Identification of "Latent Hits" in Compound Screening Collections

Jordi Mestres*,† and Gerrit H. Veeneman‡

Computational Medicinal Chemistry, Organon Laboratories Ltd., Newhouse ML1 5SH, Scotland, U.K., and Lead Discovery Unit, N.V. Organon, 5340 BH, Oss, The Netherlands

Received April 11, 2003

Abstract: The relatively low hit rates found from high-throughput screening have raised a question on whether this technology alone is sufficient to maximally exploit the full potential of current corporate screening collections. The present study introduces a knowledge-based strategy for identifying "latent hits", i.e., inactive compounds that could potentially be promoted to hits through simple chemical transformations. Examples are given of submicromolar agonist hits derived from the corresponding latent hits for the estrogen receptor.

Introduction. Over the past few years, the pharmaceutical industry has relied heavily on high-throughput screening (HTS) technology for the identification of hits. Large investments and logistic efforts have been in setting up robotic equipments and in acquiring, assembling, and maintaining big corporate compound collections. However, alongside the established value of HTS as a means for identifying hits from screening collections, it has been widely recognized that the quantity, quality, and diversity of the hits identified have been below original expectations.[1] In addition, a number of important quality issues have emerged, namely, quality of the assays (leading often to poor quality of data, large numbers of false positives, and tedious active confirmation process), quality of the collection (purity and stability of compounds, as well as origin, composition, and representation of the collection), and quality of the logistics (purchase, tracking, handling, and storage of compounds and data management). The combination of some or all of these issues is probably responsible for the relatively low hit rates found and has led to a significant increase of the cost of maintenance of HTS.

Despite all efforts, everyone is aware that any corporate screening library is intrinsically incomplete. While the estimated number of synthesizable compounds is on the order of 10^{40}, corporate screening collections contain on the order of 10^6 compounds. It seems therefore naive to think that compounds possessing the *optimal* features arranged in an *optimal* way around a core structure to bind to our protein target of interest will habitually be present in our screening collection. As shown recently, the probability of a compound having the right key functional groups arranged optimally decreases dramatically as the complexity of the molecule increases.[2] In contrast, it seems likely that compounds having *almost* the right features arranged *almost* optimally around a core structure will be indeed present in compound libraries. The subtle differences between the *optimal* and the *almost optimal* presence and arrangement of the key structural features in a compound may ultimately result in detection of activity for that compound in a HTS assay.

Part of the problem lies in the fact that a significant number of compounds present in corporate screening collections were originally acquired from external chemical suppliers with the aim of covering as much chemical space as possible within the size of the compound collection. Consequently, compound selections were mainly directed by strict diversity criteria, which meant that in most instances only the centroid compound of a cluster containing many structurally similar compounds was selected for purchase. Correspondingly, the chances of that centroid compound having the optimal features arranged in the optimal way are very slim. As recently shown, this centroid selection strategy may potentially lead to a 70% chance that the activity within the cluster will not be discovered.[3]

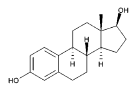
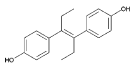
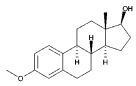
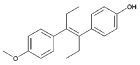
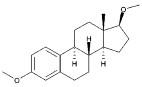
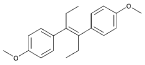
As a result of all these factors, realistically, by just screening, we seldom find optimal highly active compounds but at best suboptimal low active compounds that still require going through extensive and expensive optimization programs. However, an unknown number of nonoptimal inactive compounds that could potentially be promoted to active compounds by means of simple chemical transformations remains often entirely overlooked after completion of an HTS campaign. These "potentially active inactive compounds" will be referred to as "latent hits", and their identification is the focus of the present study. The particular knowledge-based strategy adopted for the identification of latent hits for the estrogen receptor subtype α (ERα) and examples of ERα agonist hits derived from the corresponding latent hits are presented next.

Identification of Latent Hits. The similarity-property principle states that compounds that are structurally similar are expected to have similar activities.[4] However, every medicinal chemist has often noted that some structural differences have stronger effects on the activity than others. Small modifications will produce in some cases a slight modulation of the activity, whereas in other cases they will result in a dramatic loss of activity. These effects were observed in the context of an HTS campaign aiming at identifying ERα agonists. Table 1 illustrates the effect on the HTS data of small structural changes to the ER endogenous steroid hormone 17β-estradiol. On the basis of these HTS data, some qualitative structure−activity relationships can be derived. With 17β-estradiol (**1**) as a reference, Table 1 shows that substitution of the 3-hydroxy group by a methoxy group (**2**)[5] results in a significant loss of activity whereas substitution of both the 3- and 17β-hydroxy groups by methoxy groups (**3**)[6] leads to a complete loss in activity. The relative loss in activity can be rationalized by considering the number of mutation points in the structure of each compound with respect to the key features present in 17β-estradiol, what can be referred to as "pharmacophore latency".

