JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

De Novo Design of Drug-Like Molecules by a Fragment-Based Molecular Evolutionary Approach

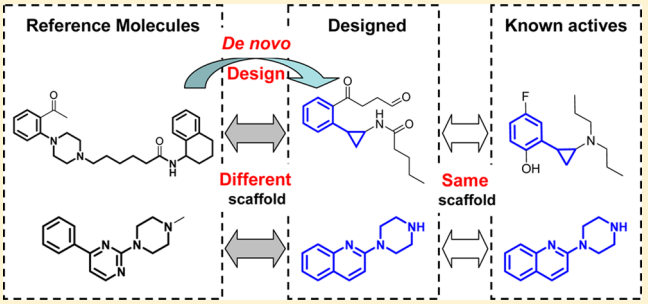
Kentaro Kawai,*[,†] Naoya Nagata,[†] and Yoshimasa Takahashi[‡]

[†]Central Research Laboratories, Kaken Pharmaceutical Co. Ltd., 14, Shinomiya Minamikawara-cho, Yamashina, Kyoto 607-8042, Japan

[‡]Department of Computer Science and Engineering, Toyohashi University of Technology, 1-1, Hibarigaoka, Tempaku-cho, Toyohashi 441-8580, Japan

Supporting Information

ABSTRACT: This paper describes a similarity-driven simple evolutionary approach to producing candidate molecules of new drugs. The aim of the method is to explore the candidates that are structurally similar to the reference molecule and yet somewhat different in not only peripheral chains but also their scaffolds. The method employs a known active molecule of our interest as a reference molecule which is used to navigate a huge chemical space. The reference molecule is also used to obtain seed fragments. An initial set of individual structures is prepared with the seed fragments and additional fragments using several connection rules. The fragment library is preferably prepared from a collection of known molecules related to the target of the reference molecule. Every fragment of the library can be used for fragment-based mutation. All the fragments are categorized into three classes; rings, linkers, and side chains. New individuals are produced by the crossover and the fragment-based mutation with the fragment library. Computer experiments with our own fragment library prepared from GPCR SARfari verified the feasibility of our approach to drug discovery.



INTRODUCTION

In the drug discovery process, various characteristics, including biological activities and ADMET properties, must be considered simultaneously. Medicinal chemists synthesize congeneric series of compounds to clarify the structure−activity relationship (SAR) of their own hits or lead series, and they use the SAR knowledge to optimize the compounds in further synthesis. In addition, drug discovery entails complicated tasks of optimization with respect to ADMET properties. Medicinal chemists are often required to change scaffolds further by so-called core hopping to address scaffold-dependent issues.[1−3] One of the good examples is GSK's B-Raf inhibitor program.[4,5] They changed the molecular frameworks (scaffolds) during both of lead generation and lead optimization stages to discover a compound for clinical trial.

Together with medicinal chemistry, computational chemistry plays an important part in the discovery of new drugs. Computational molecular design has been an active research area over the past decades. Many computerized structural design approaches have been developed, which utilize protein-structures and/or ligand-structures.[6−27] For example, LEGEND,[6] LUDI,[7] SPROUT,[8] LEA3D,[9] LigBuilder,[10,11] and SYNOPSIS[12] use protein structures; whereas TOPAS,[13] CoG,[14] and Flux[15,16] use the structures of known ligands. The former methods are referred to as the structure-base design and the latter methods to as the ligand-based design, respectively. The general advantage of ligand-based approaches is their wide range of applicability, because those approaches can be used in the case of that the three-dimensional (3D) structure of the target is not available.

Evolutionary algorithms are actively used for the computerized molecular design. They are based on concepts derived from biological evolution, including reproduction, mutation, crossover, and selection. The algorithms are widely used to solve various drug discovery problems, such as parameter optimization of QSAR/QSPR models[28] and 3D-ligand alignment[29] as well as the compound design described above. New molecules are designed by repeatedly applying the evolutionary operations to existing molecules. Among these operations, mutation and crossover are vital for generating new chemical structures. Mutation methods are roughly classified into the fragment-based mutation and the atom-based mutation. In our previous study, the atom-based method was used for the mutation, in which an atom is modified into another atom to explore the chemical space. The method often resulted in a lot of unfavorable structures that contained invalid hetero−hetero atom bonds such as O−O and N−F.[20] One of the approaches to avoiding this problem is to use exclusion rules of the substructures reported by Huang et al.[18] An alternative method

Received: July 17, 2013
Published: December 28, 2013

is to use a fragment-based mutation instead of the atom-based mutation. We also considered that the use of fragments is a good way to generate chemically feasible structures when those fragments are derived from known molecules.

Evolutionary algorithms use fitness score to select the surviving structures. For example, Molecule Evoluator[17] uses medicinal chemists' knowledge and Flux uses a similarity index (Tanimoto coefficient or Euclidean distance) as the fitness function. In addition, building structures with chemical feasibility is an important point for de novo design. Fragment-based approaches were favorably used for the purpose. For example, NovoFLAP[19] uses fragments with thirty-two chemical transformation operators to generate structures obeying valence rules. Flux builds molecules by a RECAP-based rule (11 reaction schemes) to connecting fragments. These approaches are reasonable to generate feasible structures. However, we considered that a more simplified approach with fewer connection rules has an advantage in use.

In this paper, we propose a similarity-driven fragment-based evolutionary approach to producing drug-like molecules for drug discovery. Aim of the method is to explore the candidates that are similar to a reference molecule and yet somewhat different in not only the side chains but also their scaffolds. Chemical feasibility of the candidates is also considered.

