Journal of Medicinal Chemistry

Volume 47, Number 3 January 29, 2004

Letters

Scaffold Hopping in De Novo Design. Ligand Generation in the Absence of Receptor Information

David G. Lloyd,* Christoph L. Buenemann, Nikolay P. Todorov, David T. Manallack, and Philip M. Dean

De Novo Pharmaceuticals, Compass House, Vision Park Histon, Cambridge, CB4 9ZR, UK

Received October 24, 2003

Abstract: We report here the de novo generation of chemotypes and scaffolds for the estrogen receptor, without use of the receptor structure in the assembly phase. Through use of ligand superpositions or a single bound conformation of a known active, a pseudoreceptor can be generated as a design envelope, within which novel structures are readily assembled. Many of these structures have high similarity to known chemotypes. Scaffold hopping is readily achieved within this pseudoreceptor, indicating the advantages of such an approach in discovery research.

Introduction. The use of computational de novo ligand generation software in drug design is now growing in popularity; validation cases and reviews have recently been published highlighting its utility in the discovery process.[1−7] Typically, there has been a requirement for the availability of a receptor structure, within which the ligands are assembled or grown, and with which ligand interactions are optimized as an attempt is made to tailor the molecular 'key' to its biological 'lock'.[8] Where de novo design has been successful, it was based not on the structure of an empty binding site alone, but also on knowledge about essential features of a potential ligand.[9,10] For example, the DuPont cyclic urea inhibitors for HIV protease were developed from pharmacophore features extracted from X-ray structures of complexes with other HIV protease inhibitors.[11] Likewise, knowledge of the substrate and its interactions in the binding site has facilitated structure-based design of the first neuraminidase inhibitors.[12] More recent publications have shown that structures of neuraminidase inhibitors can considerably deviate from the substrate as long as certain pharmacophore elements are fulfilled.[13−15]

Conversely, in cases where the target receptor structure is unknown, (Q)SAR and pharmacophore-based approaches often assist in the adoption of rational design strategies. However, pure computational de novo design of new ligands in such cases have rarely been reported, thus further highlighting the dependencies of ligand generation methods on the availability of receptor structure.[16−19] One significant attraction of nonreceptor based de novo design would be an application to 'scaffold hopping', moving quickly from one chemotype class to alternatives.[20] Computational de novo design could afford a rapid route to diverse, yet functionally equivalent chemotypes. If de novo design methods were not limited to structurally defined binding sites, the scope of application would be dramatically increased.

The efficacy of our de novo structure generation tool Skelgen in chemotype switching or 'scaffold hopping' has been previously illustrated when applied to a series of structure-based design challenges.[1] The utility of this software would be greatly enhanced if it were more broadly applicable to cases where receptor structure was not known. Can scaffold hopping in ligand generation be achieved where the nature of the 'lock' is unknown? To answer this question we have applied Skelgen to a ligand-based design strategy and present the design of ligands for the estrogen receptor (ER) in the absence of any consideration of the actual receptor protein structure.

When run in receptor mode (i.e. where a protein structure is available), input to Skelgen consists of a receptor pocket, defined by a protein coordinate file and a rectangular box, a set of constraints (termed *site points*[21]) that should be fulfilled by each of the generated solutions, and a set of molecular fragments that are used for the construction process. Skelgen uses a stochastic procedure for the assembly of molecular scaffolds. Validation of this method has been published

* To whom correspondence should be addressed. Tel: +44 1223 238000 Fax: +44 1223 238088, e-mail: david.lloyd@denovopharma.com.

