JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

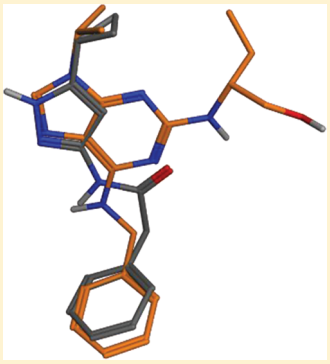
Design of Combinatorial Libraries for the Exploration of Virtual Hits from Fragment Space Searches with LoFT

Uta Lessel,*[,†] Bernd Wellenzohn,[†] J. Robert Fischer,[‡] and Matthias Rarey[‡]

[†]Department of Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG, Birkendorfer Str. 65, 88397 Biberach an der Riss, Germany

[‡]Center for Bioinformatics Hamburg, University of Hamburg, Bundesstr. 43, D-20146 Hamburg

ABSTRACT: A case study is presented illustrating the design of a focused CDK2 library. The scaffold of the library was detected by a feature trees search in a fragment space based on reactions from combinatorial chemistry. For the design the software LoFT (Library optimizer using Feature Trees) was used. The special feature called FTMatch was applied to restrict the parts of the queries where the reagents are permitted to match. This way a 3D scoring function could be simulated. Results were compared with alternative designs by GOLD docking and ROCS 3D alignments.



INTRODUCTION

At Boehringer Ingelheim fragment spaces based on reactions from combinatorial chemistry are successfully screened with feature trees[1,2] for the detection of new lead classes.[3] Once interesting scaffolds have been detected, corresponding combinatorial libraries have to be designed. At the beginning, reagent-based selections from partially enumerated sublibraries were performed:

- For each R-group a sublibrary was enumerated which contains all available and compatible reagents at one attachment point, while keeping one fixed reagent at all other R-groups.
- The partially enumerated sublibraries are scored, e.g., by docking or 3D alignments.
- Those reagents being part of the highest scoring products from each sublibrary are selected to build the final focused library.

In the meantime the software LoFT (Library optimizer using Feature Trees) was developed[4] allowing a product-based reagent selection. LoFT uses stochastic optimization, e.g., simulated annealing, which enables the design of focused combinatorial libraries optimized according to feature trees similarity to one or more queries and/or feature trees dissimilarity to one or more antiqueries. In addition, a desired range of physicochemical properties for library compounds can be taken into account. Recently two extensions of the basic program have been published: First, a modification of the feature tree descriptor allowing to distinguish different arene substitution patterns, and second, a special feature called FTMatch to restrict the parts of the queries where the reagents are permitted to match.[5] A case study was conducted to illustrate the usefulness of the latter feature. For this purpose, CDK2 was selected as the target. The LoFT results are compared with alternative designs coming from partial enumeration of the combinatorial libraries and ranking the reagents according to GOLD docking[6] and 3D alignments with ROCS.[7] Whereas the basic LoFT approach delivers products with 2D feature trees similarity to the query, the FTMatch feature helps to consider additional information, like the binding mode of the query in the selection.

METHODS

R-roscovitine with an IC_{50} of 450 nM[8] for the inhibition of CDK2 was used as a query for a feature trees fragment space search.[9] Its binding mode is shown in Figure 1. The NH group

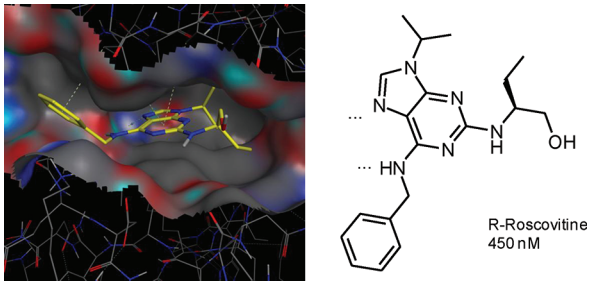


Figure 1. Binding mode of the CDK2 inhibitor R-roscovitine with an IC_{50} of 450 nM taken from PDB entry 2a4l.[8] The dotted lines in the 2D formula indicate the interactions of the ligand with the hinge region.

Special Issue: 2011 Noordwijkerhout Cheminformatics

Received: August 25, 2011
Published: December 12, 2011

 © 2011 American Chemical Society dx.doi.org/10.1021/ci2003957 | J. Chem. Inf. Model. 2012, 52, 373–379

of the benzyl amine substituent and the nitrogen in the 7-position of the purin scaffold bind to the hinge region.[8]

The hits from the feature tree fragment space search based on R-roscovitine were assessed in a postprocessing procedure.

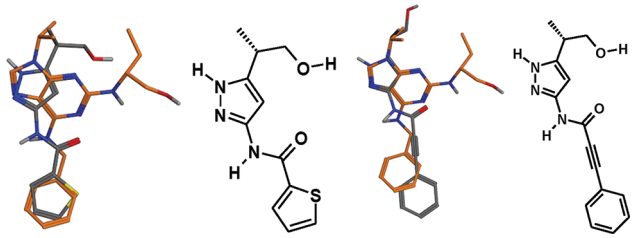


Figure 2. 3D alignments with ROCS[7] of two virtual hits from the feature trees fragment space search based on the query R-roscovitine.

