Generation of a Focused Set of GSK Compounds Biased toward Ligand-Gated Ion-Channel Ligands

Anna Maria Capelli,* Aldo Feriani, Giovanna Tedesco, and Alfonso Pozzan

Computational, Analytical, and Structural Sciences, GlaxoSmithKline S.p.A. Medicines Research Centre, Via A. Fleming, 4-37135 Verona, Italy

Received August 30, 2005

A "data mining" methodology based on substructural analysis and standard 1024 Daylight fingerprints as descriptors was applied to a set of known antagonists of a subfamily of ligand-gated ion channels comprising nicotinic acetylcholine receptors (nAChR's), 5-hydroxytryptamine, γ-amino butyric acid-A, and glycine receptors. The derived scoring function was used to generate a focused set that was screened for α7 nAChR, resulting in the identification of novel α7 ligands easily amenable to chemical modification. Finally, the same scoring function was applied retrospectively to other in-house sets screened for the same target in the same assay. The results and performance of the method are described in detail.

INTRODUCTION

Nicotinic acetylcholine receptors (nAChR's) are ligand-gated ion channels (LGICs) which belong to the receptor superfamily comprising γ-amino butyric acid (GABA-A and GABA-C), 5-hydroxytryptamine (5-HT_3), glycine, and anionic glutamate receptors.[1] In 1999, benzylidene- and cynnamylidene-anabaseine derivatives, including GTS-21, were published as partial agonists of the α7 homopentameric nAChR and claimed as therapeutic agents for neuroprotection, smoking cessation, and senile dementia.[2] Considering that biological and biochemical experiments seem to indicate the existence of a certain degree of structural commonality in the agonist binding sites of this receptor superfamily,[3] it was reasonable to conceive that their ligands could share some common structural motifs.

Several "data mining" methodologies have been recently developed to bias compound selection and library design for generic therapeutics targets (i.e., antimicrobials, anticancer agents, etc.) in order to improve the effectiveness of high-throughput screening in the discovery of novel leads.[4−8] Among them, substructural analysis (SSA) has been reported as a methodology that allows the identification of structure−activity relationships of large and disparate data sets, characterized by qualitative and quantitative activity.[9−11] In particular, in our group, SSA had been successfully applied both to antibacterials[12] and to 7-TM receptor ligands.[13]

METHODS

Substructural analysis requires the generation of two training sets of compounds classified as "actives" and "inactives". Subsequently, a weight is assigned to each substructure (usually a topological fragment or a pharmacophoric substructure) in a molecule. This weight reflects the differential occurrence of a given substructure in the two training sets. Moreover, in this approach, it is assumed that a given substructure makes an additive and constant contribution to the overall activity of a molecule. When the substructure's weights are used, it is then possible to derive a scoring function for each molecule in the set of compounds to be tested. Molecules endowed with a high score should have a higher probability to be active, and therefore, according to the proposed strategy, they should be submitted for biological activity determination (in silico screening).

The work reported in this paper has been performed with widely used and validated 2D Daylight fingerprints as descriptors whose bits do not have a 1:1 correspondence to a substructure. As our work is applicative, hence pragmatic, we have decided to use them as we were interested in broadly capturing differences of bit frequencies from the "active" and "inactive" compounds (mentioned above) for focused screening and not to analyze the results in terms of any specific fragment.

Scoring Function Generation. Both the structure and the biological data of LGIC ligands were retrieved from the World Drug Index (WDI),[14] Pharmaproject,[15] MedChem, and the Dictionary of Natural Products[16] Daylight[17] format databases generated in-house as well as from Beilstein[18] and Patent Preview[19] collections using Daylight datatype key words. Subsequently, known antagonists were extracted from this list and filtered by using a molecular weight (MW) cutoff of 800 and ad hoc internally defined structural-reactivity filters (e.g., removal of alkylating agents). As a result, 323 LGIC ligands were identified as known "actives" (active set). To identify the number of chemotypes included in the active set, LGIC ligands were clustered[20,21] at a 0.85 Tanimoto index by using 1024 Daylight fingerprints as descriptors.

In addition, 55 000 derivatives randomly selected from the corporate database and devoid of any known LGIC ligands were used as "inactives" (inactive set).

The test set was generated from 600 000 derivatives retrieved from a Daylight-formatted corporate database utilizing Merlin Control Language scripts. These compounds were reduced to 550 000 (test set) by applying MW $\leq$ 1000 and ad hoc internally defined structural-reactivity filters.

* Corresponding author e-mail: Anna-Maria.M.Capelli@gsk.com.