METHODS

Outline of Evolutionary Algorithm. The basic idea of the present method is a similarity-driven simple evolutionary approach (Figure 1). The method employs an existing active molecule of our interest. It is used as a reference molecule to navigate a chemical space to be explored. The reference molecule is also used to obtain seed fragments for making the initial population. The initial set of individual structures is prepared using the seed fragments and additional fragments from a fragment library. The procedure of the present evolutionary approach is summarized as follows:

(1) Input a reference structure.
(2) Generate seed fragments.
(3) Make an initial set of individual structures.
(4) Generate offsprings for the next generation by mutation and crossover.
(5) Evaluate fitness of the structures and select some of them to survive.
(6) Steps 4 and 5 are repeated until that the alternation of generation reaches to the specified number.

In the following, we will use the term "fragment" to mean a building block. In our approach, a molecule can be built by connecting the fragments at their connection points specified in advance. The connection points of a fragment are identified from the original molecule as shown in Figure 2.

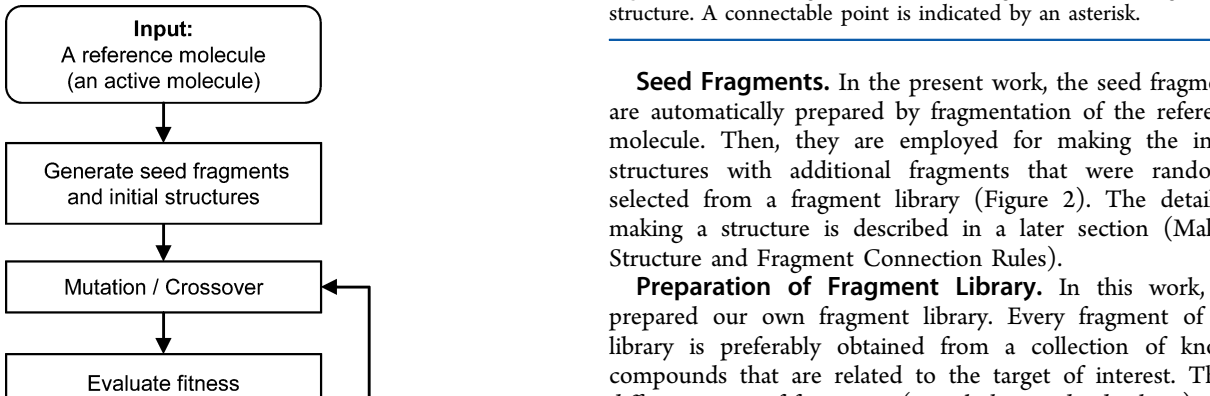


Figure 1. General scheme of the molecular evolutionary algorithm.

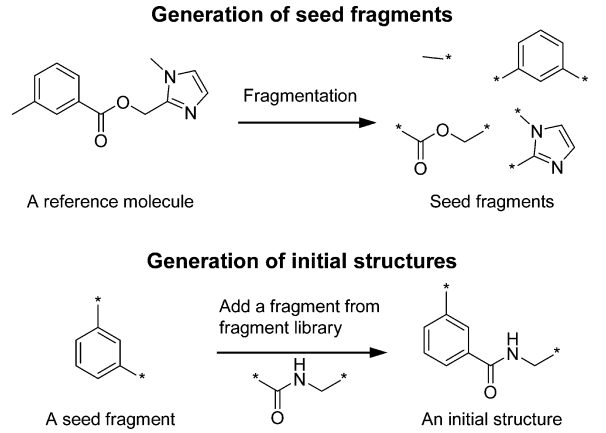


Figure 2. Example of generation of seed fragments and making a initial structure. A connectable point is indicated by an asterisk.

Seed Fragments. In the present work, the seed fragments are automatically prepared by fragmentation of the reference molecule. Then, they are employed for making the initial structures with additional fragments that were randomly selected from a fragment library (Figure 2). The detail of making a structure is described in a later section (Making Structure and Fragment Connection Rules).

Preparation of Fragment Library. In this work, we prepared our own fragment library. Every fragment of the library is preferably obtained from a collection of known compounds that are related to the target of interest. Three different types of fragments (ring, linker, and side-chain) were defined as building blocks and used to make a molecule.

Walters et al. reported that 70% of the compounds in the *Journal of Medicinal Chemistry* are composed of a relatively small number of building blocks (e.g., only 37 rings, 53 linkers, and 16 side-chains).[30] Thus, we thought that a wide variety of new molecular frameworks could be designed by the combination of the building blocks such kind.

For making a molecule from the fragments, we explicitly treated their possible connection points by labeling with dummy atoms. For example, ortho-substituted benzene and para-substituted benzene are treated as different fragments (see ring fragments in Figure 3). Every type of monocyclic rings, fused rings, or spiro rings is defined as ring fragments. It should be noted that biphenyl is not a ring fragment but a combination

dx.doi.org/10.1021/ci400418c | J. Chem. Inf. Model. 2014, 54, 49−56

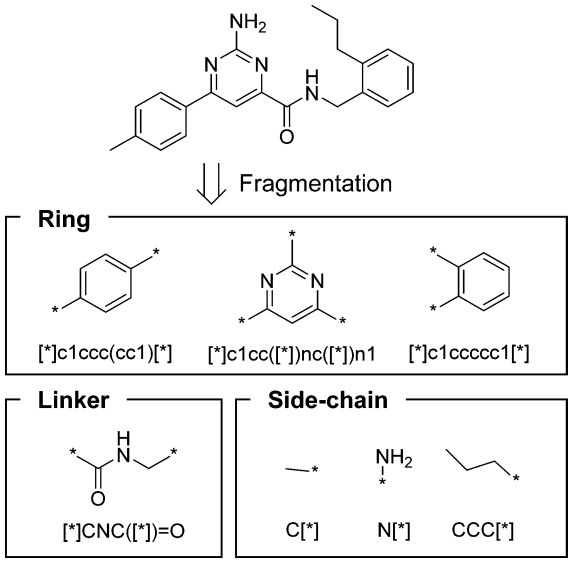


Figure 3. Preparation of fragment library. An example of fragmentation to obtain the fragments from a chemical structure. A connection point is described with a dummy atom (labeled by an asterisk).

of two ring fragments because two phenyl rings do not share any atom. A chain or an atom between rings is defined as a linker fragment, which has two or more attachment points. Every acyclic fragment with a single connection point is defined as a side-chain fragment.