* To whom correspondence should be addressed. Phone: +44 01698 736109. Fax: +44 01698 736187. E-mail: j.mestres@organon.co.uk.
† Computational Medicinal Chemistry.
‡ Lead Discovery Unit.

10.1021/jm034078c CCC: $25.00
Published on Web 07/01/2003

Table 1. Structure, Pharmacophore Latency, and HTS Data (% Change at 10^{-7} M with Respect to 17β-Estradiol) for 17β-Estradiol (**1**) and Its 3-Methyl Ether (**2**) and 3,17β-Dimethyl Ether (**3**) Analogues[a]

Compound	Latency	HTS data	Compound	Latency	RBA[7]
 (**1**)	0	100	 DES	0	2.60
 (**2**)	1	76	 DES methyl ether	1	1.31
 (**3**)	2	0	 DES dimethyl ether	2	-1.25

[a] For comparison, RBA data[7] for diethystilbestrol (DES) and its methyl ether and dimethyl ether analogues are also reported.

Interestingly, similar trends were reported recently for the corresponding analogues of diethylstilbestrol (DES),[7] a potent nonsteroidal ERα agonist.

The relevance of this simple illustrative example is that one can envisage a situation where, of the three compounds, only **3** would be present in the corporate screening collection. In such an unwanted scenario, one could miss the "perfect" molecular scaffold to interact with the receptor cavity in ERα just because **3** does not have the optimal features at the 3 and 17β positions. Compound **3** could therefore be considered a latent hit that can be promoted to a true hit by means of a simple transformation of the methoxy groups into the corresponding hydroxy groups, provided knowledge about the relevance of having hydroxy groups in those positions is available. The same argument can be applied to the dimethyl ether analogue of DES. Therefore, key to the identification of latent hits is the previous knowledge of the main pharmacophoric features required in a compound for activity, knowledge that is often available directly from ligand-bound protein X-ray crystal structures and/or indirectly from the structure−activity relationships of a different chemical series.

With this in mind, we embarked on the identification of latent hits in our corporate screening collection as a complementary means to HTS for generating nonsteroidal ERα agonists. To this aim, a subset of our screening collection containing 133 836 compounds was subjected to a filtering process to remove all compounds having a low probability of possessing the characteristics we were looking for in a nonsteroidal ERα agonist. The resulting set of 11 047 compounds was then virtually screened and ranked against DES using a ligand-based flexible superposition approach.[8] Analysis of the top-ranking compounds resultant from the virtual screen identified the two potential latent hits shown in Table 2. Both compounds did not pass the priority cutoff activity value of 70% at 10^{-6} from the HTS data and were therefore disregarded in the first instance. Note that the qualitative trend previously observed in the 17β-estradiol analogues of Table 1 is reproduced again here for these two compounds and the higher the pharmacophore latency is in the compound, the higher

Table 2. Structure, Pharmacophore Latency, HTS Data (% Change at 10^{-6} M with Respect to 17β-Estradiol), and EC_{50} Values of the Two Latent Hits Identified in the Compound Screening Collection

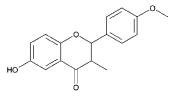
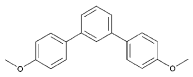
Latent Hit	Latency	HTS data	EC_{50}
 (**4**)	1	67%@10^{-6}	> 10μM
 (**5**)	2	7%@10^{-6}	> 10μM

Figure 1. Three-dimensional superposition of **4** (green) and **5** (orange) onto the reference structure of diethylstilbestrol (white). Oxygen atoms from all compounds are red. For the sake of clarity, all hydrogen atoms have been omitted.

the chances are that the compound shows poor activity despite potentially containing an interesting central scaffold. The lack of significant activity in the compounds was quantitatively confirmed from their EC_{50} values.

One of the advantages of using a three-dimensional superposition method in the last stage of a virtual screening is the possibility of visually inspecting the binding mode hypothesis derived for the compounds of interest relative to the reference compound. The superposition of **4**[9] and **5**[10] onto the structure of DES (see Table 1) is presented in Figure 1. Interestingly, the binding mode proposed for **4** is reminiscent of the binding mode observed for genistein, a structurally related compound, in the cavity of ERβ.[11] As can be observed in Figure 1, the methoxy groups in **4** and **5** are spatially found in the vicinities of the hydroxy groups in DES, thus confirming their potential as latent hits. On this basis, chemical conversion of the methoxy groups into hydroxy groups should introduce the right features in the compounds and, given their arrangement around the core structure, such transformations should result in a significant enhancement of their activity.

