Generation and Selection of Novel Estrogen Receptor Ligands Using the *De Novo* Structure-Based Design Tool, SkelGen

Stuart Firth-Clark, Henriëtte M. G. Willems,* Anthony Williams, and William Harris

De Novo Pharmaceuticals Ltd., Compass House, Vision Park, Histon, Cambridge, CB4 9ZR United Kingdom

Received July 21, 2005

A *de novo* design approach to generating novel estrogen receptor (ER) ligands is described. The SkelGen program was used to generate ligands in the active sites of seven crystal structures of ERα. Seventeen high-scoring, diverse structures were selected from the SkelGen output and synthesized without introducing any modifications to the structures. Five ligands, four of which are novel, showed ≤25 μM affinity, with the best compound displaying an IC_{50} of 340 nM. SkelGen can, therefore, be a powerful tool for designing active molecules.

INTRODUCTION

Since the early 1990s, a number of algorithms for *de novo* design—the automated, computational design of ligands to fit protein active sites—have been reported.[1] Data on how effective these algorithms are is scarce. In particular, reports on the likelihood of finding actives in the "raw" output of any of these algorithms do not exist to our knowledge.

Several *de novo* design algorithms have been validated experimentally by the synthesis of active structures. The published validation studies mostly describe how the *de novo* design program was used to generate the framework on which the active molecules were based. For example, SPROUT,[2] GrowMol,[3] and LEGEND[4] were used to generate frameworks that were changed into active molecules by medicinal chemists. In these cases, it is difficult to decide whether the de novo design algorithm or the imagination of the chemist is the greater contributor to the success. Other validation studies report the generation of active structures using large parts of known active compounds as a starting point. For example, LUDI,[6] CombiSMOG,[7] and MCDNLG[8] have been reported to increase the activity of known active ligands by suggesting modifications. In these three cases, more than half of the structure of a known active ligand was retained during the *de novo* design process. This makes it hard to assess whether these programs would be suitable for lead generation. One study that does report actives in the raw output of a de novo design program is a report by Vinkers *et al.* on SYNOPSIS.[5] Some raw output generated by SYNOPSIS was found to display activity between 10 and 80 μM, while the more-active compounds in the report were "slightly modified after additional computations".

We wanted to know if the unmodified output of our *de novo* design algorithm, SkelGen,[9−11] would contain active compounds if run without a known seed fragment. Here, we report active ligands that resulted directly and completely from *de novo* design by SkelGen, which fit the active site of the estrogen receptor (ER).

* Corresponding author phone: +44 1223 238000; fax: +44 1223 238088; e-mail: henriette.willems@denovopharma.com.

METHODS

SkelGen is an automated *de novo* design program that constructs ligands from fragments in a target protein's active site. To restrict structure generation to the active site, the target's coordinates are provided in conjunction with a cuboid that is centered around the ligand in the crystal structure to define the volume available for ligand design. The program proceeds to place a random fragment in the site and subsequently adds to, deletes, mutates, rotates, or translates the fragment while minimizing a 19-term objective penalty function, in which a perfect solution scores zero. The objective function contains weighted terms that penalize inter- and intramolecular clashing, the presence of forbidden substructures, and the violation of scoring function thresholds. A key feature of SkelGen is its ability to take pharmacophoric restraints as input, so that the user can define donor, acceptor, and hydrophobic requirements that must be satisfied to within a given tolerance before any structure is generated. Ligands incur penalties for not fulfilling the required pharmacophoric constraints.

Since our last publication detailing SkelGen's features,[9] the program's fragment set and fragment connection rules have been updated. SkelGen now contains a set of approximately 1700 fragments, which were obtained by fragmenting the World Drug Index with 14 disconnection rules, followed by automated and manual filtering to ensure that the final set is diverse and drug-like. In all fragments, bonds that were broken in the disconnection process were labeled with the disconnection rule used. A set of 25 connection rules specifies how the fragments can be recombined. Therefore, only molecules that are likely to be synthetically tractable are produced by SkelGen. To further improve synthetic tractability and drug-likeness, a substructure lookup routine filters out certain forbidden substructures. Another recent addition to the SkelGen program is the implementation of the ScreenScore scoring function, which has been shown to work well in virtual high-throughput screening exercises.[15]

Design Strategy for ER. To demonstrate that SkelGen can design novel active ligands to protein active sites, we

