Scaffold Perception, Common Pharmacophore Model Development, And Quantitative Structure-Affinity Relationships Of Sigma Site Ligands

David Enos Watson

University of Mississippi

Follow this and additional works at: https://egrove.olemiss.edu/etd

Part of the Pharmacy and Pharmaceutical Sciences Commons

Recommended Citation
https://egrove.olemiss.edu/etd/692

This Dissertation is brought to you for free and open access by the Graduate School at eGrove. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of eGrove. For more information, please contact egrove@olemiss.edu.
SCAFFOLD PERCEPTION, COMMON PHARMACOPHORE MODEL DEVELOPMENT, AND QUANTITATIVE STRUCTURE–AFFINITY RELATIONSHIPS OF SIGMA SITE LIGANDS

A Dissertation
presented in partial fulfillment of requirements
for the degree of Doctor of Philosophy in Pharmaceutical Sciences
in the Department of Medicinal Chemistry
The University of Mississippi

by
DAVID WATSON
December 2013
ABSTRACT

Sigma receptors are endogenous proteins with potential utility in treating psychological disorders, ischemia, the psychological and convulsive effects of drugs of abuse, and as an imaging agent for cancerous tissues, among others. Drug design efforts targeting these receptors have been hindered by a lack of structural information of the receptors themselves. Traditional ligand-based approaches have succeeded in generating many compounds with high affinity, and quite a few with selectivity for σ-1 receptors. There are few selective ligands for use as pharmacological probes for the σ-2 receptor. Much effort has gone into exploring the structure activity relationships of ligands targeting these receptors.

A critical review of the existing literature covering pharmacophore development for σ receptors was undertaken with the intent to develop computational models to assist in ligand-based drug design efforts. Inspired by the lack of pharmacophore models with general utility, and confronted by the obstacles of data heterogeneity, a database of σ ligands and their binding affinity data was collected. Cohorts of data collected under similar experimental methodologies were assembled and clustered by measures of scaffold dissimilarity. Multiple-Instance Learning techniques were used to train classification models that differentiated molecules as active or inactive, and to assist in the identification of relevant conformations of σ ligands at their macromolecular targets. Conformations of high-affinity ligands were then used to develop general pharmacophore models as part of a virtual screening approach. Structure-activity relationship models based on virtual screening alignments of known sigma ligands were developed in the search for selective σ-1 and σ-2 receptor probes.
DEDICATION

This work is dedicated to my sons Lucas and Xander. The vicissitudes of life can never the match the resolve you have given me to make this world a better place for you and your generation.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-HPP</td>
<td>3-3(-hydroxyphenyl)piperidine</td>
</tr>
<tr>
<td>3-PPP</td>
<td>3-(3-hydroxyphenyl)-N-(1-propyl)piperidine</td>
</tr>
<tr>
<td>5-HT</td>
<td>5-hydroxytryptamine</td>
</tr>
<tr>
<td>D₂</td>
<td>Dopamine 2 receptor</td>
</tr>
<tr>
<td>H₁</td>
<td>Histamine 1 receptor</td>
</tr>
<tr>
<td>σ-1</td>
<td>Sigma-1</td>
</tr>
<tr>
<td>σ-2</td>
<td>Sigma-2</td>
</tr>
<tr>
<td>CoMFA</td>
<td>Comparative Molecular Field Analysis</td>
</tr>
<tr>
<td>CoMSIA</td>
<td>Comparative Molecular Similarity Indices Analysis</td>
</tr>
<tr>
<td>CPU</td>
<td>Central Processing Unit</td>
</tr>
<tr>
<td>CSV</td>
<td>Comma-Separated Values</td>
</tr>
<tr>
<td>DDR</td>
<td>Double Data Rate</td>
</tr>
<tr>
<td>DTG</td>
<td>1,3-di-o-tolyl-guanidine</td>
</tr>
<tr>
<td>EA</td>
<td>External Accuracy</td>
</tr>
<tr>
<td>FPR</td>
<td>False Positive Rate</td>
</tr>
<tr>
<td>GB</td>
<td>gigabyte</td>
</tr>
<tr>
<td>GHz</td>
<td>gigahertz</td>
</tr>
<tr>
<td>HAL</td>
<td>haloperidol</td>
</tr>
<tr>
<td>IA</td>
<td>Internal Accuracy</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>kD</td>
<td>kilodalton</td>
</tr>
<tr>
<td>MCC</td>
<td>Matthews Correlation Coefficient</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics</td>
</tr>
<tr>
<td>MDMA</td>
<td>3,4-methylenedioxy methamphetamine</td>
</tr>
<tr>
<td>MHz</td>
<td>megahertz</td>
</tr>
<tr>
<td>MILES</td>
<td>Multiple Instance Learning via Embedded instance Selection</td>
</tr>
<tr>
<td>PE</td>
<td>Prediction Error</td>
</tr>
<tr>
<td>PGRMC1</td>
<td>Progesterone Receptor Membrane Component 1</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial Least Squares</td>
</tr>
<tr>
<td>PTZ</td>
<td>(+)-pentazocine</td>
</tr>
<tr>
<td>QSAR</td>
<td>Quantitative Structure–Activity Relationship</td>
</tr>
<tr>
<td>RCSB</td>
<td>Research Collaboratory for Structural Bioinformatics</td>
</tr>
<tr>
<td>RMS</td>
<td>Root-Mean-Square</td>
</tr>
<tr>
<td>SAP</td>
<td>Significance Analysis of Pharmacophores</td>
</tr>
<tr>
<td>SMILES</td>
<td>Simplified Molecular-Input Line-Entry System</td>
</tr>
<tr>
<td>SVM</td>
<td>Support-Vector Machine</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

The expertise and tutelage of Carl Raffa during my undergraduate years set me on the path that would lead to this work, and for that, I am very grateful.

Many thanks are also due to Joshua Schlesser for his companionship and financial support when there was a dearth of both, and without which my rendezvous with academia would have been a hardship.

Andrew Mullen and Shivangi Aswathi provided much-needed assistance with sketching structures and curation of binding data, which moved this project forward when the task seemed insurmountable.

Conversations with my colleagues Drs. Rohit Bhat and Sanju Narayanan planted the seeds for my inspiration to investigate σ site binding in light of their diverse tissue- and organism-dependent requirements.

Dr. Sarah Scarry was the first student I met when I came to visit the program. Her bright smile and ebullient demeanor never faltered, and her initiative in coordinating the University of Mississippi Medicinal Chemistry Journal Club helped me discover the depth of talent and insight that were latent in our department.

Dr. Mahmoud ElSohly has been a part of my life since the day of my birth; his forthright manner, charity to his community, and support towards my family is boundless.

Dr. Robert Doerksen provided much needed criticism and insight laced with his characteristic wit and tasteful double-entendres.

Dr. Christopher McCurdy has been an exceptional supportive mentor with an infectious optimism, and his unwavering advocacy through many unanticipated challenges kept me motivated when my research seemed intractable.
I cannot adequately express my gratitude for the motivational support, wise words, and opportunity Dr. Stephen Cutler has provided during this endeavor.

The scientists and support staff at Schrödinger were expedient at troubleshooting software issues and provided valuable insights into non-documented features of the software suite. In particular, Dan Schafer, Steve Dixon, Hege S. Beard, Peter S. Shenkin, Susanne Salzmann, Jeff Lowrie, Pavel Golubkov, Shane Peterson, Lynnette Sander, Mike Campbell, and Carrie Weston contributed their time and endured my questions.

I feel indebted to several publishers who graciously allowed me to include graphical material in my introductory review: Springer SBM NL for Figure 1.3; Kluwer Academic Publishers for Figure 1.12; Elsevier Science Publishers for Figure 1.4; Elsevier Science for Figure 1.10; the American Society for Pharmacology and Experimental Therapeutics Figure 1.1; NPP Books for Figures 1.2 and 1.5; and the American Chemical Society for Figures 1.6, 1.7, 1.8, and 1.9.

Samuel Stewart Watson, aside from being the best little brother anyone could ask for, was particularly adept at framing the machine learning problem in terms that a mere mortal could understand. I am edified and forever grateful.

Above all, my wife, Melody, and my sons Xander and Lucas, must be commended for tolerating my absence and distraction through this process, for growing with me over these years, and for providing a stable foundation to brace myself upon during stormy weather.
TABLE OF CONTENTS

1 Introduction 1

1.1 Therapeutic potential of σ receptor ligands .............................. 1

1.2 Sigma receptor subtypes ..................................................... 2

1.2.1 Sigma-1 receptor .......................................................... 2

1.2.2 Sigma-2 receptor .......................................................... 2

1.2.3 “Sigma-3” ................................................................. 3

1.3 Sigma pharmacophore and QSAR development: 1980–2011 .......... 3

1.3.1 Non-selective σ models .................................................... 4

1.3.1.1 Manallack topological pharmacophore model .................. 4

1.3.1.2 Gund pharmacophore (1991) ........................................ 5

1.3.1.3 Gund pharmacophore (1992) ........................................ 6

1.3.1.4 Ablordeppy pharmacophore and CoMFA model (1992) .... 6

1.3.1.5 Seri-Levy eudismic analysis ......................................... 7

1.3.1.6 Beart pharmacophore model ....................................... 8

1.3.2 Sigma-1 models ............................................................ 8

1.3.2.1 Carroll pharmacophore model ...................................... 8

1.3.2.2 Elaborated Carroll pharmacophore model ...................... 9

1.3.2.3 Gilligan pharmacophore .............................................. 10

1.3.2.4 Hudkins pharmacophore motif ..................................... 10

1.3.2.5 Glennon pharmacophore motif .................................... 10

1.3.2.6 Ablordeppey pharmacophore and CoMFA model (1998) ... 11
1.3.2.7 Huang pharmacophore and CoMFA model ........................................... 12
1.3.2.8 Cao CoMFA model ................................................................. 13
1.3.2.9 Gund pharmacophore model (2004) .............................................. 13
1.3.2.10 Jung DISCOtech pharmacophore and CoMFA model ................. 13
1.3.2.11 Laggner HypoGen pharmacophore model ................................... 14
1.3.2.12 Zampieri HypoGen pharmacophore model .................................. 15
1.3.2.13 Oberdorf Quasar pseudoreceptor model ................................... 16
1.3.2.14 Rossi Galahad pharmacophore model ........................................ 17
1.3.3 Sigma-2 models ............................................................................ 17
  1.3.3.1 Cratteri GRIND-based pharmacophore model ........................... 17
  1.3.3.2 Abate pharmacophore and CoMFA model ............................... 18
  1.3.3.3 Laurini Catalyst pharmacophore model .................................. 19
1.3.4 “Sigma-3” model ........................................................................ 20
1.3.5 Summary of pharmacophore models in the σ literature .................... 20
1.4 Overview of σ binding assay methodology ........................................... 21
1.5 Research Objectives ........................................................................ 22

2 Methods .............................................................................................. 24
  2.1 Computational equipment ............................................................... 24
  2.2 Data collection and curation ............................................................ 24
  2.3 Dataset composition ....................................................................... 25
    2.3.1 Scaffold decomposition ............................................................ 26
    2.3.2 Core fingerprints ....................................................................... 26
    2.3.3 Core clustering and network visualization ............................... 27
    2.3.4 Selection of classification model training and test sets .......... 28
  2.4 Generation of conformers ............................................................... 29
  2.5 Model development ....................................................................... 29
    2.5.1 Classification models ............................................................. 29
2.5.2 Pharmacophore fingerprints ........................................ 30
2.5.3 Implementation of SAP ........................................... 31
2.5.4 Instance-based similarity mapping ............................... 32
2.5.5 Implementation of 1-norm SVM ................................. 34
2.5.6 Selection of prototype conformers ............................... 36
2.5.7 Pharmacophore modeling ......................................... 37
2.5.8 Database screening ............................................... 38
2.5.9 Field-based QSAR ................................................. 39

3 Results ............................................................................ 40
  3.1 Competition Binding Database ..................................... 40
  3.2 Impact of Fingerprint Length ...................................... 43
  3.3 Comparison of Similarity Metrics ................................. 48
  3.4 Impact of training and test set size ............................... 48
  3.5 Development of σ-2 pharmacophore models and QSAR .... 53
    3.5.1 Development of σ-1 pharmacophore models and QSAR 63

4 Conclusions and future work ............................................. 80
  4.1 Conclusions ............................................................ 80
  4.2 Future Work ........................................................... 82

LIST OF REFERENCES .......................................................... 83

LIST OF APPENDICES .......................................................... 109

APPENDIX A. HIERARCHICAL CLUSTERING SCRIPTS .................. 110

A Hierarchical clustering scripts ........................................... 111
  A.1 hcviz directory listing .............................................. 111
  A.2 Makefile ............................................................... 111
## LIST OF FIGURES

1.1 Manallack's topological pharmacophore ............................................. 4
1.2 Gund pharmacophore (1992) ................................................................. 6
1.3 Ablordeppey pharmacophore ................................................................. 7
1.4 Beart pharmacophore model ................................................................. 8
1.5 Carroll pharmacophore ........................................................................... 9
1.6 Hudkins pharmacophore ....................................................................... 11
1.7 Glennon pharmacophore ....................................................................... 11
1.8 Laggner HypoGen pharmacophore ......................................................... 15
1.9 Zampieri HypoGen pharmacophore model ........................................... 16
1.10 Oberdorf Quasar model ....................................................................... 17
1.11 Rossi Galahad pharmacophore model ................................................. 18
1.12 Cratteri GRIND pharmacophore ......................................................... 18
1.13 Laurini Catalyst pharmacophore model .............................................. 19

2.1 Identification of significant pharmacophore features ............................ 32
2.2 SVM tuning process .............................................................................. 36

3.1 Scaffold clustering of the σ-2 dataset and selection of 54 molecules for SAP/MILES .. 44
3.2 Scaffold clustering of the σ-2 dataset and selection of 130 molecules for SAP/MILES .. 49
3.3 Sigma-2 prototype conformer heat map ................................................ 54
3.4 Problem compounds for development of Gaussian QSAR from aligned datasets .. 54
3.5 Sigma-2 3PLS regression ..................................................................... 61
3.6 Prototype ligand for the pharmacophore alignment ADHHPR.87 ............. 64
3.7 Sigma-2 virtual screening performance ........................................ 65
3.8 Gaussian steric fields presented by QSAR of ADHHPR.87 ............... 66
3.9 Gaussian electrostatic fields presented by QSAR of ADHHPR.87 ......... 66
3.10 Gaussian hydrophobic fields presented by QSAR of ADHHPR.87 ...... 67
3.11 Gaussian hydrogen bond acceptor fields presented by QSAR of ADHHPR.87 .... 67
3.12 Gaussian hydrogen bond donor fields presented by QSAR of ADHHPR.87 .... 68
3.13 RB8 in context of steric fields ............................................. 68
3.14 RB8 in context of electrostatic fields .................................... 69
3.15 RB8 in context of hydrophobic fields ..................................... 69
3.16 RB8 in context of hydrogen bond acceptor fields ....................... 70
3.17 RB8 in context of hydrogen bond donor fields ........................... 70
3.18 Scaffold clustering of the σ-1 dataset and selection of 124 molecules for SAP/MILES .... 71
3.19 Sigma-1 prototype conformer heat map .................................... 72
3.20 Sigma-1 3PLS regression ................................................ 74
3.21 Sigma-1 virtual screening performance .................................... 75
3.22 Prototype ligand for the pharmacophore alignment AAHHPR.12 ......... 77
3.23 Gaussian steric fields presented by QSAR of AAHHPR.12 ............. 77
3.24 Gaussian electrostatic fields presented by QSAR of AAHHPR.12 ....... 78
3.25 Gaussian hydrophobic fields presented by QSAR of AAHHPR.12 ....... 78
3.26 Gaussian hydrogen bond acceptor fields presented by QSAR of AAHHPR.12 .... 79
3.27 Gaussian hydrogen bond donor fields presented by QSAR of AAHHPR.12 .... 79
LIST OF TABLES

1.1 Overview of σ pharmacophore models .................................................. 22

2.1 Measures of similarity ........................................................................... 34

3.1 Breakdown of σ research articles providing competition binding data ........... 41
3.2 Breakdown of assay methodology for selective σ competition binding experiments. . 42
3.3 Statistics of the initial σ-2 classification model data set. .......................... 43
3.4 SAP and fingerprint length .................................................................... 45
3.5 SAP and fingerprint length affect MILES results ................................. 46
3.6 Confusion matrices: impact of fingerprint length without SAP .......... 47
3.7 Confusion matrices: impact of fingerprint length with SAP ............ 47
3.8 Stability of MILES performance utilizing alternative similarity metrics. .... 50
3.9 Confusion matrices: impact of similarity metric using 1,024–bit fingerprints with SAP 51
3.10 Confusion matrices: impact of similarity metric using 4,096–bit fingerprints with SAP 51
3.11 Statistics of the balanced σ-2 classification model data set. ............. 52
3.12 SAP parameters of a balanced dataset .............................................. 52
3.13 Comparison of metric performance with a more balanced molecular library .. 52
3.14 Sigma-2 pharmacophore hypotheses and related statistics ............... 56
3.15 Gaussian field fractions of hypothesis ADHHPR.87 .......................... 62
3.16 Statistics of the balanced σ-1 classification model data set. ............. 63
3.17 Sigma-1 pharmacophore hypotheses and related statistics ............... 73
3.18 Gaussian field fractions of hypothesis AAHHPR.12 .......................... 74

F.1 Sigma 1: PTZ guinea pig brain dataset .................................................. 140
F.2 Sigma 2: DTG/DXL rat liver dataset .................................................. 165
F.3 Sigma 2: DTG/PTZ rat brain dataset .................................................. 170
F.4 Sigma 2: DTG/PTZ rat liver dataset .................................................. 174
F.5 Sigma 2: DTG/PTZ guinea pig brain dataset ................................. 186
1. INTRODUCTION

Introduce such an intoxicant, and start it to ferment in humanity’s blood, and it may spread from soul to soul, until, before the world is advised of its possible results, the ever-increasing potency will gain such headway as to destroy, or debase, our civilization, and even to exterminate mankind.

Etidorpha
John Uri Lloyd

Research into \(\sigma\) receptors and their ligands began in 1976 with the proposal of a “sigma opioid” class of receptors to explain aberrant behavior upon exposure to \(N\)-allyl-normetazocine (SKF-10047, NANM) in the chronic spinal dog model of precipitated opioid abstinence.\(^1\) The use of racemate NANM in initial binding studies complicated matters, implicating binding to phencyclidine (PCP) sites,\(^2\) and to multiple opioid sites.\(^3\) In the following years, it was determined that sigma sites were distinct from opioid and PCP receptors,\(^4\) and that the psychotomimetic effects of NANM were not a result of sigma affinity.\(^5\)

1.1 Therapeutic potential of \(\sigma\) receptor ligands

Many neuroleptics and antidepressants have affinity for \(\sigma\) receptors and consequently their role in motor side effects of antipsychotics and as modulators of serotonergic and glutamatergic neurotransmission in depression have been well established.\(^6,7\) In the early 1990s, \(\sigma\) antagonists were also demonstrated to attenuate the psychostimulant effects of cocaine\(^8\) and methamphetamine,\(^9\) and more recently 3,4-methylenedioxymethamphetamine (MDMA).\(^10\) There is evidence that \(\sigma\)
ligands may also be of use in treating ischemia and certain cognitive disorders.\textsuperscript{11} Sigma sites are particularly interesting targets for cancer imaging agents due to their role in apoptosis and their overexpression in a wide range of cancers, where saturation of \( \sigma \)-2 receptors is associated with selective cytotoxicity towards several cancer cell lines.\textsuperscript{12}

1.2 Sigma receptor subtypes

1.2.1 Sigma-1 receptor

Sigma-1 receptors have been well characterized, displaying a rank order of affinity for haloperidol (HAL), 1,3-di-\( o \)-tolyl-guanidine (DTG), (+)-3-(3-hydroxyphenyl)-\( N \)-(1-propyl)piperidine (\( R \)-(+)3-PPP), (-)-cyclazocine, (+)-NANM, PCP, etoxadrol, dexoxadrol, and MK-801.\textsuperscript{13} These sites are distributed in both the central nervous system and peripheral tissues. The \( \sigma \)-1 receptor was initially cloned by Hanner et al. in 1996,\textsuperscript{14} and was demonstrated to be a 223 amino acid 25.3 kDa peptide whose closest known homologue is the yeast \( C \)\textsubscript{8} sterol isomerase. To this day, the tertiary structure of \( \sigma \)-1 receptors or close homologs have not been elucidated by crystallography. As a consequence, structure-based drug design is not currently a viable option for the design of novel \( \sigma \)-1 ligands. Although a homology model has recently been developed for the \( \sigma \)-1 receptor,\textsuperscript{15} ab initio and comparative modeling techniques are currently insufficient to make confident structural predictions for a protein of this size.

1.2.2 Sigma-2 receptor

In the late 1980s it became clear that the \( \sigma \) ligand DTG bound to a second class of receptor sites with rank order affinities to benzomorphans that contrasted themselves from classical sigma sites.\textsuperscript{16} These sites were dubbed \( \sigma \)-2 sites, and for many years now, a great challenge in \( \sigma \) research has been to discover substances with selectivity towards this subsite. Very recently, a \( \sigma \)-2 receptor was putatively identified as the 21.7 kDa Progesterone Receptor Membrane Component 1 (PGRMC1).\textsuperscript{17} PGRMC1 has a conserved cytochrome b5-like heme/sterol binding domain, and the presence of close homologs in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank.
Bank\textsuperscript{18} suggests that homology modeling may be a feasible approach for generating models for structure-based drug design. Nevertheless, a comprehensive database of ligand structures with $\sigma$-2 binding affinities and associated scaffolds will be of significant utility for ligand-based drug design and QSAR. Until a more comprehensive correlation of PGRMC1 and $\sigma$-2 ligands is carried out, ligand-based drug design strategies may well prove more meaningful.

