Improvement of Pseudo-molecule Generation on Solvent Dipole Ordering Virtual Screening (SDO-VS)

Shinya Nakamura*, Hayao Kitayoshi, Isao Nakanishi

Department of Pharmaceutical Sciences, Kindai University, 3-4-1 Kowakae, Higashiosaka, Osaka 577-8502, Japan

(Received August 23, 2017; Accepted October 11, 2017)

Solvent dipole ordering virtual screening (SDO-VS) is a virtual screening method that focuses on the shape of the SDO region at the binding site of the protein. In SDO-VS, pseudo molecules (PMs) are generated to reproduce the shape of the SDO region. Compounds that have shapes (or volumes) similar to those of the PMs are then screened from a 3D structure database. The original implementation of SDO-VS involved PMs with only sp$^3$-hybridized carbon atoms. However, utilization of sp$^2$- and sp-hybridized atoms and/or small molecular fragments, in addition to sp$^3$-hybridized atoms, is expected to provide more efficient screening. To this end, this study investigated the effect of sp$^3$-, sp$^2$-, and sp-hybridized atoms and phenyl rings as fragments for PM generation in the SDO-VS method. The screening efficiencies were compared with the original method for several drug target proteins. Overall, this new method improved screening efficiencies, as measured by the area under the curve of the corresponding receiver operating characteristic plots.

Key Words: Virtual screening, Solvent dipole ordering, Shape similarity

Introduction

Various methods exist to conduct structure-based virtual screening (VS) for the identification of novel compounds with activity against their target protein [1], such as pharmacophore search methods [2], similarity-based methods [3], de novo design [4], and docking-based method [5]. Most of the structure-based VS methods recognize ligands that are complementary to the ligand-binding pocket of the target protein. As solvent water molecules occupy the binding pocket before ligand binding, it is important to consider their dynamics. Bulk water molecules rotate freely and are oriented along various directions; consequently, their dipoles do not, on average, have a precise orientation. Conversely, because binding pockets constrain the orientations of water molecules to be ordered, strong dipoles are commonly observed. This phenomenon is known as solvent dipole ordering (SDO) [6], and the region in which highly ordered dipoles occur is known as the SDO region.

Murata et al. found that a high SDO region in the ligand-binding pocket of a protein indicates the preferred shape of the molecules that bind to the protein [7] and that its shape can be utilized for VS of active compounds. However, it is difficult to utilize the shape of the SDO
region directly for VS because of the absence of useful tools that can effectively compare the unique shape of the SDO region with molecular shapes. However, it is known that shape similarity between two molecules can be measured by comparing their volumes [8, 9], and this has been applied to VS or so-called lead hopping [10]. Therefore, if virtual molecules represent the shape of the SDO region, actual compounds with high shape similarity to virtual molecules (i.e., the SDO region) can be screened. Note that the virtual molecules do not have to be real molecules. Hence, the SDO-based method for VS (SDO-VS) was developed on the basis of this shape-similarity screening concept.

Although the driving forces of SDO have not been completely elucidated, water molecules in the ligand-binding pocket are constrained by at least two factors: (1) hydrogen bonding with the pocket residues, and (2) the hydrophobic surface structure, called hydrophobic hydration [11]. Therefore, replacement for these constrained water molecules with a ligand molecule would be expected to form hydrogen bonds between the ligand and the pocket residues. Further, entropy is gained upon desolvation of relatively unstable structured water molecules following ligand binding. On this basis, the SDO-VS method [12], which screens molecules with similar shapes to the SDO region, was developed.

The original protocol of the SDO-VS method is summarized briefly as follows (Figure 1). First, an important region for ligand binding is analyzed using the dipoles of water molecules (Figure 1(a)-(c)). Second, this (SDO) region is copied to reproduce its shape by a virtual molecule, referred to as a pseudo molecule (PM), which is composed of several sp$^3$-hybridized carbon atoms (Figure 1(d)). PMs are generated so that the atoms cover the SDO region as much as possible, because PMs are generated such that the atoms maximally cover the SDO region, which is possible because they possess shape information of the SDO region. Finally, molecules with conformations that closely match the PM shapes are retrieved from a compound database (Figure 1(e)).

Following an energetic evaluation of the retained compounds (e.g., interaction energies with proteins and deformation energies) as well as visual inspection, candidate compounds are selected for subsequent enzyme assays.