The results of introducing hydroxy groups in the two potential latent hits are shown in Figure 2. As can be observed, the removal of the one-point latency in **4** and the two-point latency in **5** resulted in ERα agonist hits with EC_{50} values of 0.004 μM (**6**)[9] and 0.8 μM (**7**),[9] respectively. The generation of these submicromolar hits provides the definite confirmation that **4** and **5** were indeed latent hits, silently disguised in our screening collection and just waiting for the right transformation

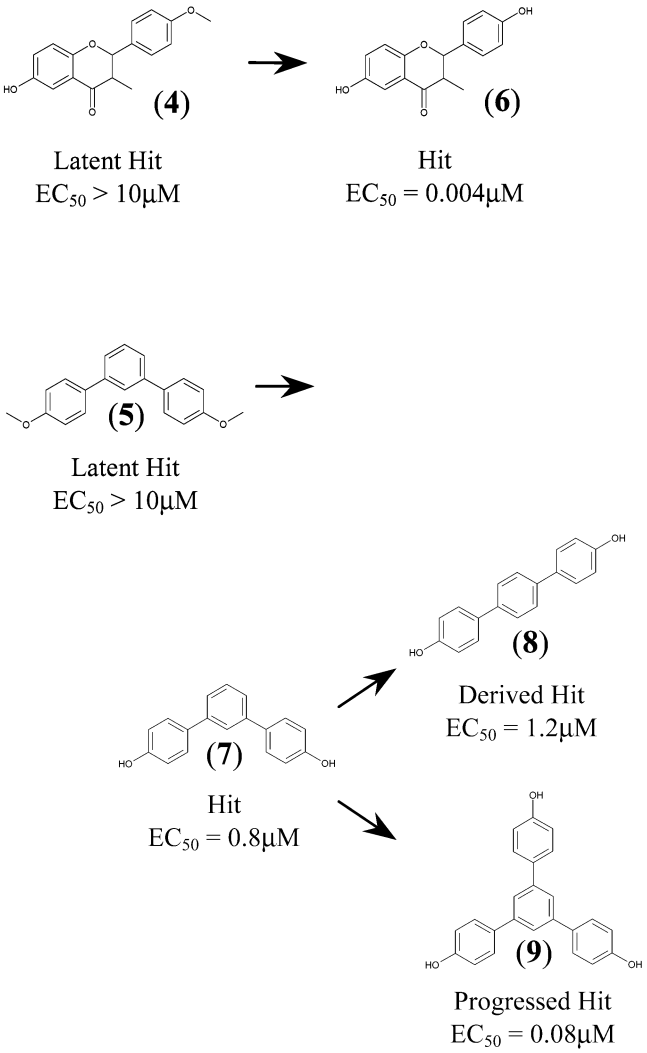


Figure 2. Promotion of latent hits (**4** and **5**) into submicromolar hits (**6** and **7**) by chemical transformation of the methoxy groups into the corresponding hydroxy groups. Compounds **8** and **9** provide evidence of further scope for optimization of **7**.

to be awakened and promoted to true hits. Furthermore, in an effort to illustrate the potential scope for progressibility of the hits generated, some analogues of **7** were synthesized, leading to the identification of compounds that either retain (**8**,[12] $EC_{50} = 1.2\ \mu M$) or improve (**9**,[13] $EC_{50} = 0.08\ \mu M$) the original activity of **7**.

Conclusions. The results presented in this work not only provide evidence of the existence of latent hits in corporate screening collections but also demonstrate that latent hits can be awakened and transformed into submicromolar hits, provided some knowledge of key pharmacophoric features is available. The concept of "latent hits" complies fully with the concept of "privileged scaffolds",[14] which may have implications for setting future screening strategies. In the context of the present work, acknowledging that all compounds belonging to a chemical series represented by an active ligand identified for a particular target can be considered latent hits for other family-related targets is equivalent to recognizing that the scaffold contained in that active compound has the potential of being a privileged scaffold for the entire target family. The combination of the "latent hit" and "privileged scaffold" concepts thus provides a rational hypothesis for pursuing hit identification strategies based on targeted libraries designed around scaffolds present in active compounds and supports current trends in the pharmaceutical industry in the application of chemogenomics approaches to drug discovery.[15]