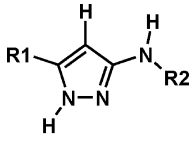


Figure 3. Scaffold of the virtual combinatorial aminopyrazole library.

For this purpose 200 OMEGA[10] conformers were generated for each hit using standard parameters. Then 3D alignments with the query were performed by ROCS.[7] The following selection of interesting scaffolds is a key step within the whole process of detecting potential new lead classes, but it is beyond the topic of this paper. It is only described to put the library design into the context of lead identification. Furthermore it is important because the 3D alignments seen during the visual inspection and the binding mode of the query were kept in mind for the design of the focused library.

In this case, two hits were detected by the feature trees fragment space search raising an interest in the amino-pyrazole class. The corresponding alignments are shown in Figure 2.

The virtual combinatorial library illustrated in Figure 3 was created from 2687 alkylhalogenides and 6368 carboxylic acids taken from external vendor catalogues. This virtual library contained 12 alkylhalogenides and 29 carboxylic acids derived from 35 active and 7 inactive aminopyrazoles described by Pevarello et al.[11] This information is only used for a plausibility check described later on.

For the reagent selection by GOLD docking[6] and by 3D alignments with ROCS,[7] two partially enumerated combinatorial aminopyrazole libraries were created: one with an isopropyl residue at R1 and all 6368 carboxylic acids and the other with a phenylacetyl moiety at R2 and all alkylhalogenides. The former

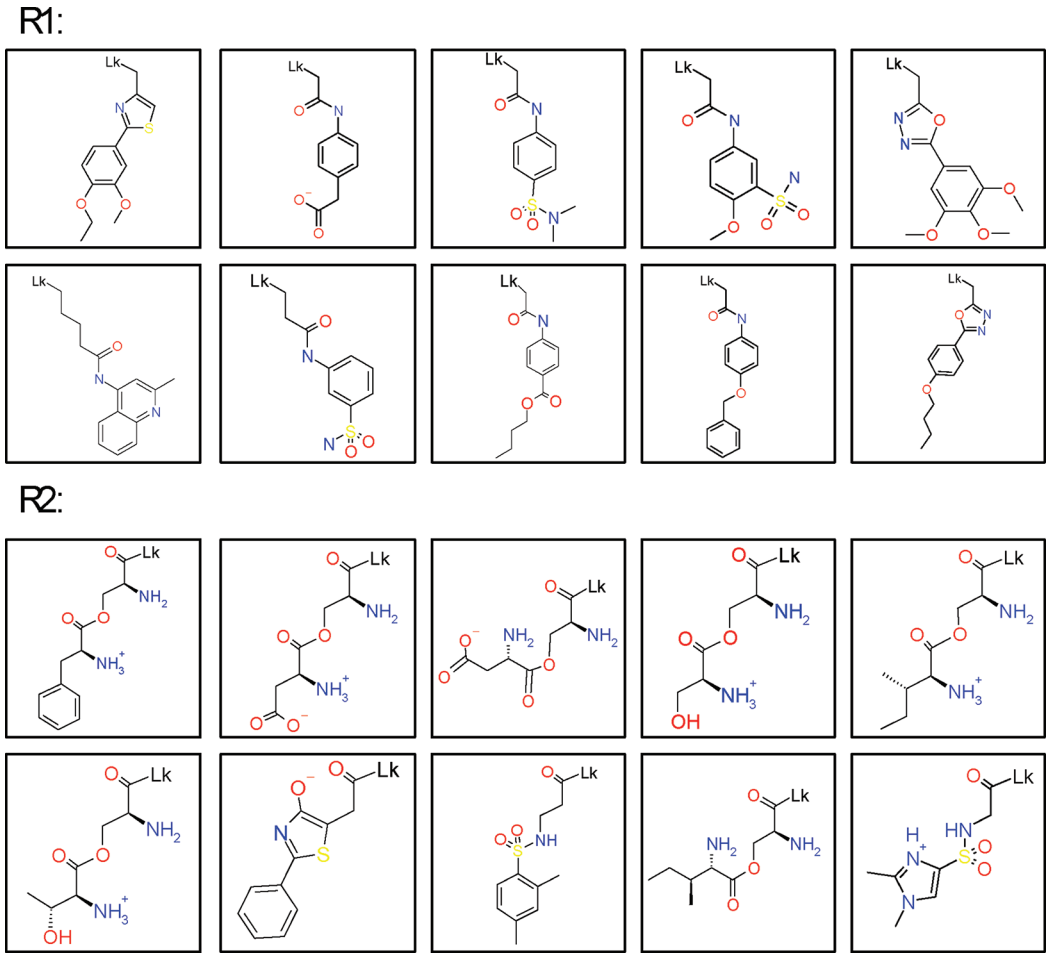


Figure 4. Reagents selected for the focused CDK2 library by GOLD docking.[6] The attachment points of the reagents to the aminopyrazole core are indicated by the symbol "Lk".

dx.doi.org/10.1021/ci2003957 | J. Chem. Inf. Model. 2012, 52, 373–379

library was used for the determination of residues R2 and the latter for the selection of the substituents at R1. For the 3D alignments with ROCS[6] 200 OMEGA[10] conformers were generated with standard parameters for each product of the partially enumerated combinatorial libraries.

