JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

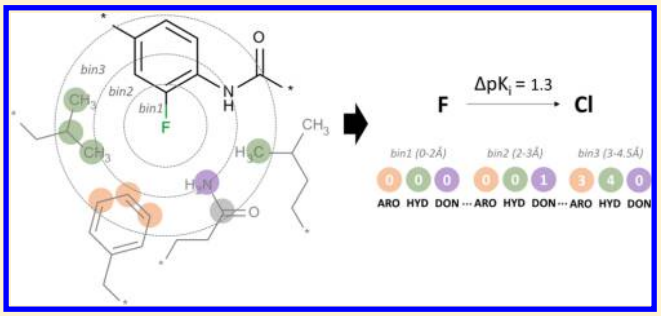
# VAMMPIRE-LORD: A Web Server for Straightforward Lead Optimization Using Matched Molecular Pairs

Julia Weber, Janosch Achenbach, Daniel Moser, and Ewgenij Proschak*

Institute of Pharmaceutical Chemistry, Goethe University, Frankfurt 60438, Germany

Supporting Information

**ABSTRACT:** VAMMPIRE-LORD (lead optimization by rational design) describes an innovative strategy to improve the binding affinity of a defined lead compound using 3D matched molecular pairs (3D-MMPs). 3D-MMPs are defined as pairs of molecules that differ in exactly one structural transformation and have a known bioactive conformation. We developed a novel atom-pair descriptor (LORD_FP) that represents the ligand—as well as the receptor environment—of a chemical transformation and built a predictive model based on 17 602 3D-MMPs. We demonstrate that the created model is able to extrapolate the knowledge of a chemical transformation and the associated effect on ligand affinity to any similar system. VAMMPIRE-LORD was implemented as a web server that guides the user step-by-step through the optimization process of a defined lead compound.




## INTRODUCTION

Small structural transformations within a molecule, along with the effect on an arbitrary molecular property sustain valuable information. Statistical models derived from these observations have been used to rationally improve lead compounds regarding a range of particular properties like solubility, plasma protein binding, or oral exposure.[1−3] The so-called matched molecular pairs (MMPs) have gained increasing interest in recent years not only in terms of optimizing physicochemical properties but also with regard to improving the binding affinity of a lead compound to a specific target.[4−7] 2D-MMP methods have been developed to identify transformations improving the binding affinity of a ligand more frequently than the average,[5,8] and it has recently been shown that considering the local environment of a transformation within the ligand can improve the predictive power of 2D-MMPs.[7,9,10] Nevertheless most of the 2D-MMP methods are designed for the characterization of specific targets or target families with a large number of known ligands and the resulting statistics are not transferable to any other target or target family. There are very few methods incorporating the protein environment into MMP calculations.[11] One of these methods is the OOMMPPAA tool,[12] a 3D-MMP method which identifies favorable positions for pharmacophoric features within a specific protein. OOMMPPAA is able to identify promising suggestions for the synthesis of novel compounds which arise from a combined analysis of structural and activity data of known ligands. However, a comprehensive data set of ligands with activity data for the considered target is necessary to generate an adequate number of MMPs for the prediction.

In order to create a target-independent data set we implemented the VAMMPIRE database,[13] a collection of 3D-MMPs in receptor context which enables us to compare the immediate amino acid environment of an arbitrary transformation independently from the rest of the protein structure. In the following we present VAMMPIRE-LORD, a Web server for rational lead optimization, which is based on the VAMMPIRE database and follows the principle that substitutions in similar chemical environments cause similar effects on the ligand affinity. We implemented a novel method to translate the context of a substitution, including the ligand- as well as the receptor environment, into an atom-pair fingerprint (LORD_FP). LORD_FP is a descriptor based on topological atom-pairs inspired by CATS3D[14] and comprises protein−ligand and protein−protein as well as ligand−ligand atom pairs. We extended VAMMPIRE database by the LORD_FP which serves as a basis for the prediction of promising substitutions on particular lead compounds. VAMMPIRE-LORD was implemented as an easy to use wizard which supports the user step-by-step through the process of lead optimization and presents the results in a 3D viewer together with the corresponding substitution information. The Web server is freely available at http://vammpire.pharmchem.uni-frankfurt.de.

## RESULTS

We obtained 17 602 MMPs which are deposited in the current version of VAMMPIRE database. The entries are subdivided into three different MMP-types (Figure 1) depending on their reliability. Type-I-MMPs represent about 10% of the database and are of the highest quality as both molecules are available as cocrystallized structures in the PDBbind.[15,16] In a Type-II-

