2nd International Electronic Conference on Synthetic Organic Chemistry (ECSOC-2), http://www.mdpi.org/ecsoc/, September 1-30, 1998

[e0003]

Predicting the Probable Receptor Targets for a Potential Drugs Based on the Assessment of Their Similarity With Endogenous Ligands

Yulia Borodina, Dmitrii Filimonov, Vladimir Poroikov

Institute of Biomedical Chemistry RAMS, Pogodinskaya Street, 10, Moscow, 119832, Russia;
E-mail: borodina@ibmh.msk.su

- Introduction
- Methods
- Results and Discussion
- Conclusions
- Acknowledgments
- References

Introduction

Todays pharmacology supposes that pharmaceutical agents interact with various physiological receptors which are the targets for neurotransmitters, hormones and other endogenous bioregulators [1]. The same physiological effects can be induced by activation/inhibition of different receptors. Contrary, interaction with several receptors can decrease the effects, induced by each receptor separately. Therefore, the prediction of probable receptor targets for substance is important for design of agent with desirable pharmacological effect.

We develop computer system SIMEST for multiple similarity assessment of a new compound with high selective small ligands of known receptors. The principal idea is that the similar compounds will interact with the same receptors.

The SIMEST includes:

- the software for similarity estimation between a *pattern molecule* and each of the ligands;
- the database of highly selective small ligands (endogenic bioregulators and their analogs).

Methods

The Similarity Estimation module is arranged as the ISIS/Base application (MDL Information Systems, Inc.). Similarity assessment is based on in-house developed topological descriptors and widely used Tanimoto coefficient.

The Database works under the ISIS/Base and integrates the structure and activity data for both endogenous ligands and highly selective agonists and antagonists for 100 receptor subtypes.

Topological Descriptors

The structure description is based on connection table (C) and table of atoms types (AT). Connection table contains the information about bonds in the molecule. We do not specify various bond types, but take into account all hydrogens congruous to the valencies and charges of atoms.

AT table includes the element types for each atom in a molecule. All chemical elements are classified according to the rules given below.

Class name	Elements
H	H
C	C
N	N
O	O
S	S
F*	F, Cl, Br, I, At
P*	P, As, Sb, Bi
Li*	Li, Na, K, Rb, Cs, Fr
Be*	Be, Zn, Cd, Hg, Si, Ni, Cu, Ge, Ru, Rh, Ag, Sn, Te, Pt, Pb, Po, B, Mg, Al, Ca, Sc, Ti, V, Cr, Mn, Fe, Co, Ga, Sr, Y, Zr, Nb, Mo, Tc, Pd, In, Ba, La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Hf, Ta, W, Re, Os, Au, Tl, Ra
He*	He, Ne, Ar, Kr, Xe, Rn, Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, Es, Fm, Md, No, Lr, Db, Jl

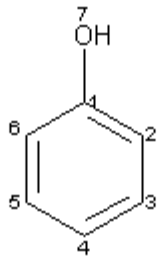
The structure of the molecule is represented as the set of multilevel topological descriptors. The first-level descriptor includes:

- class of considered atom,
- length of smallest cycle containing the atom (for atom in cycle only),
- list of nearest neighbors of the atom in certain lexicographical order.

The descriptor of higher level for an atom is generated iteratively and includes the descriptors of the previous level for an atom and its neighbours. This process can be continued up to any level substituting the list of nearest neighbor atoms by appropriate descriptors.

It is shown that inclusion of descriptors up to the 3rd level provides the satisfactory accuracy of recognition. These descriptors are reffered further as the Sub-Structure Descriptors (SSD).

Example of coding by SSD of 1^{st} and 2^{nd} levels for phenol are shown in Figure below. The descriptors of 3^{rd} level are not presented here because of their large size.



	SSD1	SSD2
1	6C(6C6CO)	6C(6C(H6C6C)6C(H6C6C)O(H6C))
2	6C(H6C6C)	6C(H(6C)6C(H6C6C)6C(6C6CO))
3	6C(H6C6C)	6C(H(6C)6C(H6C6C)6C(H6C6C))
4	6C(H6C6C)	6C(H(6C)6C(H6C6C)6C(H6C6C))
5	6C(H6C6C)	6C(H(6C)6C(H6C6C)6C(H6C6C))
6	6C(H6C6C)	6C(H(6C)6C(H6C6C)6C(6C6CO))
7	O(H6C)	O(H(O)6C(6C6CO))
8	H(6C)	H(6C(H6C6CO))
9	H(6C)	H(6C(H6C6C))
10	H(6C)	H(6C(H6C6C))
11	H(6C)	H(6C(H6C6C))
12	H(6C)	H(6C(H6C6C))
13	H(6C)	H(6C(H6C6C))
14	H(O)	H(O(H6C))