1.2.3 “Sigma-3”

For several years, Booth, Wyrick, Myers and others described a putative “$\sigma$-3” site with binding affinities different still from $\sigma$-1 or $\sigma$-2.\textsuperscript{19–22} These sites were labeled by 1-phenyl-2-amino-1,2,3,4-tetrahydronaphthalenes and were later determined to be histamine $H_1$ receptors.\textsuperscript{23} Subsequent literature dropped the reference to $\sigma$-3, but this points out the importance of scaffold recognition and diligent pharmacological classification of “novel” receptors.

1.3 Sigma pharmacophore and QSAR development: 1980–2011

Medicinal chemists often assess the performance of molecules in targeted biological assays in terms of a pharmacophore. What is meant by a “pharmacophore” can differ quite a lot from one practitioner to another. There are numerous examples of different usage of the term in the $\sigma$ literature. In order to be clear about the terminology, and to explain why many of the references to pharmacophores are not included in the following review, two usages of the term are distinguished. Pharmacophore motifs are abstract generalizations of molecular moieties or scaffolds that attempt to explain, in a retrospective sense, why certain molecules possess binding affinity at a macromolecular target. Pharmacophore models, on the other hand, are specific, well-defined combinations of features that have some predictive utility in assessing the potential of a molecule to bind to the target.

Pharmacophore motifs were quite common in the early $\sigma$ literature. To this day, ligand-based drug design efforts are couched in terms of their agreement to those motifs, and they merit review because of their longstanding influence on the interpretation of $\sigma$ ligand affinity. Several very predictive pharmacophore models have also been developed over the years. Many of these models
highlight several exemplary features possessed by almost all σ ligands developed to date; others attempt to elaborate upon putative binding site features that imply complementary features of the ligands themselves. Descriptor- or fingerprint-based QSARs are not reviewed if they do not elaborate upon a pharmacophore model (for the curious, see references 24–34).

1.3.1 Non-selective σ models

1.3.1.1 Manallack topological pharmacophore model

Manallack et al.\textsuperscript{35,36} developed the first pharmacophore models for σ ligands based on the method of Lloyd and Andrews\textsuperscript{37} in an attempt to differentiate the binding requirements of NMDA and σ ligands. This method distinguishes three receptor features that are built around each ligand, namely two “hydrophobic” features that are aligned normal to the plane of an aromatic ring at its centroid, and a third receptor feature represented by a vector from the lone pair electrons of a ligand N to a postulated H-bond acceptor belonging to the receptor. Receptor and ligand features are presented in Figure 1.1. Conformational energies, RMS distances between receptor or ligand features, and common volume overlaps to the template structure were used to guide the alignment process.

The model was designed around the crystal structure of $\text{trans-}(4aR,10bR)-9\text{-OH-}N-n\text{-propyl-octa-hydrobenzo[}f]\text{quinoline}$, a relatively rigid ligand that demonstrated a high eudismic ratio for σ
receptor sites. High-affinity ligands towards σ sites defined by the radioligand $[^3H](R)-3$-PPP were selected as active analogues, although $(R)$-3-PPP was later shown to have mixed relative affinities with respect to $\sigma$-1 and $\sigma$-2 sites. Crystal structures of $[^3H](R)$-3-PPP and $d$-NANM, and the modeled structures of DTG and HAL were used in the alignment process. Torsional analysis was required for the modeled structures and $(R)$-3-PPP, whereas the pseudochiral configuration about the N of $d$-NANM required inversion for a better fit to the template ligand.

The final model suggested a distance of 5.06 Å between the aromatic and N features. Additional electrostatic calculations were employed in their study with equivocal results depending upon the scaffold of the analogs fitted to the model. A lipophilic pocket was hypothesized to accommodate steric features of some ligands that did not detrimentally impact affinity. While the construction of a model for the “σ” receptor is useful from a conceptual standpoint, this model regrettably suffers from a paucity of ligand-based features.

1.3.1.2 Gund pharmacophore (1991)

Gund and Shukla developed a three-point pharmacophore model based upon the structures of HAL, $(R)$-3-PPP, progesterone and dextropentazocine in their neutral ionization states. Conformational analysis, molecular superposition, and electrostatics were all considered in the construction of the final model. The optimal superposition of all compounds produced 4 separate pairwise pharmacophore distance measurements.

The inclusion of progesterone as an “active” compound is questionable because of its low σ affinity, and marked differences in the pharmacophore features compared to the other active ligands used in their model development. Namely, progesterone lacks both the aromatic- and N-moieties present in the other compounds. This required special alignments of the centroid of ring B to the centroid of HAL’s fluorophenyl ring or alternatively to selected atoms of $(R)$-3-PPP and dextropentazocine aromatic rings. Oxygen lone pair electrons from ring D of progesterone were aligned to the N lone pairs of the other compounds. These special alignment considerations and the use of a non-selective radioligand restrict the general applicability of the resulting model.
1.3.1.3 Gund pharmacophore (1992)

Gund et al. reinvestigated their prior model\textsuperscript{39} after the identification of PRE-084 as a selective \( \sigma \) ligand.\textsuperscript{40} As part of their modification, molecules were superimposed upon a single ligand, HAL. Template features selected as references for superposition were the fluorophenyl moiety, the basic N-atom, and the lone-pair of the N-atom (Figure 1.2). As with the prior model, special alignment considerations were required, especially in the case of progesterone. The positioning of the cyclohexyl moiety of PRE-084 in their final model was hypothesized to explain its selectivity towards \( \sigma \) receptors.

![Figure 1.2: Gund's (1992) HAL-based pharmacophore model.\textsuperscript{40} Reprinted with permission from NPP Books.\textsuperscript{40} Copyright (1992) NPP Books.](image)

1.3.1.4 Ablordeppy pharmacophore and CoMFA model (1992)

The model of Ablordeppy et al.\textsuperscript{41} (Figure 1.3) was based upon \textit{trans-}(4aR,10bR)-7-OH-N-n-propyloctahydrobenzo[f]quinoline, similar to the template utilized in the model of Manallack and Beart.\textsuperscript{36} This was the first application of CoMFA\textsuperscript{42} to \( \sigma \) QSAR development. Training and test set data for the study were generated from a variety of assay methodologies using the non-selective
radioligand [³H]DTG without a σ-1 site masking agent. Assay variability may have been amplified in this model because of the combination of both $K_i$ and $IC_{50}$ data.

**Figure 1.3:** Ablordeppey’s template molecule *trans-*(4aR,10bR)-7-OH-N-n-propyloctahydrobenzo[f]quinoline, showing proposed molecular alignment sites. Reprinted with permission from Ablordeppey et al. Copyright (1992) Springer SMB NL.

Basic pharmacophore features identified in their model were an aromatic ring centroid separated from a positive ionizable N by a minimum of an ethylene spacer, and an pendant propyl chain on the N-atom. Training and test set ligands were aligned to a minimum of 3 of the proposed sites, although the exact details of individual alignments were not specified. In addition, some of the ligands in the dataset possessed unresolved centers of asymmetry, and the rationale for the selection of one isomer for alignment over the other(s) was not described. A CoMFA based on the alignment of training set molecules exhibited remarkable internal consistency ($r^2 = 0.979, q^2 = 0.843$; 4 PLS components), and the external test set statistics ($Q^2 = 0.88, 0.67$; two different sets) reflected a high quality model.

1.3.1.5 Seri-Levy eudismic analysis

While not a pharmacophore model *per se*, Seri-Levy et al. employed conformational analysis and superposition of a series of 3-(3-hydroxyphenyl)piperidines (3-HPPs) to investigate the stereochemical requirements of σ receptors. The original data were generated from competition binding experiments using [$³H](R)-3-PPP as the radioligand. Stereoisomers of 3-HPP were aligned in two different ways for the calculation of electrostatic potential and shape chirality indices, and the superposition of the 3-hydrophenyl rings was found to be optimal. Correlation of the shape chirality
indices to eudismic indices revealed “non-Pfeiffer behavior” for the homologues investigated and questioned the importance of an H-bond acceptor lying in proximity to these ligands within the σ receptor.

1.3.1.6 Beart pharmacophore model

Beart et al. examined the conformational requirements of a series of ifenprodil-related heterocyclic amino alcohols. Their investigation used \( ^3\text{H}\)(R)-3-PPP as a radioligand. Receptor and ligand sites (see Figure 1.4), as well as the template ligand, were defined using the same criteria proposed by Manallack et al. Low energy conformers were generated with by the application of MD or torsional analysis, and molecular models were superimposed by a three-point fit to within 0.6 Å. All of the compounds investigated fit the model well, with tight RMS fit and low energy with respect to the global minimum found through the conformer search protocols.

Figure 1.4: Beart’s pharmacophore model elaborated upon the template of Manallack, focusing on the SAR of ifenprodil analogues. Reprinted with permission from Beart et al. Copyright (1994) Elsevier Science Publishers

1.3.2 Sigma-1 models

1.3.2.1 Carroll pharmacophore model

Carroll et al. developed a σ-1 pharmacophore by modification of their previously proposed PCP pharmacophore motif. While not strictly a σ-1 model, as the activity values were derived from experiments using both \( ^3\text{H}\)(R)-3-PPP and \( ^3\text{H}\)dextropentazocine as radioligands, the features of
benzomorphan scaffolds aligned to their PCP model were the basis for pharmacophore development. Like the model of Manallack et al., three receptor features were assigned based upon a vector placed normal to the plane of the aromatic ring at its centroid along with a H-bond vector feature involving the lone pair on the N atom, although the exact placement of the N was deliberately left unspecified. An additional receptor feature was proposed to comprise a lipophilic pocket in the receptor capable of inducing selectivity for dextrorotatory benzomorphans depending upon the nature of the pendant N-alkyl substituents.

1.3.2.2 Elaborated Carroll pharmacophore model

Carroll et al. revisited their previous molecular modeling study with a deeper investigation into the flexibility of N-substituted N-normetazocine side-chains. This modification allowed the formation of additional hypotheses (see Figure 1.5) regarding the common volumes occupied by the benzomorphan side-chains and the volumes unique to the side-chains of high- and low-potency ligands.

![Figure 1.5: Carroll's N-substituted N-normetazocine based pharmacophore.](image)

1.3.2.3 Gilligan pharmacophore

Gilligan et al. provided the first model explicitly investigating the nature of σ-1 binding requirements utilizing \( d \)-NANM as a competitive binding radioligand. Based on the pharmacodynamic characteristics of 15 compounds with selectivity for σ-1 receptors over dopamine D\(_2\) and 5-HT\(_{2A}\), four pharmacophore features were identified as important elements for σ-1 affinity. These features comprise a basic N attached to two separate hydrophobic moieties, with an H-bonding feature between the nitrogen and its farthest hydrophobic partner (preferably an aromatic moiety). The distances between such features were calculated by conformational analysis, averaged and published along with their standard deviations, all of which fell within 2 Å. While this pharmacophore is historically significant, the relative contribution of each feature is not readily apparent due to the use of a fairly congeneric series of compounds for pharmacophore elucidation and the same magnitude of receptor affinity of these compounds (8–51 nM).

1.3.2.4 Hudkins pharmacophore motif

Hudkins et al. described a pharmacophore based upon a set of caramiphen analogues. They proposed a “lipophilic site 1” associated closely with a N-binding site, together corresponding to ligand features of benzomorphan and 4-aryl-piperidines. A distal “lipophilic site 2” and alternative binding mode was proposed to explain the affinity of the arylcyclopentyl ester scaffold of their designed analogues and HAL. Distances were discussed in terms of methylene spacers, and no discussion of intersite angles was presented.

1.3.2.5 Glennon pharmacophore motif

Glennon et al. proposed a general three-point motif for σ-1 binding based upon a congeneric series of relatively flexible phenylalkylamines using \( [\text{H}] \)dextropentazocine as radioligand. The salient features of their model include a hydrophobic site in the receptor, a proton-donating site corresponding to the almost requisite amine motif, and a secondary hydrophobic site with a propensity for accommodating bulky groups. While there are three site points to this model, they were
1.3.2.3 Hudkins’ 3-site pharmacophore recapitulates the presence of two hydrophobic sites separated by a positive-ionizable N-binding site. Reprinted with permission from Hudkins et al. Copyright (1994) American Chemical Society.

presented in terms of the distances of each hydrophobic site to the N-atom. The distances between the hydrophobic groups was not stated, likely because of the wide range of distances tolerated between the primary hydrophobic site and the amine-N. This pharmacophore model was revisited in a comprehensive review of their group’s σ-1 drug development strategy, and also used to rationalize the SARs of 6,8-diazabicyclo[3.2.2]nonan-2-one and 6,8-diazabicyclo[3.2.2]nonen-2-one derivatives.

Figure 1.7: Glennon’s 3-site model provided general intersite distance information and proposed a receptor region tolerating hydrophobic bulk. Reprinted with permission from Glennon et al. Copyright (1994) American Chemical Society.

1.3.2.6 Ablordepey pharmacophore and CoMFA model (1998)

As part of an investigation of the configurational requirements of the endo- and exo-stereoisomers of the high-affinity ligand SC-50691, Ablordepey et al. revisited the model of Glennon et al. with a CoMFA analysis based upon dextropentazocine as a template. After aligning 1-(5-phenylpentyl)-piperidine to the template, the remaining compounds in the dataset were aligned to either
the template or to the representative piperidine. The distance between the “phenyl-B” site and the amine-N of either the template or of the piperidines was held constant through a heuristic set of RMS fitting criteria while the remainder of each molecule was allowed to attain an extended conformation in the sterically accommodating “phenyl-A” region.\textsuperscript{56}

Significantly more care was taken in the development of this model than in the their prior non-selective σ pharmacophore.\textsuperscript{41} The training set represented alkyl- and aryl-amines, piperidines, piperazines, and a variety of compounds from different pharmacologic classes with σ-1 affinity, some with significant eudismic ratios. The test set was well sized, comprising 25\% of the total number of compounds, and was also well within the domain of applicability of the training set. This was reflected in the external test statistic \((Q^2 = 0.65)\) and the accurate prediction of the affinity constants of the resolved isomers of SC-50691. Their final model based on 64 total compounds demonstrated remarkable internal correlation and consistency \((r^2 = 0.989, q^2 = 0.732; 7\) PLS components).

1.3.2.7 Huang pharmacophore and CoMFA model

The work of Huang et al.\textsuperscript{57} focused on derivatives of \(\text{N-(1-benzylpiperidin-4-yl)-3-bromophenylacetamide (BPP)}\). Through conformational analysis of BPP and superimposing a low-energy conformer onto pentazocine with a few manual adjustments, the authors were able to align a set of 79 BPP derivatives\textsuperscript{57,58} for CoMFA studies. As part of their investigation, they sought to probe the importance of the carbonyl-to-N distances on σ-1 affinity. Intersite distances of BPP features in common with those present in the Glennon et al. pharmacophore\textsuperscript{52} were also enumerated. The full matrix of intersite distances was not disclosed.

CoMFA results for 76 molecules in the dataset provided a model with superb correlation and consistency \((r^2 = 0.91, q^2 = 0.61; 6\) PLS components)\textsuperscript{57} for the σ-1 dataset, but failed to perform suitably for the σ-2 dataset. The steric diversity around the benzyl-group corresponding to Glennon’s secondary binding site was much less than that presented by the arylacetamide moieties corresponding to the primary hydrophobic site. Because of this choice, the steric tolerance of the
secondary binding site could not be corroborated. However, the steric and electrostatic nature of
the primary hydrophobic receptor site were substantially more well defined by the CoMFA study.

1.3.2.8 Cao CoMFA model

Cao et al. used \([3-(cis-3,5-Dimethyl-4-[3-phenylpropyl]-1-piperazinyl)-propyl]-N,N-bis(4-fluo-
rophenyl)amine\) as template for the alignment of piperazinyl bis(4′-fluorophenyl)amine derivatives.
Two alignments based on alternative assignment of the piperazine N-atoms as the ubiquitous amine-
N feature present in σ pharmacophores were utilized, with the fluorophenyl-ring systems serving
as the classic σ hydrophobic groups. The preferred alignment based upon the CoMFA results
indicated that the proximal-N to the \(bis(4′-fluorophenyl)amine\) served as a better surrogate for
the requisite N-feature. Although precise intersite distances were not presented, the best CoMFA
model had a respectable level of internal correlation and consistency \((r^2 = 0.929, q^2 = 0.521; 4 \text{ PLS}
components})\).

1.3.2.9 Gund pharmacophore model (2004)

The model published by Gund et al. in 2004\(^{60}\) was templated upon PD144418, with superposition
of spipethiane, dextropentazocine, and HAL onto the general CNS motifs described by Lloyd
and Andrews\(^{37}\) along with an optional O- or S- feature present in a number of the molecules.
Electrostatic potential contours were used to characterize the placement of electronegative features,
suggesting that the secondary binding site plays an important role in ligand affinity. The final model
included a total of 3 ligand features, and the performance of σ-1 selective agents vs progesterone in
fitting the pharmacophore was notable.

1.3.2.10 Jung DISCOtech pharmacophore and CoMFA model

Jung et al. employed DISCOtech on a series of spipethiane analogues, piperidine- and piperazine-
analogues of caramiphen, benzoxazolones, benzothiazolones, and several notably σ-1-selective
molecules from the literature to develop a pharmacophore alignment and subsequent CoMFA
model. The training set comprised a total of 43 compounds and a test set of 5 compounds, and both sets represented diverse scaffolds ranging over 3 orders of magnitude in affinity for dextropentazocine defined sites. An initial 3-point pharmacophore including the aromatic ring centroid, the N-atom, and a projected H-bond from the N provided CoMFA models with poor results. They successfully overcame this obstacle by further optimization of conformers with semiemperical AM1, Hartree-Fock, DFT, or MP2 calculations, in tandem with scaling the projected H-bonding distance down to 1.4 Å. The reoptimization was followed by a very specific atom-based alignment scheme, and CoMFA analysis. AM1 charges provided better fits for electrostatic contributions, whereas the HF geometries provided more significant steric contributions to the PLS regression. The HF methodology also enhanced the external test set predictivity.

1.3.2.11 Laggner HypoGen pharmacophore model

Laggner et al. used training and test sets of diverse pharmacological classes and a wide range of activities to develop pharmacophore models for σ, emopamil binding protein, and yeast ERG2. The training set of 23 compounds was deliberately chosen to reflect diverse structures and a substantial range in affinity, as encouraged by the HypoGen documentation, although the σ-1 affinity data were not spread equally through the range. Pharmacophores for all three targets presented 5 common features; four hydrophobic ligand sites and a single positive ionizable N-atom. The top σ-1 model had high correlation (Pearson $r = 0.926$) and cost function analysis following response randomization indicated a high degree of confidence that the correlation was not spurious (95% confidence level). Estimated affinities were reported for a test set of 9 ligands, although the affinity of one of the σ-1 compounds was not determined. While the authors did not explicitly state the performance of the pharmacophore model on this external test set in terms of the Pearson correlation coefficient, the value is readily calculated from their data ($r = 0.403$). An agreement was found between their σ-1 pharmacophore and that developed by Glennon et al. Furthermore, application of their pharmacophore model to a virtual screening protocol resulted in the discovery of 5 unique hits with $K_i \leq 100$ nM.
This model was later applied to two series of alkenyl- and arylalkyl-amines where it was found to consistently overestimate affinity values. The discrepancy in predicted vs observed binding affinity was ascribed to interlaboratory variations and the inclusion of the particularly high-affinity compound fenpropimorph in the original training set.

1.3.2.12 Zampieri HypoGen pharmacophore model

Zampieri et al. used HypoGen methodology to derive an interesting pharmacophore from a series of 31 benzooxazolones. While the affinity range of training set structures and the size of the dataset fit the suggested criteria found in the HypoGen documentation, the diversity of the selected structures is questionable, as the main difference in each “series” is the length of the linker from the N-feature to the benzooxazolone-N (i.e. 3–4 methylene units), and in the case of the piperidine series, the length could be interpreted as a constrained 4-methylene unit linker. Notwithstanding this caveat, the top-ranked hypothesis included five features which were all consistent with the broader σ-1 literature, such as an aliphatic hydrophobic site, two aromatic hydrophobe sites, a positive ionizable N-site, and an H-bond acceptor. Cost function analysis indicated a high confidence in a true correlation (Pearson $r = 0.896$) between predicted and experimental activities. A test set of related benzylpiperidine-4-carboxamides and σ-1 reference ligands performed remarkably well (Pearson $r = 0.882$) when aligned to the pharmacophore model. Additionally, response randomization suggested a statistical significance of 98% for the top-ranked hypothesis. This model
was later used in the refinement of the first published σ-1 homology model.\textsuperscript{15}

![Figure 1.9: Zampieri's 5-feature benzooxazolone-based pharmacophore\textsuperscript{65} includes an H-bond acceptor site and differentiates between aromatic and aliphatic hydrophobic features. Reprinted with permission from Zampieri et al.\textsuperscript{65} Copyright (2009) American Chemical Society.](image)

1.3.2.13 Oberdorf Quasar pseudoreceptor model

Oberdorf et al. used a congeneric series of spirocyclic piperidines to develop a pseudoreceptor model of the σ-1 receptor.\textsuperscript{66} Out of the 87 total structures analyzed, only 5 were achiral, whereas the remainder comprised 41 enantiomeric pairs representing the racemate compounds tested in vitro. Development of the pseudoreceptor model required the initial preparation of a pharmacophore for structure alignment. Towards this goal, structures were protonated prior to conformational analysis and alignment, making this study distinct with respect to the other σ-1 models. A unique result of this study was the perception of an H-bond acceptor feature in the pharmacophore which complemented the positive ionizable nitrogen. According to the authors, the remaining sites presented features in line with the models of Glennon\textsuperscript{53} and Laggner et al.\textsuperscript{62} Unfortunately, the authors did not disclose the coordinates or relative distances of the features in this pharmacophore model.
1.3.2.14 Rossi Galahad pharmacophore model

Rossi et al. developed a 3-feature pharmacophore model from a set of congenic arylalke- namines. The most highly ranked hypothesis, according to their heuristics, possessed two hydrophobic features and a positive ionizable N. While the model was found to aid in classifying active vs inactive compounds, conventional methods of validating the model were not presented. Model parameters such as distances and coordinates between features were also not disclosed.