Received: August 30, 2014
Published: January 28, 2015





DOI: 10.1021/ci5005256
J. Chem. Inf. Model. 2015, 55, 207−213

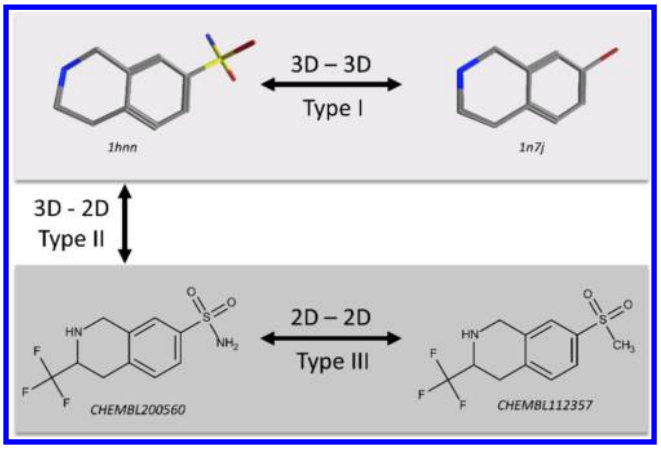



**Figure 1.** Definition of the MMP types: (Type-I-MMP) both molecules are available as cocrystallized structures in the PDBbind; (Type-II-MMP) one of the molecules is available as cocrystallized structure and serves as a basis for the prediction of the second (2D) molecule; (Type-III-MMP) both molecules are available in 2D only. The predicted coordinates of a molecule within a Type-II-MMP forms the basis for a second prediction.

MMP (20% of the database) the bioactive conformation is known for one of the molecules which forms the basis for the prediction of the conformation of the second molecule. We optimized the procedure to predict the bioactive conformation using MOE[17] (Molecular Operating Environment) docking with pharmacophore placement instead of aligning and energy minimizing the molecules (Figure 2).

In a recently published comparative assessment of docking and scoring functions,[18] it has been shown that most of the scoring functions are able to the predict 60−90% of the bioactive conformations within the top 3 ranked poses. In a previous study,[19] we could also show that docking into a cocrystallized structure containing a similar ligand improves the quality of the docking results. Therefore, molecular docking is in our opinion the most suitable method for the prediction of the bioactive conformation and at the same time serves as a filter for MMPs that most likely have different binding modes. For further improvement of the quality of 3D-MMPs we calculated the root mean square deviation between the common core atoms of the two molecules after the docking placement (core-RMSD). All MMPs with a core-RMSD greater than 1 Å were rejected. The third type of MMPs, the Type-III-MMPs, represent the major part of the database (about 70%) and are of the lowest quality as both molecules originate from ChEMBL database and are not available as cocrystallized structures. Nevertheless we accepted these pairs as Type-III-MMPs if one of the molecules was at the same time part of a Type-II-MMP and therefore had a predicted bioactive conformation. We applied the filter procedure to improve the data quality as we did for Type-II-MMPs.

The chemical environment of a substituent is defined as all non-hydrogen receptor atoms as well as all non-hydrogen ligand atoms located within a 4.5 Å radius around each non-hydrogen atom of the substituent. In case the substituent is a hydrogen atom itself, the radius is taken around its coordinates. A pharmacophore type is thereafter assigned to each atom of the receptor environment (a full listing of the atom types is given in the Supporting Information) and electrotopological state (EState) atom types[20] are assigned to each atom of the ligand environment. The atoms of the substituent itself are also typed using EState atom types. After the chemical environment is defined the LORD_FP is calculated combining ligand−ligand interactions (LLI), protein−ligand interactions (PLI), and protein−protein interactions (PPI). Interactions are defined as atom pairs (bonded or nonbonded) and their distances binned into short distance interactions (≤2 Å), medium distance interactions (>2−3 Å), and long distance interactions (>3−4.5 Å). LLIs and PPIs represent the atom type arrangement within the ligand respectively the protein environment while the PLIs describe the atom type distances between the substituent and receptor atoms. All binned atom type pairs are counted and lead to a numerical descriptor. Taking the example of PLIs, Figure 3 shows a depiction of how the SMART_FP is calculated.

For validation of LORD_FP and within the VAMMPIRE-LORD user interface we use a numeric version of the Tanimoto coefficient to compare two descriptors A and B (minimum similarity = 0, maximum similarity = 1):

$$T(A, B) = \frac{\sum_{i=1}^{n} \mathrm{common}(A_i, B_i)}{\sum_{i=1}^{n} A_i + \sum_{i=1}^{n} B_i - \sum_{i=1}^{n} \mathrm{common}(A_i, B_i)} \quad (1)$$

where $\mathrm{common}(A_i, B_i)$ is the number of common atom types in descriptor $A$ and $B$ at position $i$ and $A_i$ and $B_i$ represent the number of atom types in descriptor $A$ and $B$ at position $i$.