For the focused library design with LoFT, a corresponding feature trees fragment space was created using CoLibri.[12] It contains the aminopyrazole core and all 6338 reagents R1 and 2687 reagents R2, respectively. Ten alkylhalogenides and 10 carboxylic acids were selected by each method for the final designs.

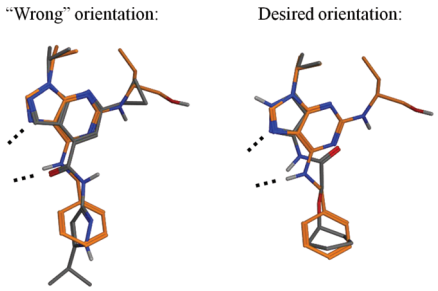


Figure 5. On the left side, a ROCS alignment of a product with the query is shown which does not allow binding to the hinge region (indicated by the dotted lines). The reagent selection was focused on products with an alignment allowing binding to the hinge region. An example is shown on the right side.

RESULTS

Selection by GOLD. Looking at the binding modes of the highest scoring products, it turned out that they do not show the desired interactions with the hinge region. Instead of selecting the first 10 compounds showing the correct binding mode, the docking run was repeated with a focus on docking poses with a similar hydrogen-bond donor and acceptor pattern, like the query (i.e., similarity constraints on the hydrogen-bond donor and acceptor patterns were applied during the GOLD run). The final design shown in Figure 4 includes the best 10 alkylhalogenides and 10 carboxylic acids from the constraint docking runs of the two partially enumerated libraries.

Selection by 3D Alignments with ROCS. The aligned products of both partially enumerated libraries were sorted according to Tanimoto combo score (shape and color field matching). As for the GOLD selection, some of the alignments do not allow binding to the hinge region. An example for a "wrong" orientation is shown in Figure 5.

The selection of the first 10 products with alignments allowing the desired binding mode for each R-group is shown in Figure 6.

