JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

LiSiCA: A Software for Ligand-Based Virtual Screening and Its Application for the Discovery of Butyrylcholinesterase Inhibitors

Samo Lešnik,[†] Tanja Štular,[‡] Boris Brus,[§] Damijan Knez,[§] Stanislav Gobec,[§] Dušanka Janežič,*[,‡] and Janez Konc*[,†]

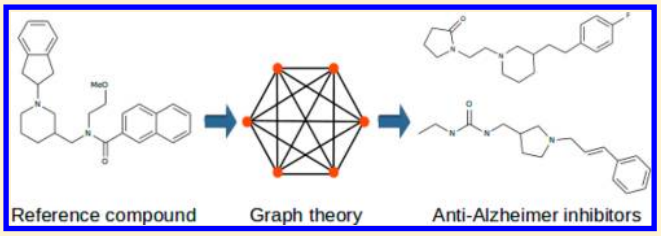
[†]National Institute of Chemistry, Hajdrihova 19, SI-1000, Ljubljana, Slovenia

[‡]Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, SI-6000, Koper, Slovenia

[§]Faculty of Pharmacy, University of Ljubljana, Aškerčeva 7, SI-1000, Ljubljana, Slovenia

Supporting Information

ABSTRACT: We developed LiSiCA (ligand similarity using clique algorithm)—ligand-based virtual screening software that uses a fast maximum clique algorithm to find two- and three-dimensional similarities between pairs of molecules and applied it to the discovery of novel potent butyrylcholinesterase inhibitors. LiSiCA, which runs in parallel on multiple processor cores, was successfully tested on the Database of Useful Decoys—Enhanced, to evaluate its ability to discriminate active molecules from decoys. We then applied LiSiCA for the discovery of novel inhibitors of human butyrylcholinesterase, a promising anti-Alzheimer target, using a known inhibitor as the reference compound. We demonstrated that LiSiCA is capable of finding potent nanomolar inhibitors, whose scaffolds differed from the reference compound, thus proving its ability for scaffold hopping and usefulness in drug discovery.



INTRODUCTION

Ligand-based virtual screening (LBVS) has become a fundamental part of the early drug discovery pipeline and is based on the assumption that similar structures have similar biochemical activity.[1] LBVS can be used to increase the structural variability of the existing crystallographically determined bioactive compounds.[2] The approach requires at least one known active compound (reference compound), which is compared against each target compound in a large molecular database. A desired property of every LBVS approach is scaffold hopping,[3] that is the ability to find bioactive target compounds with different basic scaffolds compared to the reference compound.

An approach to LBVS is to encode the compared molecules as molecular graphs, which are based on the features that are present in the molecular structures.[4] An example of such an approach is the MPHIL (mapping pharmacophores in ligands),[5] which uses a genetic algorithm on molecular graphs to identify common structural features in sets of ligands. CLIP (candidate ligand identification program)[6] is a clique based program that uses a Bron−Kerbosh clique detection algorithm[7] to find the maximum common 3D substructure in the compared molecules, based on their pharmacophoric representation. Another similar algorithm is the RASCAL algorithm,[8] which compares molecules using maximum common substructure or maximum common edge subgraph isomorphism algorithms. Approaches that are not based on graph theory are for example algorithms using MACCS keys,[9] Daylight fingerprints[10] or the 3D shape-based method ROCS.[2] Freely available servers, such as the Multi-Fingerprint Browser,[11] can be used to screen the entire ZINC database.[12]

In this article, we developed LiSiCA (ligand similarity using clique algorithm), an alignment-free LBVS algorithm and software, based on a novel two- and three-dimensional graph representation of molecules and a fast maximum clique algorithm.[13,14] LiSiCA employs all atom graph representation of compounds, in which each vertex represents one atom, in contrast to other clique-based LBVS approaches[1,15,16] that use reduced graphs, in which a vertex represents a functional group, e.g., aromatic ring, hydrogen bond donor/acceptor. Atomic-level details were achieved by using a fast maximum clique algorithm,[13,14] which enables clique searching in large graphs and is up to 2 orders of magnitude faster than the commonly used Bron−Kerbosh maximal clique algorithm. LiSiCA requires as input at least one conformation of a reference compound and a pregenerated database of conformations of target compounds for the three-dimensional screening, or alternatively a topology, that is, a list of atoms and bonds, for each target compound for the two-dimensional screening. In each step, the algorithm compares two molecules based on their two- or three-dimensional representations. Both molecules are converted to molecular graphs, from which the algorithm generates a product graph, which is then searched using the