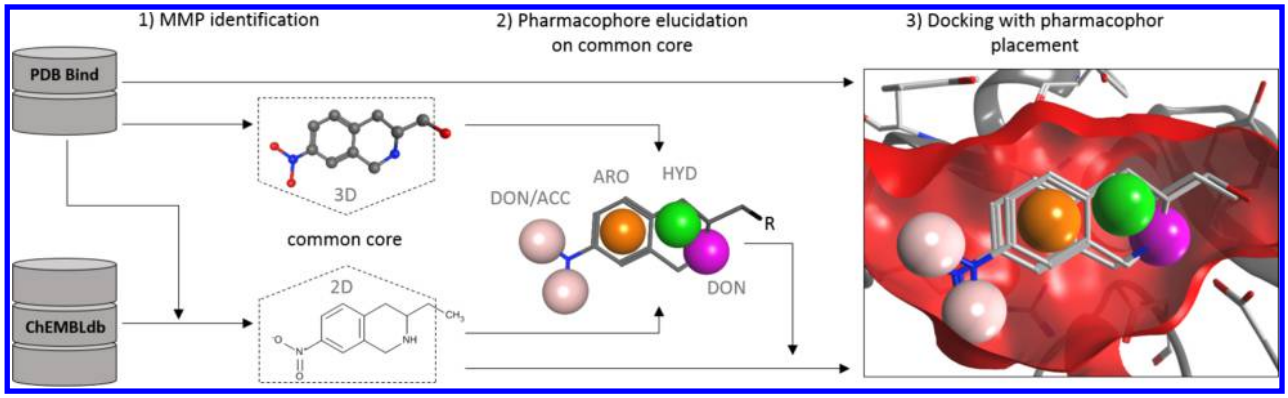



**Figure 2.** Strategy to generate Type-II-MMPs. (1) Identification of MMPs between ligands from PDBbind and ChEMBL database. (2) Extraction of the molecules common core and pharmacophore annotation on the basis of the 3D conformation of the PDBbind ligand. (3) Docking with pharmacophore placement.


DOI: 10.1021/ci5005256
J. Chem. Inf. Model. 2015, 55, 207−213

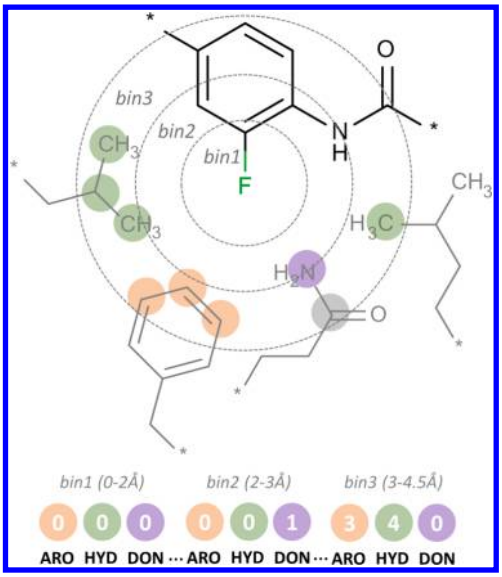



**Figure 3.** Representation of protein−ligand interactions (PLIs) in LORD_FP. A simplified depiction of a substituent in its protein environment is shown. Three distance bins (2, 3, and 4.5 Å), represented as dotted circles, are formed with their centers on the fluorine substituent (EState atom type: sF). The surrounding amino acids (side chains) are depicted in gray. The pharmacophore types assigned to the side chain atoms are marked as filled circles (hydrophobic: green, aromatic: orange, polar: gray, H-bond donor: violet). An extract of the resulting descriptor is shown below. The number of atom pairs formed between sF and the pharmacophore types within their respective bin form the numerical descriptor.

To provide an interface for medicinal chemists we implemented VAMMPIRE-LORD in the form of a wizard that guides the user step-by-step through the process of lead optimization (Figure 4).

During the steps of the wizard, the user is asked to define the target and the ligand, as well as the substituent of interest. The VAMMPIRE database will subsequently be searched for MMPs containing the desired substituent in a similar chemical environment (SMART_FP similarity ≥ threshold). The results are presented as a table storing the information about the ligands and targets (ChEMBL IDs and PDB codes), the substituents (SMILES), the publications (pubmed IDs), the type of affinity values ($K_i$, $K_d$, $IC_{50}$), the substitution effect, and the descriptor similarity. Additionally a plot, where each data point represents one MMP (Δaffinity plotted against similarity of the substitution environment), gives an overview of the results. Green dots located in the upper right corner represent MMPs with highest positive effects and at the same time highest atom type similarities. By selecting an MMP from the results page the corresponding molecules will appear in the 3D viewer in the context of the receptor environment. Additionally the composition of the receptor atom types and the amino acids surrounding the substituent can be visually inspected.

We were confident that a particular substitution observed for two different targets in a similar chemical environment (in terms of the Tanimoto coefficient between their LORD_FP descriptors) has a similar effect on ligand affinity. Therefore, we calculated the LORD_FP similarity for each target pair with different Enzyme Commission numbers (EC numbers) stored together with the same substitution. The overall distribution of similarity values is shown in Figure 5.