10.1021/jm034222u CCC: $27.50
Published on Web 12/25/2003

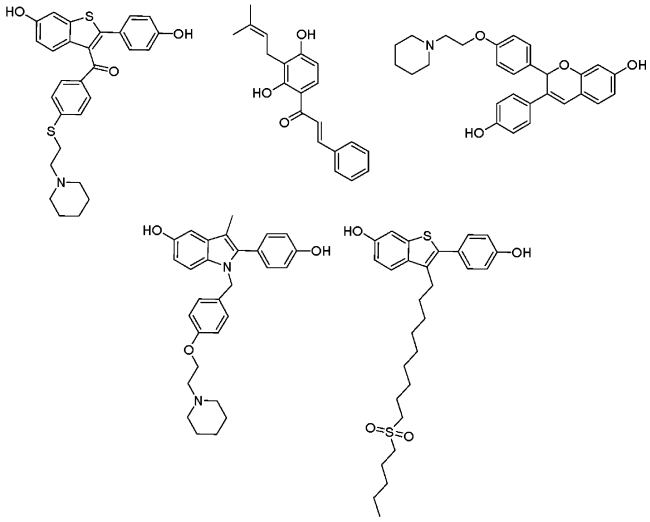

Figure 1. Ligands used in derivation of ER pharmacophoric points.

fully.[1] A single solution structure, a ligand candidate, is generated as output from a single run. The algorithm produces numerous solutions from different random starts. This provides the user with a choice of many different resultant chemotypes from multiple runs.

When operated in *ligand* mode, the need for a receptor structure is removed as Skelgen's requirement for an explicit receptor is replaced by a user-supplied inclusion shape. The inclusion shape can be defined from as little as the coordinates of a single molecule. An input atomic coordinate file defining the inclusion shape is read in where each atom point defined in the file is interpreted as the center of a featureless sphere with with an atomic van der Waals radius plus an associated user-defined radius. The union set of spheres is used as a hull (an included volume) for pharmacophore-based ligand design. A penalty, proportional to the distance to the nearest sphere is added for each non-H ligand atom lying outside the design hull. This drives the ligands to lie within the specified space. Typically, user-defined site points are also provided which constrain the search by means of pharmacophore features that must be satisfied by the designed compounds. These constraints should be well-balanced, as general as possible, and in the case of ligand mode, will stem directly from the pharmacophoric features of the available ligands for the target.

Results and Discussion. Case Study 1: Design within a Pseudoreceptor Derived from the Molecular Superposition of a Set of Ligands. A set of five known/claimed ER active training ligands were selected from review literature and the Dewent World Drug Alerts as illustrated in Figure 1.[22−24]

Using the molecular partial similarity superposition tool Quasi2, a pharmacophoric representation for these ligands was obtained allowing full ligand flexibility in the superposing process.[25,26] The key molecular features from this superposition pharmacophore were coded as site-point design constraints for use in Skelgen structure generation. A molecular hull was described for the input points with a user defined radius of 1.25 Å, as illustrated in Figure 2a. Skelgen was then run in ligand mode,

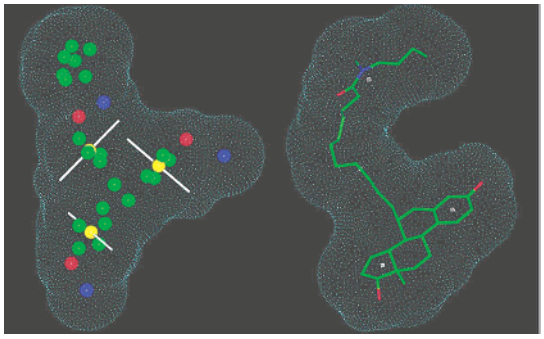

Figure 2. (a, left) ER pharmacophoric points with illustrative inclusion volume rendered as dot surface. Key: green = steric feature, red = ligand H-bond acceptor feature, blue = H-bond donor projection point, stick = aromatic vector, yellow = aromatic ring centroid. (b, right) Bound conformation of ICI 164,384 with illustrative inclusion indicated; coordinates for Skelgen pharmacophore constraints were taken from molecular conformation illustrated.

generating structures so as to satisfy the pharmacophore constraints specified within the defined inclusion volume.