Received: March 11, 2015
Published: July 9, 2015

DOI: 10.1021/acs.jcim.5b00136
J. Chem. Inf. Model. 2015, 55, 1521−1528

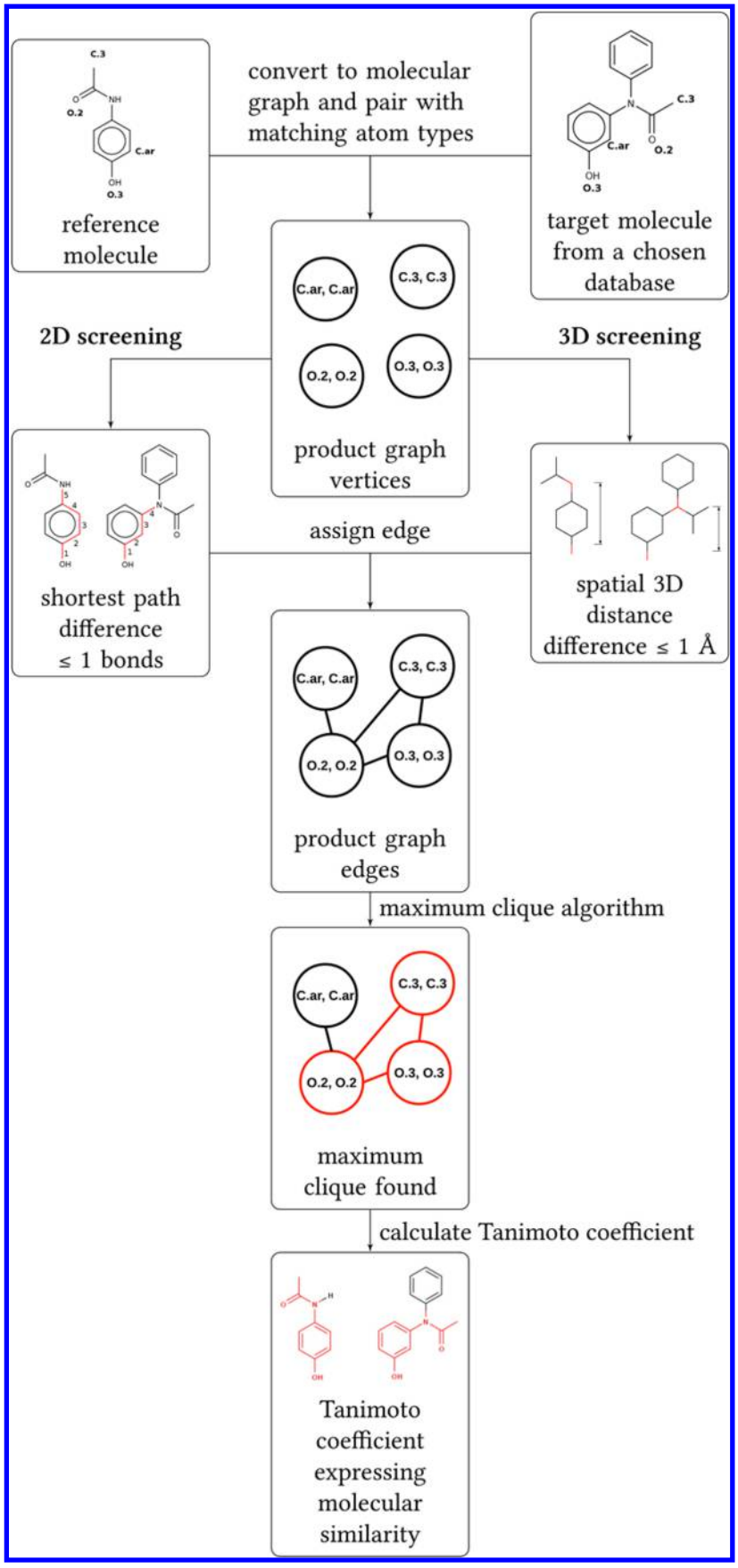


Figure 1. Schematic workflow of a single comparison step of two compounds by LiSiCA.

DOI: 10.1021/acs.jcim.5b00136
J. Chem. Inf. Model. 2015, 55, 1521−1528

maximum clique algorithm to find the largest substructure common to both molecules. Finally, the similarity between the two compared molecules is expressed using the Tanimoto coefficients, and the target compounds are ranked according to their Tanimoto coefficients.[17] We demonstrate the LiSiCA's usefulness in drug discovery by the discovery of novel, potent nanomolar inhibitors of human butyrylcholinesterase (BChE) using a recently published inhibitor[18] as a reference compound. Butyrylcholinesterase, whose level and activity increase in the late stages of Alzheimer's disease is a promising drug target,[19,20] and the novel inhibitors that we found are good candidates for hit-to-lead optimization. LiSiCA software, which has been parallelized to run on multiple CPU cores, can be downloaded from http://www.sicmm.org/konc/lisica.

METHODS

LiSiCA takes as input two compounds and converts each of them to a molecular graph, which differs in the two- and three-dimensional screening (Figure 1). In the two- and three-dimensional screening the graph vertices represent atoms; in the two-dimensional screening edges represent covalent bonds; in the three-dimensional screening edges are drawn between every pair of vertices and have no chemical meaning. From two compared molecular graphs the algorithm creates vertices of a product graph, followed by the generation of two- or three-dimensional edges of the product graph by two different procedures. The maximum clique algorithm searches the product graph and finds the largest common substructure (the subgraph that is common to both molecular graphs and contains the most vertices) present in both molecules. Finally, the algorithm calculates the Tanimoto coefficient to express the degree of similarity between the compared molecules.

Input to the LiSiCA Molecular Comparison Algorithm. The input to LiSiCA is a reference compound and a database of target compounds in the Mol2 format,[21] which can support multiple molecular models in a single file. For the three-dimensional screening, conformations of the target compounds should be generated before input to the LiSiCA using, for example, OpenBabel's obconformer tool.[22]

Generation of Molecular Graphs. The algorithm takes as input two compounds and converts each of them to a molecular graph, which differs in the two- and three-dimensional screening. In the two- and three-dimensional screening the graph vertices represent atoms; in two-dimensional screening edges represent covalent bonds, whereas in the three-dimensional screening edges are drawn between every pair of vertices and have no chemical meaning. The two-dimensional molecular graph is represented with an adjacency matrix, which carries information about the topological connectivity (bonds) of the atoms of the molecule. It is important to note that unlike in many other similar methods,[1,15,16] no graph reduction is taking place in vertex generation, as each vertex represent exactly one atom. We therefore obtain a very rigorous graph representation that captures atom-level details of the compounds.