10.1021/ci050353n CCC: $33.50
Published on Web 02/21/2006

$$Sf_i = \frac{1}{N} * \sum_{j=0}^{N} \left(\frac{bit_i}{\sum_{i=0}^{1023} bit_i} \right)$$

Sf_i = Substructure activity frequency
bit_i = i^{th} bit in the binary string
N = number of compounds in the set

$$w_i = \log \frac{Sf_i(active)}{Sf_i(inactive)}$$

Sf_i = Substructure activity frequency
w_i = weight of the i^{th} bit

Figure 1. Scheme used to calculate bit weight.[10]

Table 1. Summary of the Active, Inactive, and Test Sets Used for the Validation Experiments

scoring function	active set	inactive set	test set[a]
log_w1	223 LGIC	55 000 in-house compounds	bacd_val1 bacd_val2
log_w2	223 LGIC	100 000 derivatives from commercial sources	bacd_val1 bacd_val2

[a] bacd_val1 (10 000 compounds, 1% LGIC); bacd_val2 (10 000 compounds, 0.1% LGIC).

Following the substructural analysis methodology, the active, inactive, and test sets were described with the use of 1024 Daylight fingerprints. Then, each bit weight (R2 LOG_W)[10] was derived using the equation reported in Figure 1 (top) and used to calculate a score for each test set molecule as the sum of the bit weights (Figure 1, bottom). This score represents a sort of "overall activity" of each compound, and the test set derivatives endowed with a higher score should have a higher chance of being active.

The scoring function used to build the focused set SS609 is referred to as LOG_W.

Scoring Function Validation. In silico validation studies were performed by using the "salting approach"[4] in order to assess the performance of the scoring function and to define an appropriate score threshold to build SS609 as a subset of the 550 000 in-house derivative test set.

For this purpose, 100 LGIC ligands were randomly removed from the active set previously described and used to seed at 1% a test set of commercially available compounds; the result is a set (bacd_val1) made of 10 000 derivatives.

The remaining 223 LGIC ligands (from the original 323 "actives") were used to define the active set for the validation experiments.

Furthermore, 10 of the 100 LGIC ligands mentioned above were used to seed at 0.1% an additional test set containing commercially available compounds; the result is another set (bacd_val2) made of 10 000 compounds.

Finally, a further inactive set (100 000 compounds) was made of randomly selected commercially available derivatives.

Both the bacd_val1 and bacd_val2 sets were scored with the scoring functions log_w1 and log_w2 generated with the active and inactive sets reported in Table 1and the bit weight scheme shown in Figure 1.

For comparison, the bacd_val1 set was also ranked according to its similarity to the active set of 223 LGIC ligands mentioned in Table 1 by using a Tanimoto index and 1024 Daylight fingerprints as descriptors. This calculation was performed by using the largest single similarity.

Postprocessing of the Test Set Ranked Ligands. The top scoring compounds of the 550 000 compound test set, ranked with the LOG_W scoring function, were submitted to central nervous system druglike property cutoff filters derived from a computational analysis of 5370 WDI molecules, schematically reported in Figure 2 and subsequently clustered, giving rise to the ligand-gated ion-channel focused set SS609.

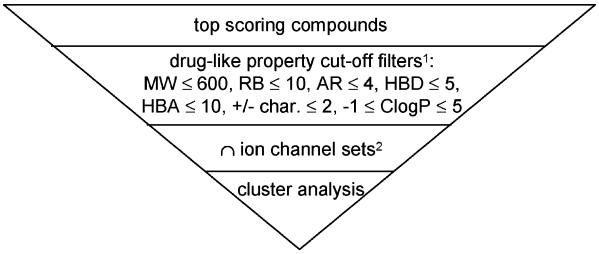


Figure 2. Workflow utilized to filter the top scoring compounds of the 550 000 in-house derivative test set. [1]MW = molecular weight; RB = number of rotatable bonds; AR = number of aromatic rings; HBD = number of hydrogen-bond donors; HBA = number of hydrogen-bond acceptors; +/− char = number of positive/negative charges; ClogP = ACD calculated lipophilicity. [2]Removal of compounds in common with SS447 and SS14, ion-channel in-house biased sets.

$$Initial\_enhancement = \frac{Actual\_A@X\%}{Random\_A@X\%}$$

Actual A@X% = no. of positives in the X% top ranking compounds
Random A@X% = no. of positives in the X% top ranking compounds assuming a random distribution

$$Global\_enhancement = \frac{A50random}{Actual\_A50}$$

A50 random, Actual A50 = no. of compounds that should be tested to find half of the positives respectively in the random case and in the ranked list.