Materials and Methods. Screening Database. Three-dimensional coordinates of a portion of our corporate screening collection containing 133 836 compounds were generated with the program CORINA.[16] Hydrogen atoms were added automatically and Gasteiger−Marsili atomic charges[17] were generated by using Sybyl 6.8.[18]

ER Pharmacophore. On the basis of a database of in-house and literature ERα agonist compounds and on structural information extracted from ligand-bound protein X-ray crystal structures, a simplistic two-point pharmacophore model was derived consisting of a six-membered ring containing a hydroxy group lying 9.2−12.6 Å from another hydroxy group. The high degree of simplicity of the pharmacophore model can be justified by the fact that it will be used only to filter out any compound in the screening collection not possessing these key features. In this respect, the presence of a "pharmacophore latency" in a molecule is defined as the absence of any of the key hydroxy groups identified in the two-point pharmacophore model, the hydroxy groups having been replaced by other functional groups from which chemical transformation to a hydroxy group is in principle synthetically feasible.

Atom-Centered Feature Fingerprints. Atoms in molecules are assigned to one or more of four pharmacophoric features, namely, hydrophobic, aromatic, hydrogen-bond acceptor, and hydrogen-bond donor. The assignment is based on Sybyl atom types.[19] Atom-centered feature fingerprints are then derived from distances between pairs of features using a 2 Å binning. Similar approaches can be found in the literature.[20,21] On the basis of the analysis of our ERα agonist database, because ERα agonists contain a fairly rigid central scaffold, feature fingerprints were directly obtained from single CORINA three-dimensional structures for all compounds in the screening database.

Filtered Screening Database. The original screening database containing 133 836 compounds was subject to a filtering process to reduce the number of compounds that will be then processed with a flexible ligand superposition method. Three different filters were applied: (i) pharmacophoric filter using feature fingerprints obtained from the two-point pharmacophore model up to a latency of 2; (ii) torsional filter using a number of rotatable bonds less than 5; (iii) steroid filter using the steroidal graph framework. A total of 11 047 compounds passed the three filters, and that would constitute our filtered screening database. As a validation, all nonsteroidal compounds present in our ERα agonist database passed the three filters.

Flexible Ligand Superposition. Three-dimensional superpositions were obtained with the program MIMIC.[22,23] The standard two types of steric and electrostatic Gaussian-based molecular fields were used, and the default weighted 2(steric):1(electrostatic) cosine-like similarity index was applied during the flexible optimizations with minimum sampling (90° search).

Diethylstilbestrol (see Table 1 for structure) in the bound conformation, found when forming a complex with ERα (PDB code 3ERD),[24] was used as the reference molecule onto which each compound in the filtered screening database was flexibly superimposed.

Chemistry. 17β-Estradiol and progesterone were purchased from Aldrich. 1,3-Bis(4-hydroxyphenyl)benzene was obtained from ChemBridge. All other compounds were prepared as described in the corresponding references.

Cell Line. Chinese hamster ovary (CHO) cells, derived from CHO K1 cells obtained from the American Type Culture Collection (Rockville, MD), contained the human ERα and the rat oxytocin promotor (RO) with firefly luciferase reporter gene (LUC) (hERα-RO-LUC (clone 2B1-1E9)). These cells were cultured in medium with 5% charcoal-treated supplemented bovine calf serum at 37 °C in Roux flasks (175 cm^3) and flushed with 5% CO_2 in air until pH 7.2−7.4 was reached. Media were renewed every 2−3 days and were all free of phenol red.

In Vitro Transactivation Data. For transactivation studies, CHO cells were stably transfected with hERα-RO-LUC. The assay was done as described previously.[25] An amount of 5×10^4 cells were seeded into a 384/96-well plate for HTS/EC_{50} data and incubated with compounds for 16 h in medium with 5% charcoal-treated supplemented bovine calf serum at 37 °C in a humidified atmosphere of air with 5% CO_2. For HTS data, of the total 90 μL incubation volume, 45 μL of LucLite was added for cell lysis and luciferase measurement. For EC_{50} data, of the total 250 μL incubation volume, 200 μL was removed and 50 μL LucLite was added for cell lysis and luciferase measurement. Luciferase activity was measured in a Topcount luminescence counter (Canberra Packard). Compounds were tested at 10^{-6} M for HTS data. Full agonistic curves and EC_{50} values were determined from 10^{-5}−10^{-12} M with a dilution factor of 3.16.

Acknowledgment. We are indebted to Rob van der Hulst, Edwin de Zwart, Bas van de Kar, Karen McGurk, and Willem Schoonen for the synthesis and testing of the compounds presented in this work.