Generation of Product Graph Vertices. LiSiCA generates a product graph from the molecular graphs of the reference and the currently selected target compound. It iterates through the vertices (representing atoms comprising the molecule) of the first and the second molecular graph and combines all the vertices that represent atoms of the same type into new, two-component vertices of the product graph, for example C.3 (sp3 hybridized carbon) type atom from the first molecular graph (representing the reference molecule) is combined with C.3 type atom from the second molecular graph (representing the target molecule); where C.3 is an example of a SYBYL atom type,[23] which are used in the Mol2 format. Each vertex of the product graph thus contains two components: (1) a vertex from the reference compound and (2) a matched (by atom type) vertex from the target compound. Therefore, product graph vertices represent every possible combination of same-type atoms from both molecules. By using vertices that encode SYBYL atom types that distinguish atoms according to their environment, we implicitly take into account the chemical and binding properties of individual atoms. For example, nitrogen can be represented as sp3-, sp2-, and sp-hybridized, aromatic, or amide type, mirroring its possible binding interactions.

Generation of Edges in 2D Product Graph. The generation of product graph vertices is followed by finding the edges that connect them. We use the following rule to generate an edge between two product graph vertices: an edge exists if the shortest path (measured in the number of bonds) between their first components is of the same length $\pm s$ as the shortest path between their second components, where s is the allowed discrepancy between the shortest path lengths, and is set by default to the value of zero. Shortest paths between each two atoms in reference and target molecular graphs are computed using the Floyd–Warshall algorithm (Figure 2).[24]

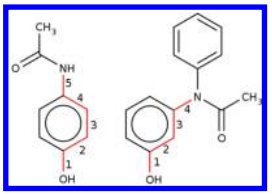


Figure 2. Example use of Floyd–Warshall algorithm on a paracetamol molecule and its analog. The shortest paths of five and four bonds in length between the oxygen–nitrogen atom pairs are highlighted in red.

Generation of Edges in 3D Product Graph. The algorithm for generating three-dimensional product graph edges takes into consideration the spatial relationships between coordinates of the reference and target molecule atoms (Figure 3). The program thus requires different conformations of target compounds, and generates a different set of edges for each conformation of the same target compound. Edges are generated as follows: (i) The algorithm calculates the spatial distance D_R between the first components (vertices of the reference molecular graph) and the distance D_T between the second components (vertices of the target molecular graph).

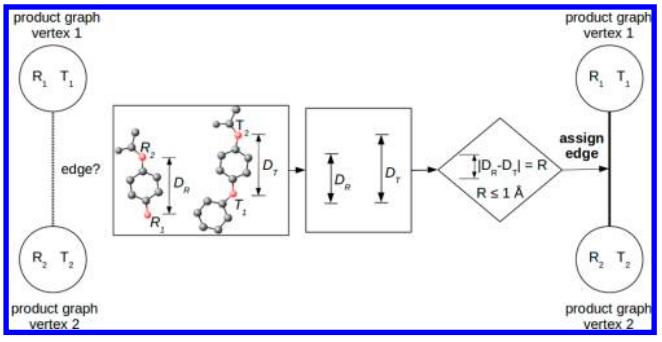


Figure 3. Generation of 3D product graph edges.

DOI: 10.1021/acs.jcim.5b00136
J. Chem. Inf. Model. 2015, 55, 1521–1528

(ii) Then it calculates the absolute difference R between D_R and D_T. (iii) If R is equal or smaller than r, where r is set to 1 Å by default, then an edge between the two corresponding product graph vertices exists, otherwise it does not.

Maximum Clique Algorithm on a Product Graph. A clique is a graph in which every pair of vertices is connected by an edge.[13] The maximum clique is the clique that contains the highest number of vertices in a given graph. The maximum clique algorithm returns the maximum common subgraph of the compared molecular graphs. The maximum common subgraph represents the maximum common substructure, defined as a similar topological (for two-dimensional screening) or spatial arrangement (for three-dimensional screening) between the reference and the target compound. The level of similarity between the compared molecules is expressed with the Tanimoto coefficient $T(R, T) = NS/(NT + NR - NS)$, where NS is the number of atoms in the common substructure, NR the number of atoms that constitute the reference molecule, and NT is the number of atoms that constitute the target molecule. Finally, the target molecules (different conformations of the same target molecule for the three-dimensional comparison) are ranked by the Tanimoto coefficient based on their similarity to the reference molecule. The output of LiSiCA is thus a ranked list (by Tanimoto coefficient) of the target compounds' best-scored conformations.