In this work, Bioactive molecules in GPCR SARfari 2.0[31] of ChEMBL[32] (141 990 molecules, 947 914 assays) were used to prepare the fragment library. First, compounds with biological activities indicated by IC_{50}, K_i, K_d, log IC_{50}, log K_i, log K_d, pIC_{50}, pK_i, or pK_d values were retrieved. The subsequent removal of large molecules (heavy atoms ≥50), duplicated molecules, and metal-containing molecules such as ferrocenes yielded 97 084 unique molecules with 313 980 assays. Then, fragments were extracted from the molecules and duplicated fragments were removed by comparing canonical SMILES generated by OpenBabel.[33] Again, phenyl rings with one, two, or three connection points in different configurations are treated as different fragments, respectively. Frequent fragments with molecular weights ≤300 were extracted, and finally, 1527 rings, 605 linkers, and 471 side-chains were prepared as a fragment library. Even in a simple estimation, the number of possible structures that consist of one ring, one linker, and one side-chain exceeded 10^8. When discriminating with respect to connection points and using multiple rings, multiple linkers, and multiple side-chains, a very large number of structures can be generated by using the present fragment library.

Making Structure and Fragment Connection Rules. Individual structures of each population are produced from the fragments according to several connection rules. We defined three allowable connections. They are ring−ring connection, ring−linker connection, and ring−side chain connection. On the other hand, three unallowable connections were defined. They are linker−linker connection, side chain−side chain connection, and linker−side chain connection. Figure 4 illustrates the connection rules employed in this work.

For making a molecule, two fragments can be connected according to the rules only if they have the same bond multiplicity at both of the connection points. If there are other connection points that are free for the connection, then implicit hydrogen atoms are attached to all the connection points to complete the structure.

Mutation. Mutation is a key operation for generating individuals for the next generation. In mutation, a parent molecule is randomly selected from existing molecules, and one of the following operations is then applied against the parent molecule.

- add a fragment
- remove a fragment
- replace a fragment

In the operation *add a fragment*, a base fragment was randomly chosen from fragments that constituted the parent molecule. Then, a new fragment to be introduced was selected from the fragment library. If the base fragment type is ring, then a new ring, linker, or side-chain fragment is selected from the library and linked with a single or a double bond that obeying valence at the attachment point (Figure 5a). If the base fragment is a linker or a side-chain, then a ring fragment was selected and added to the base fragment. In the case of *remove a fragment*, a terminal fragment was randomly removed from the parent molecule (Figure 5b). In *replace a fragment*, a base fragment was randomly selected and then replaced with a new fragment selected from the fragment library. To achieve this operation, the fragment types and bond orders at the attachment point must be consistent (Figure 5c).

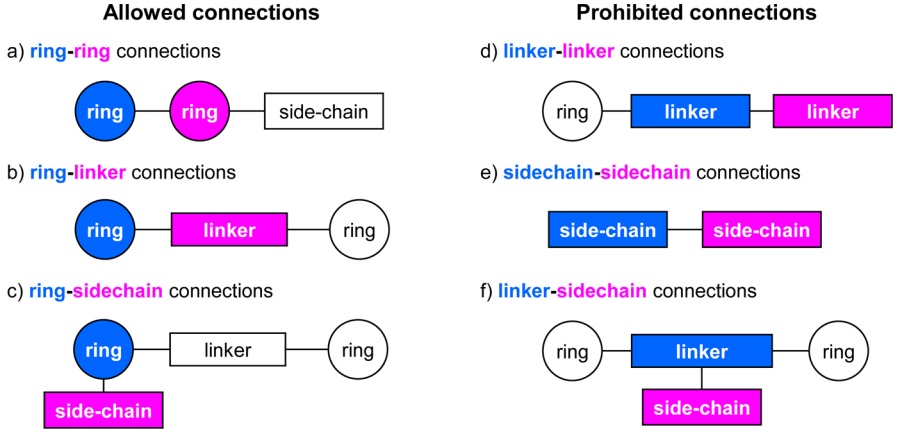


Figure 4. Fragment connection rules for generating molecules with chemical validity. Allowd connections (a−c) and prohibited connections (d−f).

dx.doi.org/10.1021/ci400418c | J. Chem. Inf. Model. 2014, 54, 49−56

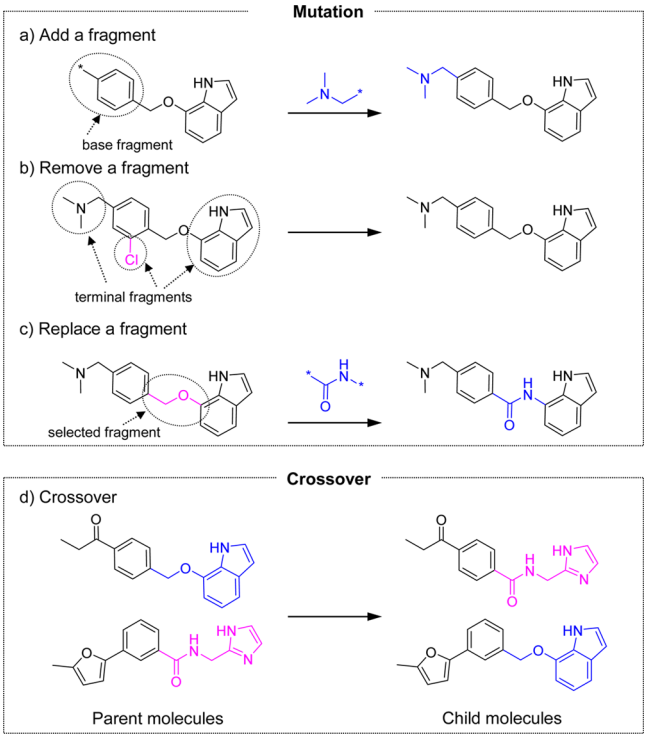


Figure 5. Examples of molecular mutations (a–c) and a crossover operation (d). Connection position is indicated by an asterisk.