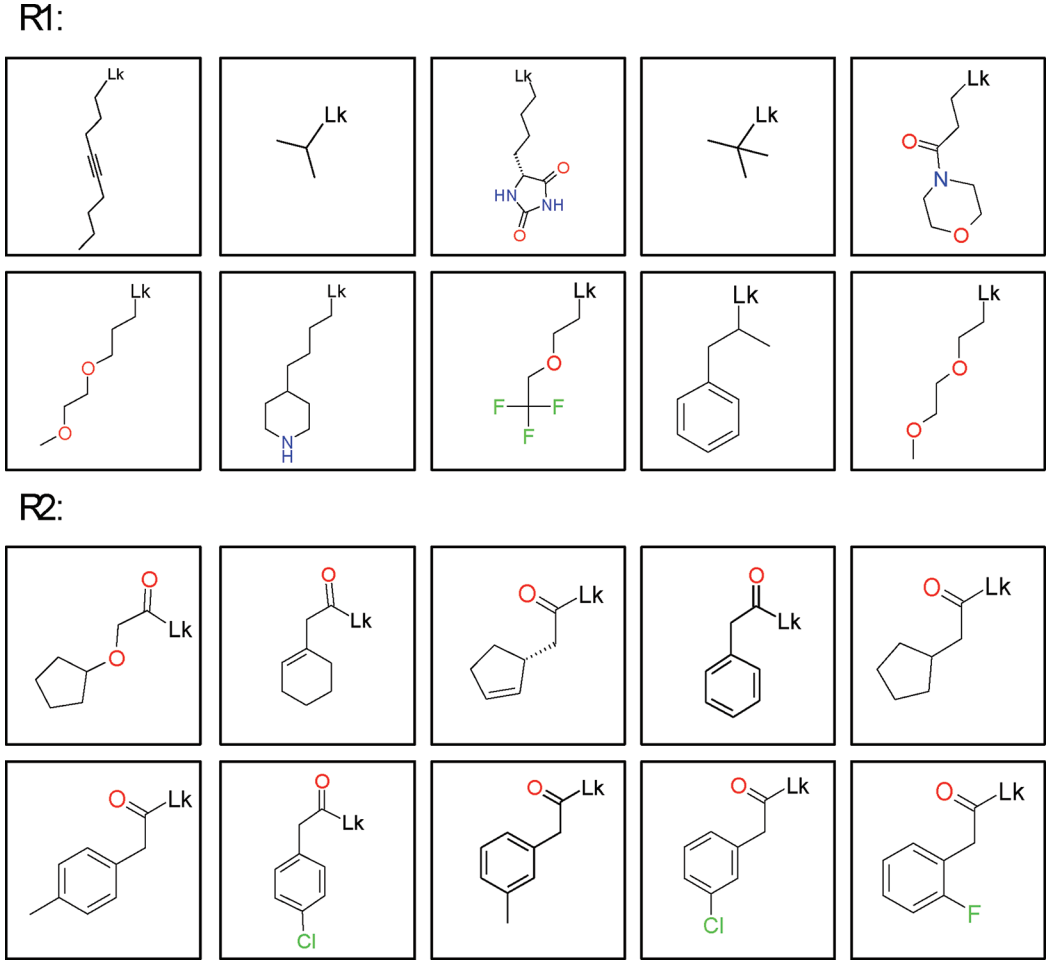


Figure 6. Reagents selected for the focused CDK2 library by 3D alignments with ROCS.[7] The attachment points of the reagents to the aminopyrazole core are indicated by the symbol "Lk".

dx.doi.org/10.1021/ci2003957 | J. Chem. Inf. Model. 2012, 52, 373–379

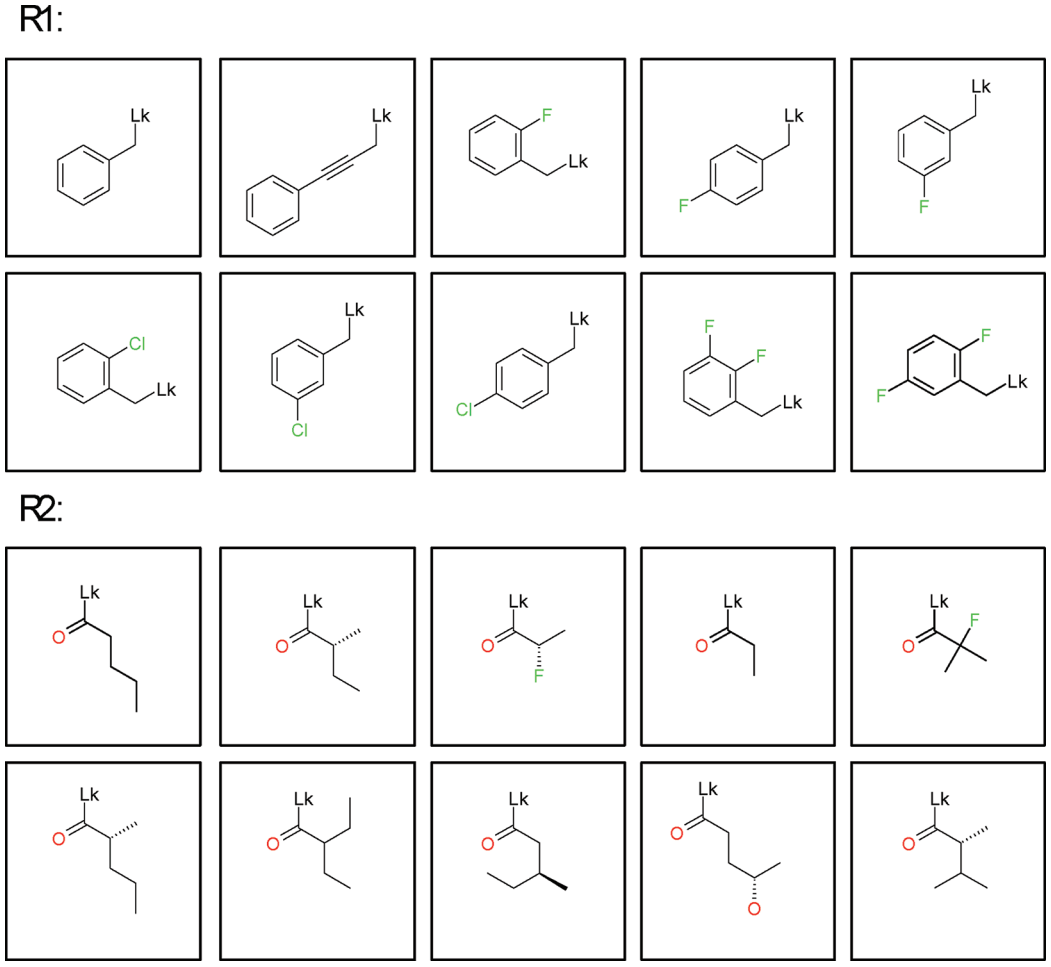


Figure 7. Reagents selected for the focused CDK2 library by LoFT optimizing the feature trees similarity to the query without any further constraint. The attachment points of the reagents to the aminopyrazole core are indicated by the symbol "Lk".

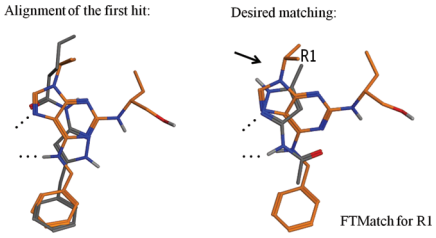


Figure 8. The 3D alignment of the first product from the first LoFT design and the desired matching of the aminopyrazole core with the query are shown. In the desired matching, the FTMatch restriction used for the next design cycle is indicated by the arrow. The dotted lines indicate the interactions of the ligand with the hinge region.