Case Study 2: Design within a Pseudoreceptor Derived from the Bound Conformation of a Single Active Ligand. To challenge notions of traditional de novo structure generation, the bound conformation of a pure steroidal antiestrogen ICI 164,384, as cocrystalized within the ER beta isoform, was extracted from its PDB entry 1HJI.[27,28] In this instance, as the investigation was to reflect a limited availability of ligand information, a looser hull was described using an inclusion volume with a user defined radius of 2.5 Å. The key functional ligand features were described as pharmacophoric-like constraints within this volume, as illustrated in Figure 2b. Again, Skelgen was run in ligand mode to generate structures that satisfied the inclusion volume boundaries and pharmacophore-like constraints provided.

Pharmacophoric Constraints Applied. To generate structures, one must supply Skelgen with sensible and appropriate pharmacophoric constraints. In our two case studies the following constraints were applied as part of the design strategy, drawing from ligand features observed in both the superposed pharmacophore set and the bound active.

- Aromatic feature in region of steroid A ring
- Aromatic or flat ring feature in region of steroid D ring
- Amphoteric feature at approximate para position for feature 1 (optional)
- Amphoteric feature at approximate para position for feature 2 (optional)
- Any two ring systems at intervals extending along the side chain volume.

Satisfaction of these constraints within a Skelgen run would produce an output ligand providing the penalty limit for the new structure was not exceeded.

Output. Our interest in this study has been to focus attention toward the generation of chemotypes that mimic the 'steroid' portion of the antagonists; we have not been particularly concerned with the side chain positioning in the ligands. Only visual inspection through a chemist's eye of the output from a de novo design program can lead to a full assessment of potentially

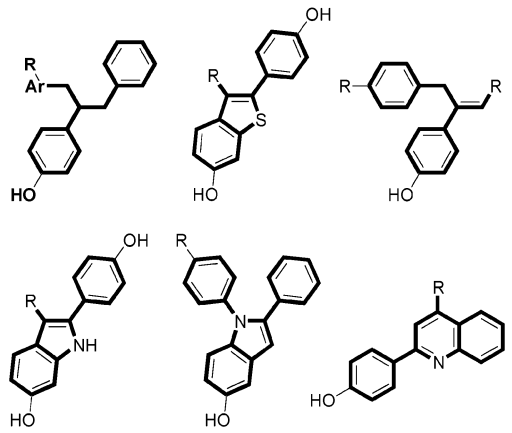

Figure 3. Representative ER scaffolds produced by Skelgen in de novo structure generation using pharmacophoric points contained within an inclusion volume described by superposed active ligands.

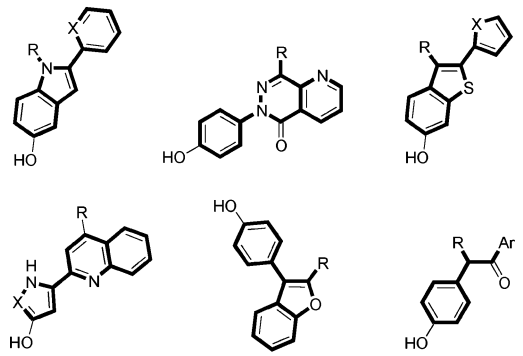

Figure 4. Representative scaffolds produced by Skelgen in de novo structure generation when applied to pharmacophoric points contained within an inclusion volume described by a single known active ligand. X denotes an arbitrary heteroatomic insertion.

interesting binding motifs and scaffolds, yet this assessment is intrinsically subjective. To reduce this subjectivity in analysis, we used a partitioning method to group output on the basis of similarity by maximum common subgraph (MCS) analysis, utilizing our Macos utility previously described.[1] MCS analysis allows for the ready identification of common substructures in de novo output and assists the user in selecting representative compounds for assessment. Within Macos, the common subgraph between two structures, A and B, is found by an exhaustive depth-first search. Common subgraphs are extended one atom at a time, alternating between atoms from A and B. This approach to assessing the similarities in generated structures allows us to focus on scaffold or core similarity, while remaining tolerant of variation in side chain composition.

