352	0.003
7	Endopeptidase (Cysteine Endopeptidase, Cysteine Protease)	0.995076	0.003
8	Sodium:Neurotransmitter Symporter (SNF)	0.994975	0.003
9	**Transient Receptor Potential Ion Channel TRP**	**0.994497**	**0.003**
10	Cannabinoid Receptor	0.993807	0.004

compound B

	Target	*Score*	*p Value*
1	Sodium:Neurotransmitter Symporter (SNF)	0.993439	0.004
2	Monoamine Oxidase	0.993232	0.004
3	Sodium Channel	0.992043	0.005
4	Dihydroorotate Dehydrogenase (Dihydroorotate Oxidase)	0.990197	0.007
5	Calcium Channel	0.988672	0.008
6	Endopeptidase (Aspartic Endopeptidase, Aspartic Protease)	0.988663	0.008
7	Endopeptidase (Cysteine Endopeptidase, Cysteine Protease)	0.987973	0.008
8	Cyclooxygenase (Prostaglandin-Endoperoxide Synthase)	0.986773	0.009
9	DNA Polymerase	0.986216	0.010
10	Tumor Necrosis Factor-Alpha	0.985507	0.010
19	**Transient Receptor Potential Ion Channel TRP**	**0.974855**	**0.018**

A-X. Total Synthesis of (−)-Englerin A

A-XX. A Prior Research for (−)-Englerin A mimics

Acerson, M. J.; Bingham, B. S.; Allred, C. A.; Andrus, M. B.; *Tetrahedron Lett.* **2015**, *56*, 3277.