1.3.3 Sigma-2 models

1.3.3.1 Cratteri GRIND-based pharmacophore model

The model presented by Cratteri et al. used grid-independent descriptors to generate a description of the putative environment surrounding α-tropanyl derivatives in their bioactive conformations. Some of the derivatives were tested as racemate mixtures, and the decision was made to develop a model based upon the (R) isomers, stemming from their modestly higher eudismic ratio. A significant result of their analysis was a description of the putative dimensions of
three σ-2 receptor regions based on the selectivity of a subset of the data: distances between two hydrophobic regions, and from one of these to a H-bond donor region were disclosed. This suggests an upper limit on the total width of the binding pocket of the σ-2 receptor.

Figure 1.12: Cratteri’s GRIND-based pharmacophore\textsuperscript{68} provided a pseudoreceptor model which places some restrictions on the overall size of the binding site. Reprinted with permission from Cratteri et al.\textsuperscript{68} Copyright (2004) Kluwer Academic Publishers.

1.3.3.2 Abate pharmacophore and CoMFA model

Abate et al. used an automated alignment of cyclohexylpiperazines and congeners to the template molecule, 1-Cyclohexyl-4-[3-(5-methoxy-1,2,3,4-tetrahydro-naphthalen-1-yl)propyl]piperazine ((R)-PB28), in their pursuit of a predictive CoMFA model for σ-2 ligands.\textsuperscript{70} Initial CoMFA perfor-
mance with the automated alignment was poor, so an alternative manual alignment of the dataset to 1-Cyclohexyl-4-[3-(naphthalen-1-yl)propyl]piperazine was undertaken, with the focus not on global alignment of all features, but rather on the overlay of the piperazine ring. To simplify the alignment, and because of electrostatic considerations, the piperazine moieties were modeled in their dibasic ionization state. The manual alignment led to CoMFA models with both internal correlation and consistency, as well as external predictivity for structurally related compounds ($r^2 = 0.95$, $q^2 = 0.73$, $Q^2 = 0.55$; 4 PLS components). The final model supported a SAR interpretation of the binding requirements of cyclohexylpiperazines to $\sigma$-2 receptors. Apart from suggesting an important $\sigma$-2 pharmacophore feature, no intersite distances between multiple features were proposed.

1.3.3.3 Laurini Catalyst pharmacophore model

The first ligand-based pharmacophore model specifically aimed at determining the features necessary for $\sigma$-2 affinity were recently reported by Laurini et al., based on the benzooxazolone motif. This set of ligands had been previously used in the development of a $\sigma$-1 pharmacophore, and share the same drawbacks with regard to the diversity represented in the training set. Affinities for $\sigma$-2 span a much narrower range than for $\sigma$-1, and are not distributed evenly through the training set.

![Figure 1.13: Laurini’s 5-feature benzoazolone-based $\sigma$-2 pharmacophore](image)

Figure 1.13: Laurini’s 5-feature benzoazolone-based $\sigma$-2 pharmacophore implicated an absolute requirement of an aliphatic hydrophobic feature for $\sigma$-2 binding and revisits the H-bond acceptor feature proposed by Zampieri. Reprinted with permission from Laurini et al. Copyright (2010) Pergamon.
Assuming that the previous considerations do not adversely affect the regression, this is a groundbreaking pharmacophore. The model implicates an absolute requirement for an aliphatic hydrophobic moiety for high affinity, although the authors admit that σ-2 selectivity is not explained by this feature. An aromatic hydrophobic site and a general hydrophobic feature build upon this pharmacophore. The ubiquitous positive ionizable N is also present. Rounding out the 5-feature pharmacophore is an H-bond acceptor complementing the carbonyl of the benzoxazolones. Response randomization studies indicate a 95% confidence level that the hypothesis is statistically significant, and the robustness of the model is backed by leave-one-out pharmacophores showing no substantive difference from the best hypothesis.

1.3.4 “Sigma-3” model

Myers et al. used DISCO (as part of the Sybyl package) to develop a CoMFA model for a series of phenylaminotetralins which displayed affinity for a putative “σ-3” receptor. Subsequently Bucholtz et al. determined that the predominant activity of these compounds was at histamine H1 receptors. Accordingly, the development of their “σ-3” model and the corresponding QSARs will not be discussed, and is mentioned only for the edification of the curious reader.

1.3.5 Summary of pharmacophore models in the σ literature

As seen in Table 1.1, a great number of pharmacophore models have been proposed over the years. Many of the early pharmacophores, and several of the more recent σ-2 pharmacophores are based upon perceptions of a virtual receptor or projected points. Until the mid-nineties, these pharmacophore models were not developed with a focus on receptor subtype selectivity. Only two of the studies involved what would be regarded today as “diverse” structures; the majority were focused on elaborating the SAR of a congenic series of molecules. Affinity prediction has also become more of a focus in the past decade. The consideration of molecular alignments has been a recurring challenge in σ pharmacophore development, requiring special consideration in many cases. When manual superposition to a template or automated alignments of congenic series
are applied in CoMFA studies, there is a likelihood that spurious correlations will be found in the SAR because the chosen alignments do not accurately reflect the conformation of the ligands to receptor sites. Additionally, a majority of the experiments carried out with diverse scaffolds were performed with only a handful of compounds. Because of the symmetric nature of many of the 3-point pharmacophores with Hydrophobe-Nitrogen-Hydrophobe features coupled with the fact that many σ ligands possess a high degree of flexibility, the issue is raised over whether multiple binding modes may exist. Further complicating matters is the possibility of subsites on σ receptor subtypes. These factors all detract from the general utility of the majority of these models for ligand-based drug design.

1.4 Overview of σ binding assay methodology

An overwhelming number of radioligands and binding assays have been developed for imaging and competition binding studies of σ receptors. Despite the preponderance of radioligands, there are few subtype selective compounds commonly used for σ-1 research, of which (+)-pentazocine is most commonly used, and only one generally-used radioligand (DTG) for σ-2 receptor assays, which is ordinarily accompanied by a modicum of PTZ to saturate residual σ-1 sites. Sigma receptors are ubiquitously expressed in all tissues with minimal specificity. While cloned σ-1 receptors have been expressed in yeast and HeLa cells, almost all competition binding experiments to date have been performed in brain or liver tissue homogenates from guinea pig or rat. The majority of σ-1 assays have been performed using guinea pig brain membranes, whereas rat liver membranes are generally preferred for σ-2. The diversity of assay methodologies presents a challenge for computational efforts aimed at generating QSARs and pharmacophore models, as the variability in assay performance under different conditions can lead to significant differences in perceived ligand affinity. This variability carries over as uncertainty in model development and negatively affects model performance.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Receptor</th>
<th>Ligand</th>
<th>Lone pairs</th>
<th>QSAR</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>σ non-selective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manallack et al.</td>
<td>1988</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Gund and Shukla</td>
<td>1991</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Gund et al.</td>
<td>1992</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>Ablordeppey et al.</td>
<td>1992</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>•</td>
<td>41</td>
</tr>
<tr>
<td>Seri-Levy et al.</td>
<td>1994</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>•</td>
<td>43</td>
</tr>
<tr>
<td>Beart et al.</td>
<td>1994</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td><strong>σ-1 selective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carroll et al.</td>
<td>1992</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td></td>
<td>47,49</td>
</tr>
<tr>
<td>Gilligan et al.</td>
<td>1992</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Hudkins et al.</td>
<td>1994</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Glennon et al.</td>
<td>1994</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Ablordeppey et al.</td>
<td>1998</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>•</td>
<td>56</td>
</tr>
<tr>
<td>Huang et al.</td>
<td>2001</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>•</td>
<td>57</td>
</tr>
<tr>
<td>Cao et al.</td>
<td>2003</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>•</td>
<td>59</td>
</tr>
<tr>
<td>Gund et al.</td>
<td>2004</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Jung et al.</td>
<td>2004</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>•</td>
<td>61</td>
</tr>
<tr>
<td>Laggner et al.</td>
<td>2005</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>•</td>
<td>62</td>
</tr>
<tr>
<td>Zampieri et al.</td>
<td>2009</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>•</td>
<td>65</td>
</tr>
<tr>
<td>Oberdorf et al.</td>
<td>2010</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>•</td>
<td>66</td>
</tr>
<tr>
<td>Rossi et al.</td>
<td>2011</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>•</td>
<td>67</td>
</tr>
<tr>
<td><strong>σ-2 selective</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cratteri et al.</td>
<td>2004</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>•</td>
<td>68</td>
</tr>
<tr>
<td>Abate et al.</td>
<td>2009</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>•</td>
<td>70</td>
</tr>
<tr>
<td>Laurini et al.</td>
<td>2010</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>•</td>
<td>71</td>
</tr>
</tbody>
</table>

1.5 Research Objectives

Given the lack of selective σ-2 probes and the lack of general utility of many of the modeling efforts to date, the field of σ research remains open to new discoveries. Modeling software and computational techniques have advanced significantly in recent years, and a fresh approach using curated data is overdue. That being said, σ ligand design has routinely focused on highly flexible molecules with a great degree of pharmacophore symmetry that pose challenges when there is a lack of crystallographic information to provide target structures. Virtual screening and affinity assessments, particularly efforts at identifying selective scaffolds, will require a more fundamental
understanding of the general binding requirements of σ ligands. As part of these efforts, the following objectives were established:

1. Create a database of binding competition experiments to identify experimental methodologies that are sufficiently similar for dataset collection.

2. Collect and curate data from the literature which utilize the most prevalent methodologies.

3. Use a combination of scaffold analysis, distance measures, and hierarchical clustering to generate representative training and test sets for classification of active and inactive molecules towards either σ-1 or σ-2, depending upon the underlying methodology.

4. Identify the most significant conformers of the active compounds for pharmacophore development.

5. Use these pharmacophores to clarify structural features necessary for affinity and selectivity.
2. METHODS

We have no future because our present is too volatile. We have only risk management. The spinning of the given moment's scenarios. Pattern recognition.

Pattern Recognition
William Gibson

2.1 Computational equipment

Database preparation and analysis was undertaken on a 2.8 GHz Intel Core 2 Duo Mac OS X laptop furnished with 8 GB 1067 MHz DDR3 RAM. Computational methods were performed on the same system, or on a 2.5 GHz Intel Core i5 Mac OS X desktop furnished with 4 GB 1333 MHz DDR3 RAM. CPU and memory intensive calculations were done on a Fedora Linux cluster comprising 8 Microway computers, four with 2.5 GHz 8-core Intel Xeon L5420 CPUs and 16 GB 667 MHz DDR2 RAM, and another four equipped with 2.27 GHz 16-core Intel Xeon L5520 CPUs and 32 GB 1333 MHz DDR3 RAM.

2.2 Data collection and curation

Articles with competition binding data were identified from the literature using a combination of electronic search and cross-referencing to related literature found in proceedings, books, and patents, to a limited extent. Additional sources of binding data included data generated by our lab and those generated by our colleagues. Bibliographic information as well as details concerning the
radioligand, masking ligands (if present), matrix (i.e. organism and tissue preparation), as well as PubMed ID (if available) were recorded in a PostgreSQL database (version 9.2). Records with competition binding data were analyzed to determine which experimental methodologies were most prevalent. An experimental methodology was defined as a combination radioligand, masking ligand, and matrix.

Following the identification of relevant data sources, those competition experiments with the most reliable data (e.g. $K_i$ in lieu of $IC_{50}$ data) were added to a project table in Maestro (version 9.3, Schrödinger, LLC, New York, NY, 2012). Molecules were sketched using the stereochemistry assigned in the original article, unless corrected for in an erratum, or unless there was evidence in the experimental section that the molecule had been incorrectly sketched in the paper. Affinity data and experimental methodology were annotated along with any assay limits, if applicable. In the case that the stereochemistry was unknown, ambiguous, or known to be tested as a racemate, the molecule was marked as racemate. Several other properties and SMILES strings were calculated using the Generate SMILES script in Maestro and QikProp (version 3.5, Schrödinger, LLC, New York, NY, 2012). Once sketched, LigPrep (version 2.5, Schrödinger, LLC, New York, NY, 2012), was used to clean up the structures and to generate relevant stereoisomers, if necessary. All molecules were double-checked for consistent stereochemistry and accurate affinity data entry. A Python script (Appendix E.1) was written to automate the processing of PubMed IDs into SMILES strings, but it was determined that the PubMed records were not reliable enough to identify molecules of interest.

2.3 Dataset composition

The initial dataset was divided into subsets representing each predominant methodology. These subsets were limited to contain molecules of unambiguous stereochemistry that were determined to have affinity within the assay limits. On occasion, a molecule would be tested more than once under a given set of experimental conditions, and thus would be replicated in the dataset. In order to deal with this situation, which was very common for reference compounds, the molecules were imported
into Canvas\textsuperscript{74} (version 1.5, Schrödinger, LLC, New York, NY, 2012), duplicate identification was performed, and statistical analysis was carried out on the duplicates. This analysis was undertaken to determine the variability of the collated experimental data. After analysis, duplicate data points were removed, retaining the highest affinity ($pK_i$) measured in any experiment. This choice was made deliberately because of a classification technique (\textit{vide infra}) used later on in the modeling process. Expurgated datasets were then exported as SMILES and CSV files.

2.3.1 Scaffold decomposition

Molecular scaffolds were decomposed by two methods implemented in Strip-It (version 1.0.1, Silicos-it, Schilde, Belgium, 2012). Initially, a canonical SMILES representation of each molecule was converted into RINGS\_WITH\_LINKERS\_2 scaffolds preserving the ring structures, any linking chains between rings, and exolinker double bonds. Aside from these linkers, all pendant groups are eliminated in this process, which allows for a mapping of the molecule to an underlying scaffold that may be shared by other analogues. All of the unique scaffolds were then subjected to a subsequent decomposition into MURCKO\_1 core scaffolds. The core scaffolds are similar to the former, with the exception that all atoms are converted to C, the bond order between connected atoms is reduced to a single bond, and exolinker double bonds are removed. This process allows for the perception of common ring and linker systems and reduces the ambiguity presented by unsaturated systems and heteroatoms. It should be noted that information about stereochemical configuration is lost in the scaffold generation process, although this is fortuitous since it allows for molecules that have identical atom connectivity to be clustered together and provides an avenue for investigating eudismic analysis.

2.3.2 Core fingerprints

A measure of the difference between scaffolds was required in order to cluster similar scaffolds. Dendritic and radial (extended-connectivity) fingerprints calculated by Canvas were evaluated for this purpose based on the moderate to high number of bits typically set “on”, and on their
Stereochemical information was ignored in the hashing of radial fingerprints in order to be consistent with the scaffold generation process. Default atom typing (Daylight invariant and Fn functional types, respectively) were used in preparing core fingerprints. The impact of atom typing was not investigated because the core scaffolds do not distinguish atoms other than carbon, and also because bond orders were reduced by the core decomposition process.

With binary bit-strings, the minimum distance provided with this metric is 0, meaning the molecules are indistinguishable by fingerprint. The distance increases as the number of bits which are different in each fingerprint increases. Several distance measures are provided in Canvas, and it was determined that the Euclidean distance measure worked very well to distinguish clusters of molecules with similar scaffolds. Euclidean distance \( D = \sqrt{A + B} \), where \( A \) is the number of bits exclusively set “on” in the bitstring of scaffold molecule A, and \( B \) is the number of bits exclusively set “on” in the scaffold of molecule B. Based on the initial results (data not shown), it was determined that radial fingerprints at the default level of 4 iterations provided a sufficient degree of distance to ensure that core scaffolds could be distinguished from one another, whereas dendritic fingerprints with the default 5-atom path length led to a small but significant proportion of cores that were not distinguished.

### 2.3.3 Core clustering and network visualization

Following the general procedure of Guiguemde et al., the McQuitty linkage method of hierarchical clustering was chosen in order to ensure that every core scaffold was connected to at least one other core by a node. Kelley’s criterion was used to select an appropriate number of clusters. This particular method is well suited for the application of common scaffold perception when paired with a suitable fingerprint and distance metric. Structures and clusters are paired based on the minimizing the average distance between members, and as clustering progresses, more distant candidates are paired incrementally in such a manner to ensure that the largest clusters of similar cores are produced. The end product of clustering is a dendrogram relating successive nodes to leaf core scaffolds and child nodes.
The visualization method of Guiguemde et al.\textsuperscript{76} was implemented with slight modifications using custom scripts to process the scaffolds generated by Strip-It and the cluster dendrograms produced by Canvas. An edge network consisting of molecule to scaffold, and scaffold to core relationships was supplemented by core to node, and node to node edge relationships from the dendrogram. When singleton molecule to scaffold relationships were identified, they were replaced by molecule to core relationships and the original molecule to scaffold relationships were deleted from the network. No attempt was made to identify common child scaffolds. The resulting network was visualized using the yFiles circular network layout in Cytoscape\textsuperscript{78} (version 2.8.3, Cytoscape Consortium). Binding affinity data for molecules were imported from the CSV file as node attributes. By applying node size and color mappings based upon binding affinity, clusters with a sufficiently large number of members and a suitable spread of affinity data could be selected and assigned cluster membership for pharmacophore and QSAR development in Phase. Suitable values for core-node distance for cluster discrimination were found at a Euclidean distance of \textasciitilde10–12, based on the classical benzomorphan and morphinan scaffolds. Scripts for scaffold decomposition, core fingerprinting, and core clustering may be found in Appendix A.

2.3.4 Selection of classification model training and test sets

Using the range of affinity values determined from the duplicate experimental points within each individual methodology, it was possible to define a range of affinity values to classify set members as active, not determined, or inactive. The choice of selecting the highest affinity value for each molecule biases the selection of the inactive set towards truly inactive compounds while placing a handful of compounds that would have been indeterminate for classification purposes into the active set. Once suitable affinity ranges were selected, the inactive and active members of each cluster were identified, and up to three members were selected from each cluster to achieve an acceptable level of diversity among scaffold classes for training and test sets. Class membership and cluster information was exported from Cytoscape to Canvas as a CSV file, so that the original datafile could be used to generate the classification model sets.
2.4 Generation of conformers

All pharmacophore modeling approaches require a set of ligand conformations to be generated either before or during the pharmacophore elucidation step. For a useful model, the set of ligand conformers must contain conformations very near to those found in the target–ligand complex. ConfGen\textsuperscript{79} (version 2.3, Schrödinger, LLC, New York, NY, 2012.) is a tool included in the Schrödinger Suite that is integrated in pharmacophore and docking protocols. Chen and Foloppe developed an optimized set of parameters for ConfGen that more frequently identifies conformers close to established bioactive states of two sets of ligands extracted from crystal structures.\textsuperscript{80} These modifications have since been incorporated into the standard protocols, with the exception of two modifications to the CGO6 and CHYD parameters which control elimination of “compact” structures and hydrogen-bonding electrostatics, respectively. To this end, the CGO6 opcodes in the Macromodel file were deleted, CHYD was added and the first argument in the opcode string was set to -1. These additional “compactness-allowed” modifications were utilized because the impact on the total number of conformers generated is minimal and also because of the extension of the conformational sampling to potentially important conformations of highly-flexible ligands, which are frequently presented in the σ-literature.

2.5 Model development

2.5.1 Classification models

Multiple Instance Learning via Embedded instance Selection (MILES) is a supervised learning technique proposed by Chen et al.\textsuperscript{81} and extended by Fu et al.\textsuperscript{82} in the context of drug activity prediction. MILES, as applied to bags of conformers and their relationships to individual conformers, is particularly useful as a means of associating the activity of a molecule to individual conformers across a set of training molecules.

We used this technique, with a few modifications, to develop classification models for σ receptor ligands, using prototype conformational instances to generate more robust pharmacophore models.
This is particularly important for $\sigma$-1 and $\sigma$-2 ligands, as very few are rigid enough to provide simple solutions, i.e., the vast majority of active and inactive compounds are highly symmetric, flexible ligands, with little bulk or chirality to assist in interpreting how pharmacophore features might correspond to one another. Another issue with $\sigma$ ligands is that no crystallographic information is presently available that might allow for structure-based perception of multiple binding modes. MILES assists in overcoming this problem by allowing for the selection of multiple significant pharmacophore configurations which can be used in traditional pharmacophore modeling applications.

2.5.2 Pharmacophore fingerprints

Pharmacophore fingerprints are a useful means of encoding the distances between potential pharmacophore features into a string of bits that represent the presence or absence of some particular combination of features. A variety of general molecular features such as hydrophobic, aromatic, positive, negative, and hydrogen bond donor or acceptor sites is typically used in tandem with binned distances to generate a hashed fingerprint. Although it is possible to calculate fingerprint data based on 3-point pharmacophores, 4-point pharmacophores, or a combination of both, only 4-point pharmacophores were used in this work. Among the reasons for selecting 4-point pharmacophores are the ability to infer the presence of chirality, and by not combining them with 3-point features the impact of correlated descriptors is reduced. Four-point pharmacophore fingerprints were generated using the default features defined by Phase (version 3.4, Schrödinger, LLC, New York, NY, 2012), along with a custom feature matching the centroid of piperidine and piperazine moieties. In some cases, no four-point pharmacophores could be generated for a structure, and these were removed from the 1n-SVM classification modeling process. Several hundred thousand fingerprint bits are typically calculated during this process, which leads to additional computational expense when similarity measurements are calculated directly within Canvas. Given a large number of conformers and fingerprint bits, the memory and computational cost can exceed the limits of contemporary workstations.
Canvas provides several options to pre-filter fingerprints after their calculation, among them the ability to discard bits that are set by less than a certain percentage of conformers, or alternatively the most informative bits can be retained. Informative bits are decided by calculating the frequency with which each bit is turned “on” based on the total number of molecules in the collection. For classification purposes, optimally informative bits in a well constructed collection will trend towards a frequency of 50%, and thus the ranking

\[ r_{\text{inf}} = |f_{\text{bit}} - 0.5|, \]

where \( r_{\text{inf}} \) is the informative bit ranking given the bit frequency \( f_{\text{bit}} \). After the bits are ranked, informative bits are retained based on those with the least deviation from optimal. The original Significance Analysis of Pharmacophores (SAP) method described by Fu et al.\(^{82}\) implemented a pre-filtering cutoff of 5% bit frequency, which can result in fingerprint lengths of several tens of thousands. Hence the option to retain informative bits was investigated to compare the utility of each approach for activity classification. One immediate benefit of the latter option is that an optimal fingerprint of arbitrary length can be determined heuristically to reduce computational overhead when applying the SAP filtering method. For comparison of fingerprint lengths and to assess the impact of SAP on the performance of MILES, a variety of fingerprint lengths were selected, including unfiltered fingerprints, those filtered by eliminating the least significant 5% of bits, and a range of informative bit string lengths from 256 to 16,384.