Tanimoto coefficient [2] is used to measure the similarity between two molecules A and B:

$$s(A,B) = \frac{\sum_{i=1}^{M} A(i)B(i)}{\sum_{i=1}^{M} A(i) + \sum_{i=1}^{M} B(i) - \sum_{i=1}^{M} A(i)B(i)}$$

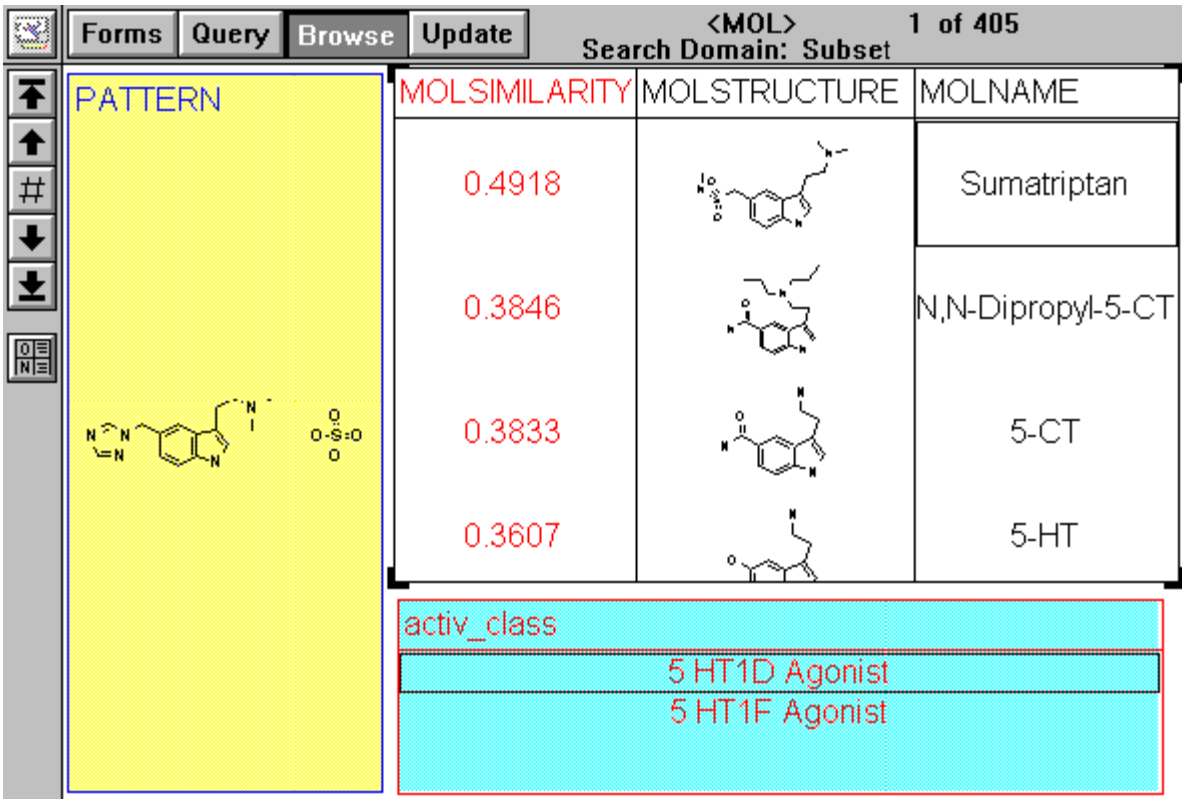
where A(i) and B(i) are equal to 1 when i-th descriptor is found in molecule A and B respectively and 0 when the i-th descriptor is absent; M is the total number of descriptors in the dataset.

How SIMEST Works

With SIMEST one can input the *pattern structure* and find the most similar ligands from the Database. The result is presented as the list of receptor subtypes arranged in descending order of corresponding similarity coefficients.

Therefore, high similarity of the compound with particular endogenous-like ligand, that is Agonist and/or Antagonist of the Receptor, lead one to the conclusion that the compound probably has the same activity.

The figure below demonstrates the result of similarity searching for compound 206830 from MDDR 96.2 database, which have 5-HT4 Agonist, 5-HT3 Antagonist activities.



Results and Discussion

The possibilities of SIMEST to distinguish the active compounds from inactive are evaluated on the 17124 compounds from MDDR 96.2 database (MDL Information Systems, Inc.). The accuracy of active compounds recognition (for each kind of activity) is calculated as [3]:

$$\frac{\frac{1}{N_b}\sum_{i=1}^{N_b} N\{s(i,1) > s(i,0)\}}{N_1 \cdot N_0}$$

where: N{s(i, 1)>s(i, 0)} is the number of cases when the active compound is more similar to i-th active ligand than inactive one, when all pairs of active and inactive compounds are compared;
N_b is the numbers of active ligands in the SIMEST/Database;
N_1 and N_0 are the numbers of active and inactive compounds in the evaluation set.

