J Comput Aided Mol Des (2012) 26:1207–1215
DOI 10.1007/s10822-012-9609-3

IADE: a system for intelligent automatic design of bioisosteric analogs

Peter Ertl · Richard Lewis

Received: 15 August 2012 / Accepted: 20 September 2012 / Published online: 28 September 2012

Abstract IADE, a software system supporting molecular modellers through the automatic design of non-classical bioisosteric analogs, scaffold hopping and fragment growing, is presented. The program combines sophisticated cheminformatics functionalities for constructing novel analogs and filtering them based on their drug-likeness and synthetic accessibility using automatic structure-based design capabilities: the best candidates are selected according to their similarity to the template ligand and to their interactions with the protein binding site. IADE works in an iterative manner, improving the fitness of designed molecules in every generation until structures with optimal properties are identified. The program frees molecular modellers from routine, repetitive tasks, allowing them to focus on analysis and evaluation of the automatically designed analogs, considerably enhancing their work efficiency as well as the area of chemical space that can be covered. The performance of IADE is illustrated through a case study of the design of a nonclassical bioisosteric analog of a farnesyltransferase inhibitor—an analog that has won a recent "Design a Molecule" competition.

Keywords IADE · Bioisosteric design · Scaffold hopping · Chemical space · Automated iterative drug design

Introduction

When faced with a task of designing novel active molecules starting from an X-ray structure of a ligand in the protein binding site, the standard method of choice for most molecular modellers is to use a process of classical design by hand. The process starts by detailed examination of the binding site in order to understand protein–ligand interactions and other factors influencing the binding. Then modellers try to modify the ligand by adding or removing substituents, fusing rings or changing atom types to alter the binding energetics, conformational preferences and physicochemical properties and thereby hopefully also enhance the potency of new analogs. This approach is working very well as documented by numerous examples from literature [1, 2] and also by many successful applications of manual molecular design in supporting in-house Novartis projects.

Classical manual molecular design also has several disadvantages. The most obvious one is that it requires the investment of a lot of time by an experienced molecular modeller. The manual editing and refining of molecules in the protein environment is not easy and is quite time consuming, since after each modification it is necessary to wait until the structure optimization is finished, the docking score is calculated, the output analyzed in terms of the starting hypothesis and so on. As a result, the throughput of manual molecular design is limited. One can perform such analyses for ten, possibly few tens of molecules, but not for hundreds. Another drawback of manual molecular design is its rather subjective character. Molecule modifications applied by different modellers depend strongly on their training and experience. It is only human that when a modeller is successful with a certain type of bioisosteric replacement in one project, he/she will try the same modification in other projects, maybe missing a different strategy that would be much better in that particular case.

To address these, one can say, natural deficiencies of molecular design by hand, we developed at Novartis a software expert system that we call IADE (from Intelligent

P. Ertl (✉) · R. Lewis
Novartis Institutes for BioMedical Research, Novartis Campus, 4056 Basel, Switzerland
e-mail: peter.ertl@novartis.com
URL: http://peter-ertl.com

Automatic DEsign). IADE can enhance work of molecular modellers by freeing them from repetitive, boring tasks, give them more time for the creative work of analyzing and evaluating analogs suggested by the program; it may even enhance their field of vision by offering broader coverage of chemical space. The authors want to stress clearly, however, that the system for automatic bioisosteric design we are presenting cannot replace the manual modelling effort. This would not be even possible with current state of scientific knowledge and the lack of accuracy associated with current scoring functions. The role of IADE is to help molecular modellers to explore better chemical space and to make their work more efficient.

The effort to support the work of molecular modellers by identifying new bioisosteric analogs by software tools has a long tradition [3]. Brute force exploration quickly becomes tedious and infeasible unless the chemical space can be reduced. Combinatorial libraries offer a way to proceed [4–6], but the central core must be maintained, a limitation if one is more interested in finding novel chemotypes. Rigorous graph theoretic procedures have been used in automated iterative design, to deconvolute from the descriptors used in a QSAR to the structures [7–11]. QSAR models do not capture all the implicit understanding of the SAR obtained by a medicinal chemist after intensive study of the chemical series and the binding site in question. Our experience is that we needed something faster, even if less rigorous.

Heuristic approaches, like the strategy presented here, have also been tried. An example is the program GROK [12], which used a genetic algorithm to manipulate a set of input structures according to a user defined fitness function. This program seems to have been very fast and efficient, but also seem to have suffered from issues of extrapolation, finding pathological conditions in many of the scoring functions examined (for example, if the scoring function contains positive term for molecular weight, or amine groups, then a larger molecule, or one covered in amines, will score best, which is a reduction ad absurdum). The Automated Iterative Design (AID) protocol of Lewis [13] made explicit provision for extrapolation and consequent prediction errors in the QSAR models by constraining exploration to areas of known/predictable SAR space. It has long been known that one of the major pitfalls of AID, especially when linked to de novo structure generation, comes from extrapolation into unreasonable areas of space. We have tried to address this issue in two ways, through the choice of fragments and the use of post processing filters.