2.5.3 Implementation of SAP

SAP was implemented using the samr package\(^{86}\) in R,\(^{87}\) closely following the methods described by Li et al.\(^{88}\) and Fu et al.\(^{82}\). Fingerprint data were read into a data table and transposed to create an appropriately formatted matrix. The “two class unpaired” response model was selected, and activity classifications were used as outcome measurements, with 500 permutations of the activity class labels used as a control. A \( \Delta \) table was computed from the data and divided into 100 intervals. Significance analysis was performed with the \( \Delta \) corresponding to an estimated false positive rate (FPR) of zero at the 90\(^{\text{th}}\) percentile. As shown in Figure 2.1, a sizable number of pharmacophore features can be
identified as lacking significance for accurate activity classification. In the initial experiments, care was taken to ensure that the fingerprints chosen for significance analysis were only selected from the conformers corresponding to training set molecules. Model sets were stratified between training and test in a 2:1 ratio on the complete dataset after sorting by activity. Further experiments were conducted with the inclusion of all fingerprints without regard to training or test set assignment. All pharmacophore fingerprint bits having a non-negative correlation to activity were retained for the entire set of conformers. Code detailing the SAP process is included in Appendix B.

2.5.4 Instance-based similarity mapping

In the MILES formulation, feature mapping can be described in terms of either a distance or a similarity mapping.\textsuperscript{82} We chose to implement similarity mapping of molecules onto conformations,
such that

\[ S(M_i, C') = \max_j S(C_{ij}, C'), \]

where each molecule \( M_i \) is evaluated against every conformer \( C' \) in the embedded feature space as the maximum similarity of any individual conformer \( C_{ij} \) in each bag to each individual conformer in every bag. By using a similarity measure, it is possible to avoid a troublesome hurdle when it comes to ranking conformers, namely that distance measures trend towards 0 if there is no perceivable difference in the pharmacophore fingerprints. When this occurs, the ranking algorithm will be unlikely to predict that the top-ranked conformations are mapped as the most active conformer, even though the weight of the support vectors may otherwise be high. This happens because the most likely conformations of an active molecule will have solutions equal to zero when mapped from their own bag. On the other hand, similarity measures trend towards higher values as the molecules share more pharmacophore fingerprint bits, even if not normalized by design or coercion, and thus the solutions have a positive correlation to the inputs in the feature mapping.

The Kulczynski similarity measure (see Table 2.1), was initially investigated as a way to ensure that the support-vector machine (SVM) correctly mapped active conformers to their parent molecules (vide infra), and to average out the similarity of each conformer to the shared pharmacophore features of both partners. This measure benefits from omitting the bits from each pharmacophore fingerprint that are set “off”, which is problematic in that activity prediction based on pharmacophore fingerprints is not formulated in a manner consistent with having any knowledge of what it means for a bit to be “off” when making comparisons between molecular conformers. Cosine, Dice, McConnaughey, Petke, Simpson, Tanimoto, and Tversky similarity metrics were also investigated for their performance in the MILES context for similar reasons. Training and external test sets were assigned using the same stratified 2:1 approach described in 2.5.5. Conformations representing the test set were removed from the instance-based embedding. Code for instance-based similarity mapping is included in Appendix C.
Table 2.1: Measures of similarity. Canvas provides the following similarity metrics that lack an explicit count of bits that are set to 0 in both pharmacophore fingerprints.

<table>
<thead>
<tr>
<th>Similarity metric</th>
<th>Definition $^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cosine</td>
<td>$\frac{c}{\sqrt{ab}}$</td>
</tr>
<tr>
<td>Dice</td>
<td>$\frac{c}{0.5(a + b)}$</td>
</tr>
<tr>
<td>Kulczynski</td>
<td>$0.5 \left( \frac{c}{a} + \frac{c}{b} \right)$</td>
</tr>
<tr>
<td>McConnaughey</td>
<td>$\frac{(c \times c) - (a - c)(b - c)}{ab}$</td>
</tr>
<tr>
<td>Petke</td>
<td>$\frac{c}{\max(a, b)}$</td>
</tr>
<tr>
<td>Simpson</td>
<td>$\frac{c}{\min(a, b)}$</td>
</tr>
<tr>
<td>Tanimoto</td>
<td>$\frac{c}{a + b - c}$</td>
</tr>
<tr>
<td>Tversky</td>
<td>$\frac{c}{\alpha(a - c) + \beta(b - c) + c}$</td>
</tr>
</tbody>
</table>

$^a$ For binary fingerprints F1 and F2, a bit is set "on" and given the value "1" if the feature represented by that bit is present in the fingerprint; otherwise, the bit is given the value "0." $a$ is the count of bits set on in F1, $b$ is the count of bits set on in F2, and $c$ is the count of bits that are set on in both fingerprints. In the Tversky metric $\alpha$ and $\beta$ are parameters used to scale the count of bits that are exclusively set on in F1 and F2, respectively.

2.5.5 Implementation of 1-norm SVM

Support vector machines are commonly used for classification purposes, and can be useful for evaluating novel compounds as “active” or “inactive” based on competition binding data. Pharmacophore fingerprints which are derived from multiple conformations of chemical entities are prone to have common spurious configurations that provide no productive information about their complexes with biological targets. In this context, 1-norm SVMs are particularly useful because they are less susceptible to over-fitting the input data, and tend to eliminate noise resulting from common but meaningless pharmacophore configurations. Linear programming (LP) techniques have previously been formulated to solve 1n-SVMs,$^{90-92}$ and have been used in the context of pharmacological classification.$^{81}$
The parametric-cost linear program of Yao and Lee\textsuperscript{92} uses the standard formulation:

\[
\begin{align*}
\text{minimize} \quad & (c + \lambda a)'z \\
\text{subject to} \quad & Az = b \\
& z \geq 0,
\end{align*}
\]

along with the following definitions:

\[
\begin{align*}
z & \equiv (\beta_0^+ \beta_0^- (\beta^+)' (\beta^-)' (\zeta^+)' (\zeta^-)' )' \\
c & \equiv (0 0 0' 0' 1' 0' 0')' \\
a & \equiv (0 0 1' 1' 0' 0' 0')' \\
A & \equiv (Y -Y \text{diag}(Y)X -\text{diag}(Y)X I' -I')' \\
b & \equiv 1.
\end{align*}
\]

The solution, \(z\), minimizes the distance of support vectors from the hyperplane given the objective function. Here \(c\) is a vector of normalized indices of the \textit{slack variables} used as part of calculating the objective function, \(b\) is a vector representing the right-hand side of the \textit{constraint constants}, \(a\) is a vector used to implement the objective function, \(Y\) are the class labels, \(X\) is the feature matrix in the context of the MILES\textsuperscript{81} formulation, and \(\lambda\) is a tunable \textit{control parameter} of the objective function. An optimal hyperplane \(\beta_0 + \beta X\) is then solved subject to the non-separable case of a 1n-SVM classifier by the introduction of the slack variable \(\zeta\) that allows for some points to be misclassified. Classification of activity is based on the sign of the output, \(\text{sgn}(\beta_0 + \beta X)\).

Tuning of the 1n-SVM (as illustrated in Figure 2.2) was accomplished using the perry package\textsuperscript{93} of R, with 5-fold cross-validation using random splits of the data for a total of 15 replicates. The \(\lambda\) parameter was initially chosen by validations over the range of \(10^{-12} - 10^4\) at each power of 10. Prediction error was assessed as the misclassification error rate. After narrowing down the range for \(\lambda\), a second cross-validation was undertaken at 21 points ranging from the nearest lower power of 10 to the higher, spaced at arithmetic intervals between each order of magnitude. An optimal tuning parameter corresponding to the least mean cross-validation prediction error was then used to generate the 1n-SVM slack and penalty values that are necessary to calculate the accuracy of the training set and the confusion matrix of the external test set. Confusion matrices were calculated.
Figure 2.2: Tuning of the 1-norm SVM. The control parameter $\lambda$ is selected as the minimum value (red point) at which the mean cross-validation prediction error (PE, misclassification error rate) is also minimized. The first round of tuning covers a wide order of magnitudes. A second round of tuning is performed to optimize this parameter.

using the caret package of R. The code used to implement this 1-norm SVM is included in Appendix D.

2.5.6 Selection of prototype conformers

Non-zero elements of $\beta$ correspond to prototype conformers for which molecular similarity measures contribute positively or negatively to the classification of molecular activity. Many of the conformations will have no impact on classification, and can be removed from the cohort of conformers taken on to traditional pharmacophore development. In practice, the indices of training set conformations from the original collection are congruent with the $\beta$ so that the non-zero elements can be collected to generate a substantially smaller subset of conformers. Ranking of all conformers could be performed at this point, but the prototype conformers are sufficient for pharmacophore model development given that there are no known crystal structures for complexes of $\sigma$ receptors and their ligands. After the prototype conformers are retrieved from the full set, the corresponding fingerprints for active ligands are used to generate a substantially smaller similarity matrix. This matrix is visualized as a clustered heat map to determine a practical number of active ligands that are needed to generate common pharmacophore hypotheses.
2.5.7 Pharmacophore modeling

Phase provides means for both manual and automated pharmacophore development. Individual hypotheses can be generated from each prototype conformer manually, but the development of common pharmacophore hypotheses from the entire set of prototype conformers allows for the perception of features common to subsets of entire cohort of selected conformers. Common pharmacophore hypotheses are generated and evaluated in a five-step process: generating ligand conformations, identification of pharmacophore sites, perception of common pharmacophore hypotheses, scoring of hypotheses, and building atom- or pharmacophore-based 3D QSAR models.

In structure-based pharmacophore development, the conformations of ligands are known with a high degree of certainty. Without such information, ligand-like conformations must be selected through other means, such as through the use of the ConfGen application. Highly flexible ligands, such as those found frequently in the σ literature are problematic because the number of potential common pharmacophore hypotheses rapidly becomes too great to discern any statistically significant hypotheses after the scoring step. In preliminary studies, it was not infrequent to generate tens of thousands of common 5-point hypotheses; analyzing the resulting QSAR models became an intractable task. By utilizing the prototype active and inactive conformers that were identified through MILES, the number of common pharmacophores is reduced by several orders of magnitude.

Phase provides a default set of pharmacophore sites including hydrogen bond acceptors (A), hydrogen bond donors (D), positive ionizable (P), negative ionizable (N), hydrophobic (H), and aromatic rings (R). This set was extended to include piperazine and piperidine rings (X) as group features, as was performed during the pharmacophore fingerprinting steps. After editing the feature definitions, sites were generated for all the input conformers. The prototype conformers that share a high pharmacophore fingerprint similarity will likely share many of the same features. Common pharmacophores were identified by determining an reasonable number of each type of feature, an optimal number of site points (from 3–7), and requiring matches to the number of active compounds identified from the heat map of active prototype conformers. Scoring of hypotheses was performed with the default settings in Phase, and the hypotheses that survived the initial scoring process were
taken on to pharmacophore screening and QSAR development without regard to score.

Typically atom- or pharmacophore-based QSAR models would be generated at this point. However, the common pharmacophore workflow makes it difficult to generate these models if the conformations of all other ligands in the database are not carried throughout the process. Atom-based models are best performed on cohorts of structures of limited diversity, and are not well suited to molecules with a high degree of rotational freedom. Pharmacophore-based models are more appropriate under these circumstances, but the resulting models cannot be used to infer activity when steric clashes are important to binding affinity, and there is no way to run this application outside of the common pharmacophore hypothesis workflow. On the other hand, the surviving hypotheses can be used outside of the workflow to screen through databases of ligands with known activity to retrieve cohorts of ligands which can then be utilized for QSAR development using alternative methodologies.

2.5.8 Database screening

The Phase Advanced Pharmacophore Screening application was used to retrieve pharmacophore hits from the database of all ligands. Decoy sets of 1000 drug-like ligands\textsuperscript{94} were combined with the curated sigma database ligands and inactive ligands from the superset for calculation of enrichment. For these purposes, an affinity at 10,000 micromolar or better was considered “active”. In the case of $\sigma$-2 ligands, a set with an average molecular weight of 400 was utilized, whereas the $\sigma$-1 database was more matched to the decoy set having an average molecular weight of 360. Databases were created using the same ConfGen parameters as were used in the fingerprinting and pharmacophore modeling steps. Searching of the database was performed using the existing conformers without refinement. Default matching criteria were used, with the exception that only 4 sites out of the total were required to match, and preference was given to conformers matching more sites. Four-site matches are optimal under these circumstances, as they naturally allow alignments based on chirality, but are not overly aggressive at screening out matches that contribute valuable QSAR information. Hits were scored based on default scoring fitness, but were rejected if vector features diverged at
an angle of 90° or more, or if the volume score was less than 0.3. These limits were determined
heuristically to generate reasonable alignments without rejecting too many ligands from the database.
Compound affinities were ranked according to a field-based QSAR, described below.

2.5.9 Field-based QSAR

Field-based QSAR based on the CoMFA (Force Field) and CoMSIA (Gaussian) approaches was
recently incorporated into the Schrödinger software suite. Both of these approaches evaluate fields
based on a rectangular grid encompassing the training set molecules. However, the Force Field
method involves a Partial Least Squares (PLS) regression based on the fields evaluated at each grid
point, and is very sensitive to the alignment of ligands. The Gaussian method evaluates fields based
on a weighted function dependent upon the distance of atoms from the grid points, and is less
sensitive to alignment artifacts. Given the diverse nature of scaffolds retrieved during the database
screening steps, the Gaussian approach was used for QSAR analysis.

Each set of aligned ligands was used to develop independent QSAR models. A random split of
the dataset into 50% training to test set was applied to the ligands. This ratio was chosen as it was
the default for Phase, and ideally should be set high enough to generate a predictive model with a
representative external test set. These sets are automatically distributed in a relatively even manner
across the range of activities present in the dataset. A maximum of 3 PLS factors at a grid spacing
of 1 Å were used in the regression. The grid was extended 3 Å beyond training set limits, and force
fields within 2 Å of any training set atom were ignored. Steric and electrostatic fields were truncated
at 30 kcal/mol. Variables with a standard deviation of less than 0.01, or with |t-value| less than
2.0 were eliminated from the regression. Cross-validation was performed with the leave-one-out
technique.
3. RESULTS

We've learned from experience that the truth will come out. Other experimenters will repeat your experiment and find out whether you were wrong or right. Nature's phenomena will agree or they'll disagree with your theory. And, although you may gain some temporary fame and excitement, you will not gain a good reputation as a scientist if you haven't tried to be very careful in this kind of work. And it's this type of integrity, this kind of care not to fool yourself, that is missing to a large extent in much of the research in cargo cult science.

Cargo Cult Science
Richard Feynman

3.1 Competition Binding Database

A total of 1,208 articles representing the σ literature covering the years 1981–2011 were identified as potential sources of competition binding data, which was confirmed for 564 articles. After characterizing the nature of the competition binding assay methodology, it was possible to identify the most prevalent combinations of hot ligands and masking agents as shown in Table 3.1. Furthermore, 7-OH-DPAT, 7-OH-PIPAT, DTG or HAL (without a masking ligand), ifenprodil, NANM and PPP are not sufficiently target or subtype selective for the purposes of pharmacophore development. We chose to collect structural and binding data for articles containing DTG with a masking agent as σ-2 selective methodology or PTZ as a σ-1 selective ligand. In order to confidently combine
Table 3.1: Breakdown of σ research articles providing competition binding data. The following hot ligand/masking ligand combinations were only represented by a single article: 2-IPB, 4-IBP, 4-IPBS, DuP734, I-benzamide, IPAB, IPEMP, ANSTO-14, clonidine, DTG/DuP734, DTG/AC915, DTG/carbetapentane, DTG/DXM, FPS, HAL/l-sulpiride, HAL/spiropersidol/BIMU-8, IPIPAG, MS377, azido-DTG, NANM/dizocilpine, NANM/etorphine, PB28/PTZ, PPP/DXL, progesterone, PTZ/Lu28-179, SA4503, SN56, and SW120. Several of the hot ligands are now known to more effectively target other receptors.

<table>
<thead>
<tr>
<th>Hot ligand</th>
<th>Masking ligands</th>
<th>No. articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>3′-iodopentazocine</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7-OH-DPAT</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>7-OH-PIPAT</td>
<td>Spiroperidol</td>
<td>2</td>
</tr>
<tr>
<td>DTG</td>
<td></td>
<td>128</td>
</tr>
<tr>
<td>DTG</td>
<td>DXL</td>
<td>55</td>
</tr>
<tr>
<td>DTG</td>
<td>NANM</td>
<td>25</td>
</tr>
<tr>
<td>DTG</td>
<td>PTZ</td>
<td>189</td>
</tr>
<tr>
<td>DXM</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>HAL</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>HAL</td>
<td>L-sulpiride</td>
<td>1</td>
</tr>
<tr>
<td>HAL</td>
<td>Spiroperidol</td>
<td>24</td>
</tr>
<tr>
<td>Ibenprofen</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>NANM</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>NANM</td>
<td>MK801</td>
<td>7</td>
</tr>
<tr>
<td>NE100</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>PIMBA</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>PPP</td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>PTZ</td>
<td></td>
<td>356</td>
</tr>
<tr>
<td>RHM-1</td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

data from different articles into sets of compounds for quantitative computational experiments, the methodologies should be as practically identical as possible. To this end, the matrix used in the σ assay is a very important consideration, so we further limited our investigations to those articles in Table 3.2 which were most widely used for competition binding experiments. Of the combinations investigated, σ-1 assays performed with PTZ on guinea pig brain tissue homogenates, and σ-2 assays run with DTG using PTZ as a masking agent on rat liver homogenates, were found to represent a sufficient number of active and inactive compounds for our computational requirements. The curated datasets included 723 unique σ-2 ligands and 1,396 curated σ-1 ligands.
Table 3.2: Breakdown of assay methodology for selective σ competition binding experiments.

<table>
<thead>
<tr>
<th>Hot ligand</th>
<th>Masking ligand</th>
<th>Organism</th>
<th>Tissue</th>
<th>No. articles</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTG</td>
<td>DXL</td>
<td>rat</td>
<td>brain</td>
<td>3</td>
</tr>
<tr>
<td>DTG</td>
<td>DXL</td>
<td>rat</td>
<td>liver</td>
<td>44</td>
</tr>
<tr>
<td>DTG</td>
<td>NANM</td>
<td>guinea pig</td>
<td>brain</td>
<td>18</td>
</tr>
<tr>
<td>DTG</td>
<td>NANM</td>
<td>rat</td>
<td>brain</td>
<td>5</td>
</tr>
<tr>
<td>DTG</td>
<td>NANM</td>
<td>rat</td>
<td>liver</td>
<td>2</td>
</tr>
<tr>
<td>DTG</td>
<td>PTZ</td>
<td>guinea pig</td>
<td>brain</td>
<td>33</td>
</tr>
<tr>
<td>DTG</td>
<td>PTZ</td>
<td>guinea pig</td>
<td>brain plus cerebellum</td>
<td>1</td>
</tr>
<tr>
<td>DTG</td>
<td>PTZ</td>
<td>human</td>
<td>MCF-7 ADR cells</td>
<td>3</td>
</tr>
<tr>
<td>DTG</td>
<td>PTZ</td>
<td>rat</td>
<td>brain</td>
<td>32</td>
</tr>
<tr>
<td>DTG</td>
<td>PTZ</td>
<td>rat</td>
<td>brain minus cerebellum</td>
<td>2</td>
</tr>
<tr>
<td>DTG</td>
<td>PTZ</td>
<td>rat</td>
<td>liver</td>
<td>111</td>
</tr>
<tr>
<td>PTZ</td>
<td>guinea pig</td>
<td>guinea pig</td>
<td>brain</td>
<td>212</td>
</tr>
<tr>
<td>PTZ</td>
<td>guinea pig</td>
<td>guinea pig</td>
<td>brain minus cerebellum</td>
<td>17</td>
</tr>
<tr>
<td>PTZ</td>
<td>guinea pig</td>
<td>guinea pig</td>
<td>brain plus cerebellum</td>
<td>2</td>
</tr>
<tr>
<td>PTZ</td>
<td>guinea pig</td>
<td>clone in E. Coli</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>PTZ</td>
<td>guinea pig</td>
<td>clone in S. Cerevisiae</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>PTZ</td>
<td>guinea pig</td>
<td>liver</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>PTZ</td>
<td>human</td>
<td>MCF-7 cells</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>PTZ</td>
<td>human</td>
<td>brain</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>PTZ</td>
<td>human</td>
<td>jurkat cells</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>PTZ</td>
<td>mouse</td>
<td>brain</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>PTZ</td>
<td>rat</td>
<td>C6 cells</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>PTZ</td>
<td>rat</td>
<td>brain</td>
<td></td>
<td>55</td>
</tr>
<tr>
<td>PTZ</td>
<td>rat</td>
<td>brain minus cerebellum</td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>PTZ</td>
<td>rat</td>
<td>liver</td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>
3.2 Impact of Fingerprint Length

Fingerprints of various lengths were calculated using the options available in Canvas. A σ-2 rat liver/DTG data set (Table 3.3 and Figure 3.1) was processed with the following treatments: no bit filtering, filtering out bits that are present in less than 5% of the conformers, or retaining informative bits with fingerprint lengths of 256, 512, 1,024, 2,048, 4,096, 8,192, and 16,384 bits.