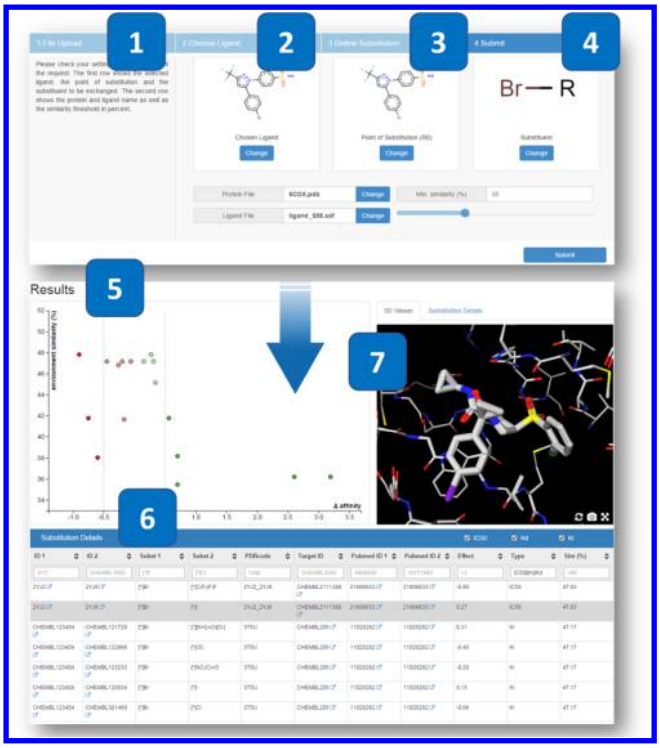



**Figure 4.** VAMMPIRE-LORD wizard. (1) Input of a PDB code or file upload. (2) Selection of the lead compound. (3) Substituent selection. (4) Summary and definition of a similarity threshold. (5) Plot showing all MMPs that match the query (Δaffinity plotted against similarity in percent). Positive effects are shown in green, and negative effects are shown in in red. (6) Table with details about molecules, substituents, targets, and publications. (7) 3D viewer showing both molecules of an MMP in the context of the receptor environment.

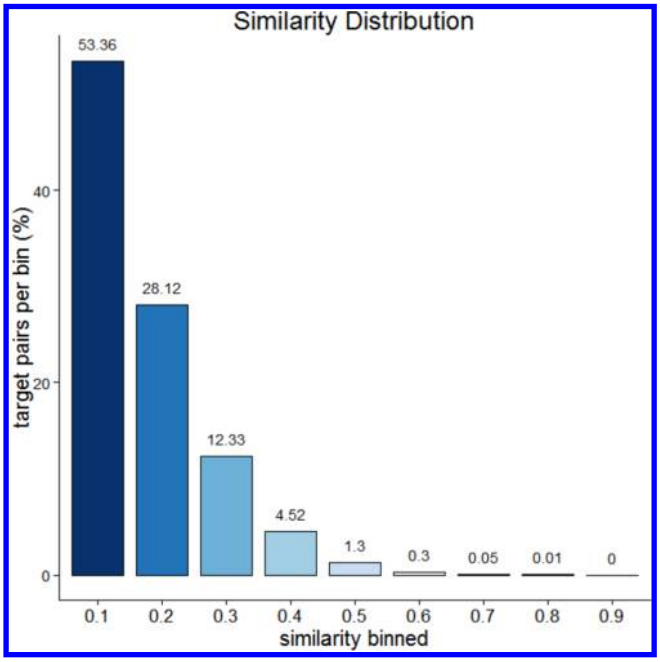



**Figure 5.** LORD_FP similarity distribution for all target pairs stored in VAMMPIRE database. The data is binned into nine groups representing the number of target pairs with a LORD_FP similarity between 0 and 1 (in steps of 0.1).

To prevent that overrepresented target pairs dominate the statistical evaluation we picked random samples for each target


DOI: 10.1021/ci5005256
J. Chem. Inf. Model. 2015, 55, 207−213

pair and repeated the calculation 20 times. We then plotted the number of matching substitution effects, which is the number of target pairs where both effects are either positive or negative, depending on the LORD_FP similarity (Figure 6). We expected the number of matches to increase with increasing similarity as well as with the strength of the substitution effects on both targets. In order to demonstrate this, we calculated the number of matches for three subsets: one including all target pairs, one including those pairs with an effect of at least 0.5 log units, and one with an effect of at least 1 log unit. Each of these subsets was subsequently analyzed regarding five defined thresholds (0, 0.3, 0.4, 0.5, 0.6) which appeared appropriate after visual inspection of the overall similarity distribution.

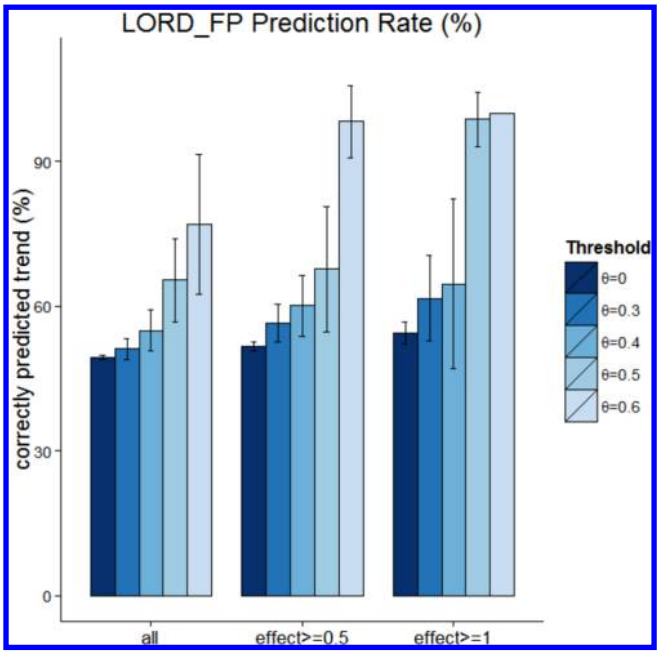



**Figure 6.** LORD_FP prediction rate. The number of target pairs where both effects are either positive or negative are shown in dependency of the LORD_FP similarity. Three subsets are shown (all: all target pairs, effect ≥ 0.5: target pairs with an effect of at least 0.5 log units, effect ≥ 1: target pairs with an effect of at least 1 log unit). The threshold ($\theta$) gives the minimum similarity of a target pair to be included in the statistics.