IADE methodology

The design of new bioisosteric analogs with IADE starts in a similar way to a classical molecular design project. The modeller needs to thoroughly examine the binding site, understand the binding and, based on this detailed analysis, prepare the command file and necessary data that IADE will use. The usual information needed to start a IADE run is: (1) the structure of ligand in active conformation that serves as a template; (2) a protein prepared for docking. This information is in most cases extracted from an X-ray structure of ligand–protein complex, although one can use also a template geometry generated by a pure ligand-based approach. After launching IADE, the program performs an iterative design and outputs the best analogs. The modeller, usually together with project chemists, analyses the results, and if the suggested molecules fulfill the necessary project criteria, they can serve as a basis for synthesis. On the other side, if the project team is not satisfied, the IADE command file may be modified and the run repeated.

The IADE process flow, shown schematically in Fig. 1, starts by fragmenting the template ligand into set of pairs consisting of substituents with rests of the molecule, and a set of linkers (fragments with 2 connection points) with respective capping Rgroups. All activated nonring single bonds [14] are fragmented in this process. Examples of fragment pairs and triplets generated by this process are shown in Fig. 2.

In the next step, the substituents in fragment pairs and the central linkers of fragment triplets are replaced by bioisosteric analogs and the molecules are reconstructed. The process of identifying these bioisosteric fragments is the heart of the IADE system. Bioisosteric fragments are identified based on their similarity in property space with the target fragments. This methodology was developed at Novartis and has been described previously [14–16]. The compatible fragments are selected from a database of more than 10,000 common substituents and linkers obtained by fragmenting the ChEMBL database [17] containing over 1 million bioactive molecules. When creating the fragment database, the extracted substituents and linkers were characterized by a

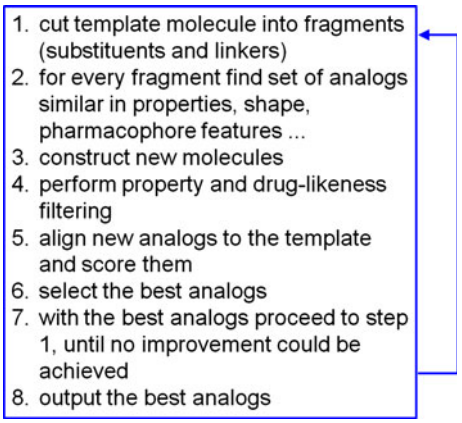

Fig. 1 Scheme of the IADE workflow

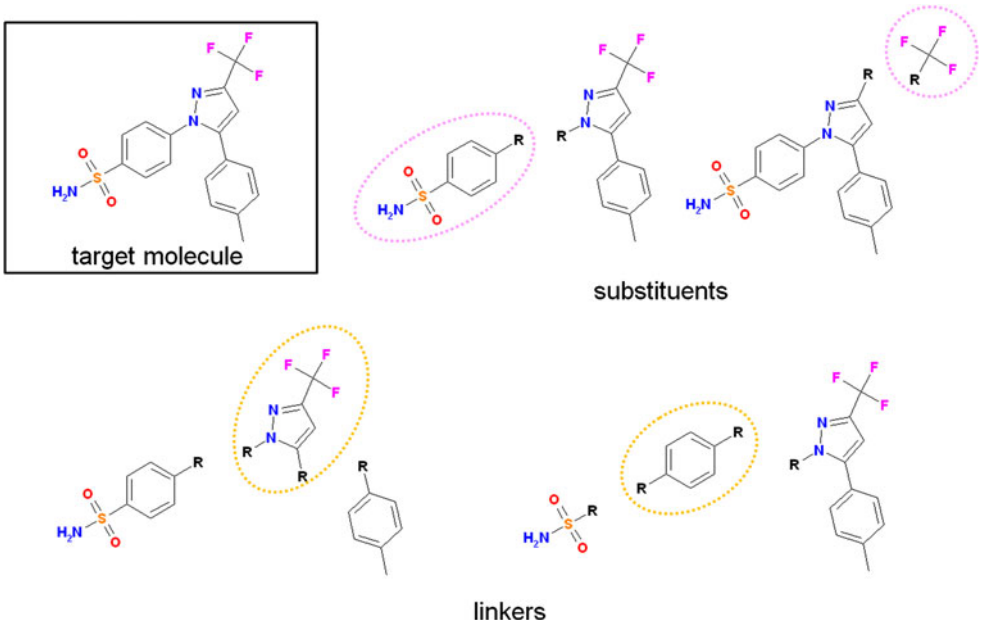

Fig. 2 Example of fragment pairs and triplets generated by fragmenting the target molecule. Only two possible fragmentations are shown for each type, the actual number is much larger

number of calculated descriptors. These descriptors represent properties that are most relevant to medicinal chemistry, describe ligand-receptor interactions and are also quick to calculate. Since the fragments have one (Rgroups) or two (linkers) connections to the main part of the molecule, it is necessary to consider descriptors characterizing the properties of these connection vectors. In characterizing fragments in the database the following descriptors were used: (1) the size of fragments is described simply by the atom count; (2) fragment shape and basic pharmacophore features are described by vectors characterizing number of hydrogen bond donors and acceptors at certain topological distances from the connection points; (3) for linkers the mutual position of Rgroups is characterized simply by the topological distances between them; (4) electron-donating or -accepting power at connection points is characterized by quantum chemical parameters developed in house [18] that are surrogates for experimental Hammett σ constants; (5) the ADME properties of fragments, such as hydrophobicity, permeability and solubility are characterized by CLOGP (calculated logP) and TPSA [19] values. Fragments are stored in the database as SMILES strings with connection points marked as R atoms, together with the calculated descriptors described above. The SMILES representations are in canonical form. When searching for replacement of linkers, one needs to consider both possible orientations of the fragments in the database (like an amide and inverse amide). The most similar fragments are identified as those where the sum of differences in their descriptors is minimal. Examples of bioisosteric substituents and linkers suggested by this procedure are shown in Fig. 3.