10.1021/ci0502956 CCC: $33.50
Published on Web 12/02/2005

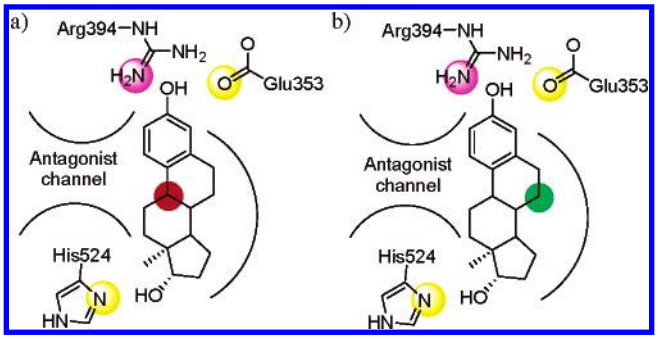


Figure 1. Pharmacophore constraints used for the design of the ERα ligands. In both strategies, hydrogen bonds must be made with two enzyme acceptors (Glu-A353 and His-A524, yellow sphere) and the specified enzyme donor (Arg-A394, magenta sphere). In addition, one ligand point must be satisfied either (a) close to the "antagonist" channel T-junction (red sphere) or (b) in a buried part of the site (green sphere).

have applied the program to ER. Modulators for this target have been developed as therapies for breast cancer, osteoporosis, and cardiovascular diseases.[12,13] Currently, there are 15 crystal structures of ER available in the protein data bank, of which 12 are ERα and three are ER*β*. Because of the larger body of information available for ERα, structure generation focused on this subtype. From the 12 available structures, seven were selected (3ert, 1l2i, 1qkt, 1gwr, 1gwq, 1err, and 1pcg) showing different conformations of active site residues, providing the opportunity for SkelGen to generate alternative ligands within different target structures. For all seven crystal structures, hydrogen atoms were added and minimized on residues located in the active site using the program, InsightII/Discover.[14] In this process, the heavy atoms were fixed to their original positions. All crystal structures were superimposed onto the one with the best resolution—3ert at 1.9 Å resolution. For each of the seven structures used, we defined the active site to include the corresponding ligand with a 5 Å boundary around it. Two sets of pharmacophoric constraints were subsequently prepared to guide the *de novo* design, which are shown in Figure 1. In the first strategy, we specify that ligand-donor atoms must satisfy Glu-A353 and His-A524, while a ligand-acceptor atom must interact with the Arg-A394 side chain. In addition to these restrictions, a hydrophobic atom was required at the T-junction between the main active site cavity and the "antagonist" channel. In a trial run, we noticed that this strategy produced many compounds that did not have surface contact with the upper part of the active site. Therefore, a second strategy was added, in which the same three hydrogen bonding requirements were applied as in the first case, but with the hydrophobic requirement located within a more buried part of the site. Each point slightly varies in position depending on the crystal structure used, since specific atoms of the cocrystallized ligands were used to define the hydrophobic points. The placement of these pharmacophore restraints, therefore, depends on the coordinates of the selected ligand atom in each crystal structure.

SkelGen Setup. The ScreenScore scoring function in SkelGen provides an "interaction energy", giving greater negative values for more favorable ligands. It is possible to set a maximum ScreenScore limit, so that all SkelGen output must have scores below a certain value. Such low values can be obtained by building larger ligands that provide a greater contact surface with the protein. To avoid excessively large structures being produced, it is possible to specify a ScreenScore per heavy atom threshold to limit the size of ligands produced, while providing greater complementarity with the active site. To set the maximum values for both these criteria, the crystal structure ligand was scored in the corresponding active site in its position in the crystal structure. A polar atom ratio constraint of 0.25 was applied to prevent the generation of ligands in which more than 25% of the heavy atoms are polar.