Compound Database Preparation and Virtual Screening. For the virtual screening of the butyrylcholinesterase inhibitor, the ZINC Drugs Now subset with 7.4 million drug-like compounds was used.[12,25] The ZINC drug-like subset was first filtered using the FILTER 2.0.2 application (OpenEye Scientific Software, Inc., Santa Fe, NM, USA; www.eyesopen.com). Known or predicted aggregators (a function developed by Shoichet)[26] and the compounds with predicted poor solubility (see Supporting Information for the configuration file) were eliminated. In the next step, an in-house Python script was used to remove Pan Assay Interference Compounds (PAINS).[27] The resulting filtered library consisted of roughly 5.52 million compounds. In the final step, the database was processed with the Omega 2.4.3 software (OpenEye Scientific Software, Inc., Santa Fe, NM, USA; www.eyesopen.com)[28,29] using the default settings, to prepare different conformations of compounds that covered as much conformational space as possible. A database of 5.52 million compounds with an average of 188 conformations per compound was then used for the ligand-based virtual screening. The virtual screening with LiSiCA was performed using the bioactive conformation of compound **1** as the reference compound (PDB code 4TPK, ligand code 3F9),[18] with the LiSiCA's 3D option (requesting the three-dimensional screening) and all other settings set to default values. Compounds in the database were ranked according to the Tanimoto coefficient, and the top 30 compounds, most similar to the reference compound **1**, were obtained from different vendors and evaluated biochemically for their inhibitory potency against the butyrylcholinesterase.

Inhibitory Potency Against the Human Butyrylcholinesterase. The inhibitory potencies of the LiSiCA's ligand-based virtual screening hits were determined using the method of Ellman.[30] 5,5′-Dithiobis (2-nitrobenzoic acid) (Ellman's reagent, DTNB), and the butyrylthiocholine iodide (BTC) were purchased from Sigma-Aldrich (Steinheim, Germany). The recombinant human BChE at the stock concentration of 4.6 mg/mL in 10 mM MES buffer (pH 6.5) was kindly donated by Florian Nachon (IBS, Grenoble). The enzyme solutions were prepared by dilution of the concentrated stocks in a phosphate-buffered solution (0.1 M, pH 8.0). The reactions were carried out in a final volume of 300 μL of 0.1 M phosphate-buffered solution, pH 8.0, containing 333 μM DTNB, 5 × 10−4 M BTC and 1 × 10−9 M human BChE. The reactions were started by addition of the substrate, at room temperature. The final content of organic solvent (dimethyl sulfoxide; DMSO) was always 1%. The formation of the yellow 5-thio-2-nitrobenzoate anion as a result of the reaction of DTNB with the thiocholine was monitored for 1 min as the change in absorbance at 412 nm, using a 96-well microplate reader (Synergy H4, BioTek Instruments, Inc., USA). To determine the blank value (b), phosphate-buffered solution replaced the enzyme solution. The initial velocity (v_0) was calculated from the slope of the linear trend obtained, with each measurement carried out in triplicate. For the first inhibitory screening, stock solutions of the test compounds (1 mM) were prepared in DMSO. The compounds were added to each well at a final concentration of 10 μM. The reactions were started by the addition of the substrate to the enzyme and inhibitor that had been preincubated for 300 s, to allow complete equilibration of the enzyme−inhibitor complexes. The initial velocity in the presence of the test compound (vi) was calculated. The inhibitory potency was expressed as the residual activity ($RA = (v_i - b)/(v_0 - b)$). For the IC_{50} measurements, seven different concentrations of each compound were used to obtain enzyme activities of between 5% and 90%. The IC_{50} values were obtained by plotting the residual enzyme activities against the applied inhibitor concentrations, with the experimental data fitted to the equation: $Y = \text{bottom} + (\text{top} - \text{bottom})/(1 + 10^{((\log IC_{50} - X) \times \text{HillSlope})})$, where X is the logarithm of the inhibitor concentration and Y is the residual activity. For the fitting procedure, the Gnuplot software and an in-house python script were used. As a positive control, we used tacrine, a known inhibitor of BChE, for which we measured an IC_{50} of 0.012 ± 0.003.

Chemistry-Hit Analysis. ^{1}H NMR spectra were recorded at 400 MHz on a Bruker Avance III NMR spectrometer at 295 K, unless otherwise stated. The chemical shifts (δ) are reported in parts per million (ppm) and are referenced to the solvent used (DMSO-d_6, 2.50 ppm). Mass spectra were recorded on a LC/MS system Agilent 6224 Acurate Mass TOF LC/MS spectrometer (IC UL FKKT). Analytical reversed-phase HPLC was performed on a Thermo Scientific system equipped with a quaternary pump system, an autosampler, a photodiode array detector set to 256.0 nm, and Chameleon 7 software. An Agilent Eclipse Plus C18 column (150× 4.6 mm, 5 μm) was used, thermostated at 25 °C, and with a flow rate of 1.0 mL/min and a sample injection volume of 10 μL. A general method with 0.1% TFA in H_2O/MeCN gradient (0−15 min, 10% MeCN → 90% MeCN; 15−19 min, 90% MeCN; 19−20 min, 90% MeCN →10% MeCN; equilibration time: 7 min) was used. The purities of the hit compounds were ≥95%, as determined by HPLC (Table S4, Supporting Information).