Selection Applying LoFT. In the first design run with LoFT only the feature trees similarity of the products to R-roscovitine was optimized. The reagent selection is shown in Figure 7.

A ROCS[7] 3D alignment of the products with roscovitine shows that they do not overlay in the desired binding mode. Consequently, the LoFT design was repeated making use of the FTMatch feature which allows to influence the alignment of the query and the products during the design process. In this case, first of all the bond to the reagents R1 was fixed to align with the bond to the substituent at the 9-position of the purin core. The alignment of the first product from the first LoFT design with the query and the desired alignment are illustrated in Figure 8.

The next design cycle with LoFT is aimed at optimizing the feature trees similarity to the query with the FTMatch restriction for R1. The results from this run are presented in Figure 9.

Now the residues at R1 resemble the isopropyl group of the query, but the residues R2 still differ from the intended corresponding part of the query. The ROCS[7] 3D alignment of the first product from this focused library demonstrates that the fit of the substituents R1 corresponds to the constraint set, but the lower part of the molecules is turned in a wrong direction. So, in this case it is necessary to introduce a second constraint for the bond to the substituents R2. The alignment of the first product from the second LoFT design with the query and the two FTMatch constraints set to get the desired alignment are illustrated in Figure 10.

Figure 11 shows the reagents selected by the next LoFT design cycle optimizing the feature trees similarity to the query with FTMatch restrictions for R1 and R2. The 3D alignments of the products from this library now correspond to the idea followed when selecting the scaffold (see Figure 12).

But the reagents selected are very similar to each other. If the idea is to broaden the aminopyrazole space which is explored, a diversity term can be included in the LoFT design. For this

dx.doi.org/10.1021/ci2003957 | J. Chem. Inf. Model. 2012, 52, 373–379

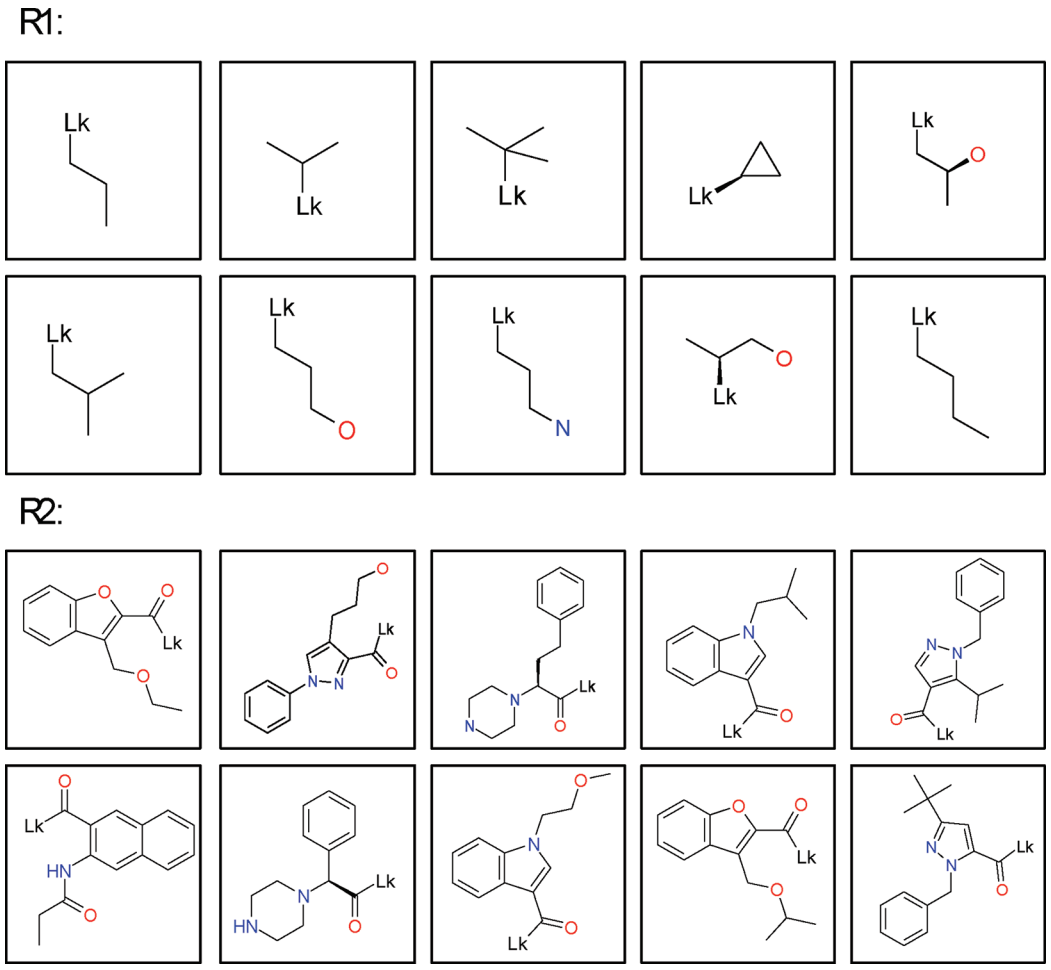


Figure 9. Reagents selected for the focused CDK2 library by LoFT optimizing the feature trees similarity to the query with an FTMatch constraint for R1. The attachment points of the reagents to the aminopyrazole core are indicated by the symbol "Lk".