Although a few applications of the SDOVS method have been reported [12, 13], its calculation protocol has not yet been fully established. Indeed, significant scope for improvement in several aspects of the procedure remains, such as more effective utilization of the SDO values, generation of PMs, and the development of databases composed of realistic compounds in multiple conformations. In this study, the aspect of generating more accurate PMs was chosen.

In the original SDO-VS method [12], PMs were generated using sp$^3$-hybridized carbon atoms to reproduce the shape of the SDO region. However, such PMs are nothing like real molecules from an energetic and topological point of view. Instead, employing carbon atoms with various hybridizations or small molecular fragments for PM generation is likely to provide a more accurate representation of PMs as actual molecules, which may well improve the screening performance of the SDO-VS method. Therefore, in this study, a comparison of the screening performance between PMs composed of only sp$^3$-hybridized carbons and PMs composed of various hybridizations (called sp$^3$PM and varPM, respectively) was performed. Because the final performance of the SDO-VS procedure depends on steps such as energy minimization of complex structures and visual inspection, this study focused on examining the screening performance to distinguish true ligands among many decoys.

### Table 1. Number of compounds and ligand components for each target.

| protein | #Ligand$^a$ | #Decoy$^b$ | %sp$^3$$^c$ | %sp$^2$$^d$ | %sp$^e$ | #Ring$^f$ | #Heavy$^g$ | Vol.$^h$ |
|---------|----------|----------|------------|------------|--------|---------|---------|---------|
| AChE    | 107      | 3892     | 66.7       | 33.3       | 0.0    | 3.2     | 25.2    | 404.8   |
| ADA     | 39       | 927      | 93.2       | 6.8        | 0.0    | 2.4     | 17.9    | 277.5   |
| ALR2    | 26       | 995      | 57.3       | 42.7       | 0.0    | 3.1     | 21.8    | 319.1   |
| CDK2    | 72       | 2074     | 59.6       | 40.4       | 0.0    | 3.7     | 24.3    | 360.5   |
| COMT    | 11       | 468      | 67.0       | 33.0       | 0.0    | 1.3     | 16.5    | 239.2   |
| COX1    | 25       | 911      | 58.5       | 41.5       | 0.0    | 2.3     | 19.9    | 301.1   |
| HIVRT   | 43       | 1519     | 62.8       | 36.2       | 1.0    | 2.6     | 22.8    | 363.0   |
| MR      | 15       | 636      | 63.4       | 36.6       | 0.0    | 4.2     | 26.0    | 426.2   |
| TK      | 22       | 891      | 66.4       | 32.2       | 1.4    | 2.2     | 18.0    | 267.2   |

$^a$: The numbers of compounds in each dataset are listed as ligands and decoys, respectively.
$^b$: The percentages of components in ligand atoms for sp$^3$, sp$^2$, and sp hybridizations, respectively.
$^c$: The average numbers of components in ligands for ring structures and heavy atoms, respectively.
$^d$: The average molecular volume (Å$^3$) of ligands. Prepared conformations in datasets were calculated.
Materials and Methods

The nine datasets listed in Table 1 were selected from "a directory of useful decoys" (DUD) dataset [14] to examine the screening performances of the methods used to generate PMs. These targets included acetylcholine esterase (AChE), adenosine deaminase (ADA), aldose reductase (ALR2), cyclin-dependent kinase 2 (CDK2), catechole O-methyltransferase (COMT), cyclooxygenase 1 (COX1), human immune deficiency virus reverse transcriptase (HIVRT), mineralocorticoid receptor (MR), and thymidine kinase (TK). Each dataset includes the 3D structure of the target protein, its ligand, the chemical structures of active ligands, and inactive decoys.

Before the SDO analysis of each target, all amino acid protonation states and the positions of the hydrogen atoms for each target protein were assigned by the Protonate-3D method [15] implemented in MOE [16]. Molecular dynamics (MD) calculations for each target protein were performed using the GROMACS software package [17]. The AMBER-ff99SB force field [18] was assigned for each protein, and then the protein structure was neutralized and placed in a TIP3P [19] water box with a margin of 10 Å between the protein and the boundaries of the periodic box using the tLEaP module of AMBER [20]. The particle mesh Ewald method [21] with a cut-off of 8 Å was employed to calculate the long-range electrostatic interactions. Before the MD simulations, an energy minimization step was run for the solvated protein structure to relax possible clashes existent in the generated structure following addition of the hydrogen atoms and solvent molecules. Initially, MD simulations were run for 100 ps to relax the system under NVT conditions at 300 K. To equilibrate the water density, an additional 400 ps of MD simulations under NPT conditions at 1 bar and 300 K were then run prior to data collection. Finally, production MD simulations of 2 ns were performed under NVT conditions at 300 K; the water structures were extracted every 5 ps (i.e., 401 conformations).