Figure 3. Initial and global enhancements used to assess the performance of SSA.

Retrospective Analysis of the Focused Sets Screened Against α7 nAChR. A retrospective analysis of SS609 and other focused sets screened in the α7 nAChR binding assay was carried out to evaluate the performance of SSA. Accordingly, the focused set derivatives were described with 1024 Daylight fingerprints and scored using the same scoring function (LOG_W) used to generate SS609. At the same time, they were ranked according to their largest single similarity with respect to the 323 LGIC ligand active set with a Tanimoto index and Daylight fingerprints as descriptors. The performance of SSA and the similarity method at ranking true α7 nicotinic receptor ligands higher than random derivatives in comparison to random selection was assessed by calculating both the initial and the global enhancements described in Figure 3. The position of the positives in the ranked list was calculated and plotted together with that predicted for random selection and for "perfect prediction" (boost plots).[4]

In the end, the positive hits of each set were clustered[21] at a Tanimoto index ≥ 0.85 using 1024 Daylight fingerprints as descriptors in order to evaluate their structural diversity. Finally, their similarity to the active set of 323 LGIC ligands was estimated with a Tanimoto index and 1024 Daylight fingerprints as descriptors.

RESULTS AND DISCUSSION

Scoring Function Generation. The sizes of the active, inactive, and test sets generated are summarized in Figure 4. A cluster analysis of the active set resulted in the identification of several chemotypes, as highlighted in Figure 4 (left).

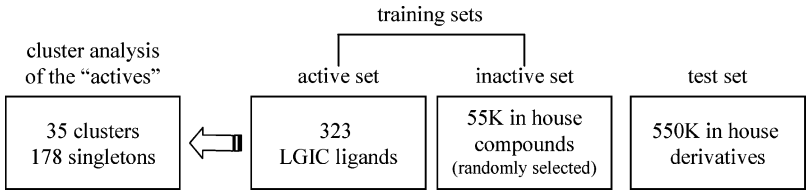


Figure 4. Sets of "actives" and "inactives" used to derive the scoring function (LOG_W), which was then used to rank the test set of 550 000 in-house derivatives.

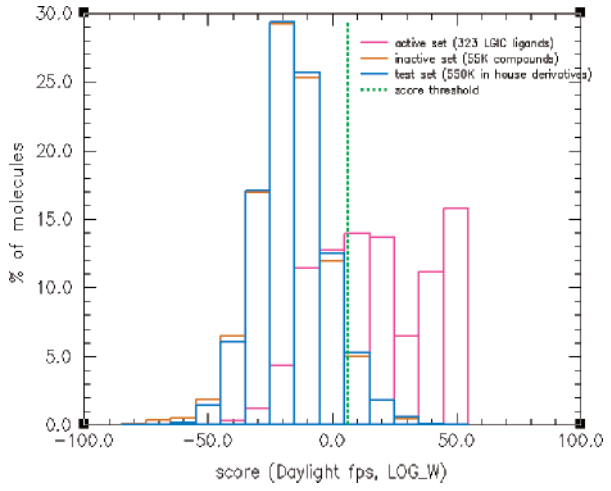


Figure 5. Score distribution of the active set (323 LGIC ligands, purple), the inactive set (55 000 in-house randomly selected derivatives, orange), and the test set (550 000 in-house derivatives, blue). The score threshold applied to the ranked test set to build the SS609 focused set is highlighted by the green line and is described in the text.

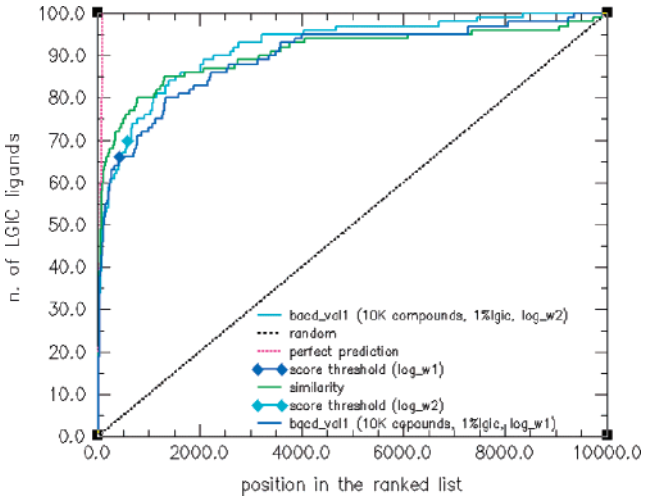


Figure 6. Cumulative recall plot for ranking the bacd_val1 set using similarity (green), log_w1 (blue), and log_w2 (cyan) scoring functions. Similarity refers to the similarity of the bacd_val1 set with respect to the active set of 223 LGIC ligands using 2D Daylight fingerprints and the Tanimoto index. The largest single similarity was utilized. The log_w1 and log_w2 scoring functions are described in Table 1.