Crossover. The crossover operation generates new two child molecules by exchanging a fragment set derived from a parent molecule with the other fragment sets derived from another parent molecule. Two parent molecules were randomly selected, and a crossover point was then randomly selected from each parent molecule to define a set of fragments to be crossed over. To achieve crossover, both the fragment type and bond order at the two crossover points must be the same (Figure 5d).

Fitness Evaluation. Our aim of the current work is to explore the candidate structures that are similar to a reference molecule and yet somewhat different in the scaffolds. For that reason, the Tanimoto coefficient was used as a fitness function to evaluate the molecular similarity to the reference molecule. Surviving compounds were then selected in accordance with the fitness scores (details are to be described in the Selection subsection). These evolutionary processes were repeated for the number of generations specified in advance to yield the designed structures as an output.

In this study, a particular descriptor (referred to as topological-fragment-descriptor; TFD) was employed to profile chemical structures. The TFD was calculated in a manner similar to that of topological-fragment-spectra method.[34,35] For the first, we enumerated all possible structural fragments that have the specified number (six atoms in this work) or less connected atoms excepting hydrogen atoms. Each fragment was characterized by its constituent atoms based on atomic type, hybridization, and whether the atom contained at least one (aromatic or aliphatic) ring. Then, each characterized fragment was hashed into a single integer. The occurrence of individual fragments with the same characteristic value was then counted to generate a numerical vector. Every chemical structure was described as a multidimensional numerical pattern vector by means of the TFD method.

Selection. Population of the next generation (new parent molecules) was selected from both the parent and the offspring. A tournament method was used to determine the surviving individuals. First, two individuals were exclusively chosen from the population of parent and offspring in a random manner. Then, the individual with the higher fitness value was selected, and another individual was discarded. This process was applied to all of the molecules, and finally, half of the population was selected as surviving individuals.

Molecular Evolution Experiment. Computer experiments for exploring GPCR ligands were carried out to verify the feasibility of the present approach. The fragment library for mutation was prepared using the fragments obtained from GPCR SARfari. We carried out molecular evolution experiments for different targets (hAA2A and r5HT1A) from GPCR-SARfari database. They are among largest classes in terms of the number of the compound entries. The reference structures for each target were randomly chosen from its own class of compounds in the database. The number of structures to be generated in each generation was set to 100. This means that in every generation, 100 offspring molecules were generated from 100 parent molecules. Among the 200 molecules, 100 molecules were selected as surviving molecules. With respect to the operators, a mutation rate of 0.8 and crossover rate of 0.2 were used for the evolution parameters. This means that each individual member of the parent population produces an offspring with a mutation rate of 80% and a crossover rate of 20%. The number of generations to be evolved was set to 500, and the individuals that survived in the final generation were referred as the "designed molecules". The values of the evolutionary parameters were determined based on our preliminary study. The result showed that higher fitness was obtained when mutation rate was set between 0.7 and 0.8. With respect to the number of generations, the fitness reached its plateau within 500 generations. Every computer experiment was repeated 10 times for each reference structure.

RESULT AND DISCUSSION

Change of the Mean Fitness. First, a computational trial of the molecular evolution was carried out to explore the candidate structures for hAA2A with a reference molecule **1**. To examine the performance of the current approach, the change of the mean fitness was measured.

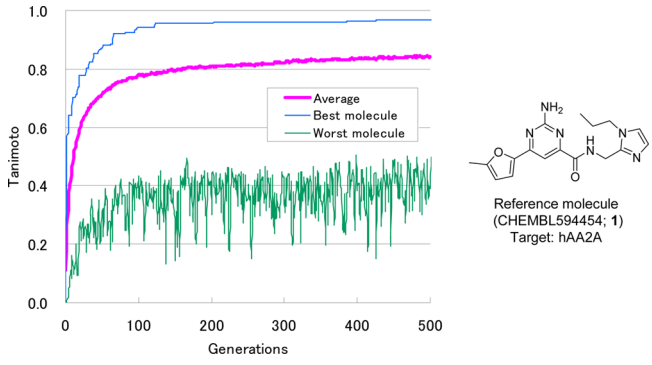


Figure 6. Change of the mean fitness for reference molecule **1**. The red shows the total average of the ten trials. The blue line shows that for the best molecule, and the green for the worst molecule.