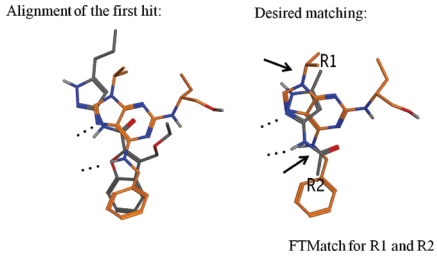


Figure 10. The 3D alignment of the first product from the second LoFT design and the desired matching of the aminopyrazole core with the query are shown. In the desired matching both FTMatch restrictions used for the next design cycle are indicated by the arrows. The dotted lines indicate the interactions of the ligand with the hinge region.

purpose the reagents were clustered with LoFT according to their feature trees similarity. Furthermore a diversity term was added to the scoring function penalizing a reagent if other members of its cluster are among the selected reagents. The corresponding library design is shown in Figure 13.

DISCUSSION

To evaluate the plausibility of the different focused library designs, it was checked how many of the building blocks leading to the 42 aminopyrazoles with known activity described by Pevarello et al[11] were selected.

In the design achieved by GOLD[6] docking, none of the building blocks was selected. The building blocks selected with the aid of GOLD[6] docking scores are much larger than the reagents, leading to the 42 known aminopyrazoles. Their scores are increased due to additional interactions compared to the building blocks being part of the known 42 aminopyrazoles. The latter are not found among the 500 top-scoring products from the partially enumerated libraries.

The reagent selection by ROCS[7] contained three of the building blocks, namely isopropyl and *tert*-butyl for R1 and phenylacetyl for R2. But the combinations of these building blocks have not been described, in other words it is not known whether they are active. Among the first 500 reagents R1 and R2 ranked by ROCS,[7] 8 alkylhalogenides and 4 carboxylic acids leading to the 42 known aminopyrazoles are detected.

In the LoFT design with two FTMatch restrictions, the first five reagents R1 in the upper row (see Figure 11) and phenylacetyl for R2 belong to the building blocks derived from the known actives. In addition to this, the combination of cyclopropyl at R1 and phenylacetyl has been described as a CDK2 inhibitor with an activity of 48 nM.[11]

This means that the focused library designed by LoFT contains at least one known active. On the one hand, the hit rate of 1% seems to be rather small. On the other hand, it has to be taken into account that the whole fragment space contains 2648 × 6101 = 16 155 448 products. Only a small fraction,