Table 3.3: Statistics of the initial σ-2 classification model data set. The active and inactive compounds were selected from the curated dataset containing 723 ligands, using p\(K_i\) cutoffs of 6.0 and 8.301 for inactive and active molecules, respectively.

| No. of molecules | Training set | | | Test Set | | | | | | Total no. of conformers |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Active | Inactive | Active | Inactive | Total | |
| | | | | 14 | 22 | 6 | 12 | 54 | 5,634 |

To determine the extent to which SAP methodology improved the calculations, each treatment was performed without SAP. SAP methodology was then used to determine a subset of significant bits, as shown in Table 3.4. Initial experiments were conducted by holding out the fingerprints from the test set, except in the case of unfiltered fingerprints. Unfiltered fingerprints could not be treated to SAP because of excessive memory resource requirements. The ratio of retained bits to informative/frequent bits following significance analysis decreases steadily from 95% down to 84% when 8,192 or more post-filtered bits are used. Retained bits also make up a very small percentage of the pre-filtered bits, ranging from 0.03% up to 2.41%. Threshold cutoffs for removal of insignificant bits tended to increase with increasing fingerprint length. Notably, repeated runs of SAP on the same fingerprints tended to generate different selections of significant bits as the fingerprint length rose above 1,024 bits (data not shown). One likely explanation for this phenomenon is that the number of random permutations of the data is fixed at 500, whereas the number of potential significant bits rises with fingerprint length. It may therefore be possible to resolve a reproducible SAP fingerprint by increasing the number of permutations if the computational cost is justifiable.

A similarity matrix using the Kulczynski metric was calculated for each treatment, and MILES
Figure 3.1: Clustered σ-2 scaffold data. Molecules classified as inactive are in blue and active molecules are colored orange; the remaining unclassified molecules are in gray. Fifty-four molecules (larger hexagons) were selected from this set for SAP and MILES performance analysis. Solid lines connect molecules and scaffolds to their cores. A clustering cutoff at a Euclidean distance of 10-12 is indicated by dashed lines. The dotted lines demonstrate the connectivity between cores and nodes in the dendrogram beyond the clustering cutoff distance.
was used to develop classification models. Confusion matrices, internal validation statistics, and cross-validation statistics were used to characterize the impact of fingerprint length on the quality of the final classification models. In order that cross-validation results remained matched across treatments, the same selection of splitting replicates were used.

Table 3.4: Significance analysis parameters after Canvas filtering and SAP: the total number of bits ($b_{tot}$) of the complete fingerprint was 696828.

<table>
<thead>
<tr>
<th>$b_{inf}$</th>
<th>$b_{freq}$</th>
<th>$b_{sap}$</th>
<th>$\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>256</td>
<td>245</td>
<td>0.416</td>
<td></td>
</tr>
<tr>
<td>512</td>
<td>479</td>
<td>0.606</td>
<td></td>
</tr>
<tr>
<td>1,024</td>
<td>937</td>
<td>1.280</td>
<td></td>
</tr>
<tr>
<td>2,048</td>
<td>1,829</td>
<td>1.528</td>
<td></td>
</tr>
<tr>
<td>4,096</td>
<td>3,576</td>
<td>1.906</td>
<td></td>
</tr>
<tr>
<td>8,192</td>
<td>6,960</td>
<td>2.638</td>
<td></td>
</tr>
<tr>
<td>16,384</td>
<td>13,910</td>
<td>3.111</td>
<td></td>
</tr>
<tr>
<td>5% (19805)</td>
<td>16,794</td>
<td>3.108</td>
<td></td>
</tr>
</tbody>
</table>

*a* Number of informative bits, or % bits filtered by minimum frequency; *b* bits remaining after SAP; *c* selected threshold for elimination of insignificant bits

As shown in Table 3.5, regardless of whether or not SAP was performed, there was no obvious trend in $\lambda$ parameter or the cross-validation prediction error. Internal accuracy of the non-SAP models was 100% (except for the 256 bit fingerprint). External accuracy and MCC were adversely affected with every treatment compared to the unfiltered fingerprint. Filtering by SAP slightly degraded internal accuracy, particularly at shorter fingerprint lengths. The external accuracy and Matthews Correlation Coefficient (MCC) measures were inconsistent, with no obvious trend towards better performance with increasing fingerprint length. At 8,192 bits with SAP filtering, the MILES classification performance was essentially identical to using raw fingerprints.

Confusion matrices and cross-validation statistics for the non- and SAP-filtered sets are presented in Tables 3.6 and 3.7, respectively. Classification of true negatives performed more poorly than that of the unfiltered fingerprint for all of the informative bit fingerprints, whereas true positive classification performed identically when the fingerprint length was set at 2,048 bits or longer. In the case of the traditional SAP method, or with a pre-filtered 8,192 bit fingerprint, the classification
Table 3.5: Tuning parameters and MILES results vary depending on the fingerprint length chosen and the application of SAP.

<table>
<thead>
<tr>
<th>Pre-filtering treatment</th>
<th>$\lambda^a$</th>
<th>PE$^b$</th>
<th>IA$^c$</th>
<th>EA$^d$</th>
<th>MCC$^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>no SAP treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>256</td>
<td>8e-01</td>
<td>0.163</td>
<td>0.972</td>
<td>0.772</td>
</tr>
<tr>
<td></td>
<td>512</td>
<td>1e-05</td>
<td>0.150</td>
<td>1.000</td>
<td>0.611</td>
</tr>
<tr>
<td></td>
<td>1,024</td>
<td>2e-05</td>
<td>0.207</td>
<td>1.000</td>
<td>0.722</td>
</tr>
<tr>
<td></td>
<td>2,048</td>
<td>7e-06</td>
<td>0.191</td>
<td>1.000</td>
<td>0.611</td>
</tr>
<tr>
<td></td>
<td>4,096</td>
<td>5e-05</td>
<td>0.196</td>
<td>1.000</td>
<td>0.722</td>
</tr>
<tr>
<td></td>
<td>8,192</td>
<td>8e-06</td>
<td>0.156</td>
<td>1.000</td>
<td>0.722</td>
</tr>
<tr>
<td></td>
<td>16,384</td>
<td>2e-06</td>
<td>0.135</td>
<td>1.000</td>
<td>0.778</td>
</tr>
<tr>
<td></td>
<td>5%</td>
<td>7e-06</td>
<td>0.135</td>
<td>1.000</td>
<td>0.778</td>
</tr>
<tr>
<td></td>
<td>raw</td>
<td>2e-06</td>
<td>0.196</td>
<td>1.000</td>
<td>0.833</td>
</tr>
</tbody>
</table>

processed with SAP

|                         | 256         | 1e02   | 0.144  | 0.944  | 0.667   | 0.316   |
|                         | 512         | 1e02   | 0.137  | 0.917  | 0.667   | 0.25    |
|                         | 1,024       | 9e01   | 0.209  | 0.944  | 0.778   | 0.5     |
|                         | 2,048       | 1e02   | 0.226  | 0.944  | 0.611   | 0.161   |
|                         | 4,096       | 1e02   | 0.254  | 0.944  | 0.667   | 0.316   |
|                         | 8,192       | 9e-05  | 0.250  | 1.000  | 0.833   | 0.614   |
|                         | 16,384      | 3e-02  | 0.252  | 1.000  | 0.778   | 0.472   |
|                         | 5%          | 5e-05  | 0.265  | 1.000  | 0.833   | 0.614   |

$^a$optimal tuning parameter; $^b$cross-validation prediction error; $^c$internal accuracy of prediction; $^d$external test set accuracy; $^e$Matthews Correlation Coefficient

performance was identical that of the unfiltered fingerprint. The use of SAP therefore allows the use of far fewer bits to achieve MILES classification results comparable to those found with raw fingerprints. Unfortunately there is no way to know in advance how many fingerprint bits to retain in the pre-filtering step. Some caveats with this interpretation are that only the Kulczynski metric was used in the MILES formulation for these experiments, and that the dataset was biased towards more inactive compounds. In this case, at least, the bias can be resolved by selecting fewer inactives, or by broadening the criteria for the inclusion of active compounds from the scaffold perception step.
Table 3.6: Confusion matrices of models at various fingerprint lengths without SAP filtering.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Unfiltered</th>
<th>5% Cutoff</th>
<th>256 bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>256 bits</td>
<td>11</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>512 bits</td>
<td>9</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1,024 bits</td>
<td>9</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>2,048 bits</td>
<td>8</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3.7: Confusion matrices at various fingerprint lengths in tandem with SAP filtering.

<table>
<thead>
<tr>
<th>Reference</th>
<th>5% Cutoff</th>
<th>256 bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>256 bits</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>512 bits</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1,024 bits</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2,048 bits</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4,096 bits</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>8,192 bits</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>16,384 bits</td>
<td>11</td>
<td>3</td>
</tr>
</tbody>
</table>
3.3 Comparison of Similarity Metrics

Given the mixed results using the Kulczynski metric, several other metrics were investigated in the MILES formulation. SAP-filtered fingerprints calculated previously using the 1,024 and 4,096 most informative bits were utilized in this approach. Similarity metrics were calculated using those available in Canvas that do not explicitly include the shared number of bits turned “off”. Thus cosine, Dice, McConnaughey, Petke, Simpson, Tanimoto, and Tversky metrics were used to calculate similarity matrices for feature mapping. Internal validation accuracy, confusion matrices, and external test set prediction statistics were used to characterize the impact of fingerprint metric, as well as to investigate the impact of fingerprint length using these metrics.

As shown in Table 3.8, every similarity measure performed better than the original Kulczynski metric in terms of internal accuracy. Interestingly, shorter fingerprint lengths were better or equivalent to the unfiltered fingerprint at classification of the external test set with every metric except McConnaughey and Simpson. Longer fingerprints led to poorer external classification results for cosine, Petke, and Simpson metrics, but better results for the McConnaughey metric. The mean cross-validation prediction error increased with the use of longer fingerprints. Matthews Correlation Coefficients indicate performance on par with the unfiltered fingerprints using the Cosine and Petke metrics, and slightly better performance using Dice, Tanimoto, or Tversky metrics. The Dice, Tanimoto, and Tversky metrics appear to perform very well and are less sensitive to changes in fingerprint length.

3.4 Impact of training and test set size

Given the previous results, it was worthwhile to investigate the performance of SAP and MILES with a more balanced dataset. A library of 130 molecules (Table 3.11 and Figure 3.2) was assembled using all of the active ligands from every possible cluster, along with inactive molecules sampled from the remaining clusters. In many projects, it is likely that there will be many more inactive molecules than active ones. Additionally, there are some clusters which simply do not possess active
Figure 3.2: Clustered σ-2 scaffold data for the balanced dataset. Visualization details are the same as in Figure 3.1. One hundred and thirty molecules (larger hexagons) were selected from this set for SAP and MILES performance analysis. All molecules classified as active were included from all available clusters. Inactive compounds were added to roughly approximate the number of actives by selecting poor affinity ligands as evenly as possible, cluster by cluster.
Table 3.8: Stability of MILES performance utilizing alternative similarity metrics. Lower stability is associated with denominators in the similarity calculations that involve products or minimum/maximum functions.

<table>
<thead>
<tr>
<th>Pre-filtering treatment</th>
<th>$\lambda^a$</th>
<th>PE$^b$</th>
<th>IA$^c$</th>
<th>EA$^d$</th>
<th>MCC$^e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,024 bit fingerprint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cosine</td>
<td>2e-07</td>
<td>0.185</td>
<td>1.000</td>
<td>0.833</td>
<td>0.614</td>
</tr>
<tr>
<td>Dice</td>
<td>8e-04</td>
<td>0.169</td>
<td>1.000</td>
<td>0.889</td>
<td>0.756</td>
</tr>
<tr>
<td>McConnaughey</td>
<td>9e-06</td>
<td>0.313</td>
<td>1.000</td>
<td>0.611</td>
<td>0.161</td>
</tr>
<tr>
<td>Petke</td>
<td>9e-04</td>
<td>0.185</td>
<td>1.000</td>
<td>0.889</td>
<td>0.756</td>
</tr>
<tr>
<td>Simpson</td>
<td>1e02</td>
<td>0.263</td>
<td>1.000</td>
<td>0.778</td>
<td>0.5</td>
</tr>
<tr>
<td>Tanimoto</td>
<td>1e-04</td>
<td>0.185</td>
<td>1.000</td>
<td>0.889</td>
<td>0.756</td>
</tr>
<tr>
<td>Tversky</td>
<td>8e-04</td>
<td>0.169</td>
<td>1.000</td>
<td>0.889</td>
<td>0.756</td>
</tr>
<tr>
<td>4,096 bit fingerprint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cosine</td>
<td>2e-03</td>
<td>0.209</td>
<td>1.000</td>
<td>0.778</td>
<td>0.5</td>
</tr>
<tr>
<td>Dice</td>
<td>4e-06</td>
<td>0.194</td>
<td>1.000</td>
<td>0.889</td>
<td>0.756</td>
</tr>
<tr>
<td>McConnaughey</td>
<td>2e-02</td>
<td>0.370</td>
<td>1.000</td>
<td>0.722</td>
<td>0.403</td>
</tr>
<tr>
<td>Petke</td>
<td>3e-03</td>
<td>0.219</td>
<td>1.000</td>
<td>0.833</td>
<td>0.632</td>
</tr>
<tr>
<td>Simpson</td>
<td>5e-03</td>
<td>0.287</td>
<td>1.000</td>
<td>0.556</td>
<td>0.</td>
</tr>
<tr>
<td>Tanimoto</td>
<td>4e-06</td>
<td>0.248</td>
<td>1.000</td>
<td>0.889</td>
<td>0.756</td>
</tr>
<tr>
<td>Tversky</td>
<td>4e-06</td>
<td>0.194</td>
<td>1.000</td>
<td>0.889</td>
<td>0.756</td>
</tr>
</tbody>
</table>

$^a$optimal tuning parameter; $^b$cross-validation prediction error; $^c$internal accuracy of prediction; $^d$external test set accuracy; $^e$Matthews Correlation Coefficient

molecules, yet it is important that the pharmacophore information in these molecules is not lost. It is noteworthy that some molecules, particularly flexible ones, will generate many more conformers than other molecules, so while the number of compounds may be balanced, it is possible for the number of conformers to remain unbalanced.

As part of this investigation, pharmacophore fingerprints from all molecules were utilized in the significance analysis. This allows a less biased comparison to the performance of the metrics with respect to raw fingerprints, since similarity metrics calculated with Canvas will use all available bits. Experiments were run using raw fingerprints, or the 1,024 or 16,384 most informative bits. A summary of the latter fingerprints is provided in Table 3.12. The Dice, Petke, Tanimoto, and Tversky metrics were chosen because of their superior performance in the prior experiments.
Table 3.9: Confusion matrices for diverse similarity metrics using 1,024–bit fingerprints in tandem with SAP filtering. The McConnaughey metric performed notably poorer than all other metrics.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Cosine</th>
<th>Dice</th>
<th>Kulczynski</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>11</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>T</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>McConnaughey</th>
<th>Petke</th>
<th>Simpson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>8</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>T</td>
<td>4</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Tanimoto</th>
<th>Tversky</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 3.10: Confusion matrices for diverse similarity metrics using 4,096–bit fingerprints and SAP filtering. With longer fingerprint lengths, the Kulczynski and McConnaughey metrics performed better. The Simpson metric performed notably poorer at this length.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Cosine</th>
<th>Dice</th>
<th>Kulczynski</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>10</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>T</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>McConnaughey</th>
<th>Petke</th>
<th>Simpson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>9</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>T</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference</th>
<th>Tanimoto</th>
<th>Tversky</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>T</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

51
Table 3.11: Statistics of the balanced σ-2 classification model data set.

<table>
<thead>
<tr>
<th>No. of molecules</th>
<th>Training set</th>
<th>Test Set</th>
<th>Total</th>
<th>Total no. of conformers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active</td>
<td>Inactive</td>
<td>Active</td>
<td>Inactive</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>44</td>
<td>21</td>
<td>22</td>
</tr>
</tbody>
</table>

Table 3.12: Significance analysis parameters after Canvas filtering and SAP with a more balanced dataset: the total number of bits \( b_{\text{tot}} \) of the complete fingerprint was 2,859,553.

<table>
<thead>
<tr>
<th>( b_{\text{inf}}^a )</th>
<th>( b_{\text{sap}}^b )</th>
<th>( \Delta^c )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,024</td>
<td>730</td>
<td>1.645</td>
</tr>
<tr>
<td>16,384</td>
<td>12,520</td>
<td>2.488</td>
</tr>
</tbody>
</table>

\(^a\)Number of informative bits; \(^b\) bits remaining after SAP; \(^c\) selected threshold for elimination of insignificant bits

Table 3.13: Comparison of metric performance with a more balanced molecular library. Classification of the external dataset improved with larger training and test sets, but only with larger fingerprints. SAP-filtered fingerprints at 16,384 bits were comparable to the performance with raw fingerprints.

<table>
<thead>
<tr>
<th>Pre-filtering treatment</th>
<th>( \lambda^a )</th>
<th>PE(^b)</th>
<th>IA(^c)</th>
<th>EA(^d)</th>
<th>MCC(^e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>unfiltered fingerprint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dice</td>
<td>1e-06</td>
<td>0.129</td>
<td>1.000</td>
<td>0.930</td>
<td>0.868</td>
</tr>
<tr>
<td>Petke</td>
<td>9e-07</td>
<td>0.138</td>
<td>1.000</td>
<td>0.884</td>
<td>0.774</td>
</tr>
<tr>
<td>Tanimoto</td>
<td>8e-07</td>
<td>0.181</td>
<td>1.000</td>
<td>0.884</td>
<td>0.788</td>
</tr>
<tr>
<td>Tversky</td>
<td>1e-06</td>
<td>0.129</td>
<td>1.000</td>
<td>0.930</td>
<td>0.868</td>
</tr>
<tr>
<td>1,024 bit fingerprint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dice</td>
<td>3e-06</td>
<td>0.221</td>
<td>1.000</td>
<td>0.767</td>
<td>0.542</td>
</tr>
<tr>
<td>Petke</td>
<td>9e-07</td>
<td>0.138</td>
<td>1.000</td>
<td>0.884</td>
<td>0.774</td>
</tr>
<tr>
<td>Tanimoto</td>
<td>4e-05</td>
<td>0.211</td>
<td>1.000</td>
<td>0.791</td>
<td>0.612</td>
</tr>
<tr>
<td>Tversky</td>
<td>3e-06</td>
<td>0.221</td>
<td>1.000</td>
<td>0.767</td>
<td>0.542</td>
</tr>
<tr>
<td>16,384 bit fingerprint</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dice</td>
<td>9e-04</td>
<td>0.194</td>
<td>1.000</td>
<td>0.884</td>
<td>0.788</td>
</tr>
<tr>
<td>Petke</td>
<td>2e-06</td>
<td>0.188</td>
<td>1.000</td>
<td>0.884</td>
<td>0.788</td>
</tr>
<tr>
<td>Tanimoto</td>
<td>4e-06</td>
<td>0.182</td>
<td>1.000</td>
<td>0.907</td>
<td>0.828</td>
</tr>
<tr>
<td>Tversky</td>
<td>8e-04</td>
<td>0.194</td>
<td>1.000</td>
<td>0.884</td>
<td>0.788</td>
</tr>
</tbody>
</table>

\(^a\)optimal tuning parameter; \(^b\)cross-validation prediction error; \(^c\)internal accuracy of prediction; \(^d\)external test set accuracy; \(^e\)Matthews Correlation Coefficient
3.5 Development of σ-2 pharmacophore models and QSAR

Prototype conformers discovered in the “balanced” 16,384 bit experiment using Tanimoto similarity metrics were taken through the Develop Pharmacophore Hypothesis workflow in Phase. The heat map of active conformers indicates several clusters containing 3 distinct ligands. Matching on 6 sites with a minimum of 3 ligands allowed for up to 3 hydrogen bond acceptor moieties, and some of the most similar conformers had pairs of acceptors in close proximity, so the feature frequency of acceptors was reduced to 1. After scoring, the complete linkage clustering method was used to select representative pharmacophores with a minimum similarity of 0.9 based on survival scores. Thirty-four hypotheses were retained and used for screening the complete conformer database.

A Gaussian Field-based QSAR analysis was performed on the subsequent alignments. Two compounds 3.4 were removed from each alignment, if present, as they consistently caused problems. Compound 1 from Choi et al. would occasionally pass the volume filter of the database screen, but was too large when assigned to the training set for the QSAR to run because points only associated with this compound fell far away from all other training set compounds, requiring more PLS factors than Phase was designed to accommodate. Compound 5 from Fontanilla et al. was always predicted to have a much higher affinity than observed by the original authors, at times up to 4 orders of magnitude greater.