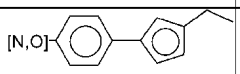
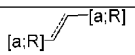
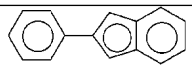
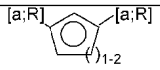
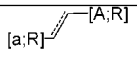
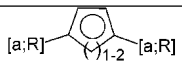
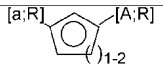
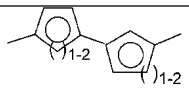
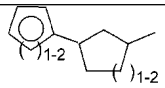
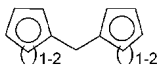
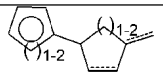
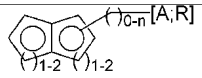
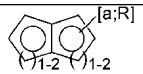
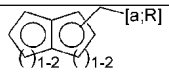
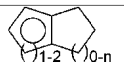
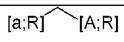
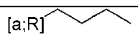
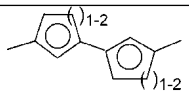
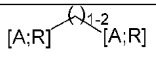
For each strategy against each crystal structure, 1000 solutions were requested from SkelGen, producing a maximum of 14 000 solutions across seven crystal structures. A total of 13 311 ligands were actually generated, as some attempts did not generate a ligand in the maximum time allowed. There were a total of 5492 unique ligands present in this set.

Compound Selection. For each crystal structure and strategy, the generated ligands were sorted in terms of their ScreenScore and duplicates were removed. The top 25 structures for each crystal-structure−strategy combination were then selected, giving a total of 350 designs to be considered for synthesis.

The selection of compounds for progression was made on the basis of synthetic tractability and exploration of the structural diversity within the set of 350 compounds. To ensure that a diverse set of compounds would be selected for synthesis, the structures were clustered. Several automated clustering procedures were applied to the data set, including Jarvis−Patrick clustering and Ward's hierarchical clustering on Daylight fingerprints and Cerius2 topological descriptors.[14] None of the above methods produced clusters with similar chemical structures. Therefore, automated clustering was abandoned in favor of manual clustering using Daylight SMARTS.[16] It was found that the data set could be split into 22 clusters with clear structural similarity by applying the SMARTS listed in Table 1 in order.

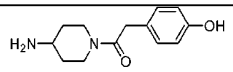
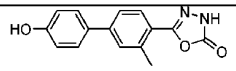
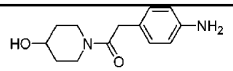
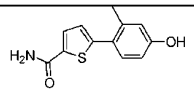
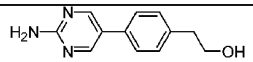
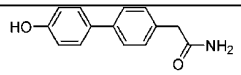
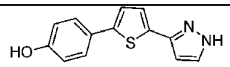
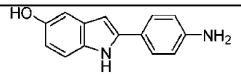
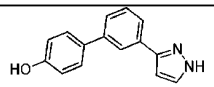
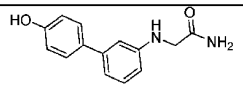
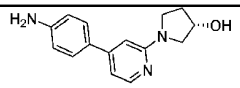
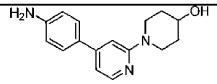
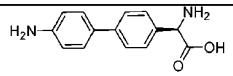
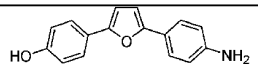
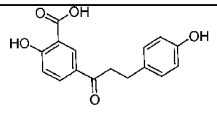
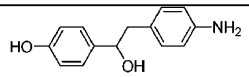
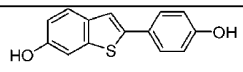
An analysis of the 22 clusters showed that cluster 20 contained exclusively very flexible molecules. This cluster was discarded as not drug-like. All remaining compounds were ranked using an automated retrosynthetic analysis function that incorporates reagent lookup in the Available Chemicals Directory. This uses a list of 107 possible chemical transformations, including simple functional group chemistry, nucleophilic substitution, aromatic substitution, heterocycle formation, and condensation reactions. The retrosynthetic analysis function has completed its task when all the reagents generated by the retrosynthetic analysis are inactive, that is, either found in the database or none of the transforms can be applied. At this point, a route fraction (RF) score is calculated for each route as the mass fraction of the target molecule represented by reagents that were found in the database. The RF score, therefore, ranges from 0.0 to 1.0, with a score of 1 indicating that all reagents are available, so the target should be synthetically accessible. Seven clusters were found to have a best RF score of less than 0.9, suggesting that none of the molecules in these clusters were synthetically accessible. However, an inspection by a chemist indicated that three clusters did contain synthetically accessible molecules (clusters 5, 7, and 17). It was found that the retrosynthetic analysis tools lacked the transforms required to analyze these molecules correctly. The other four clusters with a low RF value were discarded (clusters 9, 15, 18, and