After identification of bioisosteric substituents and linkers, every fragment in the template molecule is replaced by 10–20 compatible analogs (this number is an adjustable IADE parameter) and the molecules are reconstructed. This generates quite a large number of analogs per molecule, usually a few hundred. These analogs are now filtered to remove any with substructures incompatible with drug likeness [20] or that are possibly toxicophores. Despite the fact that fragments were extracted from a database of drug-like bioactive molecules, this step is necessary, because non-drug-like features may be introduced by joining two incompatible groups together. Analogs are also checked to have physicochemical properties, particularly logP and TPSA, within user-desired ranges (these may be again defined in the IADE command file). A synthetic accessibility score developed in-house [21], a measure of the ease of synthesis based on several simple rules and fragment contributions, is used to filter out molecules that are too complex and would cause problems by synthesis. And, of course, analogs are checked for their uniqueness using their canonical SMILES representation; all duplicates are discarded.

All the processing described so far is done on molecules described only by their atom connectivity, and no 3D information has been used. Therefore the processing is very fast. As seen from the previous description, the IADE design strategy is not transformation driven, as for example the program DOGS [22] that uses a database of common chemical reactions to design new molecules, but fragment driven. In the next step, the retained analogs are converted into 3D by CORINA [23]. Then a set of representative conformations is generated using standard methods from which the conformation most similar to the template ligand identified. The FieldAlign software from Cresset [24] is used for this purpose. FieldAlign finds the best alignment

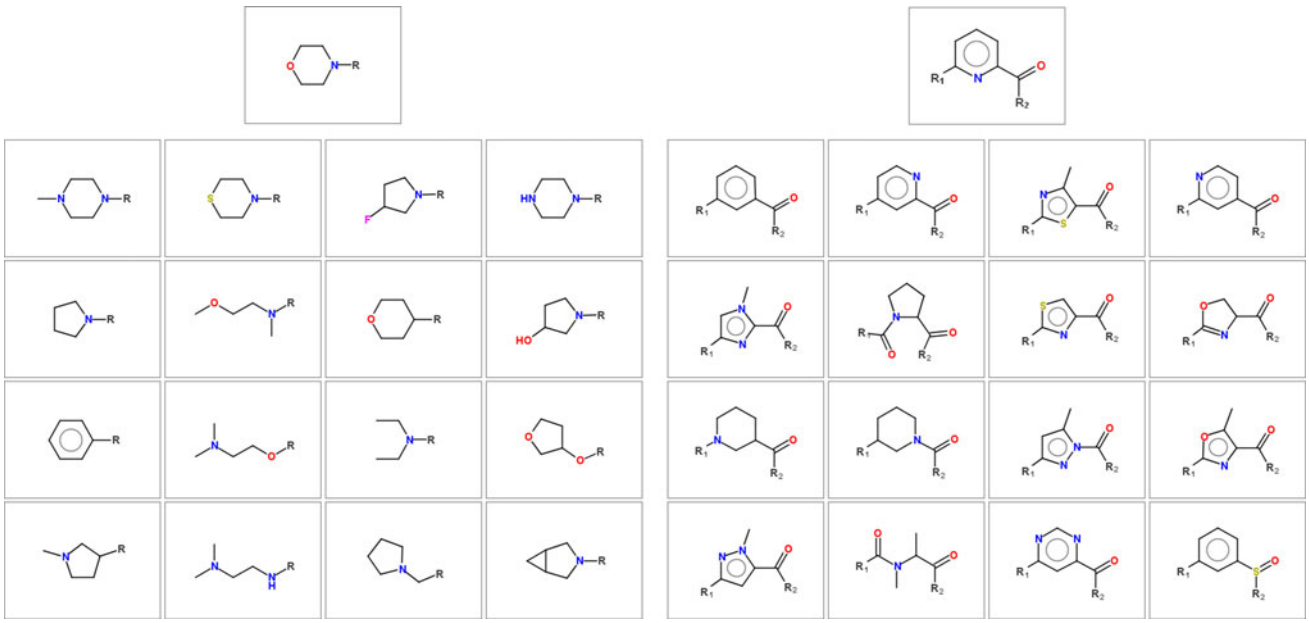

Fig. 3 Example of bioisosteric substituent and linkers identified by the property similarity search

based on superposition of positive, negative, hydrophobic and steric fields around the molecules. Alignment based on fields and not on molecular skeletons is particularly suited to identify non-classical bioisosteric analogs.

The final step, scoring, depends on the type of information that is available for the project. In case when only a ligand-based model is used and no information about the target protein is available, the Cresset similarity score (employing an equal weighting of field and shape similarities) is used. When an X-ray structure of the protein binding site is available, the Glide docking score [25] is used to rank the analogs. Only the refinement option (i.e. ligand is optimized in the protein environment starting from an initial geometry) is used, not the full flexible docking procedure. This should assure that the analogs will have the same binding mode as the original template ligand. Additional scoring methods may be added later, for example scores obtained by other docking packages, 3D QSAR models, or even simple 2D QSAR models. In the future we plan to add the possibility to use not only a score calculated by a single program but a fitness function obtained as a combination of several parameters, for example a weighted combination of docking score, synthetic accessibility score and calculated logP and TPSA values.