The region selected for analyzing SDO was set according to the ligand position with a margin of 3 Å, assuming that the molecules to be screened would bind to the same binding pocket region. The analysis space was divided into a 1 Å grid, and the SDO value for the i-th grid (SDOi) was calculated as the norm of the dipole vector averaged inside this grid (\(\langle d_i \rangle\)) (Eq. (1)):

\[
SDO_i = \left| \langle d_i \rangle \right| = \sum_{k=1}^{N} \left| d_{i,k} \right| / N_i
\]

where \(d\) is the dipole vector of a water molecule in the grid and \(N_i\) is the number of water molecules inside the i-th grid. PMs were generated using two methods to reproduce the shape of the SDO region. A schematic example is shown in Figure 2, and a flowchart for generating PMs is shown in Figure 3.

As mentioned in the Introduction, PMs were originally composed of only sp2-hybridized carbon atoms. This study adopted the use of carbon atoms with all hybridizations (i.e., sp, sp2, and sp3) to generate PMs. In the following description, PMs composed of sp2-hybridized carbon atoms and variable hybridizations are referred to as sp3PM and varPM, respectively. First, an initial sp2-hybridized carbon atom was placed at the center of a randomly selected grid whose SDO value was higher than 0.9 (Figure 2(a)). Second, the next sp3-hybridized carbon atom was placed at a distance of a C–C single bond length in the direction of a randomly chosen neighboring grid (within a distance of 3.2 Å as an approximate hydrogen bond distance) that belongs to the SDO region (Figure 2(b)). Subsequently, selection of a ‘parent’ atom in an already generated PM was performed according to a predetermined probability (Table 2) to generate a ‘child’ structure via addition of an atom or replacement into phenyl ring. Usually, the child structure is elongated by the addition of an atom with probable hybridization bonded to the parent atom (Figure 2(c)). If the generated child structure is incompatible with the SDO region, it is considered inappropriate; it would be discarded and another child structure would be generated (Figure 2(d)). In varPM, addition of not only a child atom but also placement of the parent atom with a phenyl ring was possible (Figure 2(e)(f)). Additionally, the parent atom was sometimes selected as atoms within the middle segments of the PMs, thus forming branched structures (Figure 2(g)).

In the generation of sp3PM- and varPM-type pseudo molecules, the probability and the maximum number of branches were 10% and three, respectively. In varPM, the probability of phenyl ring replacement and hybridization of generating carbon atoms were approximated from the chemical structures of an in-house compound database. The probabilities were set to 20% (at most 3 rings), 75%, 20%, and 5% for the phenyl ring, sp3, sp2, and sp hybridization for the carbon atom, respectively. In sp3PM, only sp3-hybridized carbon atoms were considered (i.e., no phenyl rings or other atom hybridizations).

Generation of these child structures was repeated until the maximum number of atoms was reached (Figure 2(h)). Then, this process was repeated until 100 unique PMs were generated after verifying the absence of duplicates (Figure 2(i)).

A shape comparison was performed by volume superposition of the ligand query molecules and the decoy database for each PM generated in the SDO region of the ligand-binding pocket using the MOE-Flexible Alignment module. During each superposition, the PM structure was held fixed and the conformation of the query molecule was flexibly aligned to maximize the overlap of both molecular volumes, regardless of their atom types or pharmacophore features. Five superposition candidates per pair were proposed to avoid initial structure
dependency. Shape similarities of five superposed pairs were estimated by shape Tanimoto (ST) values by way of the Tanimoto coefficients of their volumes [8, 9]. The conformation with the highest ST value among five hundred candidates (100 PMs with five superpositions per PM) was adopted as a query molecule’s binding configuration in the SDO region for ligand binding.