Statistical Evaluation of LiSiCA. Receiver operating characteristic (ROC) curves are a commonly used method to assess virtual screening performance.[31] A ROC curve is a plot of the true positive rate (TPR, sensitivity) versus the false positive rate (FPR, 1 − specificity). Considering a rank i, two rates can be written as following: $TPR = TP_i/(TP_i + FN_i) = Se_i$ and $FPR = FP_i/(TN_i + FP_i) = 1 - Se_i$, where TP_i, FN_i, TN_i, and FP_i are true positive, false negative, true negative, and false positive

DOI: 10.1021/acs.jcim.5b00136
J. Chem. Inf. Model. 2015, 55, 1521−1528

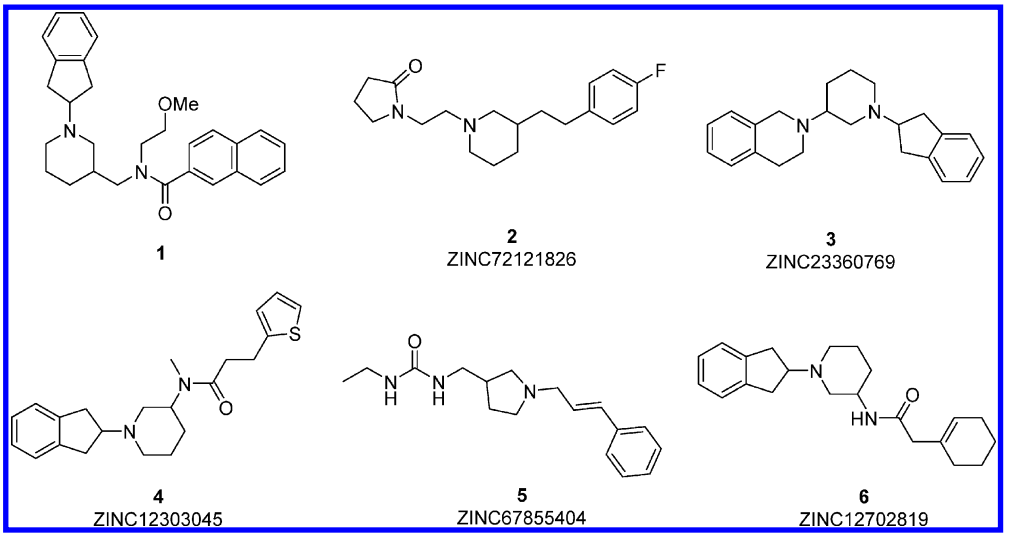


Figure 4. Structure of the reference compound **1** (1,3-disubstituted piperidine analog) along with the structures of the novel human BChE inhibitors **2**−**6** discovered by LiSiCA.

rates at threshold i, respectively, and Se_i and Sp_i are the sensitivity and the specificity. The area under the curve (AUC) of a ROC plot is a scalar representation of the overall quality of the virtual screening method and measures the rank of a randomly selected active compared to a randomly chosen decoy. The value can vary between 0 and 1, where 1 represents a perfect ranking, where all actives are ranked above decoys, whereas 0.5 corresponds to a ranking no better than random.

RESULTS AND DISCUSSION

We evaluated the ability of LiSiCA to discriminate between active and inactive compounds using the Directory of Useful Decoys—Enhanced (DUD-E) database, a standard database used to evaluate LBVS software.[32] We also screened the same database with the well-established LBVS method ShaEP[33] and compared the results to those obtained with LiSiCA. Moreover, we evaluated the parallel speedup of LiSiCA on a multi core platform; finally, we applied LiSiCA to the discovery of novel inhibitors of human butyrylcholinesterase (BChE), a promising anti-Alzheimer drug target (Figure 4).[19,20] The discovered inhibitors will contribute to the fight against this more and more prevalent disease.

Evaluation on the Directory of Useful Decoys—Enhanced. The DUD-E database contains known active and decoy compounds for 102 target proteins in addition to their corresponding crystal ligands that were used as reference compounds in our screening. For the three-dimensional screening with LiSiCA and ShaEP we generated multiple conformers of active and decoy compounds using the Balloon software[35] with the same settings as in the original ShaEP article.[33] To ensure fair comparison between methods, we first conducted a control screening on the DUD database[36] using ShaEP and were able to recreate the average AUC results reported in ref 33. (Table S2, Supporting Information). We then screened every crystal ligand of the DUD-E set against its corresponding active and decoy compounds with the two-dimensional LiSiCA and against each conformer structure with the three-dimensional LiSiCA and ShaEP, and the resulting average ROC curves are shown in Figure 5 (AUCs for each individual DUD-E case are in Table S1, Supporting Information).

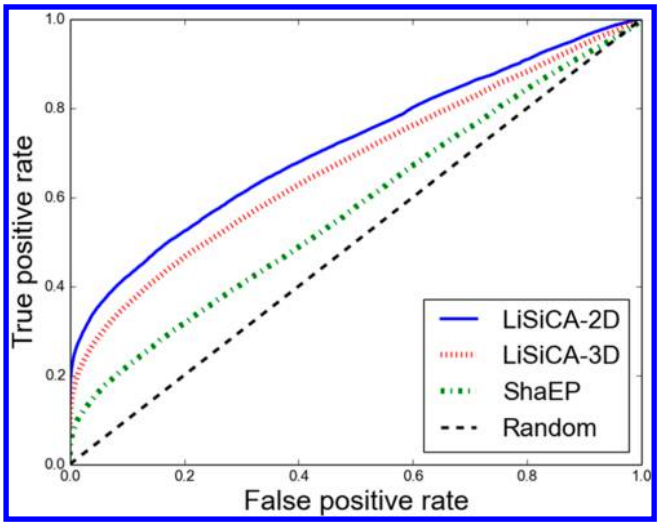


Figure 5. Average ROC curves for two- (blue solid curve) and three-dimensional (red dotted curve) LBVS with LiSiCA and ShaEP (green dashed-dotted curve) for 102 protein targets from the DUD-E database. The ROC curves were obtained by vertically averaging[34] the individual 102 ROC curves for each protein target.