These 323 LGIC ligands (active set) and the 55 000 in-house derivatives (inactive set) were used to generate the scoring function (LOG_W) utilized to rank the 550 000 in-house derivative test set.

Figure 5 shows the score distributions of the training sets (active set, purple; inactive set, orange) and the test set (blue). The plot indicates that the scoring function generated is able to effectively separate the training set of known "actives" from that of "inactives", demonstrating the ability of the method to distinguish between "LGIC" and "non-LGIC" ligands in the training set. In addition, a part of the score distribution of the test set compounds overlaps with that of the known actives.

Scoring Function Validation. The results of the validation experiments performed are summarized in Figure 6; this figure includes (1) the positions of known LGIC actives in the bacd_val1 set, ranked with the log_w1 (blue line), log_w2 (cyan line), and SSA scoring functions and similarity to the 223 LGIC ligands (green line); (2) the trace of the best possible performance (perfect prediction, purple line), where all of the actives are ranked right at the top of the list; and (3) the line of completely random performance (random, dotted black line), where all of the actives are distributed equally throughout the entire ranking.

As shown in Figure 6, SSA (cyan and blue plots) performs significantly better than random selection (black line) at pushing the positives in the bacd_val1 set toward the top of the ranked list. In addition, 2D similarity (green curve) performs slightly better than SSA. The SSA plots scored with the log_w1 and log_w2 scoring functions perform similarly to random selection when more than 70% of the known actives have been retrieved and the scores are between 3.4 and 8.2 (log_w1) and 4.9 and 6.6 (log_w2). Therefore, we empirically decided to set the score threshold at 6.2 (Figure 6, blue and cyan diamonds, respectively) and to use it as the score cutoff for the 550 000 in-house derivative test set ranked list, leading to the generation of SS609.

The performance of the SSA and similarity methods at ranking known LGIC ligands higher than random derivatives in comparison to random selections was also assessed with these validation experiments by calculating both the initial and the global enhancements (I_{en} and G_{en}, respectively) as described in Figure 3.

As reported in Table 2, SSA performs similarly to 2D similarity in terms of I_{en}, but worse in regard to G_{en}. However, I_{en} matters the most as a few hundred compounds are generally submitted for focused screening.

Furthermore, among the 73 known LGIC actives included in the 1000 top scoring bacd_val1 set ranked with log_w1 (Table 2), only five compounds are structurally similar at Tanimoto index $\geq$ 0.85 to the 223 LGIC active set ligands. On the contrary, 53 out of the 80 known LGIC actives in the 1000 top scoring bacd_val1 derivatives ranked with the similarity method are similar to the 223 LGIC active set ligands when using the same criteria. This result highlights the strength of SSA with regard to similarity at fishing out hits not similar to known actives.

Postprocessing of the Test Set Ranked Ligands. The postprocessing cascade of the top scoring 37 000 compounds of the 550 000 compound test set, ranked with the LOG_W

Table 2. Initial (I_{en}) and Global (G_{en}) Enhancements Obtained for SSA and Similarity (I_{en} and G_{en} as Described in Figure 3)[a]

test set used for validation	method	A@1K	A@1K$_{random}$	I_{en}	ActualA50	A50$_{random}$	G_{en}
bacd_val1 (10 000 compounds, 1% LGIC)	sim[b]	80	10	8	61	5000	82
	log_w1[c]	73	10	7.3	119	5000	42
	log_w2[c]	76	10	7.6	102	5000	49
bacd_val2 (10 000 compounds, 0.1% LGIC)	log_w1	8	1	8	151	5000	33
	log_w2	8	1	8	28	5000	178

[a] I_{en} was calculated at 10% of each ranked list. [b] Similarity of the test set with respect to the active set used for validation experiments using 2D Daylight fingerprints and the Tanimoto index. The large single similarity was utilized. [c] log_w1 and log_w2 as described in Table 1.