All processing steps described so far cover only one generation of a complex iterative process. At the end of each generation, the analogs are sorted according to their score and the best ones are used as starting points for the next generation. This process is repeated until no improvements in score could be obtained, or until the modeller supervising the project is satisfied with the generated analogs.

As could be seen from the description above, the IADE process flow is quite complex. Its various modules are shown in Fig. 4. The system uses several 3rd party packages including Mopac, CORINA, FieldAlign and Glide and several in-house programs written in Java and C performing mostly SMILES processing and molecule manipulation like cutting and gluing structures, and calculating the drug-likeness score. Several web services are also utilized, including a services to calculate Hammett sigma parameters, synthetic accessibility score or CLOGP and TPSA descriptors. The program connects to the databases containing substituents and linkers with calculated properties. All these modules are connected together by several Python scripts.

Despite the program complexity, its design is modular, with all modules connected through simple and well defined interfaces. So it is relatively easy to replace a module by another one with the same functionality. This allows the use of the current best-in-class method for particular task, but without too much dependency on any particular software vendor. Alternatively, local knowledge about the target can be encoded as an additional module, for example a defined exclusion space covered by an existing chemotype.

Users interact with the program by a simple command file, containing keywords that govern the IADE functionality. Example of common tasks include standard bioisosteric analog design, optimization of scaffold while keeping substituent pattern intact (=scaffold hopping [26]) or

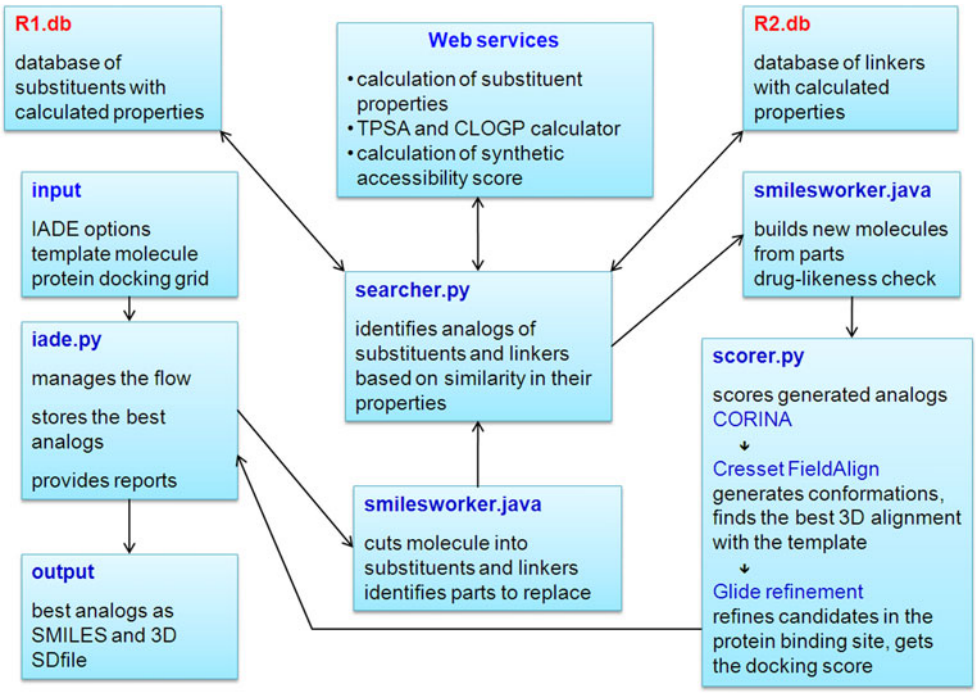

Fig. 4 Modular structure of the IADE program

optimization of one or more substituents, while keeping the rest of the molecule unchanged. IADE supports also fragment growing, i.e. designing a molecule optimally fitting the protein cavity starting from an X-ray structure of a bound fragment. The algorithm in this case is slightly modified from the procedure used by standard bioisosteric replacement. When growing a fragment, the template molecule is replaced after each generation by a structure having the best interaction energy with the protein so far. This procedure allows molecule to grow and it also supports identification of fragments providing the optimal interaction with the target protein. Substructures that should be kept intact during the optimization, or on the other hand that should be exclusively modified, can be defined via SMILES patterns. The command file contains also several other parameters, including paths to the necessary data files (for example, a protein grid needed for docking) or parameters telling IADE whether it should run locally, or on a cluster, and on how many processors.

After every iteration IADE creates two data files. In the first file all molecules processed so far, together with their scores, are stored in as canonical SMILES. This file is used for detection of structures already processed in previous generations. Newly generated molecules, also in the canonical SMILES form, are checked against this list, and when already known, are discarded. The second file contains the best 1000 analogs in 3D SDfile format, in the same coordinate system as the target protein. With help of this file, it is easy to check how the generated analogs fit the binding site and how they develop from generation to generation. By using these two files, it is also easy to restart IADE at arbitrary point, possibly with slightly modified parameters.

The time needed to run a IADE optimization depends on several factors, such as the size and complexity of the template molecule that should be replaced, the number of bioisosteric fragments in the replacement set or the number of the best analogs that are used as a starting point in the next generation. Basically, the 2D processing (i.e. molecule fragmentation, identification of bioisosteric substituents and linkers, analog construction and cleaning) is relatively fast, the same is also true for calculation of docking scores (as mentioned above, fully flexible docking is not performed, only refinement). The most time consuming task is the conformation search and the subsequent identification of the most similar conformer using the Cresset FieldAlign software. When running locally on a server with only a few processors, one generation may require several hours to complete; when running on a grid, usually an hour is sufficient for a generation. Thus even complex calculations, processing ten thousand analogs per generation, are generally finished overnight.