Using this criterion, the accuracy of active compounds recognition has been calculated for 49 different mechanisms of action. These results are given below.

The Accuracy of Agonists Recognition (estimated for 17124 comps from MDDR 96.2)

Activity	Accuracy, %	N_1	N_0	N_b
Vitamin D Agonist	100	66	17058	2
Histamine (H3) Agonist	100	1	17123	2
Adenosine (A2) Agonist	100	17	17107	1
Melatonin Agonist	99	7	17117	2
Androgen Agonist	99	1	17123	6

Estrogen Agonist	99	21	17103	2
Adenosine (A1) Agonist	98	14	17110	4
Adrenergic (beta) Agonist	97	34	17090	2
5 HT4 Agonist	95	28	17096	1
Retinoid Acid Agonist	95	49	17075	2
Dopamine (D1) Agonist	95	16	17108	1
Adrenergic (beta1) Agonist	94	3	17121	1
5 HT1D Agonist	93	55	17069	4
5 HT1 Agonist	91	22	17102	1
5 HT1B Agonist	91	3	17121	2
Dopamine (D2) Agonist	88	31	17093	3
GABA B Agonist	86	4	17120	2
Adrenergic (alpha2) Agonist	79	9	17115	7
5 HT1A Agonist	70	156	16968	5
Dopamine Agonist	65	27	17097	3
5 HT3 Agonist	59	6	17118	3
Thyroid Hormone Agonist	58	2	17122	2
Muscarinic M1 Agonist	49	138	16986	2
In Average:	87			

The Accuracy of Antagonists Recognition (estimated for 17124 comps from MDDR 96.2)

Activity	Accuracy, %	N_1	N_0	N_b
P2T Antagonist	100	2	17122	1
Adrenergic (beta1) Antagonist	99	10	17114	1
Histamine (H3) Antagonist	99	13	17111	2
Adrenergic (beta) Antagonist	96	37	17087	3
Adrenergic (alpha1) Antagonist	88	68	17056	2
5 HT2 Antagonist	86	105	17019	1
5 HT Antagonist	85	22	17102	1
Dopamine (D4) Antagonist	84	25	17099	3
5 HT3 Antagonist	82	204	16920	5
Adenosine (A1) Antagonist	80	28	17096	2
AMPA Antagonist	78	40	17084	4
5 HT1A Antagonist	78	48	17076	5
Adrenergic (alpha2) Antagonist	76	49	17075	4
Dopamine (D2) Antagonist	73	89	17035	3
Muscarinic Antagonist	70	21	17103	3
5 HT2B Antagonist	69	5	17119	1
5 HT2A Antagonist	69	3	17121	5
5 HT2C Antagonist	68	16	17108	4
Dopamine (D1) Antagonist	68	33	17091	1
GABA B Antagonist	65	6	17118	3

Muscarinic M3 Antagonist	62	3	17121	3
Adrenergic (alpha) Antagonist	59	17	17107	1
NMDA Antagonist	59	234	16890	3
Muscarinic M2 Antagonist	47	3	17121	1
Adenosine (A2a) Antagonist	22	1	17123	1
In Average:	73			

The average accuracy of recognition is significantly higher for agonists than for antagonists. The result is probably explained by the fact that antagonists have the less specific structures than the agonists.

Conclusions

- New computer system SIMEST for assessing similarity of chemical substances with endogenous-like ligands is developed.
- It is shown in experiments with MDDR database that the recognition of specific activity averages about 87% for agonists and 73% for antagonists.

Acknowledgments

We are gratefully acknowledge the MDL Information Systems, Inc. that kindly provided us with ISIS/Base, ISIS/Host and MDDR database for carrying out this work.

This work is supported in part by the Russian Ministry of Science and Technical Politics (Scientific Program "R & D of New Drugs by the Methods of Chemical and Biological Synthesis", Branch 04 "Computer Aided Drug Design", Project No 04.01.13/96).

References

---

1. Gilman A., Rall T., Nies A., Eds., The Pharmacological Basis of Therapeutics, Pergamon Press, New York and others, 1990.

2. Dean, P. M. Ed., Molecular Similarity in Drug Design, Blackie Academic & Professional, 1995, pp 111 - 137.

3. Borodina Yu., Filimonov D., and Poroikov V. Computer-Aided Estimation of Synthetic Compounds Similarity with ndogenous Bioregulators, Quantitative Structure-Activity Relationships, 1998, In press.

Comments

During 1-30 September 1998, all comments on this poster should be sent by e-mail to ecsoc@listserv.arizona.edu with e0003 as the message subject of your e-mail. After the conference, please send all the comments and reprints requests to the author.