The performance of sp3PM and varPM to classify ligands and decoys in each target protein was compared using a receiver operating characteristic (ROC) plot. Note that a large area under the curve (AUC) of a ROC plot indicates a high degree of classification between the ligands and decoys. The average ST values of the ligands and decoys for both PMs were also compared and their significant differences were confirmed by t-test analysis.

**Figure 1.** Ligand screening procedure of SDO-VS. (a) Example of a ligand-binding pocket for analysis. (b) SDO analysis of the ligand-binding pocket. The arrows designate each calculated SDO within the grid. Longer arrow lengths with increasing red color indicate larger SDO values. (c) Definition of a SDO region (e.g., $SDO > 0.7$). (d) PM (cyan) generation inside the SDO region (enlarged white frame in (c)). (e) Example of shape comparison between a PM (cyan) and a screened ligand (magenta).
Figure 2. Example of PM generation in a SDO region. The grid centers of the SDO region are shown as dots. (a) First atom generation at a grid center with a high SDO value (i.e., SDO > 0.9). Candidate center points are shown in red. (b) Second atom elongation with a single C–C bond length to the direction of a neighboring grid center (shown in red dots). (c) Addition of a child atom (e.g., sp\(^3\)-hybridized carbon) to a selected parent atom; probable directions are determined according to hybridization (shown by gray dummy atoms). (d) Example of an sp\(^2\)-hybridized child atom addition to C4. The adjacent C5 position was determined according to sp\(^2\) hybridization; an atom generated outside the SDO region (magenta) would be eliminated. (e) Continuing PM generation and preparation for dummy atoms before phenyl ring replacement. (f) Replacement of parent atom C7 with the ipso-position of a phenyl ring; a probable ortho-position atom was selected from prepared dummy atoms. (g) Branched child atom generation from a selected parent atom. (h) Child atom generation or ring replacement was repeated until the number of atoms in the PM reached a pre-defined maximum. (i) First atom generation of the next PM for another grid center, with repeated generation until the maximum number of PMs is reached.
Figure 3. Flowchart of PM generation in this study. Letters (a)–(i) correspond to the panels in Figure 2.

Table 2. Probabilities and maximum numbers for PM generation.

|                       | sp3PM | varPM |
|-----------------------|-------|-------|
| Maximum number of PMs to generate (#MaxPM) | 100   | 100   |
| Percentage of generating child atom with sp^3 carbon | 100   | 75    |
| Percentage of generating child atom with sp^2 carbon | 0     | 20    |
| Percentage of generating child atom with sp carbon | 0     | 5     |
| Percentage of phenyl ring replacement | 0     | 20    |
| Maximum number of phenyl rings in PM | 0     | 3     |
| Percentage of parent selection at branching position | 10    | 10    |
| Maximum number of branching in PM | 3     | 3     |
| Maximum number of atoms in PM (#MaxAtom) | 20    | 20    |
| Minimum number of atoms in PM | 8     | 8     |

Table 3. Molecular property comparison between sp3PM and varPM.

|       | sp3PM | varPM |
|-------|-------|-------|
|       | %sp3  | #Heavy| Vol.   | %sp3 | %sp2 | %sp  | #Ring | #Heavy| Vol.   |
| AChE  | 100.0 | 18.3  | 274.3 | 57.3 | 37.8 | 4.9  | 0.6   | 16.6  | 247.8 |
| ADA   | 100.0 | 17.3  | 260.2 | 60.5 | 33.5 | 6.0  | 0.5   | 15.2  | 227.2 |
| ALR2  | 100.0 | 17.6  | 262.7 | 58.9 | 37.4 | 3.7  | 0.6   | 16.1  | 237.3 |
| CDK2  | 100.0 | 18.1  | 269.1 | 57.3 | 38.3 | 4.5  | 0.5   | 15.8  | 234.7 |
| COMT  | 100.0 | 19.5  | 290.1 | 62.1 | 33.6 | 4.4  | 0.9   | 17.9  | 264.4 |
| COX1  | 100.0 | 18.8  | 279.7 | 58.8 | 36.1 | 5.2  | 0.6   | 17.6  | 257.9 |
| HIVRT | 100.0 | 11.7  | 181.9 | 65.7 | 31.8 | 2.5  | 0.1   | 10.5  | 164.2 |
| MR    | 100.0 | 11.2  | 173.7 | 65.3 | 30.6 | 4.2  | 0.1   | 10.3  | 160.4 |
| TK    | 100.0 | 12.6  | 195.8 | 64.5 | 31.6 | 3.9  | 0.2   | 11.3  | 174.3 |