References

(1) Lahana, R. How Many Leads from HTS? *Drug Discovery Today* **1999**, *4*, 447−448.
(2) Hann, M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856−864.
(3) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* **2002**, *45*, 4350−4358.
(4) Johnson, M. A., Maggiora, G. M., Eds. *Concepts and Applications of Molecular Similarity*; Wiley: New York, 1990.
(5) Akamanchi, K. G.; Varalakshmy, N. R. Aluminium Isopropoxide-TFA, a Modified Catalyst for Highly Accelerated Meerwein−Ponndorf−Verley (MPV) Reduction. *Tetrahedron Lett.* **1995**, *36*, 3571−3572.
(6) Node, M.; Kumar, K.; Nishide, K.; Ohsugi, S.; Miyamoto, T. Odorless Substitutes for Foul-Smelling Thiols: Synthesis and Applications. *Tetrahedron Lett.* **2001**, *42*, 9207−9210.
(7) Shi, L. M.; Fang, H.; Tong, W.; Wu, J.; Perkins, R.; Blair, R. M.; Branham, W. S.; Dial, S. L.; Moland, C. L.; Sheehan, D. M. QSAR Models Using a Large Diverse Set of Estrogens. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 186−195.
(8) Despite the wealth of structural information on the estrogen receptor from X-ray crystallography, note that docking methods are not well suited for identifying latent hits because these compounds are heavily penalized by scoring functions because of the lack of the key functionalities for interacting with the receptor cavity.
(9) Merck. German Patent NL 6600810, 1965.
(10) Saednya, A.; Hart, H. Two Efficient Routes to *m*-Terphenyls from 1,3-Dichlorobenzenes. *Synthesis* **1996**, 1455−1458.
(11) Pike, A. C. W.; Brzozowkki, A. M.; Hubbard, R. E.; Bonn, T.; Thorsell, A.-G.; Engström, O.; Ljungren, J.; Gustafsson, J.-Å.; Carlquist, M. Structure of the Ligand-Binding Domain of Oestrogen Receptor Beta in the Presence of a Partial Agonist and a Full Antagonist. *EMBO J.* **1999**, *18*, 4608−4618.
(12) Eichenauer, U. BasF A.G. German Patent DE3831712, 1990.
(13) Yamato, T.; Hideshima, C.; Tashiro, M. A Convenient Preparation and Inclusion Behaviour of 1,3,5-Tris-(Hydroxyphenyl)-benzenes. *Chem. Express* **1990**, *5*, 845−848.
(14) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887−2893.
(15) Caron, P. R.; Mullican, M. D.; Mashal, R. D.; Wilson, K. P.; Su, M. S.; Murcko, M. A. Chemogenomic Approaches to Drug Discovery. *Curr. Opin. Chem. Biol.* **2001**, *5*, 464−470.
(16) Sadowski, J.; Gasteiger, J.; Klebe, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Crystal Structures. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1000−1008.
(17) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity: A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3288.
(18) *Sybyl 6.7*; Tripos Inc.: St. Louis, MO, 2000.
(19) Sybyl Atom Types: http://tripos.com/custResources/mol2Files/atom_types.html.
(20) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(21) Horvath, D. High Throughput Conformational Sampling and Fuzzy Similarity Metrics: A Novel Approach to Similarity Searching and Focused Combinatorial Library Design and Its Role in the Drug Discovery Laboratory. In *Combinatorial Library Design and Evaluation: Principles, Software Tools and Applications*; Ghose, A., Viswanadhan, V., Eds.; Marcel Dekker: New York, 2001; pp 429−472.
(22) Mestres, J.; Rohrer, D. C.; Maggiora, G. M. MIMIC: A Molecular Field-Based Similarity Program. Exploiting the Applicability of Molecular Similarity Approaches. *J. Comput. Chem.* **1997**, *18*, 934−954.
(23) Mestres, J.; Knegtel, R. M. A. Similarity versus Docking in 3D Virtual Screening. *Perspect. Drug Discovery Des.* **2000**, *20*, 191−207.
(24) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen. *Cell* **1998**, *95*, 927−937.
(25) Schoonen, W. E. G. J.; Dijkema, R.; de Ries, R. J. H.; Wagenaars, J. L.; Joosten, J. W. H.; de Gooyer, M. E.; Deckers, G. H.; Kloosterboer, H. J. Human Progesterone Receptor A and B Isoforms in CHO Cells II. Comparison of Binding, Transactivation and ED50 Values of Several Synthetic (Anti)progestagens in Vitro in CHO and MCF-7 Cells and in Vivo in Rats. *J. Steroid Biochem. Mol. Biol.* **1998**, *64*, 157−170.

JM034078C