The mean fitness of each population was calculated for the 100 surviving molecules. As mentioned above, every experi-

dx.doi.org/10.1021/ci400418c | J. Chem. Inf. Model. 2014, 54, 49–56

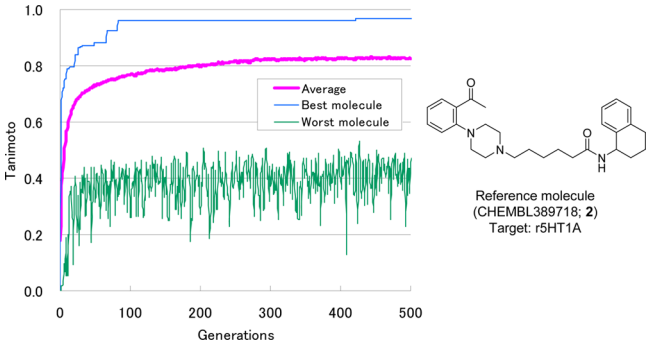


Figure 7. Change of the mean fitness for the reference molecule **2**. The red shows the total average of the ten trials. The blue line shows that for the best molecule, and the green for the worst molecule.

ment was repeated 10 times for each reference structure. Figure 6 shows the total average of them obtained from those trials. The mean fitness favorably increased as the number of generations increased, and the fitness reached its plateau within 500 generations. For example, fitness increased from initial structures of 0.11 to the final molecules of 0.84 for the reference molecule **1**.

We tried another computational experiment to explore the candidate structures for the other target, r5HT1A. The mean fitness curve is similar to the above case was observed for the reference molecule **2** (Figure 7). Fitness of the best molecules and worst molecules are also plotted in Figures 6 and 7. The fitness of the best molecules in each generation also increased as the number of generations increased. These results suggest that our evolutionary approach was successful for exploring the candidate molecules in a huge chemical space. For the computational trial, creating 500 generations of an evolutionary experiment with one CPU core required approximately 75 min.

Ligand Design for hAA2A. Figure 8 shows part of the results of the hAA2A ligand design, which were obtained by the current evolutionary approach. We investigated the individual scaffolds of the designed molecules in terms of chemical graph (CG) expression for nonterminal vertex graph (NTG).[36] An NTG is defined as a graph that has no terminal vertex and no isolated vertex. In the design experiment, 100 molecules with 11 NTG scaffolds including molecules **3**–**6** are designed using the seed fragments derived from the reference molecule **1**. Among them, many molecules (44 molecules including **3** and

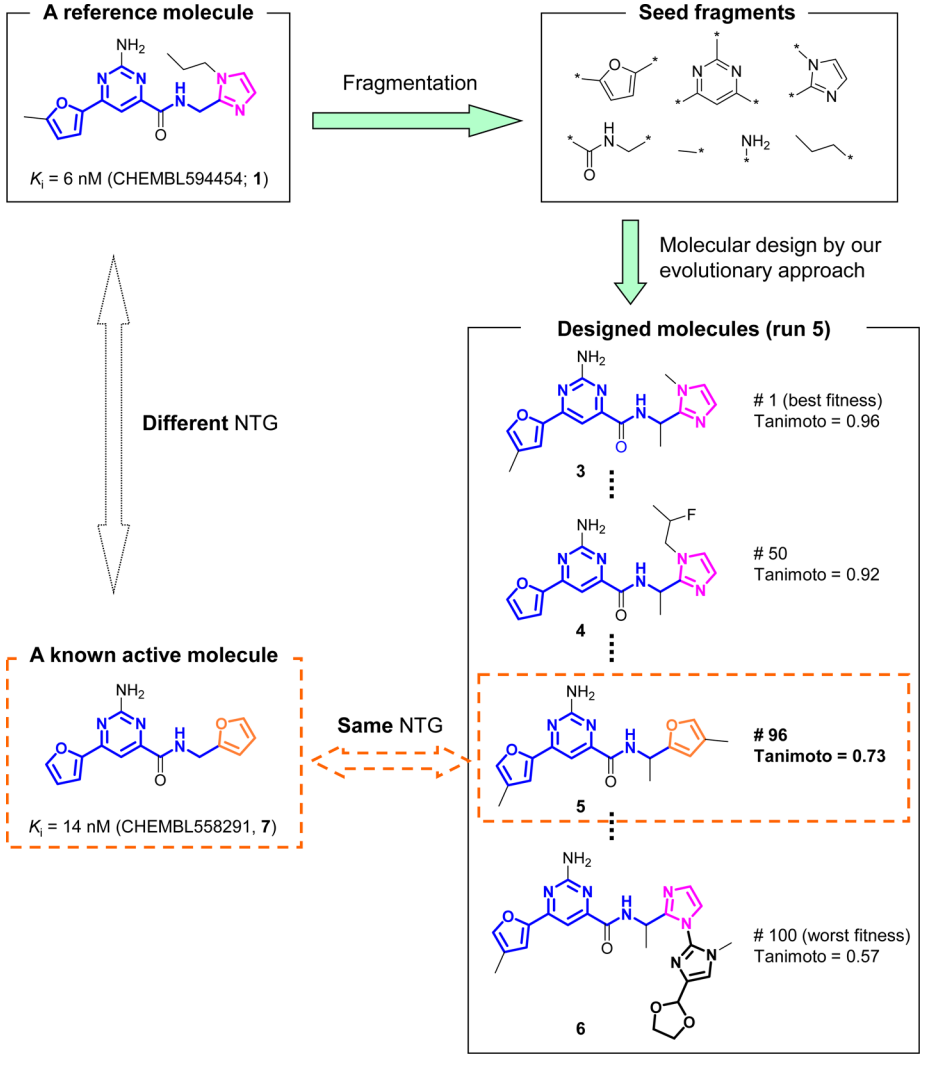


Figure 8. Result of hAA2A ligand design. All of the designed molecules are taken from the 5th trial. The NTG (NTG/CG) of a molecule is specified by bold atoms and bold bonds.