dx.doi.org/10.1021/ci2003957 | J. Chem. Inf. Model. 2012, 52, 373−379

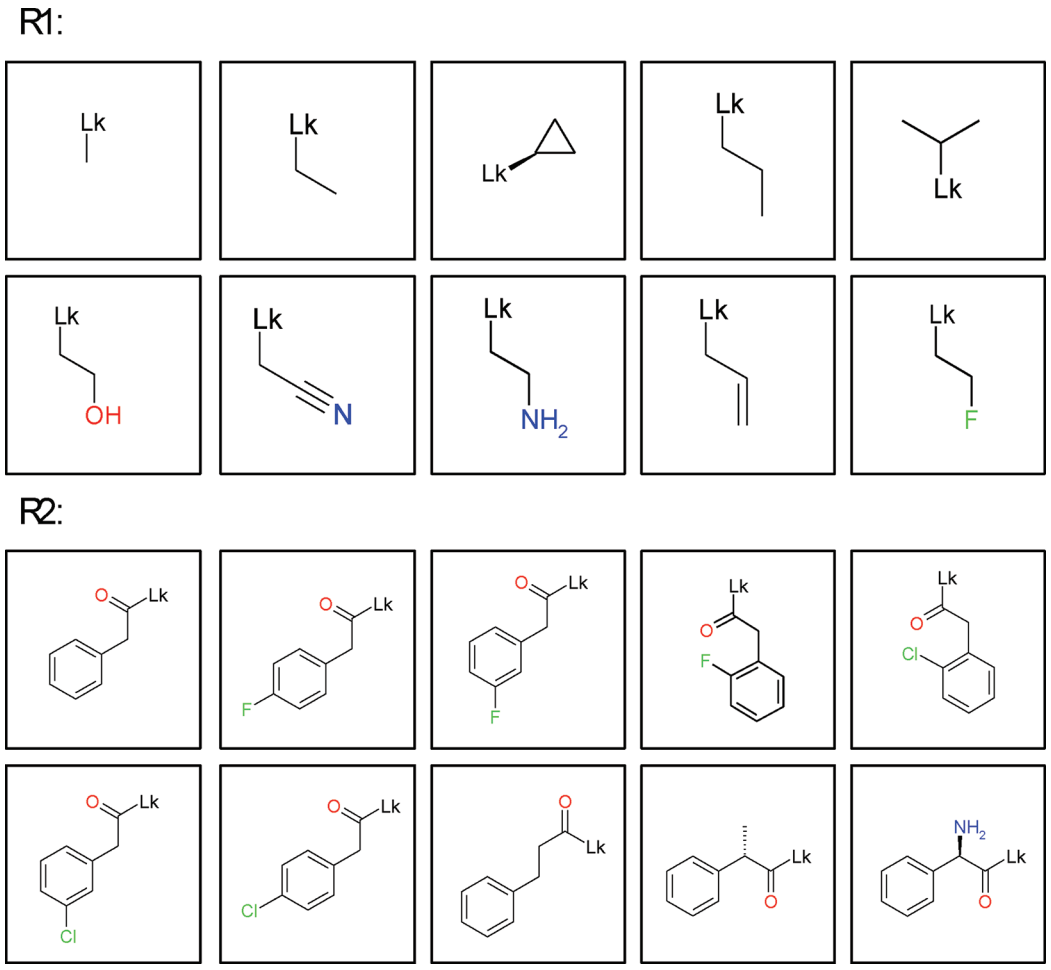


Figure 11. Reagents selected for the focused CDK2 library by LoFT optimizing the feature trees similarity to the query with FTMatch constraints for R1 and R2. The attachment points of the reagents to the aminopyrazole core are indicated by the symbol "Lk".

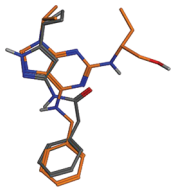


Figure 12. ROCS[7] 3D alignment of one of the products from the final LoFT design with the query.

namely 42 compounds, have a known activity, and only 35 are known to be active. This means that statistically one hit can be expected in about 460 000 compounds. Furthermore, it is not known whether the other products selected are active, and several compounds, especially those containing the building blocks, are very similar to the known actives and might be worth testing. Anyway, the last two reagents chosen for R2 in the final LoFT selection have been described to be active in combination with cyclopropyl at R1 in another publication of Pevarello et al,[13] which was not included in the plausibility check. These additional actives are shown in Figure 14.

In the final LoFT design the diversity of the selected reagents is increased. Consequently less of the building blocks leading to the 42 known aminopyrazoles, namely two alkylhalogenides and one carboxylic acid, are detected, because some of them cluster together and are deprioritized due to their similarity.

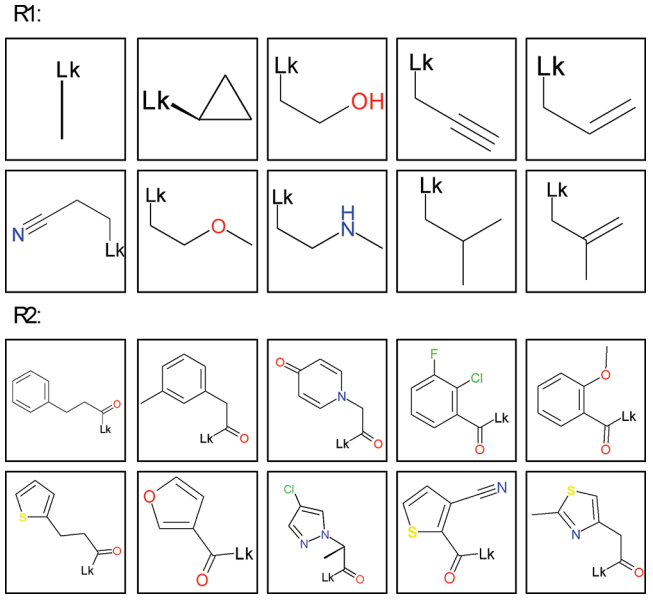


Figure 13. Reagents selected for the focused CDK2 library by LoFT optimizing the feature trees similarity to the query. Additionally, FTMatch constraints for R1 and R2 and a diversity term penalizing reagents taken from the same cluster were used. The attachment points of the reagents to the aminopyrazole core are indicated by the symbol "Lk".

dx.doi.org/10.1021/ci2003957 | J. Chem. Inf. Model. 2012, 52, 373−379

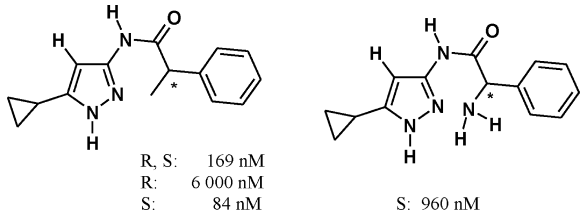


Figure 14. Known CDK2 inhibitors published by Pevarello et al[13] which were included in the focused CDK2 library designed by LoFT with FTMatch restrictions for R1 and R2.