Table 1. Clustering Results

Cluster	SMARTS	Depiction[a]	Size (excluding duplicates)	Best Route Fraction	Selected	Synthesized
1	a1aa([!R][!R])aa1a2aaa([N,O])aa2		25 (15)	1	2	1
2	[!R][!R]a1aaa(aa1)a2a[a;R1][a;R1]aa2		90 (62)	1	7	3
3	[a;R][!R]~[!R][a;R]		24 (18)	1	3	1
4	a1aaa(aa1)a2aa3aaaaa3a2		10 (8)	1	3	2
5	[a;R]-[a;R][a;R][a;R]-[a;R]		12 (11)	0.7	3	3
6	[a;R]-[!R]~[!R]-[A;R]		12	1	3	2
7	[a;R]-[a;R][a;R][a;R][a;R]-[a;R]		5	0.37	1	1
8	[a;R]-[a;R][a;R][a;R]-[A;R]		9	1	2	2
9	[a;R]-[!R]~[!R]~[!R]-[A;R]		9	0.25		
10	[a;R]-[!R]~[!R]~[!R]-[a;R]		7	1	2	1
11	[!R]-a[a;R1][a;R1]a-[a;R][a;R][a;R]-[!R]		23 (21)	1	2	1
12	[a;R1][a;R1][a;R1][a;R1]-[A;R][A;R][A;R]-[!R]		16	1[b]		
13	[a;R1][a;R1][a;R1][a;R1][a;R1]-[!R]-[a;R1][a;R1][a;R1][a;R1[a;R1]		17	1	1	
14	[a;R1][a;R1][a;R1][a;R1][a;R1]-[A;R][A;R]~[A;R][A;R]~[!R]		19 (18)	1	1	
15	[a;R1][a;R2]([a;R2])[a;R1].[A;R]		12 (11)	0.19		
16	[a;R1][a;R2]([a;R2])[a;R1].[a;R]-[a;R]		8	1	2	
17	[R1][a;R2]([a;R2])[R1].[a;R]-[!R]-[a;R]		9 (8)	0.45	1	
18	[A;R1][a;R2]([a;R2])[R1]		12	0.32		
19	[a;R]-[!R]-[A;R]		14 (13)	1	1	
20	[a;R]-[!R]~[!R]~[!R]~[!R]		7	1		
21	[!R]aaa-aaa[!R]		4 (3)	1	1	
22	[A;R]-[!R]-[A;R] and [A;R]-[!R][!R]-[A;R]		5	0.07		

[a] Any atom in a substructure can be any element, apart from one atom in the substructure for cluster 1, which has an "N or O" requirement. Circles in rings indicate that the ring has to be aromatic; rings without circles have to be aliphatic. Bonds drawn with a single and a dotted line can be any bond type. [a;R] refers to any aromatic atom that is in a ring of any size. [A;R] refers to any aliphatic atom that is in a ring of any size. [b] The single compound with RF = 1 had two stereocenters and was, therefore, regarded as not synthetically tractable. The rest of the compounds in this cluster had RF values < 0.3.

Table 2. Compound Activity in the ERα Ligand-Binding Assay

DNP No.	Origin	Strategy[a]	Cluster No	Structure	Inhibition @ 10μM[b]	IC_{50} (μM)[b]
17β-estradiol	Reference					0.027
DNP-13794	1qkt	b	6		1%	
DNP-13795	3ert	a	7		55%	25
DNP-13796	1gwr	b	6		0%	
DNP-13797	1gwr	a	1		2%	
DNP-13798	1gwr	a	2		4%	
DNP-13799	1qkt	b	2		25%	
DNP-13800	1qkt	a	5		46%	24
DNP-13801	1qkt	a	4		16%	
DNP-13802	1err	a	5		41%	18
DNP-13805	3ert	b	11		4%	
DNP-13806	3ert	a	8		1%	
DNP-13807	3ert	a	8		9%	
DNP-13808	1gwr	a	2		9%	
DNP-13809	1qkt	b	5		94%	4.1
DNP-13810	1qkt	a	10		11%	
DNP-13811	1l2i	a	3		7%	
DNP-13813	1gwq	a	4		110%	0.34

[a] Please refer to Figure 1 for definitions of strategies a and b. [b] Biological screening was performed at Cerep using the fluorescence polarization assay, 813-ah, with human recombinant ERα.

22). In addition, cluster 12 had to be discarded because the only molecule with an RF of 1 contained two stereocenters that could not be introduced easily.

From the 16 remaining clusters, 35 compounds were selected for synthesis. Selection was based on synthetic tractability, cost, and diversity within the cluster. A total of 17 compounds from 10 clusters were synthesized within our chemistry time frame.