Table 3. Biological Profile of the Most Interesting Compound Obtained from the Screening of SS609

pIC$_{50}$ (binding)			pIC$_{50}$ (FLIPR Ca^{2+} assay)					pIC$_{50}$ (e-phys)
rat α7	5-HT$_3$	5-HT$_4$	α7 (GH4C1)	α1 (TE671)	α3 (IMR32)	5-HT$_3$ (CHO)	α4β2 (HEK293)	α7 (GH4C1)
6.98	4.81	5.71	8.37	4.86	4.59	5.17	5.15	9.4

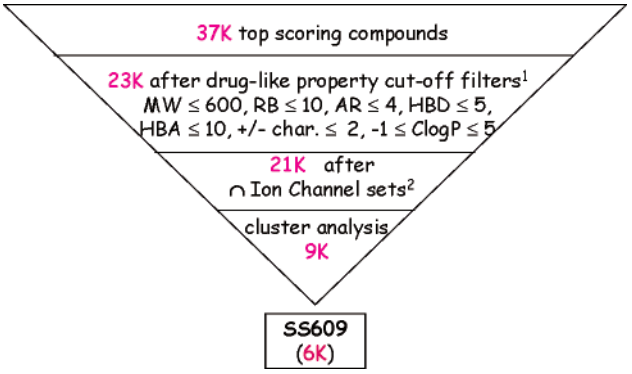


Figure 7. Overview of the postprocessing cascade of the 37 000 top scoring compounds of the 550 000 in-house test set, ranked with the LOG_W scoring function. [1]MW = molecular weight; RB = number of rotatable bonds; AR = number of aromatic rings; HBD = number of hydrogen-bond donors; HBA = number of hydrogen-bond acceptors; +/− char = number of positive/negative charges; ClogP = ACD calculated lipophilicity. [2]Removal of compounds in common with SS447 and SS14, ion-channel in-house biased sets.

scoring function, is reported in Figure 7. About 6 000 compounds were finally cherry picked (SS609).

Screening Results of SS609 Focused Set. Screening of the SS609 focused set in the in-house α7 nicotinic receptor binding assay resulted in the identification of hits easily amenable to structural modifications. A total of 10 out of 15 are antagonists in the FLIPR Ca^{2+} assay,[22] and in particular, one of them is endowed with an interesting biological profile, shown in Table 3, in particular, in terms of potency (pIC$_{50e-phys}$ = 9.4) and selectivity against other nicotinic subtypes. Moreover, one of the actives is a close analogue of tropisetron, a well-known 5-HT$_3$ receptor antagonist recently reported as an α7 partial agonist.[23]

Table 4. Sets Screened for α7 nAChR

set[a]	*N*	number of positive hits	hit rate (%)	number of chemotypes[b]	hit similarity[c]	number of novel chemotypes[d]
5-HT$_3$	297	10	4.3	6	6	0
SS447	10 302	28	0.27	14	5	11
SS609	5923	15	0.25	15	1	14
Core98	28 435	29	0.1	29	2	27[e]
SS14	5045					

[a] Set descriptions: 5-HT$_3$, in-house 5-HT$_3$ ligands; SS447 and SS14, in-house ion-channel biased sets; Core98, in-house collection representative set; SS609, ligand-gated ion-channel biased set. [b] Sum of centroids and singletons obtained with cluster algorithm[21] and 1024 Daylight fingerprints as descriptors. [c] Number of the positive hits similar to 323 known LGIC antagonists. Similarity was established at Tanimoto index ≥ 0.85 (1024 Daylight fingerprints as descriptors). [d] Number of quality hits after removal of compounds similar to known LGIC antagonists. [e] The majority of the positives obtained are not easily amenable to structural modifications.