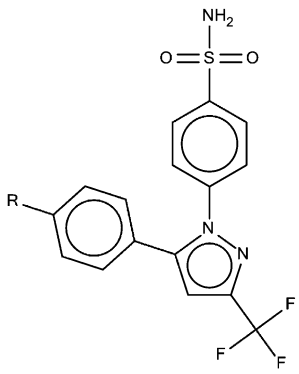
The intrinsic validation of the LORD_FP on VAMMPIRE database shows that the number of correctly predicted trends is highly dependent on the similarity as well as on the strength of the substitution effects. However, it should be noted that the subset of substitutions with effects greater than 1 order of magnitude (for both targets) is comparatively small. In each of the 20 iterations an average of 3655 target pairs are detected of which 21 have a similarity of at least 0.5 and only 5 have a similarity of at least 0.6.

To demonstrate the usefulness of VAMMPIRE-LORD we reproduced an extract of the SAR (structure activity relationship) of cyclooxygenase-2 (COX-2) inhibitors using the SAR of Celecoxib derivatives as a reference[21] (Table 1). This example is especially interesting because the different substituents cover 6 orders of magnitude in inhibitory activity.

All molecules where docked using MOE docking to predict the bioactive conformations (results can be found in the Supporting Information). For each substituent R the LORD_FP was calculated and VAMMPIRE database was then searched for substitutions containing R and one of the other substituents of the COX-2 SAR. Only those substitutions with an effect of at least 0.5 log units and a LORD_FP similarity of at least 0.5 were considered since a mean prediction rate of almost 70% is expected (according to Figure 6). As a result we got a directed substitution network with green and red edges representing matches and mismatches respectively (Figure 7). Eighty-three substitutions with an effect of at least 0.5 log units were found in VAMMPIRE database of which 25 had a LORD_FP similarity of at least 0.5. Eighteen of the 25 substitutions matched the effects of the COX-2 SAR while 7 substitutions were mismatches which corresponds to the expected prediction rate.

The environment with the highest similarity (0.63) was found for the substitution of $-CF_3$ to $-Cl$ together with a positive effect on ligand affinity on both targets (Figure 8). The associated protein is the Endothelial PAS domain-containing protein 1 (EPAS1) and is a result of a Type-III-MMP (CHEMBL2311960 → CHEMBL2311959) placed into the crystal structure with the ligand *N*-(3-chloro-5-fluorophenyl)-4-nitro-2,1,3-benzoxadiazol-5-amine (PDB code: 4GHI). The

**Table 1. SAR of Celecoxib Derivatives Forming MMPs**

| | R-group [a] | $IC_{50}$ (μM) [b] | R-group [a] | $IC_{50}$ (μM) [b] |
|---|---|---|---|---|
|  | $-NMe_2$ | 0.005 | -NH2 | 0.34 |
| | -OMe | 0.008 | -OEt | 0.64 |
| | -SMe | 0.009 | -Et | 0.86 |
| | -Cl | 0.010 | $-NO_2$ | 2.63 |
| | -NHMe | 0.016 | $-CF_3$ | 8.23 |
| | -H | 0.032 | $-CO_2H$ | 11.2 |
| | -Me | 0.040 | $-CH_2OH$ | 93.3 |
| | -F | 0.041 | -OH | >100 |

[a]Substitutions at R-group. [b]$IC_{50}$ values determined in a recombinant human COX-2 assay.[21]



DOI: 10.1021/ci5005256
J. Chem. Inf. Model. 2015, 55, 207−213

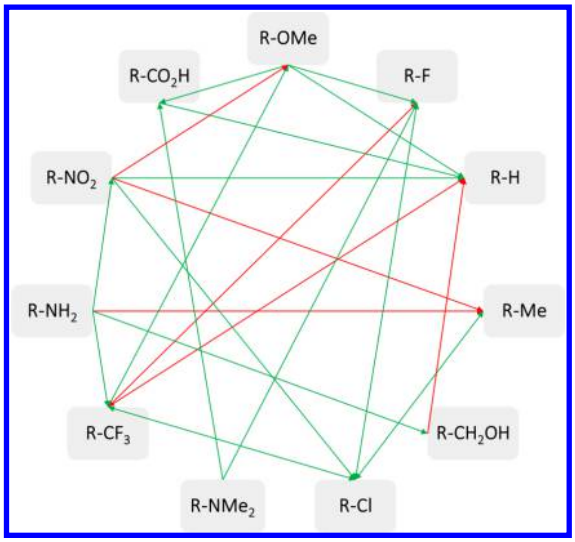



**Figure 7.** Substitution network representing the results of the VAMMPIRE-LORD search algorithm. All substitutions within the COX-2 SAR (Table 1) were searched in the VAMMPIRE database (effect of at least 0.5 log units and a LORD_FP similarity of at least 0.5). The VAMMPIRE substitutions found are represented as directed edges and colored green when the effect matched the COX-2 SAR and red if the effect did not match.

ligand environment is aromatic in both cases, moreover the amino acids around the substitution are mostly aromatic and hydrophobic. In both cases methionine, phenylalanine, and tyrosine are part of the chemical environment. Although the arrangement and the entire composition of the amino acids is different, the LORD_FP fingerprint similarity is high. Considering the sequence identity of the two proteins (4.3%) this is a good example for the generalizing power of the LORD_FP approach.