Column field names are same as Table 1.
### Results and Discussions

The properties of the generated PMs for each target are shown in Table 3. The numbers of heavy atoms and the volumes were ca. 10% smaller for varPM than for sp3PM. This difference is caused by the linear structure resulting from sp hybridization. To illustrate, consider the simple example of butane and 2-butyne; the lengths of both terminal carbon atoms with respect to their length parameters are 3.80 and 4.22, respectively. The linear structure is approximately 10% longer than the zigzag structure. This means that the number of atoms required to fill an SDO region of the same length is smaller for varPM than for sp3PM, and the volume becomes smaller as the number of atoms decreases.

The number of atoms of the generated PM was set to 20 atoms to be approximately equal to the number of heavy atoms of the actual ligand (Table 1). However, the actual number of generated atoms was smaller than this specified number. For instance, there is more than a two-fold difference between HIVRT and MR. The reason for this difference is that most of these ligands contain a fused polycyclic ring, such as the dipyriddiazepinone scaffold ([6,7,6]-fused ring) in ZINC00004778 (nevirapine, Figure 4(a)) of the HIVRT ligand and the steroid scaffold in ZINC03814383 (progesterone, Figure 4(b)) of the MR ligand. If these fused structures are included, the number of atoms and the ligand volumes become larger than those of the PMs, even though the compounds occupy the same binding pocket region. In this study, phenyl rings were included in the generated PMs. However, the generation of another ring such as cyclohexane or a naphthyl ring would be desirable in future studies to further maximize the similarity between PMs and ligands.

Table 4 and Figure 5 compare the screening performances of both PM approaches (sp3PM and varPM). For most targets, the ROC curve of varPM demonstrated superior performance than that of sp3PM. For instance, significant improvement was observed in COMT and TK (△AUC). Most COMT ligands are very small compounds; for example, ZINC00330141 (pyrogallol, Figure 4(c)) has only three hydroxyl groups on its phenyl ring. The maximum values of the shape similarities of ZINC00330141 to sp3PMs and varPMs were 0.637 (very low among the compounds set for COMT) and 0.885 (very high), respectively. Furthermore, although the TK screening by sp3PM failed to identify any ligands within the top 100 ranks, ZINC02046905 (Figure 4(d)) was found at the 8th rank in the screening by varPM; its shape similarity to varPM was 0.746, which is higher than the best similarity to sp3PM. This ligand is a thymine derivative with a small planar structure. Because such small planar ligands are difficult to screen by aliphatic sp3PMs, generation of the phenyl ring in varPM effectively enhanced the resultant similarity.

However, the ROC curves for HIVRT and MR by varPM were slightly worse than sp3PM, because of the characteristic shapes of the ligand compounds. As described above, ZINC00004778 (nevirapine), as a HIVRT inhibitor, has a [6,7,6]-fused ring with a “butterfly-like shape” conformation [22]. It would be easier to reproduce such a complicated shape using sp3PM, which is more flexible than varPM to form various conformations. Furthermore, the MR ligand ZINC03814383 (progesterone) has a steroid ring; this very bulky shape would be generated less easily using varPM, which has a flat sp² hybridization and a phenyl ring or linear sp hybridization. Since the shape similarity of both PM approaches to bulky ligands is not sufficiently