In case study 1, a total of 446 ligands were generated which satisfied the volume and pharmacophoric constraints applied. MCS grouping of the output identified 144 output classes at a similarity cutoff of 0.5. When multimember class groupings were visually assessed, Skelgen-generated common substructures were readily identified. Representatives of scaffold types generated are illustrated in Figure 3.

It is immediately apparent that many of the major known ER chemotypes are readily generated de novo within the prescribed pharmacophoric constraints including phenylindolyl, phenylbenzothiophenyl, phenylbenzofuranyl (not shown), phenylquinolinyl, and triarylethyl derivative scaffold cores.

In case study 2, a total of 127 ligands were generated that satisfied the applied volume and pharmacophoric constraints. MCS similarity grouping was again employed to identify common substructures, in conjunction with visual analysis of the output. Representatives of the core scaffold types generated using this steroidal input ligand are illustrated in Figure 4. Again, starting from within an inclusion volume described by a steroidal lead compound, and with only consideration of the pharmacophoric features of this known active, Skelgen has produced many of the known ER chemotypes.

In both of these design challenges, the majority of solutions observed consist of a bicyclic ring system connected to another ring via a single bond, as is the case with most known ER inhibitors. Larger ring systems are not contained in the Skelgen template library, nor does Skelgen close rings by introducing additional bonds or linker fragments. Therefore, larger polycyclic ring systems are not observed in the solution set. In the absence of receptor structural information to 'guide' ligand formation through appropriate interaction penalties with amino acids, a higher instance of arbitrary heteroatomic insertions is recorded. Where appropriate, we have rendered such insertions as generic X elements in our representation of the core scaffolds produced. Previous synthetic studies have highlighted and reiterated the eclectic binding preferences of the ER.[29] Comparison of the breadth of scaffolds retrieved in this investigation to those generated in previous work[1] where the ER crystal structure *was* utilized indicates that ligand information alone will suffice to facilitate the generation of multiple chemotypes, although receptor information, by virtue of its higher intrinsic feature content, will return more knowns than when employing ligand derived site models. The success of moving from steroid to nonsteroid chemotypes while using a single compound as a constraint indicates the minimal input requirement possible for de novo design in Skelgen. We find it extremely encouraging that Skelgen, running in the complete absence of receptor information, can reproduce not only so many known chemotypes but also suggests numerous novel structures through heteroatomic and ring substitutions to complement and expand on the existing scaffolds.

Conclusion. The vast nature of input amenable to and available for such ligand-derived de novo design runs indicates this technique's robust and varied application in rational drug discovery. A small number of active ligands could easily be superposed using any appropriate pharmacophore modeling tool or QSAR technique to provide an overall inclusion shape; many such alternate methods exist.[30] Pharmacophoric constraints would then be taken directly from the output. In cases where fewer, or single, active ligand(s) are known, output from a comprehensive conformational search could be used to define the inclusion shape, with pharmacophoric constraints taken from the known active. In ideal cases, some knowledge of a ligand binding mode would be available such as the information gleaned from solid-state NMR studies on GPCR ligands or peptides. This 'bound' conformation of the ligand would provide an excellent starting point for the

rational de novo design of alternate chemotypes with equivalent functionality and allow for rapid generation of follow-up screening candidate libraries in these most challenging of design cases.

References

(1) Stahl, M.; Todorov, N. P.; James, T.; Mauser, H.; Boehm, H. J.; Dean, P. M. A validation study on the practical use of automated de novo design. *J. Comput. Aided Mol. Des.* **2002**, *16* (7), 459−478.

(2) Schneider, G.; Lee, M. L.; Stahl, M.; Schneider, P. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput. Aided Mol. Des.* **2000**, *14* (5), 487−494.

(3) Mancera, R. L. De novo ligand design with explicit water molecules: an application to bacterial neuraminidase. *J. Comput. Aided Mol. Des.* **2002**, *16* (7), 479−499.

(4) Takano, Y.; Koizumi, M.; Takarada, R.; Takimoto Kamimura, M.; Czerminski, R.; Koike, T. Computer-aided design of a factor Xa inhibitor by using MCSS functionality maps and a CAVEAT linker search. *J. Mol. Graph. Model.* **2003**, *22* (2), 105−114.