Statistical results from these calculations are presented in Table 3.14. Standard deviations of regression were less than the average standard deviation of activity values under all circumstances, which suggests that these models are not over-fit. Higher $R^2$ and $Q^2$ statistics are preferred, as long as the RMSE is not too much greater than the SD. High leave-one-out cross-validated $R^2$ is no guarantee of a predictive model, although an old rule of thumb is that a model is generally only useful if $R^2$ CV is at least 0.5. The Stability metric more precisely estimates the effect of removing molecules from the training set. In Phase, random subsets of 10% are left out and the predictions (not the observed activities) are compared to the full model; the higher this value, the less sensitive the model to changes in the training set. Very few of the resulting models show
Figure 3.3: Sigma-2 prototype conformer heat map

![Sigma-2 prototype conformer heat map]

Figure 3.4: Problem compounds for development of Gaussian QSAR from aligned datasets

![Problem compounds]

Chol 1
Nucl. Med. Biol. 28, 657

Fontanilla 5
Biochemistry 47, 7205
appreciable sensitivity when increasing the number of PLS factors. Scrambling of activity labels should also not produce a coefficient of determination comparable to $R^2$, as that would indicate that the model is no better than one generated at random. Pearson-$r$ is a measure of the ability of the model to predict the relative rank correctly in the external test set.
Table 3.14: Sigma-2 pharmacophore hypotheses and related statistics for regressions of 1–3 PLS factors. Each hypothesis is labeled with the nature of the pharmacophore features (A = H-bond acceptor, D = H-bond donor, H = hydrophobic, R = aromatic, N = negative ionizable, P = positive ionizable). The number of database screening matches within the σ-2 database is shown in parentheses. A total of 723 ligands representing the curated σ-2 data were screened.

<table>
<thead>
<tr>
<th>No. Factors</th>
<th>SD</th>
<th>$R^2$</th>
<th>$R^3$ CV</th>
<th>$R^2$ Scramble</th>
<th>Stability</th>
<th>$P^i$</th>
<th>$P^s$</th>
<th>RMSE</th>
<th>$Q^2$</th>
<th>Pearson-r</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHHPR.15 (470)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.8564</td>
<td>0.4186</td>
<td>0.2916</td>
<td>0.1568</td>
<td>0.971</td>
<td>167.1</td>
<td>3.82e-29</td>
<td>0.96</td>
<td>0.2381</td>
<td>0.5076</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7584</td>
<td>0.5460</td>
<td>0.3484</td>
<td>0.2855</td>
<td>0.944</td>
<td>138.9</td>
<td>2.45e-40</td>
<td>0.91</td>
<td>0.3160</td>
<td>0.5809</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6570</td>
<td>0.6608</td>
<td>0.3636</td>
<td>0.4169</td>
<td>0.879</td>
<td>149.3</td>
<td>9.98e-54</td>
<td>0.88</td>
<td>0.3484</td>
<td>0.6149</td>
<td></td>
</tr>
<tr>
<td>ADHHPR.17 (472)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.8098</td>
<td>0.4337</td>
<td>0.3308</td>
<td>0.1207</td>
<td>0.983</td>
<td>178.5</td>
<td>1.33e-30</td>
<td>0.98</td>
<td>0.2110</td>
<td>0.4782</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7284</td>
<td>0.5439</td>
<td>0.3846</td>
<td>0.2306</td>
<td>0.962</td>
<td>138.3</td>
<td>2.83e-40</td>
<td>0.94</td>
<td>0.2807</td>
<td>0.5409</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6342</td>
<td>0.6557</td>
<td>0.4180</td>
<td>0.3365</td>
<td>0.908</td>
<td>146.7</td>
<td>3.23e-53</td>
<td>0.92</td>
<td>0.3010</td>
<td>0.5576</td>
<td></td>
</tr>
<tr>
<td>ADHHPR.28 (471)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.8405</td>
<td>0.4074</td>
<td>0.2708</td>
<td>0.1333</td>
<td>0.971</td>
<td>160.2</td>
<td>2.75e-28</td>
<td>0.87</td>
<td>0.3334</td>
<td>0.5785</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7271</td>
<td>0.5584</td>
<td>0.3541</td>
<td>0.2572</td>
<td>0.941</td>
<td>146.7</td>
<td>6.66e-42</td>
<td>0.84</td>
<td>0.3737</td>
<td>0.6131</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6124</td>
<td>0.6881</td>
<td>0.3937</td>
<td>0.3486</td>
<td>0.872</td>
<td>169.9</td>
<td>3.68e-58</td>
<td>0.84</td>
<td>0.3819</td>
<td>0.6248</td>
<td></td>
</tr>
<tr>
<td>ADHHPR.3 (485)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.8033</td>
<td>0.4425</td>
<td>0.3035</td>
<td>0.1265</td>
<td>0.969</td>
<td>191.3</td>
<td>2.05e-32</td>
<td>0.91</td>
<td>0.2355</td>
<td>0.5100</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7120</td>
<td>0.5639</td>
<td>0.3709</td>
<td>0.2309</td>
<td>0.949</td>
<td>155.1</td>
<td>5.68e-44</td>
<td>0.87</td>
<td>0.3012</td>
<td>0.5732</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6227</td>
<td>0.6678</td>
<td>0.4260</td>
<td>0.3391</td>
<td>0.909</td>
<td>160.1</td>
<td>6.51e-57</td>
<td>0.84</td>
<td>0.3575</td>
<td>0.6055</td>
<td></td>
</tr>
<tr>
<td>ADHHPR.39 (483)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.7866</td>
<td>0.4560</td>
<td>0.3574</td>
<td>0.1346</td>
<td>0.983</td>
<td>200.3</td>
<td>1.94e-33</td>
<td>0.87</td>
<td>0.2942</td>
<td>0.5509</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.6899</td>
<td>0.5833</td>
<td>0.4188</td>
<td>0.2671</td>
<td>0.96</td>
<td>166.6</td>
<td>5.77e-46</td>
<td>0.83</td>
<td>0.3590</td>
<td>0.6158</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6093</td>
<td>0.6764</td>
<td>0.4312</td>
<td>0.3741</td>
<td>0.909</td>
<td>165.1</td>
<td>8.89e-58</td>
<td>0.81</td>
<td>0.3835</td>
<td>0.6356</td>
<td></td>
</tr>
<tr>
<td>ADHHPR.43 (509)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.8055</td>
<td>0.4300</td>
<td>0.3095</td>
<td>0.1276</td>
<td>0.979</td>
<td>190.1</td>
<td>1.33e-32</td>
<td>0.89</td>
<td>0.2983</td>
<td>0.5477</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.6872</td>
<td>0.5868</td>
<td>0.3740</td>
<td>0.2553</td>
<td>0.94</td>
<td>178.2</td>
<td>6.78e-49</td>
<td>0.93</td>
<td>0.2409</td>
<td>0.5140</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6081</td>
<td>0.6777</td>
<td>0.4157</td>
<td>0.3623</td>
<td>0.896</td>
<td>175.3</td>
<td>3.49e-61</td>
<td>0.92</td>
<td>0.2605</td>
<td>0.5363</td>
<td></td>
</tr>
<tr>
<td>ADHHPR.57 (482)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.7906</td>
<td>0.4654</td>
<td>0.3419</td>
<td>0.1351</td>
<td>0.976</td>
<td>208.1</td>
<td>2.35e-34</td>
<td>0.88</td>
<td>0.3031</td>
<td>0.5721</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.6836</td>
<td>0.6020</td>
<td>0.3891</td>
<td>0.2575</td>
<td>0.938</td>
<td>180.0</td>
<td>2.41e-48</td>
<td>0.81</td>
<td>0.4080</td>
<td>0.6483</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6100</td>
<td>0.6844</td>
<td>0.3729</td>
<td>0.3630</td>
<td>0.863</td>
<td>171.3</td>
<td>4.48e-59</td>
<td>0.82</td>
<td>0.3893</td>
<td>0.6395</td>
<td></td>
</tr>
<tr>
<td>ADHHPR.83 (555)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.8241</td>
<td>0.4213</td>
<td>0.3240</td>
<td>0.1268</td>
<td>0.985</td>
<td>200.9</td>
<td>1.21e-34</td>
<td>0.88</td>
<td>0.3026</td>
<td>0.5531</td>
<td></td>
</tr>
</tbody>
</table>

56
2 0.7482 0.5247 0.3578 0.2477 0.964 151.8 3.83e-45 0.86 0.3253 0.5785
3 0.6656 0.6252 0.3567 0.3436 0.903 152.4 4.21e-58 0.80 0.4181 0.6492

ADHPRR.87 (490)
1 0.7962 0.4861 0.3848 0.1306 0.983 229.8 5.45e-37 0.86 0.3519 0.5958
2 0.6884 0.6174 0.4596 0.2716 0.962 189.5 4.07e-63 0.76 0.4979 0.7126
3 0.6085 0.7023 0.4735 0.3702 0.925 152.4 3.26e-51 0.86 0.3253 0.5785

ADHPRR.30 (483)
1 0.8250 0.3970 0.2357 0.1570 0.961 158.0 3.56e-28 0.97 0.1038 0.4020
2 0.7062 0.5601 0.3278 0.2785 0.925 152.1 2.43e-43 0.92 0.2054 0.4853
3 0.6346 0.6462 0.3239 0.3638 0.866 144.9 2.01e-53 0.88 0.2735 0.5474

ADHPRR.34 (469)
1 0.8177 0.4255 0.3110 0.1264 0.979 172.6 7.21e-30 0.90 0.2531 0.5136
2 0.7206 0.5557 0.3631 0.2715 0.948 145.1 1.34e-41 0.85 0.3470 0.5995
3 0.6277 0.6644 0.3867 0.3716 0.887 152.4 1.71e-54 0.83 0.3677 0.6229

ADHPRR.36 (485)
1 0.8527 0.3627 0.2797 0.1378 0.988 140.0 7.18e-26 0.96 0.1851 0.4455
2 0.7131 0.5561 0.4011 0.2385 0.965 153.5 6.22e-44 0.90 0.2767 0.5458
3 0.6237 0.6619 0.4524 0.3412 0.936 159.2 3.62e-57 0.86 0.3342 0.5881

ADHPRR.38 (494)
1 0.8519 0.3378 0.2691 0.1345 0.991 124.4 1.27e-23 0.92 0.2700 0.5197
2 0.6968 0.5587 0.4061 0.2333 0.957 153.8 6.85e-44 0.90 0.2938 0.5583
3 0.5909 0.6840 0.4550 0.3385 0.911 174.6 3.02e-60 0.89 0.3088 0.5789

ADHPRR.40 (499)
1 0.8094 0.4487 0.3511 0.1398 0.983 201.0 8.73e-34 0.93 0.2372 0.5107
2 0.7405 0.5404 0.3925 0.2712 0.97 144.6 2.96e-42 0.92 0.2562 0.5250
3 0.6506 0.6467 0.3703 0.3711 0.903 149.5 4.49e-55 0.89 0.2999 0.5654

ADHPRR.41 (506)
1 0.8684 0.3677 0.2640 0.1164 0.98 146.0 8.51e-27 0.93 0.2383 0.4939
2 0.7713 0.5032 0.3323 0.2436 0.958 126.6 1.05e-38 0.90 0.2854 0.5438
3 0.6787 0.6169 0.3573 0.3404 0.911 133.6 1.33e-51 0.87 0.3373 0.5850

AHHHPR.109 (485)
1 0.8807 0.3760 0.2821 0.1540 0.984 144.6 2.22e-26 0.93 0.2702 0.5201
2 0.7654 0.5306 0.3470 0.2704 0.951 135.1 5.63e-40 0.91 0.3036 0.5559
3 0.6580 0.6545 0.3831 0.3686 0.895 150.3 1.18e-54 0.83 0.4205 0.6485

AHHHPR.111 (482)
1 0.8684 0.3986 0.3088 0.1335 0.986 158.4 3.31e-28 0.94 0.2609 0.5198

57
<table>
<thead>
<tr>
<th></th>
<th>0.7702 0.5289 0.3619</th>
<th>0.2694 0.961 133.6</th>
<th>1.26e-39 0.91 0.3132</th>
<th>0.5752</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.6472 0.6687 0.4091</td>
<td>0.3602 0.901 159.5</td>
<td>1.4e-56 0.85 0.3924</td>
<td>0.6441</td>
</tr>
</tbody>
</table>

**AHHHPR.113 (574)**

| 1 | 0.8764 0.3373 0.2245 | 0.1272 0.98 145.1 | 2.78e-27 0.96 0.1525 | 0.4349 |
| 2 | 0.7838 0.4717 0.3109 | 0.2363 0.963 126.8 | 4.45e-40 0.94 0.1936 | 0.4903 |
| 3 | 0.6485 0.6396 0.3853 | 0.3530 0.909 167.4 | 2.07e-62 0.88 0.3008 | 0.5741 |

**AHHHPR.114 (575)**

| 1 | 0.8061 0.4241 0.2603 | 0.1274 0.96 210.6 | 3.9e-36 0.87 0.3084 | 0.5633 |
| 2 | 0.7414 0.5146 0.3224 | 0.2276 0.951 151.1 | 1.86e-45 0.85 0.3264 | 0.5778 |
| 3 | 0.6687 0.6065 0.3603 | 0.3276 0.923 145.9 | 3.25e-57 0.81 0.3891 | 0.6305 |

**AHHHPR.46 (546)**

| 1 | 0.8127 0.4173 0.2958 | 0.1397 0.977 193.3 | 1.63e-33 0.89 0.2708 | 0.5285 |
| 2 | 0.6899 0.5816 0.3799 | 0.2482 0.945 187.0 | 1.27e-51 0.82 0.3821 | 0.6269 |
| 3 | 0.5596 0.7258 0.4567 | 0.3516 0.883 236.5 | 5.6e-75 0.84 0.3485 | 0.6236 |

**AHHHPR.51 (538)**

| 1 | 0.8486 0.3625 0.2867 | 0.1268 0.99 151.2 | 8.07e-28 0.95 0.2298 | 0.4843 |
| 2 | 0.7565 0.4952 0.3078 | 0.2515 0.951 130.0 | 4.62e-40 0.95 0.2235 | 0.4880 |
| 3 | 0.6589 0.6185 0.3396 | 0.3372 0.884 142.6 | 5.95e-55 0.93 0.2706 | 0.5391 |

**AHHHPR.53 (571)**

| 1 | 0.8344 0.3880 0.2941 | 0.1206 0.987 179.4 | 5.07e-32 0.91 0.2397 | 0.5051 |
| 2 | 0.7453 0.5135 0.3628 | 0.2205 0.969 148.8 | 7.61e-45 0.90 0.2663 | 0.5447 |
| 3 | 0.6105 0.6747 0.4098 | 0.3182 0.895 194.3 | 3.33e-68 0.87 0.3030 | 0.5836 |

**AHHHPR.90 (584)**

| 1 | 0.8362 0.3927 0.3097 | 0.1301 0.989 186.9 | 3.72e-33 0.97 0.2044 | 0.4688 |
| 2 | 0.7405 0.5254 0.3587 | 0.2499 0.964 159.4 | 2.45e-47 0.92 0.2883 | 0.5492 |
| 3 | 0.6531 0.6321 0.3793 | 0.3546 0.912 164.4 | 5.15e-62 0.91 0.3010 | 0.5746 |

**AHPHR.48 (588)**

| 1 | 0.7977 0.4207 0.3186 | 0.1338 0.98 211.4 | 2.26e-36 0.96 0.1798 | 0.4468 |
| 2 | 0.7291 0.5178 0.3739 | 0.2216 0.967 155.7 | 1.17e-46 0.94 0.2261 | 0.4890 |
| 3 | 0.6018 0.6726 0.4553 | 0.3286 0.927 197.9 | 9.5e-70 0.87 0.3275 | 0.5808 |

**AHPHR.49 (584)**

| 1 | 0.8112 0.3999 0.2979 | 0.1074 0.983 192.6 | 6.64e-34 0.93 0.2246 | 0.4829 |
| 2 | 0.7324 0.5125 0.3743 | 0.2069 0.972 151.4 | 1.18e-45 0.93 0.2312 | 0.5031 |
| 3 | 0.6483 0.6193 0.4111 | 0.3108 0.942 155.6 | 6.94e-60 0.91 0.2628 | 0.5358 |

**AHPHR.50 (579)**

<p>| 1 | 0.8690 0.3173 0.2414 | 0.1330 0.99 133.4 | 1.36e-25 1.02 0.0990 | 0.3498 |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7513</td>
<td>0.4915</td>
<td>0.3575</td>
<td>0.2337</td>
<td>0.972</td>
<td>138.2</td>
<td>9.87e-43</td>
<td>0.95</td>
<td>0.2168</td>
<td>0.4783</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6119</td>
<td>0.6638</td>
<td>0.4187</td>
<td>0.3286</td>
<td>0.911</td>
<td>187.6</td>
<td>3.8e-67</td>
<td>0.93</td>
<td>0.2439</td>
<td>0.5172</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AHHPRR.54 (618)

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9070</td>
<td>0.2783</td>
<td>0.2250</td>
<td>0.1168</td>
<td>0.995</td>
<td>118.0</td>
<td>1.82e-23</td>
<td>0.93</td>
<td>0.2711</td>
<td>0.5214</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7586</td>
<td>0.4968</td>
<td>0.3613</td>
<td>0.2134</td>
<td>0.974</td>
<td>150.5</td>
<td>3.31e-46</td>
<td>0.85</td>
<td>0.3804</td>
<td>0.6172</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6252</td>
<td>0.6592</td>
<td>0.4302</td>
<td>0.3069</td>
<td>0.924</td>
<td>196.0</td>
<td>9.66e-71</td>
<td>0.79</td>
<td>0.4642</td>
<td>0.6818</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AHHPRR.55 (623)

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9338</td>
<td>0.2729</td>
<td>0.2197</td>
<td>0.1073</td>
<td>0.995</td>
<td>116.3</td>
<td>3.05e-23</td>
<td>0.97</td>
<td>0.1672</td>
<td>0.4226</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7792</td>
<td>0.4954</td>
<td>0.3601</td>
<td>0.2035</td>
<td>0.972</td>
<td>151.7</td>
<td>1.28e-46</td>
<td>0.94</td>
<td>0.2209</td>
<td>0.5057</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6848</td>
<td>0.6115</td>
<td>0.3724</td>
<td>0.3144</td>
<td>0.929</td>
<td>161.6</td>
<td>6.52e-63</td>
<td>0.91</td>
<td>0.2657</td>
<td>0.5515</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AHHPRR.61 (414)

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8200</td>
<td>0.4307</td>
<td>0.3573</td>
<td>0.1587</td>
<td>0.991</td>
<td>155.1</td>
<td>7.08e-27</td>
<td>1.02</td>
<td>0.0551</td>
<td>0.3543</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.6735</td>
<td>0.6178</td>
<td>0.4198</td>
<td>0.2760</td>
<td>0.94</td>
<td>164.8</td>
<td>2.5e-43</td>
<td>0.98</td>
<td>0.1263</td>
<td>0.4496</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.5777</td>
<td>0.7202</td>
<td>0.4620</td>
<td>0.3813</td>
<td>0.89</td>
<td>174.1</td>
<td>7.05e-56</td>
<td>0.98</td>
<td>0.1208</td>
<td>0.4656</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AHHPRR.62 (411)

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7923</td>
<td>0.4314</td>
<td>0.3427</td>
<td>0.1304</td>
<td>0.987</td>
<td>154.0</td>
<td>1.09e-26</td>
<td>0.92</td>
<td>0.2729</td>
<td>0.5237</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.6740</td>
<td>0.5906</td>
<td>0.4379</td>
<td>0.2531</td>
<td>0.967</td>
<td>145.7</td>
<td>6.76e-40</td>
<td>0.91</td>
<td>0.2914</td>
<td>0.5427</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6138</td>
<td>0.6621</td>
<td>0.4400</td>
<td>0.3653</td>
<td>0.938</td>
<td>131.3</td>
<td>4.02e-47</td>
<td>0.89</td>
<td>0.3253</td>
<td>0.5723</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

AHHPRR.63 (406)

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8221</td>
<td>0.3896</td>
<td>0.2747</td>
<td>0.1600</td>
<td>0.979</td>
<td>127.7</td>
<td>3.26e-23</td>
<td>0.91</td>
<td>0.2838</td>
<td>0.5344</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7077</td>
<td>0.5500</td>
<td>0.3558</td>
<td>0.2834</td>
<td>0.948</td>
<td>121.6</td>
<td>3.1e-35</td>
<td>0.88</td>
<td>0.3295</td>
<td>0.5788</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.5995</td>
<td>0.6787</td>
<td>0.3835</td>
<td>0.3854</td>
<td>0.888</td>
<td>139.4</td>
<td>1.43e-48</td>
<td>0.87</td>
<td>0.3442</td>
<td>0.6088</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HHHPRR.4 (559)

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8985</td>
<td>0.2975</td>
<td>0.2088</td>
<td>0.1162</td>
<td>0.987</td>
<td>117.8</td>
<td>4.16e-23</td>
<td>0.98</td>
<td>0.1176</td>
<td>0.3704</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7198</td>
<td>0.5508</td>
<td>0.3758</td>
<td>0.2340</td>
<td>0.954</td>
<td>169.8</td>
<td>7.24e-49</td>
<td>0.90</td>
<td>0.2626</td>
<td>0.5267</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6119</td>
<td>0.6766</td>
<td>0.4459</td>
<td>0.3547</td>
<td>0.919</td>
<td>192.4</td>
<td>2.46e-67</td>
<td>0.87</td>
<td>0.3173</td>
<td>0.5746</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HHHPRR.5 (562)

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9323</td>
<td>0.2411</td>
<td>0.1670</td>
<td>0.1360</td>
<td>0.99</td>
<td>88.3</td>
<td>2.15e-18</td>
<td>0.97</td>
<td>0.1366</td>
<td>0.3781</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7843</td>
<td>0.4649</td>
<td>0.2634</td>
<td>0.2576</td>
<td>0.946</td>
<td>120.3</td>
<td>2.46e-38</td>
<td>0.88</td>
<td>0.2836</td>
<td>0.5347</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6650</td>
<td>0.6166</td>
<td>0.2712</td>
<td>0.3687</td>
<td>0.862</td>
<td>148.0</td>
<td>3.66e-57</td>
<td>0.83</td>
<td>0.3629</td>
<td>0.6087</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

HHHPRR.6 (561)

<p>| | | | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.8535</td>
<td>0.3443</td>
<td>0.2547</td>
<td>0.1175</td>
<td>0.987</td>
<td>146.0</td>
<td>2.7e-27</td>
<td>0.97</td>
<td>0.1534</td>
<td>0.4115</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.7553</td>
<td>0.4883</td>
<td>0.3412</td>
<td>0.2285</td>
<td>0.971</td>
<td>132.2</td>
<td>4.95e-41</td>
<td>0.99</td>
<td>0.1137</td>
<td>0.3932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.6312</td>
<td>0.6440</td>
<td>0.3795</td>
<td>0.3440</td>
<td>0.905</td>
<td>166.4</td>
<td>1.34e-61</td>
<td>0.96</td>
<td>0.1787</td>
<td>0.4606</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Standard deviation of the regression; coefficient of determination; leave-one-out cross-validated coefficient of regression; coefficient of determination using scrambled activity data; stability of the model to changes in training set composition; ratio of model variance to activity variance; significance level of $F$; Root-Mean-Square Error in test set predictions; coefficient of determination for the external test set; correlation between predicted and observed external test set activity.
Based on the battery of statistics, QSAR model ADHHPR.87 comprising either 2- or 3-PLS factors performed significantly better than the remaining models. Plots of the regression for training and tests sets are provided in Figure 3.5. This pharmacophore corresponded to a conformer of a molecule developed by Mach et al. (see Figure 3.6). The Pearson-\(r\) of the 3-PLS factor QSAR model indicates modest relative accuracy, and the remaining statistics are acceptable. Virtual screening of the curated dataset spiked with known inactives and the set of 1,000 drug-like decoys led to a recovery of 57.6% of known active compounds. Receiver-operator characteristic and performance as percentage of the database screened are provided in Figure 3.7. After ranking based on the QSAR, enrichment factors \(\text{EF}_1\), \(\text{EF}^*_1\), and \(\text{EF}'_1\) were determined to be 2.6, 42, and 65, respectively. The latter two measures indicate a robust early enrichment despite the inability to recover all known actives.