The mean LiSiCA's AUC for the two-dimensional virtual screening test was 0.71 ± 0.15, whereas the mean AUC for the three-dimensional screening was 0.67 ± 0.16. Both approaches, the two- and three-dimensional screening, were equally efficient ($p > 0.05$) at discriminating actives from decoys within the DUD-E database, which is in correlation with previous findings that the two-dimensional virtual screening performs equally well if not better than three-dimensional virtual screening.[37] Using ShaEP we obtained an average AUC of 0.57 ± 0.11, therefore indicating that LiSiCA was overall better ($p < 0.01$ for both the 2D and the 3D method) at discriminating actives from decoys. In addition, LiSiCA was also faster, with an average CPU time per conformation of 8 ± 7 ms (2D) and 27 ± 30 ms (3D) against the ShaEP's 32 ± 16 ms (Table S1, Supporting Information).

For most DUD-E protein targets the deviations of LiSiCA's AUCs from average were small; however, there were cases where AUCs were nearly one (Table S1, Supporting

DOI: 10.1021/acs.jcim.5b00136
J. Chem. Inf. Model. 2015, 55, 1521−1528

Table 1. Activities and Binding-Efficiencies of the Reference Compound and the Novel Human BChE Inhibitors Discovered with LiSiCA

compound number	ZINC ID	IC_{50} ± SEM (μM) huBChE	Tanimoto score	binding-efficiency index pIC_{50}/MWT (kDa)
1	reference	0.021 ± 0.002	1	17.34
2	ZINC72121826	0.08 ± 0.004	0.4268	22.28
3	ZINC23360769	0.23 ± 0.03	0.4458	19.97
4	ZINC12303045	0.37 ± 0.02	0.4070	17.45
5	ZINC67855404	0.38 ± 0.01	0.4304	22.34
6	ZINC12702819	0.84 ± 0.1	0.4186	17.95

Information), and contrarily cases where they were below 0.50. The presence of extreme AUC values in both directions is most likely due to the software's rigorous atomic recognition method. A product graph vertex is created only if the two compared atoms are of the exact same SYBYL atom type. From visually comparing active and decoy compounds, we found good performance if the reference crystal structure contained a large representative scaffold that was also present in most compounds from the active data set, but was rare in the decoy data set. In these cases the actives share with the reference structure many atoms of the same type that are also similarly topologically and spatially arranged, and contribute considerably to the high value of the Tanimoto coefficient. On the contrary, for 10 (two-dimensional screening) and 17 (three-dimensional screening) DUD-E targets our method performed no better or worse than random. In most of these cases the reference crystal structure has a scaffold that is not present, or is present in a very few active compounds in the corresponding database.

Inhibition of Human Butyrylcholinesterase. The 30 top hits from ligand-based virtual screening were purchased from different vendors and evaluated in vitro for their inhibition of recombinant human BChE, using the method of Ellman.[30] Among the purchased compounds, 16 (53%) showed >50% inhibition at 10 μM (Table S3, Supporting Information), further demonstrating LiSiCA's ability to score active compounds significantly higher than inactive. The most effective among the purchased inhibitors were demonstrated to be compounds **2**–**6** (Figure 4, Table 1) with IC_{50} values in the nanomolar range and were considered as hit compounds. Further characterization revealed that compound **2** (1-(2-(3-(4-fluorophenethyl)piperidin-1-yl)ethyl)pyrrolidin-2-one) is the most potent human BChE inhibitor in this series with an IC_{50} value of 80.3 nM; its superimposition with the reference compound in the butyrylcholinesterase binding site is shown in Figure 6. Also encouraging is the observed scaffold hopping and increased binding efficiency[38] that can be observed when examining the hit compounds **2**–**6**. Especially interesting for further research and drug development is compound **5**, which has a significantly different scaffold than the original reference molecule (MACCS scores in Table S3, Supporting Information), as well as having the highest binding efficiency index among the novel inhibitors found.

Speed Evaluation on Multiple Cores. We evaluated the speed performance of LiSiCA by screening around one-fifth of the DUD-E decoys against their corresponding reference compounds. Our goal was to measure the total screening time required and also the speedup of the algorithm on multiple processor cores. Execution times were measured for two- and three-dimensional virtual screening, with different number of processor threads; the algorithm was run 5 times to obtain an average time for each number of threads. Speedups of

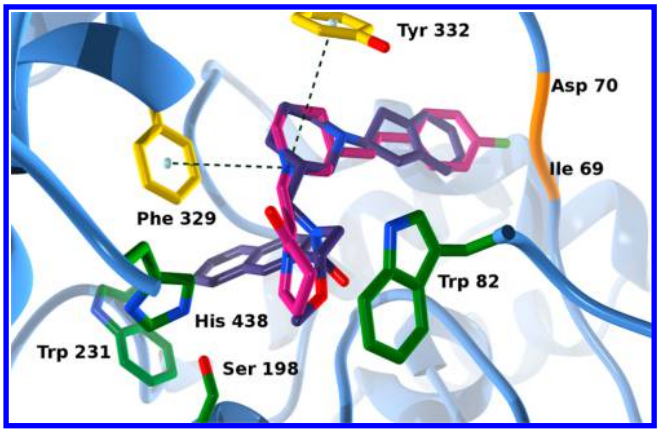


Figure 6. Superimposition of the reference compound **1** (purple) and the best inhibitor, compound **2**, (pink) in the butyrylcholinesterase binding site (green and yellow stick models, PDB code 4TPK). Possible cation-π interactions[39] between the binding site's phenyl rings (yellow stick models) and the quaternary ammonium cation of compound **1** are shown as green dashed lines. The residues that may, through their backbone atoms, interact hydrophobically with the fluorophenyl ring of compound **2** are shown as orange ribbons.