(5) Schmidt, J. M.; Mercure, J.; Tremblay, G. B.; Page, M.; Kalbakji, A.; Feher, M.; Dunn-Dufault, R.; Peter, M. G.; Redden, P. R. De novo design, synthesis, and evaluation of novel nonsteroidal phenanthrene ligands for the estrogen receptor. *J. Med. Chem.* **2003**, *46* (8), 1408−1418.

(6) Waszkowycz, B. Structure-based approaches to drug design and virtual screening. *Curr. Opin. Drug Discovery Dev.* **2002**, *5* (3), 407−413.

(7) Bogacewicz, R.; Trylska, J.; Geller, M. Ligand design package (Ludi- -MSI) applied to known inhibitors of the HIV-1 protease. Test of performance. *Acta Pol. Pharm.* **2000**, *57* Suppl, 25−28.

(8) Kuhn, P.; Wilson, K.; Patch, M. G.; Stevens, R. C. The genesis of high-throughput structure-based drug discovery using protein crystallography. *Curr. Opin. Chem. Biol.* **2002**, *6* (5), 704−710.

(9) Hicks, S.; Assefa, H.; Sindelar, R. Computer-aided design of enzyme inhibitors: recent studies. *Curr. Opin. Drug Discuss. Dev.* **1999**, *2*, 223−234.

(10) Klebe, G. Recent developments in structure-based drug design. *J. Mol. Med.* **2000**, *78*, 269−281.

(11) De Lucca, G. V.; Erickson-Viitanen, S.; Lam, P. Y. S. Cyclic HIV protease inhibitors capable of displacing the active site structural water molecule, *Drug Discovery Today* **1997**, *2* (1), 6−18.

(12) von Itzstein, M.; Wu, W.-Y.; Kok, G. B.; Pegg, M. S.; Dyason, J. C.; Jin, B.; Phan, T. V.; Smythe, M. L.; White, H. F.; Oliver, S. W.; Colmant, P. M.; Varghese, J. N.; Ryan, D. M.; Woods, J. M.; Bethell, R. C.; Hotham, V. J.; Cameron, J. M.; Penn, C. R. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* **1993**, *363*, 418−423.

(13) Babu, Y. S.; Chand, P.; Bantia, S.; Kotian, P.; Dehghani, A.; El-Kattan, Y.; Lin, T−H.; Hutchinson, T. L.; Elliott, A. J.; Parker, C. D.; Ananth, S. L.; Horn, L. L.; Larver, G. W.; Montgomery, J. A. BCX-1812 (RWJ-270201): Discovery of a novel, highly potent, orally active, and selective influenza neuraminidase inhibitor through structure-based drug design. *J. Med. Chem.* **2000**, *43*, 3482−3486.

(14) Wang, G. T.; Chen, Y.; Wang, S.; Gentles, R.; Sowin, T.; Kati, W.; Muchmore, S.; Giranda, V.; Stewart, K.; Sham, H.; Kempf, D.; Laver, W. G. Design, synthesis, and structural analysis of influenza neuraminidase inhibitors containing pyrrolidine cores. *J. Med. Chem.* **2001**, *44*, 1192−1201.

(15) Chand, P.; Kotian, P. L.; Deghani, A.; El-Kattan, Y.; Lin, T. H.; Hutchinson, T. L.; Babu, Y. S.; Bantia, S.; Elliott, A. J.; Montgomery, J. A. Systematic structure-based design and stereoselective synthesis of novel multisubstituted cyclopentane derivatives with potent antiinfluenza activity. *J. Med. Chem.* **2001**, *44*, 4379−4392.

(16) De Esch, I. J.; Mills, J. E.; Perkins, T. D.; Romeo, G.; Hoffmann, M.; Wieland, K.; Leurs, R.; Menge, W. M.; Nederkoorn, P. H.; Dean, P. M.; Timmerman, H. Development of a pharmacophore model for histamine H3 receptor antagonists, using the newly developed molecular modeling program SLATE. *J. Med. Chem.* **2001**, *44* (11), 1666−1674.