## DISCUSSION AND CONCLUSION

VAMMPIRE-LORD was implemented as a tool for straightforward lead optimization built up on the idea to use experimental affinity data to create a target-independent model. Based on the assumption that substitutions cause similar effects in similar chemical environments, VAMMPIRE-LORD can be very useful as it is indicating appropriate substitution options. We were able to show that a target independent model to predict a substitution effect (in terms of a trend) is possible by only taking the immediate chemical environment of a substitution into account. With a prediction rate of 70% we could show in the example of the COX-2 SAR that substitutions in similar environments cause similar effects even when the proteins are not closely related. A limiting factor of VAMMPIRE-LORD is certainly the number of MMPs as well as the protein diversity. A simplification of the substitutions like it was done in the recently published Fuzzy matched pairs approach[10] could significantly increase the number of MMPs and may lead to a wider choice of suggestions generated by VAMMPIRE-LORD. However, the permanent increase in the number of protein−ligand complexes in the Protein Data Bank (PDB)[22] as well as the affinity data in ChEMBL database[23] will contribute to future enhancement of the prediction power of this tool.

The confidence of the experimental data deposited in the ChEMBL database is an additional unavoidable problem. Translation or assignment errors of the automatically extracted experimental data are not uncommon.[24,25] Furthermore, the comparison of affinity values determined in different laboratories and assays is far from ideal, especially regarding the small substitution effects. The evaluation of the proposed results by an expert is therefore reasonable to whom VAMMPIRE-LORD can serve as a valuable idea generator for structure-based drug design.

## MATRIALS AND METHODS

**Database Preparation.** We optimized the procedure building VAMMPIRE database to receive a more trustworthy prediction of the 3D conformations. The latest version of VAMMPIRE database was created using the following strategy: The PDBbind v2014, a comprehensive data set of protein−ligand complexes with annotated affinity data obtained from the PDB was the starting point for our data processing. The ChEMBL database v19 expanded the knowledge base by 277 964 compound records with affinity values annotated for the targets deposited in the PDBbind. All procedures were implemented as automatized workflows within the data mining tool KNIME.[26] The MMP identification was carried out using the *Matched Pairs Detector*[7,27,28] available as a node (provided by Erl Wood Cheminformatics) in KNIME. MMPs were detected between the molecules stored in the PDBbind and those molecules stored in the ChEMBL database that have affinity data ($K_i$, $K_d$, and $IC_{50}$ values) assigned to one of the PDBbind targets. We implemented a few restriction rules for MMPs including the maximum size of a substituent to be limited to nine non-hydrogen atoms and the molecules common core to be twice as big as the substituents.

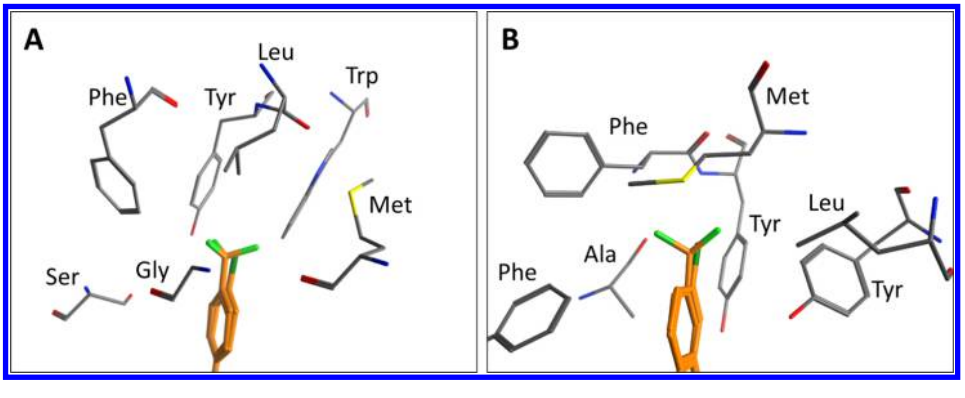



**Figure 8.** Comparison of the chemical environments of the substitution of $-CF_3$ to $-Cl$ in within COX-2 (A) and EPAS1 (B). All residues within a 4.5 Å radius around the $-CF_3$ group were selected for visualization.