Field fractions from the 2- and 3-PLS regressions (Figure 3.15) indicate a considerable amount of steric (bulk) and hydrophobic (grease) contributions to the QSAR, followed by hydrogen bond acceptor contributions. Electronic and hydrogen bond donor contributions were minimal.

Visualization of the Gaussian fields superimposed on the pharmacophore and representative ligand are provided in Figures 3.8–3.12. Positive steric effects are indicated at the hydrophobic and ring aromatic sites on the east and west ends of the ligand, respectively. The impact of negative
Table 3.15: Gaussian field fractions of hypothesis ADHHPR.87

<table>
<thead>
<tr>
<th>No. Factors</th>
<th>Steric</th>
<th>Electrostatic</th>
<th>Hydrophobic</th>
<th>Hbond Acceptor</th>
<th>Hbond Donor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.407</td>
<td>0.054</td>
<td>0.314</td>
<td>0.177</td>
<td>0.047</td>
</tr>
<tr>
<td>2</td>
<td>0.390</td>
<td>0.069</td>
<td>0.306</td>
<td>0.168</td>
<td>0.068</td>
</tr>
<tr>
<td>3</td>
<td>0.328</td>
<td>0.081</td>
<td>0.309</td>
<td>0.190</td>
<td>0.091</td>
</tr>
</tbody>
</table>

steric contributions is negligible. An interesting combination of negative and positive electrostatic contributions demonstrates unfavorable interactions with positive partial charge surrounding the aromatic ring and continuing towards the linker oxygen presented by the carbamate. Another unfavorable electrostatic positive partial charge field is present over the tertiary amine. Favorable positive electrostatics extend over the top of the aromatic ring and through the carbonyl of the carbamate and around the amine in a snake-like manner. Positive hydrophobic contours are present throughout the majority of the scaffold with the exception of the tertiary amine, and a small portions of the chain at the east end. Positive hydrogen bond acceptor projected point fields are expressed at the back of the aromatic ring and nestled between the amide and ester linkages of the carbamate, whereas negative contributions are made around the carbonyl and at the east end in the linker chain near the amine. Hydrogen bond donor projected point fields have an almost exclusively negative impact on the QSAR model with the exception of the back side of the 9-azabicyclo[3.3.1]nonan-3α-yl moiety.

Additional visualizations of the best alignment of one of our laboratory’s active and nonselective benzo[d]thiazol-2(3H)ones is provided for context of how this particular QSAR model could be used for ligand-based drug development. Due to the bent alignment, this particular conformation of RB8 is well superimposed on the positive steric and hydrophobic fields. The thiazolone moiety is also well situated within the positive electrostatic region, whereas the tertiary amine is located within the negative electrostatic region. This agrees with the relative partial charges assigned by the OPLS_2005 force field. The thiazolone carbonyl nestles a region of slightly negative hydrogen bond acceptor contours, yet the second, and particularly third carbon down the linker from the thiazolone seem particularly suited for replacement with an acceptor moiety. While RB8, aligned in this way
does not suffer from the negative hydrogen bond donor fields, one hypothetical improvement to \(\sigma\)-2 affinity would be to place a donor at the pro-S hydrogen 2 carbons away from the amine on the internal linker.

### 3.5.1 Development of \(\sigma\)-1 pharmacophore models and QSAR

The \(\sigma\)-1 database of 1396 compounds was treated to the same workflow as described for \(\sigma\)-2 ligands. A subset of active and inactive compounds (Table 3.16 and Figure 3.18) was selected using up to 3 active compounds from each cluster along with sufficient inactives (3 or more) to balance the classification set. SAP was performed using 16,384 informative fingerprints bits. The \(\Delta\) cutoff was determined to be 0.761, resulting in a fingerprint length of 13,334. The Tanimoto metric was used to generate a fingerprint matrix from this file for instance-based feature mapping. An optimal \(\lambda\) tuning parameter of 0.006, corresponding to a prediction error of 0.224 was used to train the SVM. An internal accuracy of 1 was found, with external accuracy and an MCC of 0.732 and 0.460, respectively. In total, 52 prototype conformers were found, comprising 30 active and 22 inactive conformers.

These conformers were taken through the develop common pharmacophore process, using feature definitions including the piperidine and piperazine moieties. A maximum of 6 sites matching a minimum of three ligands, based on the heat map (Figure 3.19), produced 4 representative common pharmacophores which were used to screen the entire database of \(\sigma\)-1 conformers.

<table>
<thead>
<tr>
<th>Training set</th>
<th>Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Active</td>
<td>Inactive</td>
</tr>
<tr>
<td>39</td>
<td>44</td>
</tr>
</tbody>
</table>
Figure 3.6: Prototype ligand for the pharmacophore alignment ADHHPR.87
Figure 3.7: Sigma-2 virtual screening performance. Receiver-operator characteristic and screening performance as percentage of the database screened. The 45° line corresponds to the expected performance if the screening were no better than random.
Figure 3.8: Gaussian steric fields presented by QSAR of ADHHPR.87. Positive steric field contours are shown in dark green. Negative contours are shown in yellow.

Figure 3.9: Gaussian electrostatic fields presented by QSAR of ADHHPR.87. Positive electrostatic field contours are shown in blue. Negative contours are shown in red.
Figure 3.10: Gaussian hydrophobic fields presented by QSAR of ADHHPR.87. Positive hydrophobic field contours are shown in yellow. Negative contours are shown in white.

Figure 3.11: Gaussian hydrogen bond acceptor fields presented by QSAR of ADHHPR.87. Positive H-bond acceptor field contours are shown in red. Negative contours are shown in magenta.
Figure 3.12: Gaussian hydrogen bond donor fields presented by QSAR of ADHHPR. Positive H-bond donor field contours are shown in blue-violet. Negative contours are shown in cyan.

Figure 3.13: Positive steric field contours are shown in dark green. Negative contours are shown in yellow.
Figure 3.14: Positive electrostatic field contours are shown in blue. Negative contours are shown in red.

Figure 3.15: Positive hydrophobic field contours are shown in yellow. Negative contours are shown in white.
Figure 3.16: Positive H-bond acceptor field contours are shown in red. Negative contours are shown in magenta.

Figure 3.17: Positive H-bond donor field contours are shown in blue-violet. Negative contours are shown in cyan.
Figure 3.18: Clustered σ-2 scaffold data for the balanced dataset. Visualization details are the same as in Figure 3.1. One hundred and thirty molecules (larger hexagons) were selected from this set for SAP and MILES performance analysis.
Figure 3.19: Sigma-1 prototype conformer heat map
Table 3.17: Sigma-1 pharmacophore hypotheses and related statistics for regressions of 1–3 PLS factors. Each hypothesis is labeled with the nature of the pharmacophore features (A = H-bond acceptor, D = H-bond donor, H = hydrophobic, R = aromatic, N = negative ionizable, P = positive ionizable). The number of database screening matches within the σ-1 database is shown in parentheses. A total of 1396 ligands representing the curated σ-1 data were screened.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>No. Factors</th>
<th>SDa</th>
<th>R2b</th>
<th>R2 CVc</th>
<th>R2 Scrambled</th>
<th>Stabilityf</th>
<th>Fe</th>
<th>Fc</th>
<th>RMSEh</th>
<th>Q2i</th>
<th>Pearson-rj</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAHHPR.12 (758)</td>
<td>1</td>
<td>1.0369</td>
<td>0.3573</td>
<td>0.2807</td>
<td>0.1096</td>
<td>0.988</td>
<td>209.6</td>
<td>4.37e-38</td>
<td>1.13</td>
<td>0.2357</td>
<td>0.4949</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.8772</td>
<td>0.5412</td>
<td>0.4082</td>
<td>0.2175</td>
<td>0.97</td>
<td>221.8</td>
<td>2.41e-64</td>
<td>1.08</td>
<td>0.3065</td>
<td>0.5672</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.7703</td>
<td>0.6472</td>
<td>0.4555</td>
<td>0.3127</td>
<td>0.942</td>
<td>229.3</td>
<td>1.84e-84</td>
<td>1.01</td>
<td>0.3876</td>
<td>0.6298</td>
</tr>
<tr>
<td>AAHHPR.16 (680)</td>
<td>1</td>
<td>0.9336</td>
<td>0.4356</td>
<td>0.3491</td>
<td>0.1246</td>
<td>0.984</td>
<td>250.9</td>
<td>2.83e-42</td>
<td>1.10</td>
<td>0.2056</td>
<td>0.4692</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.8293</td>
<td>0.5561</td>
<td>0.4319</td>
<td>0.2421</td>
<td>0.974</td>
<td>202.9</td>
<td>7.38e-58</td>
<td>1.06</td>
<td>0.2622</td>
<td>0.5262</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.7484</td>
<td>0.6396</td>
<td>0.4404</td>
<td>0.3456</td>
<td>0.943</td>
<td>191.1</td>
<td>3.04e-71</td>
<td>1.02</td>
<td>0.3124</td>
<td>0.5715</td>
</tr>
<tr>
<td>AAHHPR.4 (795)</td>
<td>1</td>
<td>0.9810</td>
<td>0.3996</td>
<td>0.3106</td>
<td>0.1081</td>
<td>0.986</td>
<td>255.5</td>
<td>1.87e-44</td>
<td>1.09</td>
<td>0.2405</td>
<td>0.5036</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.8617</td>
<td>0.5379</td>
<td>0.4003</td>
<td>0.2185</td>
<td>0.972</td>
<td>222.9</td>
<td>6.23e-65</td>
<td>1.00</td>
<td>0.3620</td>
<td>0.6080</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.7738</td>
<td>0.6283</td>
<td>0.4467</td>
<td>0.2978</td>
<td>0.951</td>
<td>215.2</td>
<td>1e-81</td>
<td>1.01</td>
<td>0.3439</td>
<td>0.6015</td>
</tr>
<tr>
<td>AAHHPR.8 (757)</td>
<td>1</td>
<td>1.0347</td>
<td>0.3422</td>
<td>0.2501</td>
<td>0.1087</td>
<td>0.982</td>
<td>196.1</td>
<td>3.6e-36</td>
<td>1.12</td>
<td>0.2246</td>
<td>0.4897</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.9197</td>
<td>0.4817</td>
<td>0.3326</td>
<td>0.2367</td>
<td>0.964</td>
<td>174.7</td>
<td>2.21e-54</td>
<td>1.08</td>
<td>0.2777</td>
<td>0.5506</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.8125</td>
<td>0.5965</td>
<td>0.3842</td>
<td>0.3171</td>
<td>0.928</td>
<td>184.8</td>
<td>1.51e-73</td>
<td>1.01</td>
<td>0.3654</td>
<td>0.6178</td>
</tr>
</tbody>
</table>

*a* standard deviation of the regression; b* coefficient of determination; c* leave-one-out cross-validated coefficient of regression; d* coefficient of determination using scrambled activity data; e* stability of the model to changes in training set composition; f* ratio of model variance to activity variance; g* significance level of F; h* Root-Mean-Square Error in test set predictions; i* coefficient of determination for the external test set; j* correlation between predicted and observed external test set activity.

Gaussian field-based QSARs were developed for alignments to all hypotheses, and statistics for the chosen 3 PLS-factor model are given in Table 3.17. Plots of the regression for training and tests sets are provided in Figure 3.20.

Hypothesis AAHHPR.12 performed slightly better than the remaining hypotheses, all of which shared the same number of features. The Gaussian field fractions (Table 3.18), much like those of the σ-2 QSAR, indicate a high level of steric, hydrophobic character, although the hydrogen bond field fraction is noticeably more predominant. Virtual screening of the curated dataset spiked
with known inactives and the set of 1,000 drug-like decoys led to a recovery of 54.4% of known active compounds. After ranking based on the QSAR, enrichment factors EF, EF*, and EF′ were determined to be 1.8, 12, and 23, respectively. The latter two measures indicate a robust early enrichment despite the inability to recover all known actives. Receiver-operator characteristic and performance as percentage of the database screened are provided in Figure 3.21.

Table 3.18: Gaussian field fractions of hypothesis AAHHPR.12

<table>
<thead>
<tr>
<th>No. Factors</th>
<th>Steric</th>
<th>Electrostatic</th>
<th>Hydrophobic</th>
<th>Hbond Acceptor</th>
<th>Hbond Donor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.399</td>
<td>0.042</td>
<td>0.229</td>
<td>0.252</td>
<td>0.079</td>
</tr>
<tr>
<td>2</td>
<td>0.401</td>
<td>0.045</td>
<td>0.276</td>
<td>0.208</td>
<td>0.069</td>
</tr>
<tr>
<td>3</td>
<td>0.377</td>
<td>0.054</td>
<td>0.265</td>
<td>0.215</td>
<td>0.090</td>
</tr>
</tbody>
</table>

Visualization of the Gaussian fields superimposed on the pharmacophore and representative ligand are provided in Figures 3.23–3.27. Interestingly, although the feature types of the underlying pharmacophore are very similar to the σ-2 pharmacophore, there are some important differences in the Gaussian fields. Sterically, the only favorable interaction is located very near to the positive ionizable feature. The remaining steric interactions form a pocket around the reference ligand of Hudkins et al. 51 (see Figure 3.22). Again, negative electrostatics are favored around the positive ionizable feature. Hydrophobic features extend through much of the scaffold except for the location
Figure 3.21: Sigma-1 virtual screening performance. Receiver-operator characteristic and screening performance as percentage of the database screened.
of the positive ionizable feature and the ring aromatic at the west end. Acceptor fields are favored near the ester oxygens, in the same areas which are disfavored sterically. Donor contours show a strong unfavorable interaction on the west end in the vicinity of the ring aromatic pharmacophore site.
Figure 3.22: Prototype ligand for the pharmacophore alignment AAHHRPR.12

Figure 3.23: Gaussian steric fields presented by QSAR of AAHHRPR.12. Positive steric field contours are shown in dark green. Negative contours are shown in yellow.
Figure 3.24: Gaussian electrostatic fields presented by QSAR of AAHHP.R.12. Positive electrostatic field contours are shown in blue. Negative contours are shown in red.

Figure 3.25: Gaussian hydrophobic fields presented by QSAR of AAHHP.R.12. Positive hydrophobic field contours are shown in yellow. Negative contours are shown in white.
Figure 3.26: Gaussian hydrogen bond acceptor fields presented by QSAR of AAHHPR.12. Positive H-bond acceptor field contours are shown in red. Negative contours are shown in magenta.

Figure 3.27: Gaussian hydrogen bond donor fields presented by QSAR of AAHHPR.12. Positive H-bond donor field contours are shown in blue-violet. Negative contours are shown in cyan.
4. CONCLUSIONS AND FUTURE WORK

4.1 Conclusions

Computational techniques for ligand-based drug design based on pharmacophore modeling, virtual screening, and QSARs have met with many challenges over the years when applied to σ receptor targets. One of the major hurdles, having a database of the extensive number of ligands presented in the literature since σ research began, will undoubtedly help to overcome the cause of many inconsistencies in computational analyses, namely, data heterogeneity. Curated set of σ data, based on $K_i$ values, and using comparable analytical methods is now available.

Clustering of the datasets by scaffold allowed for a sensible way of generating diverse training and test sets of active and inactive compounds in an efficient manner. This technique, pioneered by Guiguemde et al.\textsuperscript{76} has the advantage of hierarchically assembling a dendrogram of structures by related scaffold, and provides an intuitive way of visualizing properties associated with structural changes at the core of a database of ligands.

The modeling approach described here has met with a modicum of success, particularly for σ-2 classification techniques. Pharmacophore fingerprints performed very well when used for the classification of active and inactive compounds through the Multiple-instance learning via embedded subset selection protocol. Significance Analysis of Pharmacophores has the potential to
allow for a greatly reduced fingerprint bit length, although care must be taken to pick an appropriate metric, whether it be a distance or similarity measure. In particular, implementing MILES with measures which use products of bits from each instance or that use a maximum or minimum similarity are problematic if SAP has been utilized to reduce the fingerprint length. While the pharmacophore features which lead to better SAP classification accuracy can readily be calculated at some computational expense, the original implementation was designed for microarray data. The number of wells in a microarray is much smaller than the number of instances in a SAP analysis, and the SAM method which SAP is based upon uses the \( \Gamma \) function for permutation analysis. As the number of instances grows, this function causes an overflow and triggers an error, calling some question into the validity of the statistical analysis.

Aggressive reduction in fingerprint length tends to reduce classification accuracy, particularly with larger numbers of instances. A cursory analysis of the results of this research suggests that no fewer “informative bits” are used than the total number of conformers. The elimination of bits with a frequency of less than 5\%, as implemented by Fu et al.\textsuperscript{82} produced slightly more bits than the levels we investigated. This may partially explain the comparable classification performance with the 54-molecule dataset used at higher fingerprint lengths.

An issue with purely ligand-based design projects is how to identify active conformers of ligands for pharmacophore development, shape-screening, virtual screening, and other methods that depend on good knowledge of the 3D coordinates of active molecules at the biomolecular target. In preliminary experiments using a complete set of conformers, the number of pharmacophores surviving the common pharmacophore perception step was overwhelming, and model assessment at this point was unjustifiable. The MILES method, as implemented here, provided a set of prototype conformers that was useful in reducing the overhead of pharmacophore generation with very flexible molecules. Consequentially, number of common pharmacophores generated incorporating the MILES approach was orders of magnitude smaller than when using a complete set of conformers. Virtual screening of with the surviving pharmacophore hypotheses allowed for the retrieval of much of the original datasets, indicating a general utility for screening \( \sigma \)-like molecules. Quantitative
structure-activity relationships built upon the aligned hits provided virtual screening rankings that retrieved over 50% of the known actives for both \( \sigma-1 \) and \( \sigma-2 \) targets. While this level of performance is not ideal, it does suggest that multiple binding modes or binding sites are present on both receptors. It may be possible to eliminate the retrieved actives from the pharmacophore modeling process and build new models which capture the pharmacophore features of these alternate binding sites. These pharmacophore models and their related QSARs will be useful tools for further virtual screening projects targeted at discovering more diverse scaffolds from which to build upon.

4.2 Future Work

Identification of alternative pharmacophore models based on the set of ligands not retrieved with the current pharmacophore models is needed for a comprehensive description of the \( \sigma \) binding motifs. As it stands, the virtual screening performance is suitable for screening a large database of commercially available compounds. Careful consideration of the hits retrieved in such a screen will be useful for identifying more diverse scaffolds for the exploration of \( \sigma \) site probes. Small modifications of the procedures used in this work will also be useful for developing selectivity-rather than affinity-based models. It will also be interesting to expand this work to identify binding requirements of agonists versus antagonists, although the existing literature is sometimes vague or contradictory when it comes to determining the pharmacological outcomes that are used for such a classification.
LIST OF REFERENCES
LIST OF REFERENCES


Yao, Y.; Lee, Y. *Another Look at Linear Programming for Feature Selection via Methods of Regularization*; 2010.

Alfons, A. perry: Resampling-based prediction error estimation for regression models. 2013; R package version 0.2.0.


[127] Ronsisvalle, G.; Prezzavento, O.; Pasquinucci, L.; Pappalardo, M. S.; Marrazzo, A.; Vittorio, F.; Carboni, L.; Spampinato, S. Synthesis of (+)-(1'R,2'S) and (1'S,2'R)-6,11-dimethyl-1,2,3,4,5,6-hexahydro-3-[[2'-(alkoxycarbonyl)-2'-phenylcyclopropyl]methyl]-2,6-methano-3-benzazocin-8-ol. Comparison of the affinities for sigma 1 and opioid receptors with in the diastereoisomeric MPCB and CCB. *Farmaco* 1997, 52, 471–476.


[133] Vilner, B. J.; Bowen, W. D. Sigma receptor-active neuroleptics are cytotoxic to C6 glioma cells in culture. *Eur J Pharmacol* 1993, 244, 199–201.


95


Berardi, F.; Loidice, F.; Fracchiolla, G.; Colabufo, N. A.; Perrone, R.; Tortorella, V. Synthesis of chiral 1-[ω-(4-chloro phenoxo) alkyl]-4-methylpiperidines and their biological evaluation at σ1, σ2, and sterol Δ8-Δ7 isomerase sites. *J. Med. Chem.* 2003, 46, 2117–2124.


[219] Data generously provided by Dr. Jonathan Katz.