dx.doi.org/10.1021/ci400418c | J. Chem. Inf. Model. 2014, 54, 49–56

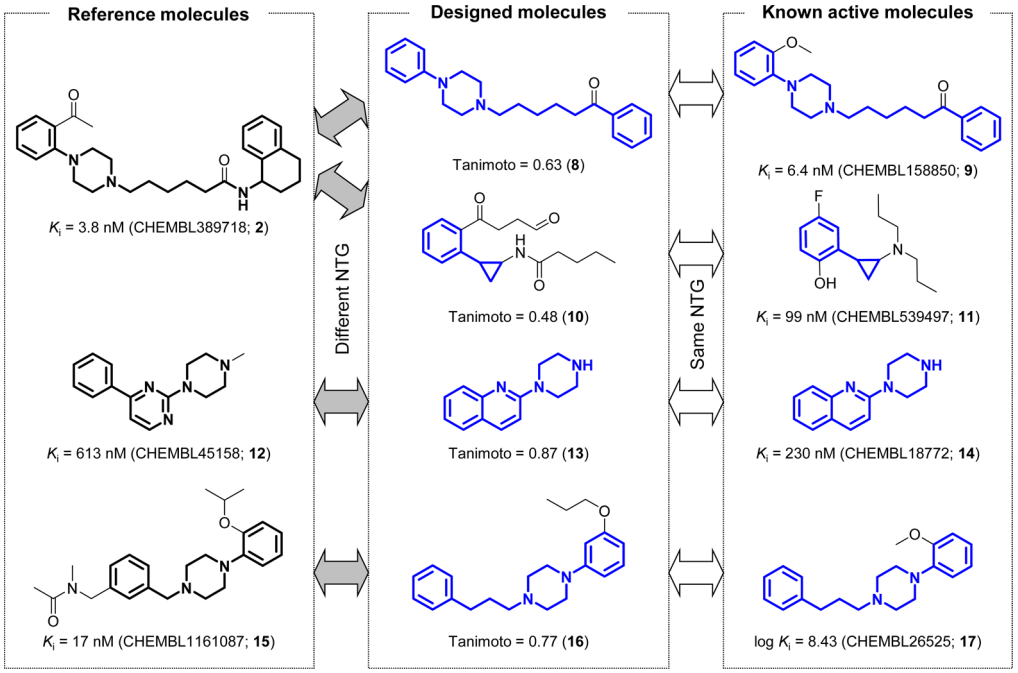


Figure 9. Successful examples of r5HT1A ligand design. Structures with known NTG (NTG/CG) were successfully designed from seed fragments. The NTG moiety is specified by bold atoms and bold bonds.

Table 1. Result of Scaffold Analysis Using NTG (NTG/CG)

	GPCR SARfari		designed[a]		
target	active molecules[b]	NTGs[b]	molecules[b]	NTGs[b]	NTG-shared[c]
hAA2A	2214	811	2975	610	6
r5HT1A	3195	1260	4063	865	10

[a]Five reference molecules were used for the design. [b]Duplicates were excluded. [c]Number of NTG/CGs that are shared in both of the entries of GPCR SARfari and the designed molecules.

4) shared the same NTG of reference **1**, demonstrating the evolutionary direction of the similarity-based approach. The notable points of the design are the following: (1) NTG scaffolds of the designed molecule **5** and a known hAA2A ligand **7** are same. (2) NTG scaffolds of **5** and the reference molecule **1** are different. The structural difference between **5** and **7** is only three methyl groups on the furan rings and the amide linker. As mentioned, a new molecule with similar but different scaffold could be successfully designed from simple seed fragments. It should also be noted that molecules with lower fitness are to be worthy of remark. For example, the fitness of the molecule **5** is 0.73.

Ligand Design for r5HT1A. Figure 9 shows part of the results of the r5HT1A ligand design. The designed molecules for r5HT1A were compared with the active molecules of GPCR SARfari as well. Some successful examples are shown in Figure 9. When **2** was used as a reference, molecule **8** was obtained as one of the designed molecules. NTG scaffolds of **2** and **8** are different from each other. Compound **9** was identified from GPCR SARfari as a similar molecule of **8**. The structural difference between **8** and **9** was only a methoxy group of the phenyl ring. Compound **10** was also designed from the reference **2**. Compound **11** was identified from the database that has the same NTG of **10**. Compound **13** was designed from a reference of **12**, and we were able to find **14** that perfectly matched a molecule in the database. Compound **16** was designed from a reference of **15**, and compound **17** that shares same NTG was found as a known active molecule. These design examples show the applicability of our proposed method.

Chemical Feasibility of Designed Molecules. In Figure 8, molecules **2**−**5** are the designed molecules with the highest fitness, the lowest fitness, and in-between molecules collected from the fifth trial (run 5). The reference molecule used in this case was **1**. Chemical feasibility (or chemical validity) of the designed molecules was examined because the candidate structures should not include unfavorable structures such as invalid heterohetero atom bonds that often appeared in our previous work. In this work, we introduced a fragment library for the mutation operation to avoid the problem. The connection rules for the fragments defined in the present method may also play an important role to improve the performance. The designed molecules are highly similar to the reference molecule. The matter is obvious from the visual inspection of Figure 8 as well. In particular, the scaffolds of compounds **3** and **4** (shown by bold atoms and bold bonds) are the same as the scaffold of the reference molecule.

Scaffold Variation of Designed Molecules. We investigated scaffold variation of the designed molecules obtained from the molecular evolution experiment. Again, the chemical graph (CG) representation of the nonterminal vertex graph (NTG) was used to define the scaffold. As shown in Figure 8, the scaffold of the designed molecule **3** is the same as that of the reference molecule **1**, but the scaffold of the molecule **5** is different from **1**. The number of unique molecules and unique scaffolds are summarized in Table 1. The results clearly show that a large number of unique molecules with a variety of the scaffolds were produced by the current molecular evolutions. The ratios of the number of unique molecules to the number of unique scaffolds were 4.88 for hAA2A and 4.70 for r5HT1A, respectively. This means that the designed molecules that shared the same scaffold are less

dx.doi.org/10.1021/ci400418c | J. Chem. Inf. Model. 2014, 54, 49−56

than five in average. For a comparison, the number of unique NTGs that appeared in the active molecules in GPCR SARfari is shown in Table 1, too.