### Table 4. Performance comparison between sp3PM and varPM.

|         | sp3PM | varPM | Δ(varPM - sp3PM) |
|---------|-------|-------|-----------------|
|         | AUC   | ST<sup>(L)</sup> | ST<sup>(D)</sup> | ΔST  | ΔAUC | ΔST  |
| AChE    | 0.567 | 0.48  | 0.46  | +0.02<sup>**</sup> | 0.623 | 0.47† | 0.41† | +0.06<sup>**</sup> | +0.056 | +0.04 |
| ADA     | 0.662 | 0.57  | 0.55  | +0.02<sup>**</sup> | 0.769 | 0.55† | 0.51† | +0.04<sup>**</sup> | +0.107 | +0.02 |
| ALR2    | 0.508 | 0.54  | 0.51  | +0.03<sup>**</sup> | 0.601 | 0.53† | 0.51† | +0.02<sup>**</sup> | +0.093 | -0.01 |
| C DK2   | 0.494 | 0.50  | 0.50  | -0.00 | 0.631 | 0.48† | 0.45† | +0.03<sup>**</sup> | +0.137 | +0.03 |
| COMT    | 0.171 | 0.51  | 0.52  | -0.01 | 0.434 | 0.52  | 0.48† | +0.04* | +0.263 | +0.05 |
| COX1    | 0.617 | 0.49  | 0.47  | +0.02 | 0.649 | 0.51† | 0.48† | +0.03 | +0.032 | +0.01 |
| HIVRT   | 0.618 | 0.44  | 0.44  | +0.00 | 0.558 | 0.40† | 0.43† | -0.03<sup>**</sup> | -0.060 | -0.03 |
| MR      | 0.694 | 0.41  | 0.43  | -0.02<sup>*</sup> | 0.600 | 0.40† | 0.42† | -0.02<sup>**</sup> | -0.094 | +0.00 |
| TK      | 0.404 | 0.54  | 0.53  | +0.01 | 0.691 | 0.53  | 0.50† | +0.03<sup>*</sup> | +0.287 | +0.02 |

<sup>a</sup>AUC of ROC curve (Figure 5).
<sup>b,c</sup>Average ST values of PMs (L) and decoys (D), respectively. Values with dagger for varPM are significantly different (p < 0.01) from each corresponding value in sp3PM.
<sup>d</sup>Difference of the average ST values between ligands and decoys; values with asterisk/s are significantly different (**, p < 0.01; *, p < 0.05).
high (Figure 4(b)), other methods must be investigated in the future.

Finally, the overall similarities of each PM to ligands and decoys were compared (Table 4). It is expected that the similarities to ligands (ST$^{13}$) are higher than those to decoys (ST$^{13}$). Although the differences in both similarities ($\Delta$ST) of sp3PM were small, those of varPM were improved to almost the same level, except for HIVRT. This improvement was caused by the difference of similarities to decoys; in particular, the ST$^{13}$ values in varPM were smaller than those in sp3PM. Some HIVRT decoys were difficult to identify unless the PMs had high flexibility, as in sp3PM. However, if the PM shape diversity becomes too high due to flexibility, then the decoys with various shapes would have been screened in sp3PM, leading to unexpectedly high ST$^{13}$ values for sp3PM. Thus, moderate flexibility of varPM resulted in a decline of the similarities to decoys, and the ability to distinguish between ligands and decoys was improved for most targets using varPM.

Conclusion

SDO-VS is a virtual screening method that focuses on the shape of the SDO region at the ligand-binding site of proteins. In SDO-VS, PMs are generated to reproduce the shape of the SDO region; compounds with conformations similar to those of the PMs are then screened from a structure database. Although PMs composed of only sp$^3$-hybridized carbon atoms were used in the original SDO-VS study, PMs comprising moieties with shapes more similar to actual compounds are expected to deliver more efficient screening.

On the basis of comparing the two PM approaches, the screening performance of varPM was found to be superior to that of sp3PM for most targets. To further improve this procedure, higher structural flexibility is required by constructing more realistic ligands with complex shapes or greater bulkiness. Such shapes may be included if more PMs were generated by the varPM approach; however, the probability of generating these shapes by varPM is not higher than that of sp3PM. Broadly, it would be preferable to generate more PMs for shape diversity, but this approach requires a tradeoff between accuracy and calculation time. In this study, the ligand characteristics in each target for PM generation were not considered at all. Although it is an advantage of SDO-VS to identify active compounds with various scaffolds to maximize efficiency at the PM screening stage, optimizing the generation conditions, such as the probability of sp$^3$ hybridization according to the shape of known active ligands, would be indispensable. Although the AUCs of ROC were low, better performance of the SDO-VS method is possible. Indeed, further improvement of the method for screening interactions such as hydrogen bonding with the ligand-binding pocket of each target protein is desirable. Nevertheless, the performance in screening using PMs was improved by varPM.

Overall, the SDO-VS protocol is still under development, with scope for further improvement beyond PM generation, including utilization of the SDO values, definition of the SDO region, and clustering and evaluating the generated PMs. Undoubtedly future research initiatives will improve many of these aspects.

Acknowledgement

This study was supported by the MEXT-Supported Program for the Strategic Research Foundation at Private Universities, 2014–2018(S1411037).