Biological Testing. Biological screening was performed at Cerep using the fluorescence polarization assay, 813-ah, with human recombinant ERα, using 17β-estradiol as a reference. IC_{50}'s were determined for compounds with >40% inhibition at 10 μM.

RESULTS AND DISCUSSION

From the synthesized set of 17 compounds, five were found to show >40% inhibition at 10 μM; this represents a 30% success rate. Actives came from three clusters, 4, 5, and 7, which all contain rigid structures with three aromatic rings. Actives were generated from both strategies a and b, and from four different ERα crystal structures. The remaining data set of 12 inactive structures is dominated by species with (a) the hydroxyl substituent on an aliphatic ring (selected from clusters 6 and 8, e.g., DNP-13807), (b) a terminal interacting group not connected to or incorporated in a ring (selected from clusters 1, 2, and 11, e.g., DNP-13798), or (c) a structure that contains a nonrigid connection between rings (selected from clusters 3 and 10, e.g., DNP-13811). The inactivity of the compounds in groups a and b could be the result of insufficient acidity of the interaction group. A limitation of the ScreenScore scoring function is that it does not take into account acidity. Results for all 17 compounds are shown in Table 2.

A preliminary prior art search of the initial 35 structures selected for synthesis had already indicated that DNP-13813 and its indole analogue belong to a known ER chemical class.[17,18] The four remaining active compounds were subjected to further prior art searches using the CAS Registry,[19] MARPAT,[20] MARPATPREV,[21] and Beilstein[22] databases. These prior art searches indicate that we are the first to publicly disclose that structures DNP-13795, DNP-13800, DNP-13802, and DNP-13809 have ER activity.

Relevance of Duplicates. Having obtained affinity data for 17 of the ligands SkelGen suggested, we reviewed our data hoping to gain some insight into the relevance of the frequency with which ligands were generated. SkelGen generates many duplicates if it is allowed sufficient simulated annealing attempts. We investigated whether the frequency of occurrence of a ligand is correlated to its activity. The most frequently occurring ligand that was synthesized is DNP-13798, which was generated 82 times over all crystal-structure−strategy combinations. Yet, this molecule has no affinity for the ERα receptor. On the other hand, micromolar ERα ligands DNP-13795, DNP-13802, and DNP-13809 are quite rare occurrences, appearing only once or twice in the course of 14 000 SkelGen attempts. Finally, the ligand with the highest affinity in the synthesized set, DNP-13813, occurred 30 times. Clearly, the frequency of occurrence cannot be used as an indicator for receptor affinity.

CONCLUSION

In conclusion, using the *de novo* structure-based design tool SkelGen, we designed several thousand unique ligands to fit the ERα active site. The 350 solutions with the best ScreenScore were clustered manually, and from this reduced set, 35 diverse, synthetically tractable compounds were selected for synthesis. Of the 17 SkelGen solutions that were synthesized in the available time, five show affinity at 25 μM or better against ERα. Four of these actives are novel. This demonstrates that the unmodified output of our *de novo* design algorithm, SkelGen, when applied to the ERα receptor, contains active compounds and that actives can be obtained without using a known seed fragment.

ACKNOWLEDGMENT

We thank Dr. Per Kallblad and Dr. Philip Dean for helpful discussions and Dr. Nikolay Todorov for providing the software.