LiSiCA on multiple cores are shown in Figure 7, where we can observe that for a commonly used 8 threads, a good speedup is evident (approximately 6.5x for both two- and three-dimensional screening). The total execution time on one thread was, for approximately three hundred thousand comparisons, on

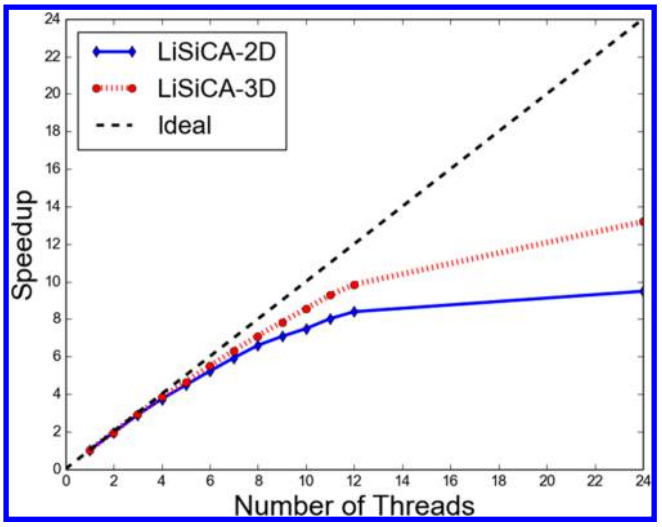


Figure 7. Speedups of the parallel LiSiCA algorithm. Threads 1–12 use physical CPU cores; the 24 threads use 12 physical and 12 virtual cores.

DOI: 10.1021/acs.jcim.5b00136
J. Chem. Inf. Model. 2015, 55, 1521–1528

average 50 min for the two-dimensional screening and 150 min for the three-dimensional screening.

CONCLUSIONS

We developed LiSiCA, a fast, parallel, freely available software for two- and three-dimensional ligand-based virtual screening of large compound databases using a graph theoretical approach, and applied it to an anti-Alzheimer target, the human butyrylcholinesterase. We discovered potent inhibitors of butyrylcholinesterase with novel scaffolds and significantly higher binding efficiencies compared to the reference compound. LiSiCA showed high enrichment ability as 53% of the tested compounds showed >50% inhibition at 10 μM. LiSiCA was able to predict novel inhibitors, extremely suitable for further experimental research and optimization, and thus proved its usefulness in the early stages of the drug discovery process.

ASSOCIATED CONTENT

Supporting Information

AUCs under the ROC curves for the DUD-E based screening and speeds measured in milliseconds per compound (conformer) for each method, AUCs under the ROC curves for the DUD based screening, purchased highest scoring compounds, filter configuration file, dose response curves, and hit purity. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00136.

AUTHOR INFORMATION

Corresponding Authors

*Phone: +386-5-611-76-59. E-mail, dusanka.janezic@upr.si.
*Phone: +386-1-4760-273. E-mail, konc@cmm.ki.si.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support through Grants P1-0002, P1-0208, J1-6743, and L1-6745 of the Ministry of Higher Education, Science and Technology of Slovenia and Slovenian Research Agency is acknowledged. We thank the groups of Florian Nachon and Martin Weik (IBS, Grenoble, FRA) for providing us the BChE.