References and Notes

[1] E. Lioanta, G. Spyrou, D. K. Vassilatis, Z. Cournia, Curr. Top. Med. Chem., 14, 3923–1938, (2014).
[2] D. Horvath, Methods Mol Biol., 672, 261–98, (2011).
[3] T. G. Kristensen, J. Nielsen, C. N. Pedersen, Comput. Struct. Biotechnol. J., 5, e201302009, (2013).
[4] G. Schneider, U. Fechner, Nat. Rev. Drug Discov., 4, 649–663, (2005).
[5] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, Nat. Rev. Drug Discov., 3, 935–949, (2004).
[6] J. Higo, H. Kono, H. Nakamura, A. Sarai, Protein Struct. Funct. Genet., 40, 193–206, (2000).
[7] K. Murata, N. Nagata, I. Nakanishi, K. Kitaura, J. Comput. Chem., 31, 791–796, (2010).
[8] J. A. Grant, M. A. Gallardo, B. T. Pickup, J. Comput. Chem., 17, 1653–1666, (1996).
[9] F. Fontaine, E. Bolton, Y. Borodina, S. H. Bryant, Chem. Cent. J. 1, 12, (2007).
[10] T. S. Rush III, J. A. Grant, L. Mosyak, A. Nicholls, J. Med. Chem., 48, 1489–1495, (2005).
[11] Y. A. Mikheev, L. N. Guseva, E. Y. Davydov, Y. A. Ershov, Russ. J. Phys. Chem. A, 81, 1897–1913, (2007).
[12] K. Murata, N. Nagata, I. Nakanishi, K. Kitaura, J. Comput. Chem., 31, 2714–2722, (2010).
[13] I. Nakanishi, K. Murata, N. Nagata, M. Kurono, T. Kinoshita, M. Yasue, T. Miyazaki, Y. Takei, S. Nakamura, A. Sakurai, N. Iwamoto, K. Nishiwaki, T. Nakaniwa, Y. Sekiguchi, A. Hirasawa, G. Tsujimoto, K. Kitaura, Eur. J. Med. Chem., 96, 396–404, (2015).
[14] N. Huang, B. K. Shoichet, J. J. Irwin, J. Med. Chem., 49, 6878–6801, (2006).
[15] P. Labute, Proteins, 75, 187–205, (2009).
[16] MOE Ver. 2013 & 2016, Chemical Computing Group Inc., Montreal, Canada.
[17] M. J. Abraham, T. Murtola, R. Schulz, S. Pall, J. C. Smith, B. Hess, E. Lindahl, SoftwareX, 1–2, 19–25, (2015).
[18] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Proteins, 65, 712–725, (2006).
[19] W. L. Jorgensen, J. Am. Chem. Soc., 103, 335–340, (1981).
[20] D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K. M. Merz, G. Monard, H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev, D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling, W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, L. Xiao, P. A. Kollman, AMBER 2016, University of California, San Francisco, (2016).
[21] T. Darden, D. York, L. Pedersen, J. Chem. Phys., 98, 10089–10092, (1993).
[22] J. Ding, K. Das, C. Tantillo, W. Zhang, A.D. Clark Jr, S. Jessen, X. Lu, Y. Hsiou, A. Jacobo-Molina, K. Andries, R. Pauwels, H. Moereels, L. Koymans, P.A.J. Janssen, R.H. Smith Jr, M. Kroeger Koepeke, C.J. Michejda, S.H. Hughes, E. Arnold, Structure, 3, 365–379, (1995).

Figure 4. Molecular volumes of sp3PM (left), ligand with chemical structure (middle), and varPM (right). (a) ZINC00004778 (nevirapine) as a HIVRT ligand, (b) ZINC03814383 (progesterone) as a MR ligand, (c) ZINC00330141 (pyrogallol) as a COMT ligand, and (d) ZINC02046905 as a TK ligand.
Figure 5. ROC curves for screening by shape similarity to PMs. True positive rate (sensitivity) and false positive rate (1 - specificity) are shown in the vertical and horizontal axes, respectively. The varPM and sp3PM curves are shown in bold and dotted lines, respectively. (a) AChE, (b) ADA, (c) ALR2, (d) CDK2, (e) COMT, (f) COX1, (g) HIVRT, (h) MR, and (i) TK.