Such a scenario can be applied to test a broader set of compounds. If actives are detected within this library, LoFT allows a further exploration of the reagents from the corresponding clusters.

CONCLUSIONS

In this case study, different focused CDK2 libraries have been designed using GOLD[6] docking, ROCS[7] 3D alignments, and LoFT for the reagent selection. The quality of the resulting libraries was assessed by the identification of building blocks derived from known actives. The focused library designed by LoFT optimizing the feature trees similarity to R-roscovitine and with FTMatch constraints for R1 and R2 delivered known actives and compounds worth testing due to their high similarity to known actives. The feature trees descriptor detects only 2D similarity, but via the FTMatch feature in LoFT, a kind of 3D scoring function is simulated. This way it is possible to direct the alignments to the pose which led to the selection of the scaffold at the beginning of the whole process.

The aim of this study was to demonstrate the applicability of feature trees as a reduced graph descriptor for focused library design. The structure of the feature tree descriptor allows an extremely fast comparison of combinatorial libraries to single compounds, like known actives.

AUTHOR INFORMATION

Corresponding Author

*E-mail: Uta.Lessel@boehringer-ingelheim.com. Telephone: +49-7351-543062.

ACKNOWLEDGMENTS

The authors thank Holger Claußen, Markus Lilienthal, and Christian Lemmen from BiosolveIT, St. Augustin, Germany, and Matthias Hilbig from the Center for Bioinformatics at the University of Hamburg for their support and fruitful discussions.

REFERENCES

(1) Rarey, M.; Dixon, J. S. Feature trees: A new similarity measure based on tree matching. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 471−490.

(2) Rarey, M.; Stahl, M. Similarity Searching in Large Combinatorial Chemistry Spaces. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 497−520.

(3) Lessel, U.; Wellenzohn, B.; Lilienthal, M.; Claussen, H. Searching Fragment Spaces with Feature Trees. *J. Chem. Inf. Model.* **2009**, *49*, 270−279.

(4) Fischer, J. R.; Lessel, U.; Rarey, M. LoFT: Similarity-Driven Multiobjective Focused Library Design. *J. Chem. Inf. Model.* **2009**, *50*, 1−21.

(5) Fischer, J. R.; Lessel, U.; Rarey, M. Improving Similarity-Driven Library Design: Customized Matching and Regioselective Feature Trees. *J. Chem. Inf. Model.* **2011**, *51*, 2156−2163.

(6) *GOLD Suite*, version 5.0.1; CCDC Software Limited: Cambridge, UK, 2010.

(7) *ROCS*, version 3.1.0; OpenEye Scientific Software, Inc: Santa Fe, NM, 2010.

(8) De Azevedo, W. F.; Leclerc, S.; Meijer, L.; Havlicek, L.; Strnad, M.; Kim, S. Inhibition of cyclin-dependent kinases by purine analogues. Crystal structure of human cdk2 complexed with roscovitine. *Eur. J. Biochem.* **1997**, *243*, 518−526.

(9) *FTrees*, version 2.3.0, BiosolveIT GmbH: St. Augustin, Germany, 2010.

(10) *Omega*, version 2.4.3; OpenEye Scientific Software, Inc: Santa Fe, NM, 2010.

(11) Pevarello, P.; Brasca, M. G.; Amici, R.; Orsini, P.; Traquandi, G.; Corti, L.; Piutti, C.; Sansonna, P.; Villa, M.; Pierce, B. S.; Pulici, M.; Giordano, P.; Martina, K.; Fritzen, E. L.; Nugent, R. A.; Casale, E.; Cameron, A.; Ciomei, M.; Roletto, F.; Isacchi, A.; Fogliatto, G.; Pesenti, E.; Pastori, W.; Marsiglio, A.; Leach, K. L.; Clare, P. M.; Fiorentini, F.; Varasi, M.; Vulpetti, A.; Warpehoski, M. A. 3-Aminopyrazole Inhibitors of CDK2/Cyclin A as Antitumor Agents. 1. Lead Finding. *J. Med. Chem.* **2004**, *47*, 3367−3380.

(12) *CoLibri*, version 1.3.1, BiosolveIT GmbH: St. Augustin, Germany, 2010.

(13) Pevarello, P.; Brasca, M. G.; Orsini, P.; Traquandi, G.; Longo, A.; Nesi, M.; Orzi, F.; Piutti, C.; Sansonna, P.; Varasi, M.; Cameron, A.; Vulpetti, A.; Roletto, F.; Alzani, R.; Ciomei, M.; Albanese, C.; Pastori, W.; Marsiglio, A.; Pesenti, E.; Fiorentini, F.; Bischoff, J. R.; Mercurio, C. 3-Aminopyrazole Inhibitors of CDK2/Cyclin A as Antitumor Agents. 2. Lead Optimization. *J. Med. Chem.* **2005**, *48*, 2944−2956.