Example of bioisosteric design by IADE

In autumn 2011, when development of IADE program was just finished, the 3rd round of well-known "Design a Molecule" competition organized by Cresset was announced. Contestants were asked to design an analog of

a farnesyltransferase inhibitor from the PDB structure 3E32 and to effectively replace the core of the ligand with a new framework. We considered this to be an excellent opportunity to test the IADE on an interesting antimalarial target in a renowned international competition. After performing the bioisosteric design supported by IADE and deciding upon the best analog without intimate knowledge of the target, the authors were glad to learn that this design was selected by a panel of judges as the winner of the "Design a Molecule" competition [27]. In the next paragraph, details of the design procedure leading to the winning analog are described, providing a good illustration of a typical IADE workflow.

The IADE command file was prepared including the structure of ligand from the 3E32 X-ray [28] as a template; the protein itself was used as an excluded volume by the FieldAlign. Since our goal was to design analogs not too similar (in the standard substructural sense) to the template ligand, certain substructures present in the template, including the sulfonamide moiety and the central ethylenediamine linker were "forbidden" in the designed analogs. Several IADE runs were performed in a semiautomatic mode, i.e. the output was checked after every generation and small adjustments were made to the command file, including exclusion of certain substructure features or focusing on the optimization of only a part of the molecule to steer IADE into desired direction. The speed of these runs allows the modeler to frame hypotheses quickly in terms of parameter adjustments. The program was thus used exactly in the way it was designed for, to provide assistance in the manual design by automatically exploring large areas of chemical space and offering novel ideas to be evaluated by modelling experts. The fitness score, in this case the Cresset similarity with the template ligand, increased continuously during the optimization (Fig. 5) until after the 7th generation it reached an optimum and did not improve any more.

The sequence of molecules leading to the winning analog is shown in the Fig. 6. The molecules with the best score in every generation are not shown here, rather the sequence leading to the winning analog. In every generation 20 molecules with the best scores were used as input structures for the next generation, so it sometimes happened that the next molecule in the sequence had actually a worse score than its parent. This clearly shows that it is necessary to use more molecules as starting points in every generation to prevent trapping in local minima and to cover more diversity in chemical space. This is in contrast to methods based on probabilistic selection.

The optimization started by replacing a large part of the template containing the substituted sulfonamide and benzyl moieties by a phenyl substituted pyrazolo-pyrimidine ring that itself was replaced by a diphenylether unit in the next generation. These replacements are not obvious and it is

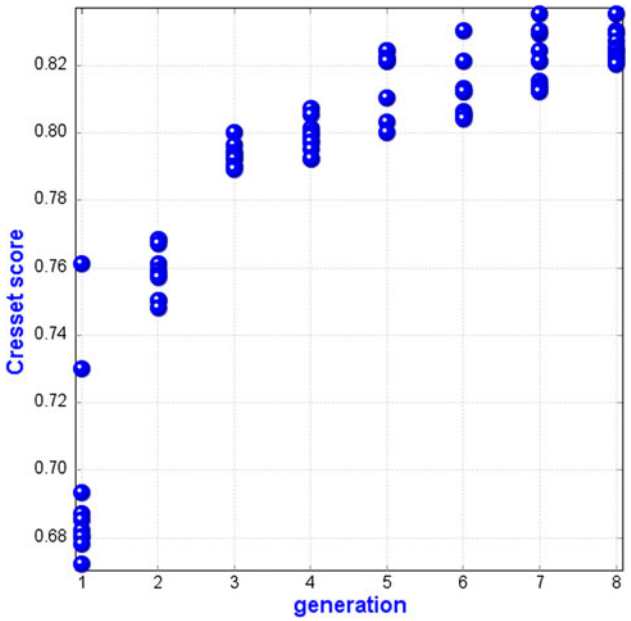

Fig. 5 Development of the Cresset score during the IADE optimization

doubtful whether even the most experienced medicinal chemists would come with such ideas within two design cycles. The program identified these replacements by matching properties of the leaving and new parts. In the 3rd step, the right-hand part of the molecule was replaced by an isosteric substituent. It is interesting to note that both the cyano-substituted phenyl and the imidazole ring remained, but they exchanged the methylene linker. After these large changes that considerably modified the molecule skeleton, several smaller substitutions were performed to improve the field similarity between suggested analogs and the template ligand. In the 4th generation, the nitrogen atom was introduced in the central phenyl. More precisely the phenyl linker was replaced by the pyridine linker, because IADE does not perform an atom-for-atom "nitrogen scan" as for example the program MORPH [29] does, but introduces such changes by the replacement of the entire linker or substituent. The advantage of this approach is that rings with several heteroatoms, or a heterocycle containing also small optimal substituents, may be introduced in a single step. In the 5th generation, the imidazole substituent was replaced by the triazole, and finally the fluorine substituent was introduced on the left-hand side phenyl to fine-tune the molecule's conformation and optimize the field match.