REFERENCES AND NOTES

(1) Bohacek, R. S.; McMartin, C. Modern Computational Chemistry and Drug Discovery: Structure Generating Programs. *Curr. Opin. Chem. Biol.* **1997**, *1*, 157−161.
(2) Han, Q.; Dominguez, C.; Stouten, P. F. W.; Park, J. M.; Duffy. D. E.; Galemmo, R. A.; Rossi, K. A.; Alexander, R. S.; Wong, P. C.; Wright; M. M.; Luettgen, J. M.; Knabb, R. M.; Wexler, R. R. Design, Synthesis and Biological Evaluation of Potent and Selective Amidino Bicyclic Factor Xa Inhibitors. *J. Med. Chem.* **2000**, *43*, 4398−4415.
(3) Ripka, A. S.; Satyshur, K. A.; Bohacek, R. S.; Rich, D. H. Aspartic Protease Inhibitors Designed from Computer-Generated Templates Bind As Predicted. *Org. Lett*. **2001**, *3*, 2309−2312.
(4) Honma, T.; Hayashi, K.; Aoyama, T.; Hashimoto, N.; Machida, T.; Fukasawa, K.; Iwama, T.; Ikeura, C.; Ikuta, M.; Suzuki-Takahashi, I.; Iwasawa, Y.; Hayama, T.; Nishimura, S.; Morishima, H. Structure-Based Generation of a New Class of Potent Cdk4 Inhibitors: New *de novo* Design Strategy and Library Design. *J. Med. Chem.* **2001**, *44*, 4615−4627.
(5) Vinkers, M. H.; de Jonge, M. R.; Daeyaert, F. F. D.; Heeres, J.; Koymans, L. M. H.; van Lenthe, J. H.; Lewi, P. J.; Timmerman, H.; Van Aken, K.; Janssen, P. A. J. SYNOPSIS: SYNthesize and OPtimize System in Silico. *J. Med. Chem.* **2003**, *46*, 2765−2773.
(6) Pisabarro, T. M.; Ortiz, A. R.; Palomer, A.; Cabré, F.; García, L.; Wade, R. C.; Gago, F.; Mauleón, D.; Carganico, G. Rational Modification of Human Synovial Fluid Phospholipase A_2 Inhibitors. *J. Med. Chem.* **1994**, *37*, 337−341.
(7) Grzybowski, B. S.; Ishchenko, A. V.; Kim, C. Y.; Topalov, G.; Chapman, R.; Christianson, D. W.; Whitesides, G. M.; Shaknovich, E. I. Combinatorial computational method gives new picomolar ligands for a known enzyme. *PNAS* **2002**, *99*, 1270−1273.
(8) Gehlhaar, D. K.; Moerder, K. E.; Zichi, D.; Sherman, C. J.; Ogden, R. C.; Freer, S. T. *De Novo* Design of Enzyme inhibitors by Monte Carlo Ligand Generation *J. Med. Chem.* **1995**, *38*, 466−472.
(9) 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*, 459−478.
(10) Todorov, N. P.; Dean, P. M. A branch-and-bound method for optimal atom-type assignment in de novo ligand design. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 335−349.
(11) Todorov, N. P.; Dean, P. M. Evaluation of a method for controlling molecular scaffold diversity in de novo ligand design. *J. Comput.-Aided Mol. Des*. **1997**, *11*, 175−192.
(12) Meegan, M. J.; Lloyd, D. G. Advances in the science of estrogen receptor Modulation. *Curr. Med. Chem.* **2003**, *10*, 181−210.
(13) Sato, M.; Grese, T. A.; Dodge, J. A.; Bryant, H. U.; Turner, C. H. Emerging Therapies for the prevention or treatment of postmenopausal osteoporosis. *J. Med. Chem.* **1999**, *42*, 1−24.
(14) *InsightII/DISCOVER*; *Cerius2*; Accelrys: San Diego, CA.
(15) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.
(16) *SMARTS*; *Merlin*; Daylight Chemical Information Systems: Mission Viejo, CA.

(17) Schopfer, U.; Schoeffter, P.; Bischoff, S. F.; Nozulak, J.; Feuerbach, D.; Floersheim, P. Toward Selective ERα Agonists for Central Nervous System Disorders: Synthesis and Characterization of Aryl Benzthiophenes. *J. Med. Chem.* **2002**, *45*, 1399−1401.

(18) von Angere, E.; Prekajac, J.; Strohmeier, J. 2-Phenylindoles. Relationship between structure, estrogen receptor affinity, and mammary tumor inhibiting activity in the rat. *J. Med. Chem.* **1984**, *27*, 1439−1447.

(19) Registry: highest RN 790189−55−8; Chemical Abstracts Service: Columbus, Ohio. http://www.cas.org/ONLINE/DBSS/registryss.html (accessed June 2004).

(20) MARPAT: Vol. 141, Iss. 22; Chemical Abstracts Service: Columbus, Ohio. http://www.cas.org/ONLINE/DBSS/marpatss.html (accessed June 2004).

(21) MARPATPREV: 20041130/ED; Chemical Abstracts Service: Columbus, Ohio. http://www.cas.org/ONLINE/DBSS/marpatprevss.html (accessed June 2004).

(22) Beilstein: 2004/11/03/ED; Chemical Abstracts Service: Columbus, Ohio. http://www.cas.org/ONLINE/DBSS/Beilsteinss.html (accessed June 2004).

CI0502956