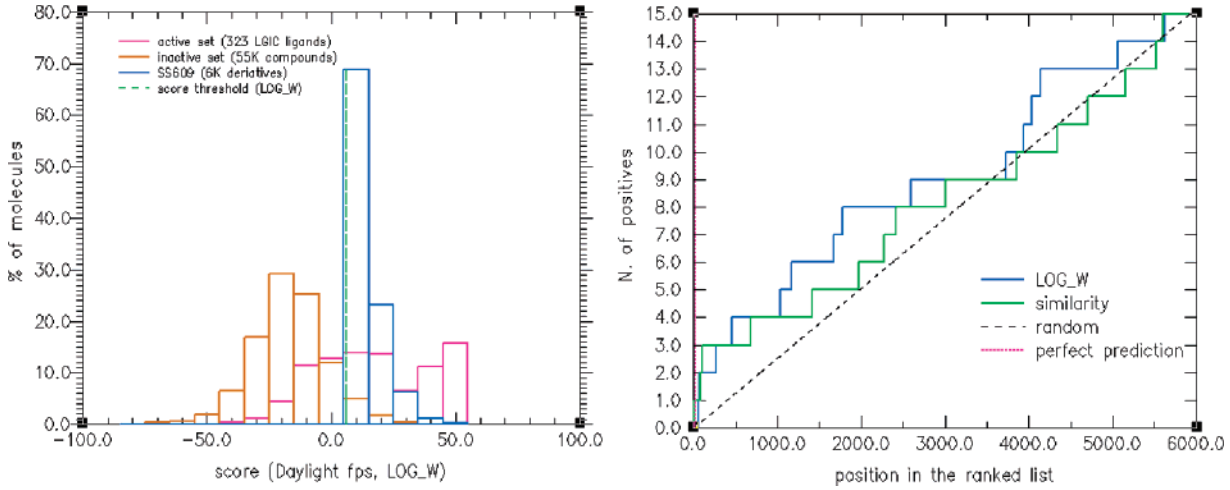
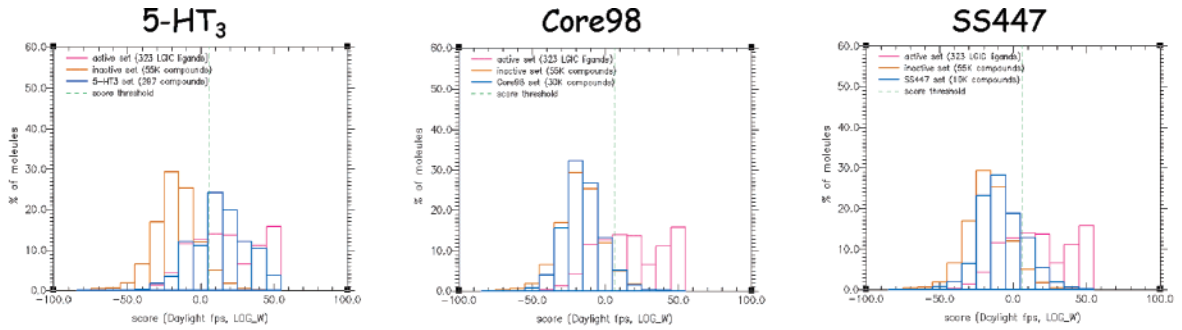
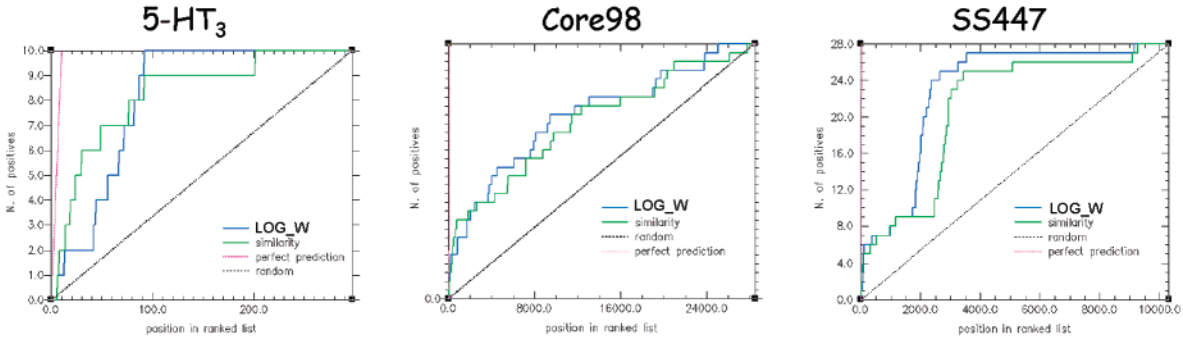


Figure 8. Left: Histograms of scores obtained for the training set of "actives" (323 LGIC ligands, purple), "inactives" (55 000 in-house derivatives, orange), and test set (SS609 screened compounds, blue). The green line represents the score threshold used to build SS609. Right: Distributions of the SS609 positive hits using SSA (blue plot) and similarity (green). Both perfect prediction (purple) and random selection (black) plots are reported for comparison.



Histograms of the scores obtained for the training set of "actives" (323 LGIC ligands, purple), "inactives" (55K in house compounds, orange) and the ranked test sets: 5-HT$_3$, Core98 and SS447 (blue).



Positive hits distributions in the ranked test sets: 5-HT$_3$, Core98 and SS447 using SSA (log_w, blue) and similarity (green).

Figure 9. Score histograms (top) and "boost" plots (bottom) obtained after applying SSA LOG_W bit weights to 5-HT$_3$ (left), Core98 (center), and SS447 (right) sets screened in the α7 rat binding assay.