DOI: 10.1021/ci5005256
J. Chem. Inf. Model. 2015, 55, 207−213

Additionally the measured affinity data within an MMP had to be of the same experimental type whereby MMPs with annotated $IC_{50}$ values were only accepted, if they were obtained from the same publication. In case more than one affinity value was stored for a molecule to a specific target the median value was taken. If a value was redundantly represented in the ChEMBL database we took the earliest publication as a reference. The substitution effect was then calculated by subtraction of the $\log_{10}$ affinity values (e.g., an effect of 1.0 represents an increase of ligand affinity by 1 order of magnitude).

**3D-MMP Generation.** The common core of an MMP is provided by the *Matched Pairs Detector* in terms of a 2D molecule representation. To create a pharmacophore model on the 3D coordinates of the common core, a mapping was performed using the Small Molecule Subgraph Detector (SMSD) toolkit.[29] The resulting 3D-common core was then typed by the ph4 annotation function ("Unified" scheme) provided by MOE as SVL snippet. The annotation points were then translated to a pharmacophore model by adding a 1 Å radius to the annotation points. The created pharmacophore model then was used as a placement method within MOE docking using *Pharmacophore* as a placement method within the *Docking Placement* node implemented in KNIME. Ten poses were created and refined using the *Pose Refinement* node which is also provided by MOE. The grid based minimization with default setup was used for the refinement. The pose with the smallest Root Mean Square Deviation (RMSD) between the common core of the placed molecule and the template was selected as "best pose" for further processing.

**Web Server.** The web server was implemented in Python using Flask version 0.10.1 (http://flask.pocoo.org/). Client-side 3D visualization is done with a custom GL mol (0.47) build (http://www.glmol.com/). As a database server PostgreSQL 9.3 (http://www.postgresql.org/) with the RDKit Cartridge is used (http://www.rdkit.org). The RDKit is also employed in server-side calculations. Client-side structure input is handled by Marvin4JS (http://www.chemaxon.com/).

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

Assignment of the pharmacophore types and the docking procedure for the COX-2 SAR. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: proschak@pharmchem.uni-frankfurt.de. Mailing address: Institute of Pharmaceutical Chemistry, Goethe University, Max-von-Laue-Str.9, D-60438 Frankfurt am Main.



### Funding

This research was supported by the Deutsche Forschungsgemeinschaft (DFG, Sachbeihilfe PR 1405/2-1, SFB 1039 A07) and LOEWE-Schwerpunkt: Anwendungsorientierte Arzneimittelforschung.


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS


J.W. thanks the Beilstein-Institut zur Förderung der Chemischen Wissenschaften (Beilstein Institute for the Advancement of Chemical Sciences), J.A., the Else Kröner Fresenius Foundation (EKFS), Research Training Group Translational Research Innovation−Pharma (TRIP), and D.M., the Deutsches Konsortium für Translationale Krebsforschung (DKTK) for a fellowship.


## ■ REFERENCES


(1) Leach, A. G.; Jones, H. D.; Cosgrove, D. A.; Kenny, P. W.; Ruston, L.; MacFaul, P.; Wood, J. M.; Colclough, N.; Law, B. Matched Molecular Pairs as a Guide in the Optimization of Pharmaceutical Properties; a Study of Aqueous Solubility, Plasma Protein Binding and Oral Exposure. *J. Med. Chem.* **2006**, *49*, 6672−6682.

(2) Griffen, E.; Leach, A.; Robb, G.; Warner, D. Matched Molecular Pairs as a Medicinal Chemistry Tool. *J. Med. Chem.* **2011**, *54*, 7739−7750.

(3) Dossetter, A. G.; Griffen, E. J.; Leach, A. G. Matched Molecular Pair Analysis in Drug Discovery. *Drug Discovery Today* **2013**, *18*, 724−731.

(4) O'Boyle, N. M.; Boström, J.; Sayle, R. a; Gill, A. Using Matched Molecular Series as a Predictive Tool to Optimize Biological Activity. *J. Med. Chem.* **2014**, *57*, 2704−2713.

(5) Wassermann, A. M.; Bajorath, J. Large-Scale Exploration of Bioisosteric Replacements on the Basis of Matched Molecular Pairs. *Future Med. Chem.* **2011**, *3*, 425−436.

(6) Posy, S. L.; Claus, B. L.; Pokross, M. E.; Johnson, S. R. 3D Matched Pairs: Integrating Ligand- and Structure-Based Knowledge for Ligand Design and Receptor Annotation. *J. Chem. Inf. Model.* **2013**, *53*, 1576−1588.

(7) Papadatos, G.; Alkarouri, M.; Gillet, V. J.; Willett, P.; Kadirkamanathan, V.; Luscombe, C. N.; Bravi, G.; Richmond, N. J.; Pickett, S. D.; Hussain, J.; Pritchard, J. M.; Cooper, A. W. J.; Macdonald, S. J. F. Lead Optimization Using Matched Molecular Pairs: Inclusion of Contextual Information for Enhanced Prediction of HERG Inhibition, Solubility, and Lipophilicity. *J. Chem. Inf. Model.* **2010**, *50*, 1872−1886.

(8) Wirth, M.; Zoete, V.; Michielin, O.; Sauer, W. H. B. SwissBioisostere: A Database of Molecular Replacements for Ligand Design. *Nucleic Acids Res.* **2013**, *41*, 1137−1143.