In Fig. 7, the 2D and 3D structures of the template ligand and the winning analog are compared. Despite the fact that the underlying bonding patterns of the structures differ considerably, the Cresset fields shown in Fig. 7 are very similar. The suggested analog is therefore a true nonclassical bioisosteric analog of the template ligand. It

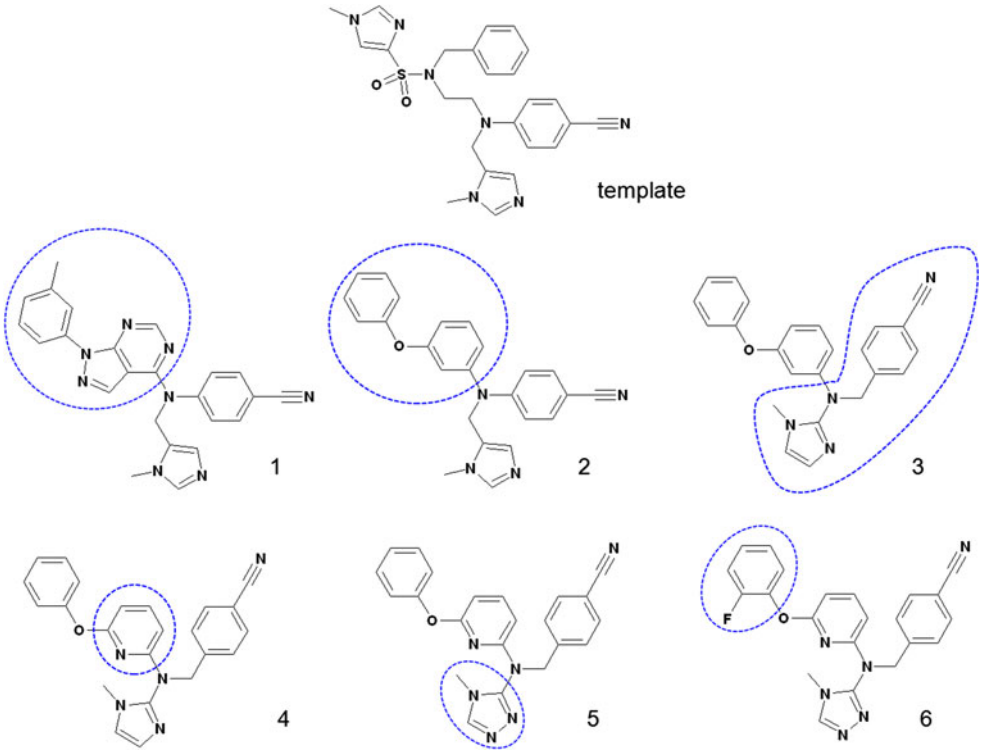

Fig. 6 Evolution of the molecule during the optimization process. The newly introduced parts are highlighted

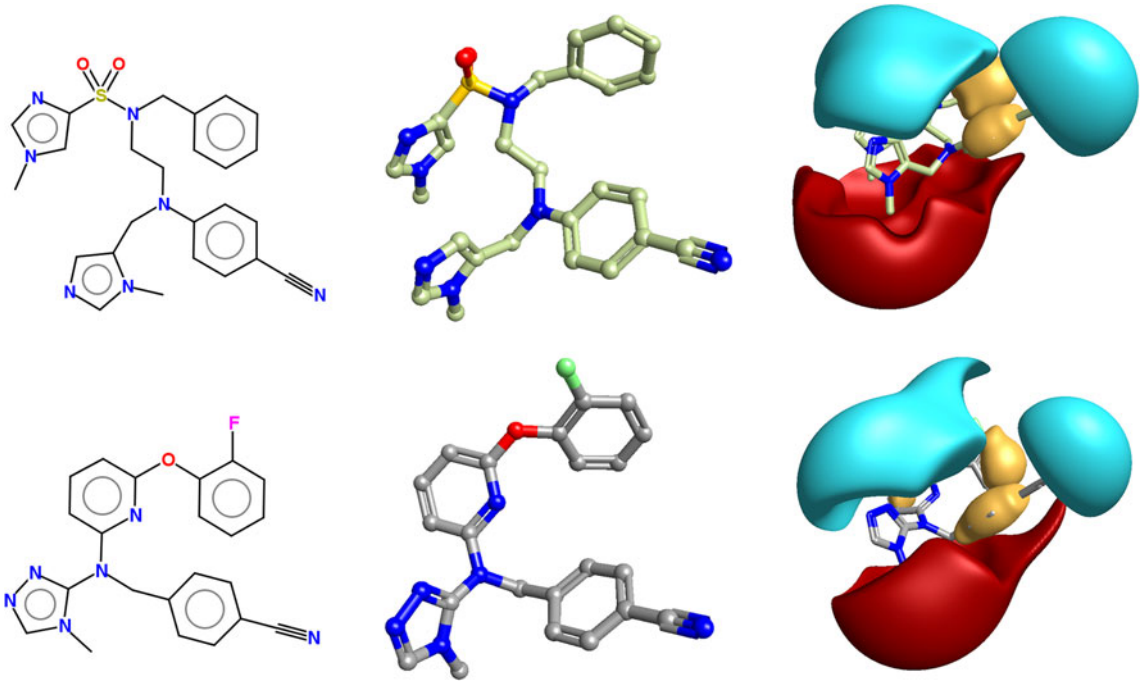

Fig. 7 Comparison of structures and Cresset fields of the template ligand (*top*) and the winning bioisosteric analog (*bottom*)

should be noted that the winning analog was not the structure with the best similarity score suggested by IADE, but the molecule selected manually from the top scoring hits taking into account also its good physicochemical properties (CLOGP = 3.4, TPSA = 80) and synthetic accessibility score.

