

K201

Scaffold Hopping Using Inductive Logic Programming

Kazuhisa TSUNOYAMA^{*1}, Ata AMINI^{*2},
Michael J. E. STERNBERG^{*3}, Stephen H. MUGGLETON^{*4}

- 1) *Molecular Medicine Research Laboratories, Drug Discovery Research, Astellas Pharma Inc.*, 21 Miyukigaoka, Tsukuba, Ibaraki 305-8585, Japan
- 2) *Equinox Pharma Ltd.*, Bessemer Building, Prince Consort Road, London SW7 2BP, United Kingdom
- 3) *Structural Bioinformatics Group, Center for Bioinformatics, Division of Molecular Biosciences, Imperial College London*, London SW7 2AZ, United Kingdom
- 4) *Computational Bioinformatics Laboratory, Department of Computing, Imperial College London*, 180 Queen's Gate, London SW7 2AZ, United Kingdom
kazuhisa.tsunoyama@jp.astellas.com

1. Introduction

Similarity search of compounds is often applied in drug discovery with the aim of finding molecules which have similar properties to an initial compound of interest. An important application of similarity search is scaffold hopping which aims to find compounds that are structurally diverse, while sharing a biological activity [1]. Several methods of scaffold hopping have been proposed so far [2]. Here we propose a new scaffold hopping method based on Inductive Logic Programming (ILP) [3].

2. Method

ILP algorithms are designed specifically to learn human-readable rules from observed relationships in positive and negative examples. The inter-relationships of the component atoms in active and inactive scaffolds were used by the learning algorithm as part of the background knowledge. We assigned pharmacophore types to each atom of every scaffold to develop the background knowledge. The coordinate information of atoms was also included along with pharmacophore types. A typical piece of background knowledge could be as follows.

Scaffold 1000 is active.

Atom a10 in scaffold 1000 is a lipophilic atom.

The x, y and z coordinate of atom a10 are respectively 5.40, 1.21, -1.82.

In the above, '1000' is a scaffold name and 'a10' is a unique label for an atom in scaffold 1000.

ILP learns rules describing combinations of spatial relationships between pharmacophore types. An example of such a rule might be as follows.

Scaffold X is active if

X contains atoms A, B and C, and

the distance between atom A and B is $8.8 \pm 0.5 \text{ \AA}$, and

the distance between atom B and C is $4.8 \pm 0.5 \text{ \AA}$, and

the distance between atom A and C is $12.9 \pm 0.5 \text{ \AA}$, and

A and C are lipophilic atoms, and B is an acceptor atom.

The predictions of these rules can be represented as a binary string. Each bit in a binary string corresponds to an individual rule and has the value 1 when the rule covers the scaffold and has the value 0 when the rule does not. The Jaccard coefficient was used for calculating the similarity between binary strings of scaffolds.

This ILP-based scaffold hopping method was compared to two previous algorithms (CATS [1], and CATS3D [4]) on ten datasets with diverse scaffolds.

3. Results and Discussion

Figure 1 shows the total numbers of retrieved scaffolds for the ten datasets. For all the values of x considered in the top x% of the ranked test dataset, more active scaffolds are retrieved by the ILP-based method compared to CATS and CATS3D. The ILP-based method is significantly better than random selection while the other two algorithms are not. In addition, the ILP-based method retrieved new active scaffolds which were not found by CATS and CATS3D. We conclude that our method is at least as good as the other methods in this study.

ILP produces human-readable rules, which makes it possible to identify the three-dimensional features that lead to scaffold hopping (Figure 2). A minor variant of a rule learnt by ILP for scaffold hopping was subsequently found to cover an inhibitor identified by an independent study. This provides a successful result in a blind trial of the effectiveness of ILP to generate rules for scaffold hopping.

The future study would include using new atom types or different similarity measures, learning more diverse rules, and, combining learnt rules with other algorithms.

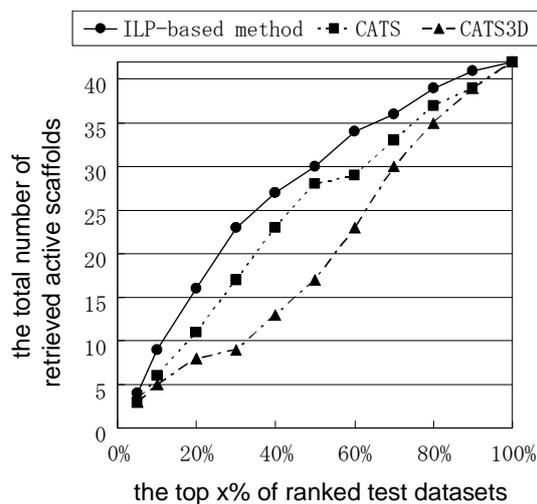


Fig. 1. The total number of retrieved active scaffolds in the top x% of the ranked test dataset.

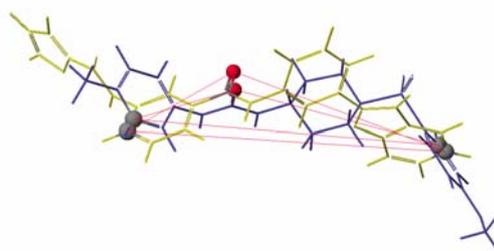


Fig. 2. 3D depiction of an example rule learnt by the ILP-based method (triangles).

References

1. Schneider G, Neidhart W, Giller T, Schmid G. *Angew. Chem. Int. Ed. Engl.*, **1999**, *38*, 2894-2896.
2. Brown N, Jacoby E. *Mini. Rev. Med. Chem.*, **2006**, *6*, 1217-1229.
3. Tsunoyama K, Amini A, Sternberg MJE, Muggleton SH. *J. Chem. Inf. Model.*, **2008**, *48*, 949 – 957.
4. Fechner U, Franke L, Renner S, Schneider P, Schneider G. *J. Comput.-Aided. Mol. Des.*, **2003**, *17*, 687-698.