Table 5. Enrichments Obtained by Retrospectively Applying SSA to SS609 and the Other Focused Sets Screened in the Same Assay[a]

set	method	A@2575	occ.[b]	A@2575$_{random}$	I_{en}	ActualA50	A50$_{random}$	G_{en}
SS447	LOG_W	24	10	7	3.4	1901	5151	2.7
10302	sim	11	9	7	1.6	2668	5151	1.9
set	method	A@1481	occ.	A@1481$_{random}$	I_{en}	ActualA50	A50$_{random}$	G_{en}
SS609	LOG_W	6	6	3.7	1.6	2000	2961	1.5
5923	sim	5	5	3.7	1.3	2419	2961	1.2
set	method	A@7109	occ.	A@7109$_{random}$	I_{en}	ActualA50	A50$_{random}$	G_{en}
Core98	LOG_W	16	16	7.2	2.2	4520	14 218	3.1
28435	sim	14	14	7.2	1.9	7182	14 218	2.0
set	method	A@74	occ.	A74$_{random}$	I_{en}	ActualA50	A50$_{random}$	G_{en}
5-HT$_3$	LOG_W	7	4	2.5	2.8	56	149	2.7
297	sim	7	4	2.5	2.8	24	149	6.2

[a] I_{en} and G_{en} as described in Figure 1. I_{en} was calculated at 25% of the ranked list. [b] Sum of centroids and singletons obtained with cluster algorithm[21] and 1024 Daylight fps as descriptors.

Retrospective Analysis of the Focused Sets Screened Against α7 nAChR. SSA was retrospectively applied to the SS609 focused set compounds screened in the α7 binding assay using the same bit weights previously used to build it (LOG_W). As clearly shown in Figure 8, there is a moderately good overlap of the SS609 screened derivatives (test set) with the active set of 323 LGIC ligands (left) and the method is able to push true actives at the top of the ranked list (right). In addition, its performance is comparable to 2D similarity and better than random selection for α7 nAChR.

The SS609 hit rate and the quality of the positive hits obtained were then compared with the hits generated by screening other in-house focused sets comprising (1) a set of in-house 5-HT$_3$ ligands (5-HT$_3$), (2) two in-house ion-channel biased sets (SS447 and SS14), and (3) a representative set of the in-house collection (Core98). As shown in Table 4, the hit rate of SS609 is similar to that of the other focused sets. Nonetheless, all of the SS609 positive hits are relatively different from each other in terms of chemotypes and with respect to the hits in the active set as only one compound is similar to known α7 nicotinic ligands.

Finally, SSA LOG_W bit weights were retrospectively used to score the other in-house focused sets mentioned above and screened in the α7 nicotinic receptor binding assay. The histograms of their score distributions (Figure 9, top) clearly show that the method is able to separate "LGIC" from "non-LGIC" derivatives. Furthermore, substructure analysis performs significantly better than random selection

in pushing true α7 nicotinic receptor actives at the top of the ranked lists (Figure 9, bottom).

As reported in Table 5, the performance of the substructural analysis is comparable to 2D similarity in regard to both I_{en} and G_{en}. However, SSA is able to fish out α7 nicotinic receptor ligands that are not structurally similar to known LGIC ligands (Table 4), as already observed in the validation experiments.

CONCLUSIONS

A fast and easily accessible "data mining" methodology based on a substructural analysis and 1024 Daylight fingerprints was successfully used to build a set of in-house derivatives (SS609) biased toward LGIC ligands. This methodology, previously applied in our group to other ligand classes such as antibacterials and 7-TM receptor ligands, has proved to be promising in identifying novel α7 nicotinic receptor ligands easily amenable to structural modifications. The use of pharmacophore fingerprints[24,25] as descriptors in a future exercise might provide additional structural novelty.

ACKNOWLEDGMENT

The authors thank colleagues in the U.K. Automated Liquid Store (ALS) for "cherry picking" and colleagues in Discovery Research Systems Biology for screening SS609.