Discussion

The use of bioisosteric replacement has a long history within medicinal chemistry, and several groups have described approaches to generate replacements relevant to drug discovery [30] and to represent them in a manner appropriate for lead optimization [31]. This approach extends this work by using a more generalized fragment definition, an automated replacement algorithm, and a scoring function that can be adapted according to local knowledge of the target.

In design programs of this type, there are always balances to be struck between the granularity of exploration and the speed of execution. At one end of the spectrum, there is the universe of compounds we could make but haven't. An example would be a combinatorial chemistry library, from which only a small subset has been synthesized. In principle, any other member of the library is also synthetically accessible. The size of the universe is the product of the numbers of R-groups and can quickly reach infeasible large numbers (scientists at Pfizer estimate that this accessible universe just within Pfizer is 10^{12} compounds, based on the Colibri approach [32], which encodes a universe via fragments and linking rules built on known synthetic protocols). It is impractical to explicitly enumerate all these compounds and to search them, so methods have been developed to narrow down the search space, based on representing the molecules as combinations of feature trees. Lessel et al. [33] have described further improvements to the Colibri approach. This universe can be searched very quickly using feature trees, and in a retrospective study, some known hits could be retrieved from a universe of 10^{11} compounds with only 4 query structures as input. While the search space is very large, this approach, to our thinking, sacrifices the quality of model that can be used to search the space.

At the other extreme, one can make small-scale changes to an intact core ("walking the methyl"). The BOMB [34] software is based on growing around a core fragment in the binding site. The incremental growing of a core structure rapidly lead to analogues that were 5,000-times more potent against HIV reverse transcriptase [35]. In addition to mutating groups at the periphery, heterocycle scans can also be performed. This strays into regions of greater error, but still some improvements could be predicted and later validated. This form of compound generation strategy is possible using IADE, but would preclude the speed of interaction that we were looking for, given the relatively high cost of the FEP studies required for each analog.

The use of an entire fragment strategy gives the necessary speed for initial exploration of a hypothesis at a fairly coarse level; we have also shown that later generations do not jump out into new areas, but instead focus on much more local improvements. This is a result of taking just the best scoring solutions. An atom-for-atom strategy would allow a much finer exploration but possibly at the expense of spending time generating infeasible molecules. However, we do not exclude the possibility of using a combined strategy for fine-polishing a hypothesis.

The case study presented here was based on a protein–ligand complex. Here we may be reasonable comfortable about interpreting the output. There is a danger inherent to any workflow based on scores. Most models have only a limited accuracy, even for docking. Even though the binding pose is probably right, the score (or relative score) is not. Similar statements can be made about the relative energies of conformations, the rankings or classifications derived from QSAR models. QSAR models also have the issue of domain applicability, which we do not allow for here. We suggest that the distance to nearest neighbors in the training set be used as a quick exclusion heuristic [36]. It is the duty of the modeller to interpret such numbers in the context of prior knowledge and experience, and to give an honest assessment of the robustness of the calculation. This is why the protocol was designed to give quick suggestions that can be analyzed in the context of local knowledge.

Conclusions

A program for automatic identification of bioisosteric analogs, scaffold hopping and fragment growing called IADE has been developed at Novartis. The system frees modellers from repetitive and time consuming tasks and allows them spend more time on creative work: evaluating and selecting the best ideas from automatically designed structures. IADE compensates for otherwise rather subjective character of design by experts through unbiased exploration and contributes non-obvious bioisosteric ideas. The modular architecture of the program, in which one can simply exchange modules from various software vendors, allows the use of best-in-class software for particular task without creating dependencies. The framework also has allowed us to use in-house tools, including substituent, linker and scaffold matchers [16], Hammett σ descriptor [18], TPSA [19, 37], synthetic accessibility score [21] and most importantly local QSAR models to tailor the optimization process. The IADE program is still in development,[1] but it already established itself as a useful help to modellers supporting medicinal chemistry projects,

[1] The authors recognize that there may be certain implementation details that were obvious to us but may be unclear from our description. We are willing to work with scientists trying to re-implement the code if there are sins of omission concerning the methodology.

delivering several very useful novel ideas for the design of novel bioactive molecules.

Acknowledgments The authors would like to acknowledge help of colleagues Peter Gedeck, Nikolaus Stiefl and Marc Litherland who helped with the implementation of IADE. Several colleagues at Novartis acted as test users and early adopters, particularly Simona Cotesta, Paulette Greenidge, Hanneke Jansen, Anna Vulpetti, Bernard Pirard and Finton Sirockin. We would like to thank also Cresset Group for help with the use of command line FieldAlign and for organizing the "Design a Molecule" competition.