(9) Warner, D. J.; Griffen, E. J.; St-Gallay, S. A. WizePairZ: A Novel Algorithm to Identify, Encode, and Exploit Matched Molecular Pairs with Unspecified Cores in Medicinal Chemistry. *J. Chem. Inf. Model.* **2010**, *50*, 1350−1357.

(10) Geppert, T.; Beck, B. Fuzzy Matched Pairs: A Means to Determine the Pharmacophore Impact on Molecular Interaction. *J. Chem. Inf. Model.* **2014**, *54*, 1093−1102.

(11) Posy, S. L.; Claus, B. L.; Pokross, M. E.; Johnson, S. R. 3D Matched Pairs: Integrating Ligand- and Structure-Based Knowledge for Ligand Design and Receptor Annotation. *J. Chem. Inf. Model.* **2013**, *53*, 1576−1588.

(12) Bradley, A. R.; Wall, I. D.; Green, D. V. S.; Deane, C. M.; Marsden, B. D. OOMMPPAA: A Tool To Aid Directed Synthesis by the Combined Analysis of Activity and Structural Data. *J. Chem. Inf. Model.* **2014**, *54*, 2636−2646.

(13) Weber, J.; Achenbach, J.; Moser, D.; Proschak, E. VAMMPIRE: A Matched Molecular Pairs Database for Structure-Based Drug Design and Optimization. *J. Med. Chem.* **2013**, *56*, 5203−5207.

(14) Fechner, U.; Franke, L.; Renner, S.; Schneider, P.; Schneider, G. Comparison of Correlation Vector Methods for Ligand-Based Similarity Searching. *J. Comput. Aided. Mol. Des.* **2003**, *17*, 687−698.

(15) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein-Ligand Complexes with Known Three-Dimensional Structures. *J. Med. Chem.* **2004**, *47*, 1675−1679.

(16) Wang, R.; Fang, X.; Lu, Y.; Yang, C.-Y.; Wang, S. The PDBbind Database: Methodologies and Updates. *J. Med. Chem.* **2005**, *48*, 4111−4119.

(17) Chemical Computing Group Inc. *Molecular Operating Environment (MOE), 2013.08*, 2013.




DOI: 10.1021/ci5005256
*J. Chem. Inf. Model.* 2015, 55, 207−213

(18) Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results. *J. Chem. Inf. Model.* **2014**, *54*, 1717−1736.
(19) Weber, J.; Rupp, M.; Proschak, E. Impact of X-Ray Structure on Predictivity of Scoring Functions: PPARγ Case Study. *Mol. Inform.* **2012**, *31*, 631−633.
(20) Hall, L. H.; Kier, L. B. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *J. Chem. Inf. Model.* **1995**, *35*, 1039−1045.
(21) Penning, T. D.; Talley, J. J.; Bertenshaw, S. R.; Carter, J. S.; Collins, P. W.; Docter, S.; Graneto, M. J.; Lee, L. F.; Malecha, J. W.; Miyashiro, J. M.; Rogers, R. S.; Rogier, D. J.; Yu, S. S.; Anderson, G. D.; Burton, E. G.; Cogburn, J. N.; Gregory, S. A.; Koboldt, C. M.; Perkins, W. E.; Seibert, K.; Veenhuizen, A. W.; Zhang, Y. Y.; Isakson, P. C. Synthesis and Biological Evaluation of the 1,5-Diarylpyrazole Class of Cyclooxygenase-2 Inhibitors: Identification of 4-[5-(4-Methylphenyl)-3-(trifluoromethyl)-1H-Pyrazol-1-Yl]benze Nesulfonamide (SC-58635, Celecoxib). *J. Med. Chem.* **1997**, *40*, 1347−1365.
(22) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(23) 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.* **2011**, *44*, 1−8.
(24) Kramer, C.; Kalliokoski, T.; Gedeck, P.; Vulpetti, A. The Experimental Uncertainty of Heterogeneous Public K I Data. *J. Med. Chem.* **2012**, *55*, 5165−5173.
(25) Kalliokoski, T.; Kramer, C.; Vulpetti, A.; Gedeck, P. Comparability of Mixed $IC_{50}$ Data - a Statistical Analysis. *PLoS One* **2013**, *8*, e61007.
(26) Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. {KNIME}: The {K}onstanz {I}nformation {M}iner. In *Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007)*; Springer: Heidelberg-Berlin, 2007; pp 319−326.
(27) Wagener, M.; Lommerse, J. P. M. The Quest for Bioisosteric Replacements. *J. Chem. Inf. Model.* **2006**, *46*, 677−685.
(28) Hussain, J.; Rea, C. Computationally Efficient Algorithm to Identify Matched Molecular Pairs (MMPs) in Large Data Sets. *J. Chem. Inf. Model.* **2010**, *50*, 339−348.
(29) Rahman, S. A.; Bashton, M.; Holliday, G. L.; Schrader, R.; Thornton, J. M. Small Molecule Subgraph Detector (SMSD) Toolkit. *J. Cheminform.* **2009**, *1*, 12−25.


DOI: 10.1021/ci5005256
*J. Chem. Inf. Model.* 2015, 55, 207−213