(17) Bureau, R.; Daveu, C.; Lemaitre, S.; Dauphin, F.; Landelle, H.; Lancelot, J. C.; Rault, S. Molecular design based on 3D-pharmacophore. Application to 5-HT4 receptor. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (4), 962−967.

(18) Waszkowycz, B.; Clark, D. E.; Frenkel, D.; Li, J.; Murray, C. W.; Robson, B.; Westhead, D. R. PRO_LIGAND: an approach to de novo molecular design. 2. Design of novel molecules from molecular field analysis (MFA) models and pharmacophores. *J. Med. Chem.* **1994**, *37* (23), 3994−4002.

(19) Manetti, F.; Forli, S.; Maccari, L.; Corelli, F.; Botta, M. 3D QSAR studies of the interaction between beta-tubulin and microtubule stabilizing antimitotic agents (MSAA). A combined pharmacophore generation and pseudoreceptor modeling approach applied to taxanes and epothilones. *Farmaco.* **2003**, *58* (5), 357−361.

(20) Schneider, G.; Neidhart, W.; Giller, T.; Schmid, G. "Scaffold-Hopping" by Topological Pharmacophore Search: A Contribution to Virtual Screening. *Angew. Chem. Int. Ed. Engl.* **1999**, *38* (19), 2894−2896.

(21) Mills, J. E.; Dean, P. M. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput. Aided Mol. Des.* **1996**, *10* (6), 607−622.

(22) Derwent World Drug Alerts, 2002, Derwent Information, London.

(23) Lloyd, D. G.; Meegan, M. J. Recent advances in estrogen receptor antagonists, *Drugs* **2000**, *6* (3) 632−642.

(24) Meegan, M. J.; Lloyd, D. G. Advances in the science of estrogen receptor modulation. *Curr. Med. Chem.* **2003**, *10*, 181−210.

(25) Lloyd, D. G.; Perry, N. P.; Todorov, N. P. T.; Alberts, I. L.; De Esch, I. J. P. Virtual site model derivation and application to lead identification. In *Proceedings of 224th American Chemical Society**CS** National Meeting*, Boston, MA, Aug 18−22, 2002. http://www.lib.uchicago.edu/cinf/224nm/presentations/224 nm015.pdf.

(26) Perry, N. P.; Lloyd, D. G.; Todorov, N. P.; Alberts, I. L. The use of ligand-based site models in virtual high-throughput screening. In *QSAR Proceedings 2002: Euro QSAR 2002: Designing drugs and crop protectants*; Blackwell Publishing Ltd.: Malden, MA, 2003..

(27) Van Den Bemd, G. J.; Kuiper, G. G.; Pols, H. A.; Van Leeuwen, J. P. Distinct effects on the conformation of estrogen receptor alpha and beta by both the antiestrogens ICI 164, 384 and ICI 182, 780 leading to opposite effects on receptor stability. *Biochem. Biophys. Res. Commun.* **1999**, *261* (1), 1−5.

(28) Pike, A. C.; Brzozowski, A. M.; Walton, J.; Hubbard, R. E.; Thorsell, A. G.; Li, Y. L.; Gustafsson, J. A.; Carlquist, M. Structural insights into the mode of action of a pure antiestrogen. *Structure* **2001**, *9* (2), 145−153.

(29) Meegan, M. J.; Hughes, R. B.; Lloyd, D. G.; Williams, D. C.; Zisterer, D. M. Flexible estrogen receptor modulators: design, synthesis, and antagonistic effects in human MCF-7 breast cancer cells. *J. Med. Chem.* **2001**, *44* (7), 1072−1084.

(30) Guner, O. F. History and evolution of the pharmacophore concept in computer-aided drug design. *Curr. Top. Med. Chem.* **2002**, *2* (12), 1321−1332.

JM034222U