REFERENCES

(1) Barker, E. J.; Buttar, D.; Cosgrove, D. A.; Gardiner, E. J.; Kitts, P.; Willett, P.; Gillet, V. J. Scaffold Hopping Using Clique Detection Applied to Reduced Graphs. *J. Chem. Inf. Model.* **2006**, *46*, 503−511.
(2) Rush, T. S.; Grant, J. A.; Mosyak, L.; Nicholls, A. A Shape-Based 3-D Scaffold Hopping Method and Its Application to a Bacterial Protein-Protein Interaction. *J. Med. Chem.* **2005**, *48*, 1489−1495.
(3) Böhm, H. J.; Flohr, A.; Stahl, M. Scaffold Hopping. *Drug Discovery Today: Technol.* **2004**, *1*, 217−224.
(4) Raymond, J. W.; Willett, P. Maximum Common Subgraph Isomorphism Algorithms for the Matching of Chemical Structures. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 521−533.
(5) Holliday, J. D.; Willett, P. Using a Genetic Algorithm to Identify Common Structural Features in Sets of Ligands. *J. Mol. Graphics Modell.* **1997**, *15*, 221−232.
(6) Rhodes, N.; Willett, P.; Calvet, A.; Dunbar, J. B.; Humblet, C. CLIP: Similarity Searching of 3D Databases Using Clique Detection. *J. Chem. Inf. Model.* **2003**, *43*, 443−448.
(7) Bron, C.; Kerbosch, J. Finding All Cliques of an Undirected Graph. *Commun. ACM* **1973**, *16*, 575−577.
(8) Raymond, J. W.; Gardiner, E. J.; Willett, P. RASCAL: Calculation of Graph Similarity Using Maximum Common Edge Subgraphs. *Comput. J.* **2002**, *45*, 631−644.
(9) Xue, L.; Godden, J. W.; Stahura, F. L.; Bajorath, J. Design and Evaluation of a Molecular Fingerprint Involving the Transformation of Property Descriptor Values into a Binary Classification Scheme. *J. Chem. Inf. Model.* **2003**, *43*, 1151−1157.
(10) Daylight Manual. http://www.daylight.com/dayhtml/doc/theory/theory.finger.html (accessed May 22, 2015).
(11) Awale, M.; Reymond, J. L. A Multi-Fingerprint Browser for the ZINC Database. *Nucleic Acids Res.* **2014**, *42*, W234−W239.
(12) Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* **2012**, *52*, 1757−1768.
(13) Konc, J.; Janežič, D. An Improved Branch and Bound Algorithm for the Maximum Clique Problem. *MATCH Commun. Math. Comput. Chem.* **2007**, *58*, 569−590.
(14) Depolli, M.; Konc, J.; Rozman, K.; Trobec, R.; Janežič, D. Exact Parallel Maximum Clique Algorithm for General and Protein Graphs. *J. Chem. Inf. Model.* **2013**, *53*, 2217−2228.
(15) Gillet, V. J.; Willett, P.; Bradshaw, J. Similarity Searching Using Reduced Graphs. *J. Chem. Inf. Model.* **2003**, *43*, 338−345.
(16) Barker, E. J.; Gardiner, E. J.; Gillet, V. J.; Kitts, P.; Morris, J. Further Development of Reduced Graphs for Identifying Bioactive Compounds. *J. Chem. Inf. Model.* **2003**, *43*, 346−356.
(17) Willett, P. Similarity-Based Approaches to Virtual Screening. *Biochem. Soc. Trans.* **2003**, *31*, 603−606.
(18) Brus, B.; Košak, U.; Turk, S.; Pišlar, A.; Coquelle, N.; Kos, J.; Stojan, J.; Colletier, J. P.; Gobec, S. Discovery, Biological Evaluation, and Crystal Structure of a Novel Nanomolar Selective Butyrylcholinesterase Inhibitor. *J. Med. Chem.* **2014**, *57*, 8167−8179.
(19) Perry, E. K.; Perry, R. H.; Blessed, G.; Tomlinson, B. E. Changes in Brain Cholinesterases in Senile Dementia of Alzheimer Type. *Neuropathol. Appl. Neurobiol.* **1978**, *4*, 273−277.
(20) Giacobini, E. *Butyrylcholinesterase: Its Function and Inhibitors*; Martin Dunitz: London, 2003.
(21) Tripos Mol2 File Format. http://www.tripos.com/data/support/mol2.pdf (accessed Jan 7, 2015).
(22) 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.
(23) SYBYL Atom Types. http://www.tripos.com/mol2/atom_types.html (accessed Jan 14, 2015).
(24) Kepner, J.; Gilbert, J. *Graph Algorithms in the Language of Linear Algebra*; SIAM: Philadelphia, 2011.
(25) Lipinski, C. A. Drug-like Properties and the Causes of Poor Solubility and Poor Permeability. *J. Pharmacol. Toxicol. Methods* **2000**, *44*, 235−249.
(26) Shoichet, B. K. Interpreting Steep Dose-Response Curves in Early Inhibitor Discovery. *J. Med. Chem.* **2006**, *49*, 7274−7277.
(27) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53*, 2719−2740.
(28) Hawkins, P. C.; Nicholls, A. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *J. Chem. Inf. Model.* **2012**, *52*, 2919−2936.
(29) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572−584.

DOI: 10.1021/acs.jcim.5b00136
J. Chem. Inf. Model. 2015, 55, 1521−1528

(30) Ellman, G. L.; Courtney, K. D.; Andres, V., Jr.; Featherstone, R. M A New and Rapid Colorimetric Determination of Acetylcholinesterase Activity. *Biochem. Pharmacol.* **1961**, *7*, 88−95.
(31) Karaboga, A. S.; Petronin, F.; Marchetti, G.; Souchet, M.; Maigret, B. Benchmarking of HPCC: A Novel 3D Molecular Representation Combining Shape and Pharmacophoric Descriptors for Efficient Molecular Similarity Assessments. *J. Mol. Graphics Modell.* **2013**, *41*, 20−30.
(32) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.
(33) Vainio, M. J.; Puranen, J. S.; Johnson, M. S. ShaEP: Molecular Overlay Based on Shape and Electrostatic Potential. *J. Chem. Inf. Model.* **2009**, *49*, 492−502.
(34) Provost, F.; Fawcett, T.; Kohavi, R. The Case Against Accuracy Estimation for Comparing Induction Algorithms. In *Fifteenth International Conference on Machine Learning*, Madison, Wisconsin, U.S.A., 1998; The International Machine Learning Society: Princeton, NJ, 1998; pp 445−553.
(35) Vainio, M. J.; Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462−2474.
(36) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(37) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening−an Overview. *Drug Discovery Today* **1998**, *3*, 160−178.
(38) Abad-Zapatero, C.; Metz, J. T. Ligand Efficiency Indices as Guideposts for Drug Discovery. *Drug Discovery Today* **2005**, *10*, 464−469.
(39) Gallivan, J. P.; Dougherty, D. A. Cation-Pi Interactions in Structural Biology. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 9459−9464.

DOI: 10.1021/acs.jcim.5b00136
J. Chem. Inf. Model. 2015, 55, 1521−1528