The number of NTG scaffolds that shared in both of the designed molecules and those of GPCR SARfari are also summarized in Table 1. It shows that, for the case of hAA2A, 6 of the 811 NTGs appeared in GPCR SARfari's molecules were successfully designed by our approach without any special consideration. In other words, six known (validated) scaffolds and 604 new scaffolds were produced during the current molecular evolution for the ligand design. For the case of r5HT1A, 10 known scaffolds and 855 new scaffolds were produced during the molecular evolution with the reference molecules.

Comparison with Other Methods. The performance of the current approach was compared with other methods. We compared with two recent works, NovoFLAP[19] and Flux,[15] because they reported the chemical structures of both of reference molecules and designed molecules. Here, we focused on molecular similarity and medicinal chemistry viewpoint. First, the study was performed using the reference molecules of CP99994 (**s1**) and ICI (**s6**).[19] Then, another study was performed using the reference molecules of Gleevec (**s11**) and a Factor Xa inhibitor (**s14**).[15] The designed molecules with the highest fitness are shown in Supporting Information Figures S1 and S2. When CP99994 (**s1**) was used as a reference, **s2** was designed as the best molecule and **s3** was designed as the second best molecule. The designed molecule **s2** is very similar to **s1**; the difference is only the substitution position of the methoxy group. The difference between **s1** and **s3** was the size of the central heteroring, in which such a design is not shown in the literature. This type of designed molecule is medicinally relevant because of an empirical knowledge that reducing the ring size may improve metabolic stability.[37] In the case of ICI (**s6**), candidate molecules which have new molecular frameworks (**s7**, **s8**) were produced by connecting the known fragments in novel ways. Although it is difficult to strictly compare the performance or the quality of different methods, the result shows at least that the similar and medicinally relevant analogues were successfully designed by our method.

■ CONCLUSIONS

We reported a similarity-driven simple evolutionary approach to producing candidate molecules for drug design and discovery. The method makes it possible to produce candidate molecules that are similar to the reference molecule and yet somewhat different in not only side chains but also their scaffolds. And it is also expected that those candidate structures are chemically feasible. The method was implemented on a software tool and validated with the computer experiments for the GPCR-related ligand design using our own fragment library prepared from GPCR SARfari.

■ ASSOCIATED CONTENT

Supporting Information

Results of comparison study (Figures S1 and S2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +81-75-594-0787. Fax: +81-75-594-0790. E-mail: kawai_kentaro@kaken.co.jp.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Teruki Honma (RIKEN) for his helpful comments and discussions. We also thank the reviewers for their helpful comments to improve the manuscript.

■ ABBREVIATIONS

QSAR, quantitative structure−activity relationship; QSPR, quantitative structure−property relationship; GPCR, G protein-coupled receptor; r5HT1A, rat 5-hydroxytryptamine receptor 1A; hAA2A, human adenosine receptor A2a

■ REFERENCES

(1) Jenkins, J. L.; Glick, M.; Davies, J. W. A 3D similarity method for scaffold hopping from known drugs or natural ligands to new chemotypes. *J. Med. Chem.* **2004**, *47*, 6144−6159.

(2) Oyarzabal, J.; Howe, T.; Alcazar, J.; Andrés, J. I.; Alvarez, R. M.; Dautzenberg, F.; Iturrino, L.; Martínez, S.; Van der Linden, I. Novel approach for chemotype hopping based on annotated databases of chemically feasible fragments and a prospective case study: new melanin concentrating hormone antagonists. *J. Med. Chem.* **2009**, *52*, 2076−2089.

(3) Bahl, A.; Barton, P.; Bowers, K.; Caffrey, M. V.; Denton, R.; Gilmour, P.; Hawley, S.; Linannen, T.; Luckhurst, C. A.; Mochel, T.; Perry, M. W.; Riley, R. J.; Roe, E.; Springthorpe, B.; Stein, L.; Webborn, P. Scaffold-hopping with zwitterionic CCR3 antagonists: identification and optimization of a series with good potency and pharmacokinetics leading to the discovery of AZ12436092. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 6694−6699.

(4) Stellwagen, J. C.; Adjabeng, G. M.; Arnone, M. R.; Dickerson, S. H.; Han, C.; Hornberger, K. R.; King, A. J.; Mook, R. A., Jr.; Petrov, K. G.; Rheault, T. R.; Rominger, C. M.; Rossanese, O. W.; Smitheman, K. N.; Waterson, A. G.; Uehling, D. E. Development of potent B-RafV600E inhibitors containing an arylsulfonamide headgroup. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 4436−4440.

(5) Rheault, T. R.; Stellwagen, J. C.; Adjabeng, G. M.; Hornberger, K. R.; Petrov, K. G.; Waterson, A. G.; Dickerson, S. H.; Mook, R. A.; Laquerre, S. G.; King, A. J.; Rossanese, O. W.; Arnone, M. R.; Smitherman, K. N.; Kane-Carson, L. S.; Han, C.; Moorthy, G. S.; Moss, K. G; Uehling, D. E. Discovery of dabrafenib: a selective inhibitor of Raf kinases with antitumor activity against B-Raf-driven tumors. *ACS Med. Chem. Lett.* **2013**, *4*, 358−362.

(6) Nishibata, Y.; Itai, A. Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation. *Tetrahedron* **1991**, *47*, 8885−8990.

(7) Bohm, H. J. The computer program LUDI: A new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61−78.