REFERENCES AND NOTES

(1) Le Novere, N.; Changeux, J.-P. The ligand gated ion channel database. *Nucleic Acids Res.* **1999**, *27*, 340−342.
(2) Meyer, E.; Kem, W. R.; Van Haaren, F.; Zoltewicz, J. A.; Defiebre, C. M.; Papke, R.; Day, A. L. Synthesis, methods of use and compns. for benzylidene- and cynnamylidine -anabaseines for use in nicotine addiction and neuropathological disorders. *PCT Int. Appl.* **1999**, *WO 99/10338*, 94.
(3) Brejc, K.; van Dijk, W. J.; Klaassen, R. V.; Schuurmans, M.; van der Oost, J.; Smit, A. B.; Sixma, L. K. Crystal structure of an ACh-binding protein reveals the ligand-binding domain of nicotinic receptors. *Nature* **2001**, *411*, 269−276.
(4) Gillet, V. J.; Willett, P.; Bradshaw, J. Identification Of Biological Activity Profiles Using Substructural Analysis And Genetic Algorithms. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 165−179.
(5) Shah, A. V.; Walters, W. P.; Murcko, M. A. Can We Learn To Distinguish between "Drug-like" and "Nondrug-like" Molecules? *J. Med. Chem.* **1998**, *41*, 3314−3324.
(6) Sadowski, J.; Kubinyi, H. A scoring scheme for discriminating between drugs and nondrugs. *J. Med. Chem.* **1998**, *41*, 3325−3329.
(7) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1*, 55−67.
(8) Harper, G.; Bradshaw, J.; Gittins, J. C.; Green, D. V. S.; Leach, A. R. The Prediction of Biological Activity for High-Throughput Screening Using Binary Kernel Discrimination. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1295−1300.
(9) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of Fingerprint-Based Methods for Virtual Screening Using Multiple Bioactive Reference Structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177−85.
(10) Ormerod, A.; Willet, P.; Bawden D. Comparison of Fragment Weighting Schemes for Substructural Analysis. *Quant. Struct.-Act. Relat.* **1989**, *8*, 115−129.
(11) Ormerod, A.; Willet, P.; Bawden, D. Further Comparative Studies of Fragment Weighting Schemes for Substructural Analysis. *Quant. Struct.-Act. Relat.* **1990**, *9*, 302−312.
(12) Feriani, A.; Pozzan, A.; Tedesco, G.; Capelli. A. M. A fast datamining method using 2/3D Fingerprints. *Book of Abstracts*, PA-19, 129, XVIth International Symposium on Medicinal Chemistry, September 18−22, 2000, Bologna, Italy.
(13) Tedesco, G.; Feriani, A.; Pozzan, A.; Capelli, A. M. A fast datamining method based on substructural analysis and different descriptors as applied to antibacterials and HTS of 7TM targets. EuroMUG2002, September 24−26 2002, Cambridge, U.K.
(14) *World Drug Index*; Derwent Publications Ltd.: London.
(15) *Pharmaproject*; PJB Publications Ltd.: London.
(16) *MedChem*; *Dictionary of Natural Products*; Chapman & Hall: New York.
(17) *Daylight*, version 4.51; Daylight Chemical Information Systems Inc.: Mission Viejo, CA.
(18) *Beilstein Handbook of Organic Chemistry*, BS9901; Springer-Verlag: Berlin.
(19) *Patents Preview*, Derwent Information Ltd.: London, 1999.
(20) Taylor, R. Simulation Analysis of Experimental Design Strategies for Screening Random Compounds as Potential New Drugs and Agrochemicals. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 59−67.
(21) Butina, D. Unsupervised data base clustering based on Daylight's fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 747−750.
(22) Kuntzweiler, T. A.; Arneric, S. P.; Donnelly-Roberts, D. L. Rapid assessment of ligand actions with nicotinic acetylcholine receptors using calcium dynamics and FLIPR. *Drug Dev. Res.* **1998**, *44*, 14−20.
(23) Macor, J. E.; Gurley, D.; Lanthorn, T.; Loch, J.; Mack, R. A.; Mullen, G.; Tran, O.; Wright, N.; Gordon, J. C. The 5-HT3 antagonist tropisetron (ICS 205−930) is a potent and selective α7 nicotinic receptor partial agonist. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 319−321.
(24) Pozzan, A.; Feriani, A.; Tedesco, G.; Capelli, A. M. Pharmacophoric 3D mill hashed fingerprints: Add another dimension to your similarity searching. *Abstracts of Papers*, 223rd ACS National Meeting, Orlando, FL, April 7−11, 2002; American Chemical Society: Washington, DC, 2002; COMP-092.
(25) Pozzan, A.; Feriani, A.; Tedesco, G.; Capelli, A. M. 3D pharmacophoric hashed fingerprints. Rational Approaches to Drug Design. *Proceedings of the 13th European Symposium on Quantitative Structure−Activity Relationships*, Duesseldorf, Germany, Aug. 27−Sept. 1, 2000.

CI050353N