References

1. Kubinyi H (2006) Success stories of computer-aided design. In: Ekins S, Wang B (eds) Computer applications in pharmaceutical research and development. Wiley, Hoboken, pp 377–424
2. Lewis RA (2012) The development of molecular modelling programs: the use and limitations of physical models. In: Livingstone DJ, Davis AM (eds) Drug design strategies: quantitative approaches. RSC Publishing, Cambridge, pp 88–107
3. Pirard B (2011) The quest for novel chemical matter and the contribution of computer-aided de novo design. Expert Opin Drug Discov 6:225–231
4. Venkatasubramanian V, Chan K, Caruthers JM (1995) Evolutionary design of molecules with desired properties using the genetic algorithm. J Chem Inf Comput Sci 35:188–195
5. Sheridan RP, Kearsley SK (1995) Using a genetic algorithm to suggest combinatorial libraries. J Chem Inf Comput Sci 35:310–320
6. Weber L, Almstetter M (1999) Diversity in very large libraries. In: Dean PM, Lewis RA (eds) Molecular diversity in drug design. Kluwer, pp 93-114
7. Skvortsova MI, Baskin II, Slovokhotova OL, Palyulin VA, Zefirov NS (1993) Inverse problem in QSAR/QSPR studies for the case of topological indices characterizing molecular shape (Kier indices). J Chem Inf Comput Sci 33:630–634
8. Kier LB, Hall LH (1993) The generation of molecular structures from a graph-based QSAR equation. Quant Struct-Act Relat 12:383–388
9. Hall LH, Fisk JB (1994) Computer generation of vertex degree sets for chemical graphs from a number of vertices and rings. J Chem Inf Comput Sci 34:1184–1189
10. Kvasnicka V, Pospichal J (1996) Simulated annealing construction of molecular graphs with required properties. J Chem Inf Comput Sci 36:516–526
11. Faulon J-L, Visco DP Jr, Pophale RS (2003) The signature molecular descriptor. 1. Using extended valence sequences in QSAR and QSPR studies. J Chem Inf Comput Sci 43:707–720
12. Weininger D (1995) Method and apparatus for designing molecules with desired properties by evolving successive populations. U.S. Patent No. 5,434,796
13. Lewis RA (2005) A general method for exploiting QSAR models in lead optimization. J Med Chem 2005(48):1638–1648
14. Ertl P (2003) Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. J Chem Inf Comput Sci 43:374–380
15. Ertl P (1998) World Wide Web-based system for the calculation of substituent parameters and substituent similarity searches. J Mol Graph Mod 16:11–13
16. Ertl P (2012) Database of bioactive ring systems with calculated properties and its use in bioisosteric design and scaffold hopping. Bioorg Med Chem 20:5436–5442
17. Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:D1100–D1107
18. Ertl P (1997) Simple quantum-chemical parameters as an alternative to the Hammett sigma constants in QSAR studies. Quant Struct-Act Relat 16:377–382
19. Ertl P, Rohde B, Selzer P (2000) Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. J Med Chem 43:3714–3717
20. Walters WP, Murcko MA (2002) Prediction of 'drug-likeness'. Adv Drug Del Rev 54:255–271
21. Ertl P, Schuffenhauer A (2009) Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. J Cheminf 1:8
22. Hartenfeller M, Zettl H, Walter M, Rupp M, Reisen F, Proschak E, Weggen S, Stark H, Schneider G (2012) DOGS: reaction-driven de novo design of bioactive compounds. PLoS Comput Biol 8:e1002380
23. CORINA version 3.20, Molecular Networks GmbH, Erlangen, Germany
24. FieldAlign version 3.01, Cresset Group, Welwyn Garden City, UK
25. Glide, version 5.8, Schrödinger, Inc., New York, NY, USA
26. Langdon S, Ertl P, Brown N (2010) Bioisosteric replacement and scaffold hopping in lead generation and optimization. Mol Inform 29:366–385
27. http://www.cresset-group.com/2011/10/design-a-molecule-competition-fall-2011-winner-announced/
28. Hast MA, Fletcher S, Cummings CG, Pusateri EE, Blaskovich MA, Rivas K, Gelb MH, Van Voorhis WC, Sebti SM, Hamilton AD, Beese LS (2009) Structural basis for binding and selectivity of antimalarial and anticancer ethylenediamine inhibitors to protein farnesyltransferase. Chem Biol 16:181–192
29. Beno BR, Langley DR (2010) MORPH: a new tool for ligand design. J Chem Inf Model 50:1159–1164
30. Raymond JW, Watson IA, Mahouni A (2009) Rationalizing lead optimization by associating quantitative relevance with molecular structure modification. J Chem Inf Model 49:1952–1962
31. Lewell XQ, Jones AC, Bruce CL, Harper G, Jones MM, Mclay IM, Bradshaw J (2003) Drug rings database with web interface. A tool for identifying alternative chemical rings in lead discovery programs. J Med Chem 46:3257–3274
32. Boehm M, Wu T-Y, Claussen H, Lemmen C (2008) Similarity searching and scaffold hopping in synthetically accessible combinatorial chemistry spaces. J Med Chem 51:2468–2480
33. Lessel U, Wellenzohn B, Lilienthal M, Claussen H (2009) Searching fragment spaces with feature trees. J Chem Inf Model 49:270–279
34. Jorgensen WL (2009) Efficient drug lead discovery and optimization. Acc Chem Res 42:724–733
35. Jorgensen WL, Ruiz-Caro J, Tirado-Rives J, Basavapathruni A, Anderson KS, Hamilton AD (2006) Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. Bioorg Med Chem Lett 16:663–667
36. Sheridan RP, Kearsley SK (2002) Why do we need so many chemical similarity search methods? Drug Discov Today 7:903–911
37. Ertl P (2007) Polar surface area. In: Mannhold R (ed) Molecular drug properties. Wiley, Hoboken, pp 111–126