(8) Gillet, V.; Johnson, A. P.; Mata, P.; Sike, S.; Williams, P. SPROUT: A program for structure generation. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 127−153.

(9) Douguet, D.; Munier-Lehmann, H.; Labesse, G.; Pochet, S. LEA3D: a computer-aided ligand design for structure-based drug design. *J. Med. Chem.* **2005**, *48*, 2457−2468.

(10) Wang, R.; Gao, Y.; Lai, L. A multi-purpose program for structure-based drug design. *J. Mol. Model.* **2000**, *6*, 498−516.

(11) Yuan, Y.; Pei, J.; Lai, L. LigBuilder 2: A practical de novo drug design approach. *J. Chem. Inf. Model.* **2011**, *51*, 1083−1091.

(12) Vinkers, H. M.; de Jonge, M. R.; Daeyaert, F. F.; Heeres, J.; Koymans, L. M.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Van Aken, K.; Janssen, P. A. SYNOPSIS: synthesize and optimize system in silico. *J. Med. Chem.* **2003**, *46*, 2765−2773.

(13) Schneider, G.; Lee, M. L.; Stahl, M.; Schneider, P. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 487−494.

(14) Brown, N.; McKay, B.; Gilardoni, F.; Gasteiger, J. A graph-based genetic algorithm and its application to the multiobjective evolution of median molecules. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1079−1087.
(15) Fechner, U.; Schneider, G. Flux (1): A virtual synthesis scheme for fragment-based de novo design. *J. Chem. Inf. Model.* **2006**, *46*, 699−707.
(16) Fechner, U.; Schneider, G. Flux (2): Comparison of molecular mutation and crossover operators for ligand-based de novo design. *J. Chem. Inf. Model.* **2007**, *47*, 656−667.
(17) Lameijer, E. W.; Kok, J. N.; Bäck, T.; Ijzerman, A. P. The molecule evoluator. An interactive evolutionary algorithm for the design of drug-like molecules. *J. Chem. Inf. Model.* **2006**, *46*, 545−552.
(18) Huang, Q.; Li, L. L.; Yang, S. Y. PhDD: a new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility. *J. Mol. Graph. Model.* **2010**, *28*, 775−787.
(19) Damewood, J. R., Jr.; Lerman, C. L.; Masek, B. B. NovoFLAP: A ligand-based de novo design approach for the generation of medicinally relevant ideas. *J. Chem. Inf. Model.* **2010**, *50*, 1296−1303.
(20) Kawai, K.; Yoshimaru, K.; Takahashi, Y. Generation of target-selective drug candidate structures using molecular evolutionary algorithm with SVM classifiers. *J. Comput. Chem. Jpn.* **2011**, *10*, 79−87.
(21) Gillet, V.; Johnson, A. P.; Mata, P.; Sike, S.; Williams, P. SPROUT: A program for structure generation. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 127−153.
(22) Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. HOOK: A program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins* **1994**, *19*, 199−221.
(23) Bohacek, R. S.; McMartin, C. Multiple highly diverse structures complementary to enzyme binding sites: Results of extensive application of de novo design method incorporating combinatorial growth. *J. Am. Chem. Soc.* **1994**, *116*, 5560−5571.
(24) Pearlman, D. A.; Murcko, M. A. CONCERTS: Dynamic connection of fragments as an approach to de novo ligand design. *J. Med. Chem.* **1996**, *39*, 1651−1663.
(25) Beccari, A. R.; Cavazzoni, C.; Beato, C.; Costantino, G. LiGen: A High Performance Workflow for Chemistry Driven de Novo Design. *J. Chem. Inf. Model.* **2013**, *53*, 1518−1527.
(26) Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **2005**, *4*, 649−663.
(27) Dey, F.; Caflisch, A. Fragment-based de novo ligand design by multiobjective evolutionary optimization. *J. Chem. Inf. Model.* **2008**, *48*, 679−690.
(28) Wang, J.; Krudy, G.; Xie, X. Q.; Wu, C.; Holland, G. Genetic Algorithm-Optimized QSPR Models for Bioavailability, Protein Binding, and Urinary Excretion. *J. Chem. Inf. Model.* **2006**, *46*, 2674−2683.
(29) Jones, G.; Gao, Y.; Sage, C. R. Elucidating molecular overlays from pairwise alignments using a genetic algorithm. *J. Chem. Inf. Model.* **2009**, *49*, 1847−1855.
(30) Walters, W. P.; Green, J.; Weiss, J. R.; Murcko, M. A. What do medicinal chemists actually make? A 50-year retrospective. *J. Med. Chem.* **2011**, *54*, 6405−6416.
(31) GPCR SARfari. https://www.ebi.ac.uk/chembl/sarfari/gpcrsarfari (accessed Mar 8, 2013).
(32) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−1107.
(33) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
(34) Takahashi, Y.; Ohoka, H.; Ishiyama, Y. Structural Similarity Analysis Based on Topological Fragment Spectra. In *Advances in Molecular Similarity*; Carbo, R., Mezey, P., Eds.; JAI Press: Greenwich, CT, 1998; Vol. 2, pp 93−104.
(35) Kawai, K.; Fujishima, S.; Takahashi, Y. Predictive activity profiling of drugs by topological-fragment-spectra-based support vector machines. *J. Chem. Inf. Model.* **2008**, *48*, 1152−1160.
(36) Takahashi, Y. Chemical data mining based on non-terminal vertex graph. In *2004 IEEE International Conference on Systems, Man and Cybernetics*, The Hague, Oct 10−13; IEEE: Piscataway, NJ, 2004; Vol. *5*, pp 4583−4587.
(37) St. Jean, D. J., Jr.; Fotsch, C. Mitigating heterocycle metabolism in drug discovery. *J. Med. Chem.* **2012**, *55*, 6002−6020.