[221] Moebius, F. F.; Reiter, R. J.; Hanner, M.; Glossmann, H. High affinity of sigma 1-binding sites for sterol isomerization inhibitors: evidence for a pharmacological relationship with the yeast sterol C8–C7 isomerase. *Br J Pharmacol* 1997, 121, 1–6.


[224] Novelli, F.; Sparatore, F. Spiro[1,2,4-benzotriazine-3(4H),4′-(1′-substituted)piperidines] and related compounds as ligands for sigma receptors. *Farmaco* 2002, 57, 871–882.


[272] Data generously provided by Dr. Christopher R. McCurdy.

[273] Data generously provided by Dr. Rae R. Matsumoto.

[274] Data generously provided by Dr. Christopher R. McCurdy.


[284] Data generously provided by Drs. James A. Fishback and Rae R. Matsumoto.


LIST OF APPENDICES
APPENDIX A. HIERARCHICAL CLUSTERING SCRIPTS
APPENDIX A. HIERARCHICAL CLUSTERING SCRIPTS

The following directory listings, Makefile, and python files implement hierarchical clustering. Prerequisites for this process are Canvas, Strip-It, Python, and Make. A SMILES input file named SMILES.smi which contains only unique structures is required. An output file cytoscape.net that contains the requisite network table data for visualization with Cytoscape is generated by the Makefile.

A.1 hcviz directory listing

xibalba:hcviz dewatson$ ls -lR
total 8
-rw-r--r-- 1 dewatson staff 1567 Jun 22 15:25 Makefile
drwxr-xr-x 6 dewatson staff 204 Sep 12 19:47 bin
drwxr-xr-x 4 dewatson staff 136 Jun 22 14:15 lib

./bin:
total 32
-rwxr-xr-x 1 dewatson staff 1378 Jun 22 13:04 network_cores.py
-rwxr-xr-x 1 dewatson staff 1341 Jun 22 13:52 network_linkage.py
-rwxr-xr-x 1 dewatson staff 1826 Jun 22 14:43 network_scaffolds.py
-rwxr-xr-x 1 dewatson staff 1414 Jun 22 13:04 singletons.py

./lib:
total 16
-rw-r--r-- 1 dewatson staff 9 Jun 22 13:02 MURCKO1.def
-rw-r--r-- 1 dewatson staff 21 Jun 22 13:03 RWL2.def

A.2 Makefile

##
# Makefile
#
# Driver to control the processing of molecular SMILES into
# a complete edge network
#
# author David Watson
# email dewatson@icloud.com
# copyright Copyright (c) 2013, David Watson
##

all: stripit canvas network

stripit:
    @echo Running Strip-It and creating scaffold files
strip-it --input SMILES.smi --output RWL2.strip  
   --scaffolds ./lib/RWL2.def
grep -v -e "-$$" RWL2.strip > RWL2.scaf
awk '{ print $$3 " " $$3 }' RWL2.scaf | grep -v "RINGS" | 
   sort -u > MURCKO1.smi
strip-it --input MURCKO1.smi --output MURCKO1.scaf  
   --scaffolds ./lib/MURCKO1.def
awk '{ print $$3 " " $$3 }' MURCKO1.scaf | grep -v "MURCKO" | 
   sort -u > RADIAL.smi

network:
./bin/network_scaffolds.py -i RWL2.scaf -o radial.net
strip-it --input singletons.smi --output singletons.scaf  
   --scaffolds ./lib/MURCKO1.def
./bin/network_cores.py -i MURCKO1.scaf -o radial.net
./bin/network_linkage.py -i radial.tree -o radial.net
./bin/singletons.py -i singletons.scaf -o radial.net
sed 's/\[/\[\]/g' singletons.sed | sed 's/\]/\]/g' > singletons.grep
grep -v -f singletons.grep radial.net > cytoscape.net

canvas:
$$SCHRODINGER/utilities/canvasFPGen -ismi RADIAL.smi -o radial.fp  
   -fptype radial -nostereo
$$SCHRODINGER/utilities/canvasFPMatrix -ifp radial.fp  
   -ocsv radial.csv -metric euclidean
$$SCHRODINGER/utilities/canvasHC -im radial.csv -linkage mcquitty  
   -ot radial.tree -kelley

mostly-clean: stripit-clean network-clean canvas-clean

stripit-clean:
   -rm -f RWL2.strip RWL2.scaf MURCKO1.smi MURCKO1.scaf RADIAL.smi

network-clean:
   -rm -f radial.net singletons.smi singletons.sed singletons.scaf  
      singletons.grep
canvas-clean:
   -rm -f radial.fp radial.csv radial.tree

all-clean: mostly-clean  
   -rm -f cytoscape.net

A.3 network_cores.py

#!/usr/bin/env python

112
network_cores.py

Creates network core node entries based on a Strip-it scaffold file

author     David Watson
email      dewatson@icloud.com
copyright  Copyright (c) 2013, David Watson

import sys
import argparse

parser = argparse.ArgumentParser(description =
    'Create a hierarchy of RWL2_Murcko1 interactions from a murko_1 strip-it' \
    ' file.

parser.add_argument('-i', metavar = 'infile', type = argparse.FileType('r'),
    help = 'Strip-It murcko1 file', required = True,
    dest = 'murcko')
parser.add_argument('-o', metavar = 'outfile', type = argparse.FileType('a'),
    help = 'Cytoscape network table', required = True,
    dest = 'cytoscape')

try:
    args = parser.parse_args()
except IOError, msg:
    parser.error(str(msg))

cytoscape = '%s %s %s %s

try:
    stripped = next(args.murcko)
except StopIteration:
    sys.exit('Strip-it file appears empty')

if not stripped.startswith('NAME'):
    sys.exit('Strip-it file appears to be invalid')

murcko = {}

while True:
    try:
        murckoline = next(args.murcko).rstrip().split("\t")
    except StopIteration:
        break

    while not murckoline[2] in murcko:
        murcko[murckoline[2]] = [murckoline[0]]
        break
    else:
murcko[murckoline[2]].append(murckoline[0])

for smiles in murcko.keys():
    for molecule in range(0, len(murcko[smiles]), 1):
        args.cytoscape.write(cytoscape % (murcko[smiles][molecule], 'S_C', smiles, '10.0'))

A.4 network_linkage.py

#!/usr/bin/env python

"""
network_linkage.py
"""

Converts a Canvas dendrogram to a Cytoscape edge network

author David Watson
email dewatson@icloud.com
copyright Copyright (c) 2013, David Watson

"""

import sys
import argparse

parser = argparse.ArgumentParser(description =
    'Create dendrogram leaf-node and node-node linkages from a Canvas tree ' \
    'file.'
)
parser.add_argument('-i', metavar = 'infile', type = argparse.FileType('r'),
    help = 'Canvas tree file', required = True, dest = 'tree')
parser.add_argument('-o', metavar = 'outfile', type = argparse.FileType('a'),
    help = 'Cytoscape network table', required = True,
    dest = 'cytoscape')

try:
    args=parser.parse_args()
except IOError, msg:
    parser.error(str(msg))

cytoscape = '%s %s %s %s %s %s

try:
    stripped = next(args.tree)
except StopIteration:
    sys.exit('Tree file appears empty')

if not stripped.startswith('0'):
    sys.exit('Tree file appears to be invalid')

args.tree.seek(0)
while True:
    try:
        nodeline = next(args.tree).rstrip().split()
    except StopIteration:
        break

node = 'N' + nodeline[0]
weight = nodeline[1].rstrip()

for linked in range(2):
    childline = next(args.tree).rstrip().split()
    if childline[0].startswith('C'):
        prefix = 'N'
    else:
        prefix = ''
    args.cytoscape.write(cytoscape % (prefix + childline[1], childline[0],
        '_N', node, weight))

### A.5 network_scaffolds.py

```
#!/usr/bin/env python

network_scaffolds.py

Create Cytoscape network node entries based on scaffolds

author    David Watson
email     dewolfson@icloud.com
copyright Copyright (c) 2013, David Watson

import sys
import argparse

parser = argparse.ArgumentParser()
    description = 'Create a network of molecule-scaffold from ' \
                   'Strip-It RINGS_WITH_LINKERS_2 output.'
parser.add_argument('-i', metavar = 'infile', type = argparse.FileType('r'),
    help = 'Strip-It rwl2 file', required = True, dest = 'rwl')
parser.add_argument('-o', metavar = 'outfile', type = argparse.FileType('w'),
    help = 'Cytoscape network table', required = True,
    dest = 'cytoscape')

try:
    args = parser.parse_args()
except IOError, msg:
    parser.error(str(msg))

cytoscape = '%s %s %s %s
'
singletons = open('singletons.smi', 'w')
patterns = open('singletons.sed', 'w')

try:
    stripped = next(args.rwl)
except StopIteration:
    sys.exit("Strip-it file appears empty")

if not stripped.startswith('NAME'):
    sys.exit("Strip-it file appears to be invalid")

rings_with_linkers = {}
args.cytoscape.write(cytoscape % ("source", "interaction", "target", "weight"))

while True:
    try:
        rwlline = next(args.rwl).rstrip().split("\t")
    except StopIteration:
        break

    while not rwlline[2] in rings_with_linkers:
        rings_with_linkers[rwlline[2]] = [rwlline[0]]
        break
    else:
        rings_with_linkers[rwlline[2]].append(rwlline[0])

for smiles in rings_with_linkers.keys():
    for molecule in range(0, len(rings_with_linkers[smiles]), 1):
        if len(rings_with_linkers[smiles]) == 1:
            singletons.write("%s %s
" % (smiles, rings_with_linkers[smiles][0]))
            patterns.write("^%s \n" % (smiles))
        else:
            args.cytoscape.write(cytoscape % (rings_with_linkers[smiles][molecule], "M_S", smiles, "10.0"))

A.6 singletons.py

#!/usr/bin/env python

singletons.py

Replace molecule-scaffold and scaffold-core mappings when there is only a single molecule representing a core

author       David Watson
import sys
import argparse

description = 'Create an edge network of molecule-core interactions ' \\
' from a Strip-It MURCKO_1 file containing singletons.'

parser.add_argument('-i', metavar = 'infile', type = argparse.FileType('r'), 
    help = 'Strip-It MURCKO_1 file', required = True, 
    dest = 'murcko')
parser.add_argument('-o', metavar = 'outfile', type = argparse.FileType('a'), 
    help = 'Cytoscape network table', required = True, 
    dest='cytoscape')

try:
    args = parser.parse_args()
except IOError, msg:
    parser.error(str(msg))

cytoscape = '%s %s %s %s

try:
    stripped = next(args.murcko)
except StopIteration:
    sys.exit("Strip-it file appears empty")

if not stripped.startswith('NAME'):
    sys.exit("Strip-it file appears to be invalid")

murcko = {}

while True:
    try:
        murckoline = next(args.murcko).rstrip().split("	")
        except StopIteration:
        break

    while not murckoline[2] in murcko:
        murcko[murckoline[2]] = [murckoline[0]]
        break
    else:
        murcko[murckoline[2]].append(murckoline[0])

for smiles in murcko.keys():
    for molecule in range(0,len(murcko[smiles]),1):
        args.cytoscape.write(cytoscape % (murcko[smiles][molecule], "M_C", 

117
A.7  MURCKO1.def

Configuration file for Strip-it

MURCKO_1

A.8  RWL2.def

Configuration file for Strip-it

RINGS_WITH_LINKERS_2
APPENDIX B. SAP IMPLEMENTATION
APPENDIX B. SAP IMPLEMENTATION

B.1 Makefile

Given a set of conformations produced by ConfGen, the following Makefile will attempt to generate fingerprints for all conformers, and if unsuccessful, will deal with removing failed fingerprint properties.

```bash
# #
# Makefile
#
# Prepare Canvas fingerprints for SAP analysis
#
# author David Watson
# email dewatson@icloud.com
# copyright Copyright (c) 2013, David Watson
#
#
fingerprints:
$$SCHRODINGER/utilities/canvasPharmFP \ 
   -fp sigmaPforeFeatures.def -imae superset.maegz \ 
   -odata conformers.fp -4pt -mostSig 16384 1>&2 2> pharmFPerrors.txt
$$SCHRODINGER/utilities/canvasFPBinary2CSV \ 
   -i conformers.fp -o conformers.csv -off 0 -notot
-grep index pharmFPerrors.txt >duds.txt
-awk '{ print $$9 }' duds.txt >duds.conformers
$$SCHRODINGER/utilities/proplister -c -noheader -p s_m_title \ 
   -p i_user_model_set -p i_user_activity_class superset.maegz -o props.csv
python bin/strip_3pt_pfores.py
```

B.2 bin/strip_3pt_pfores.py

This script takes care of removing the entries of failed fingerprint generation from the properties file.

```python
strip_3pt_pfores.py

Removes property information from the property file if Canvas is unable to generate a fingerprint for some reason

author David Watson
email dewatson@icloud.com
copyright Copyright (c) 2013, David Watson
```
import subprocess

# The following function was contributed by Olafur Waage to the website
# This uses a subprocess to execute the Unix command "wc -l" to determine
# the file length

def file_len(fname):
    p = subprocess.Popen(['wc', '-l', fname], stdout = subprocess.PIPE,
                         stderr=subprocess.PIPE)
    result, err = p.communicate()
    if p.returncode != 0:
        raise IOError(err)
    return int(result.strip().split()[0])

myFile = "props.csv"
totallength = file_len(myFile)
totalRange = range(1, totallength + 1)
duds = [int(line.rstrip()) for line in open("duds.conformers", "r")]

for myDud in duds:
    totalRange.remove(myDud)

propsFile = open("propsSubset.csv", "w")
props = [line for line in open('props.csv', 'r')]
for conformer in totalRange:
    propsFile.write(props[conformer - 1])
propsFile.close()

B.3 significanceAnalysis.R

This script is a driver for converting the Canvas fingerprints into a subset of significant fingerprints.

# #
# significanceAnalysis.R
#
# Implements SAP as described by Fu, et al. in BMC Bioinformatics 13, S3
#
# author David Watson
# email dewatson@icloud.com
# copyright Copyright (c) 2013, David Watson
#
##

# begin with a set of fingerprints and properties extracted from the superset
# load the SAP driver code
source("bin/sap.R")

# read in the raw fingerprints
FPs <- read.csv("conformers.csv",header=TRUE)

# read in the activity assignments: 1 := active, 2 := inactive
ACs <- read.csv("props.csv")
colnames(ACs) <- c("s_m_title", "i_user_model_set", "i_user_activity_class")

# determine the indices of the training set
attach(ACs)
modelSets <- 1:length(i_user_model_set) * (i_user_model_set != 2)
detach(ACs)

# generate a table containing the bags
uniqueActivities <- unique(ACs)

# classify bags as active and inactive
acActivity <- assignActivities(uniqueActivities)

# divide up the bags
intTrainClass <- acActivity[uniqueActivities$i_user_model_set == 1]
extTestClass <- acActivity[uniqueActivities$i_user_model_set == 2]

# generate significant fingerprints ("sigFPs.csv")
determineSAP(FPsubset, ACsubset)

B.4 sap.R

# #
# sap.R
#
# Implements SAP as described by Fu, et al. in BMC Bioinformatics 13, S3
#
# author David Watson
# email dewatson@icloud.com
# copyright Copyright (c) 2013, David Watson
#
##

library("matrixStats")
library("samr")
determineSAP <- function(rawFPs, ACs) {
  # transpose the fingerprint
  tFPs <- as.matrix(t(subset(rawFPs,select=-name)))
  attach(ACs)
  SAPFdata <- list(x = tFPs, y = i_user_activity_class,
  ...}
geneid = as.character(1:nrow(tFPs)),
genename = colnames(subset(rawFPs,select = -name)),
logged2 = FALSE)
detach(ACs)
SAP <- samr(SAPFdata,
  resp.type = "Two class unpaired",
nperms = 500)
save(SAP, file = "SAP.RData")
DELTAs <- samr.compute.delta.table(SAP, nvals = 100)
save(DELTAs, file = "DELTAs.RData")
DELTArow <- sum(DELTAs[1:nrow(DELTAs), 3] != 0) + 1
DELTA <- DELTAs[DELTArow, 1]
write.csv(DELTA, file = "DELTA.csv",
  quote = FALSE, row.names = FALSE)
par(pch = ".")
# Be sure a proper graphics terminal is open or comment the next line
samr.plot(SAP, DELTA)
SAPresults <- samr.compute.siggenes.table(SAP, DELTA, SAPFdata, DELTAs)
keepbits <- !colnames(FPs) %in% SAPresults$genes.lo[,2]
sigFPs <- subset(FPs, select=colnames(FPs[keepbits == TRUE]))
write.csv(sigFPs, file = "sigFPs.csv", quote = FALSE, row.names = FALSE)
cat("Significant fingerprint bits were saved to sigFPs.csv")

assignActivities <- function (acTable) {
  attach(acTable)
  activeBits <- 1*s_m_title %in% s_m_title[i_user_activity_class == 1]
inactiveBits <- -1!*s_m_title %in% s_m_title[i_user_activity_class == 1]
detach(acTable)
  activityAssignment <- activeBits + inactiveBits
activityAssignment
}
APPENDIX C. FEATURE MAPPING IMPLEMENTATION
APPENDIX C. FEATURE MAPPING IMPLEMENTATION

C.1 Makefile

Canvas is used to convert the significant fingerprints in CSV form into a native format, and then a fingerprint matrix is generated. Suitable metrics are cosine, dice, kulczynski, mcconnaughey, petke, simpson, tanimoto, and tversky. When working from unfiltered fingerprints, only the matrix step needs to be run.

```makefile
# #
# Makefile
# # Convert fingerprints into binary form and calculate similarity matrix
# # author David Watson
# email dewatson@icloud.com
# copyright Copyright (c) 2013, David Watson
#
##

features:
  $$\$SCHRODINGER/utilities/canvasCSV2FPBinary -icsv sigFPs.csv -o sigFPs.fp$$
  $$\$SCHRODINGER/utilities/canvasFPMatrix \$$
  -ifp sigFPs.fp -ocsv FPmatrix.csv -metric tanimoto

C.2 features.R

```r
# #
# features.R
# # Map bags to instances and return a feature matrix
# # author David Watson
# email dewatson@icloud.com
# copyright Copyright (c) 2013, David Watson
#
##

featureMap <- function(FPmatrix, activityClasses) {
  trainSet <-
    unique(subset(activityClasses, i_user_model_set == 1, select=s_m_title))
  featureVector <-
    vector(mode = "numeric", length = (length(unique(canvas)) *
      nrow(activityClasses)))
  count <- 0
  moleculeCount <- 0
for ( i in unique(canvas)) {
    moleculeCount <- moleculeCount + 1
    message("Evaluating molecule ", moleculeCount)
    for ( j in 1:nrow(activityClasses)) {
        count <- count + 1
        featureVector[count] <-
            max(subset(FPmatrix, canvas == as.character(i), select=j+1) )
    }
}
interimMatrix <-
    matrix(featureVector, nrow = length(unique(canvas)),
          ncol = length(canvas),
          dimnames = list(unique(canvas),canvas), byrow = TRUE)
keepFeatures <- colnames(interimMatrix) %in% trainSet$s_m_title
subset(interimMatrix, select=keepFeatures)
}

conformerMap <- function(FPmatrix, activityClasses) {
    trainSet <- unique(subset(activityClasses,
                          i_user_model_set == 1, select = s_m_title))
    conformerCount <- 0
    count <- 0
    featureVector <- vector(mode = "numeric", length = (nrow(activityClasses)^2))
    for ( i in 1:nrow(FPmatrix)) {
        conformerCount <- conformerCount + 1
        message("Evaluating conformer ", conformerCount)
        for ( j in 1:nrow(activityClasses)) {
            count <- count + 1
            featureVector[count] <- FPmatrix[i, j + 1]
        }
    }
    interimMatrix <-
        matrix(featureVector, nrow = nrow(activityClasses), ncol = nrow(activityClasses),
          dimnames = list(FPmatrix$canvas, FPmatrix$canvas), byrow = TRUE)
    keepFeatures <- colnames(interimMatrix) %in% trainSet$s_m_title
    subset(interimMatrix, select = keepFeatures)
}

C.3 mapping.R

This is the driver for the instance-based feature mapping

# #
# mapping.R
# #
# Interactive workflow for feature mapping
# #
# author David Watson

126
# read in the fingerprint similarity matrix and should be run directly after the SAP
# code otherwise read in the properties as in the SAP driver
FPmatrix <- read.csv("FPmatrix.csv", header = TRUE)
attach(FPmatrix)

# perform the instance-based feature mapping
FPfeatures <- featureMap(FPmatrix, ACs)
detach(FPmatrix)
APPENDIX D. 1-NORM SVM IMPLEMENTATION
APPENDIX D. 1-NORM SVM IMPLEMENTATION

D.1 bin/1norm.R

rconsole

# #
# 1norm.R
#
# Implementation of Yao and Lee’s parametric 1-norm SVM as an LP
# Another Look at Linear Programming for Feature Selection
# via Methods of Regularization, Technical Report No. 800r,
# Department of Statistics, Ohio State University, 2010.
# See section 4.2, equation 17
#
# author David Watson
# email dewatson@icloud.com
# copyright Copyright (c) 2013, David Watson
#
##

library("Rglpk")
library("perry")
library("caret")

oneNormSVM <- function (trialMatrix, activityAssignment, lambda) {
  trialRows <- nrow(trialMatrix)
  trialCols <- ncol(trialMatrix)
  cost <- c(rep(0, 2), rep(0, trialCols*2),
            rep(1/trialRows, trialRows), rep(0, trialRows))
  a <- c(rep(0, 2), rep(1, trialCols*2), rep(0, trialRows), rep(0, trialRows))
  A <- cbind(activityAssignment, -activityAssignment,
              activityAssignment*trialMatrix, -activityAssignment*trialMatrix,
              diag(rep(1, trialRows)), -diag(rep(1, trialRows)))
  b <- c(rep(1, trialRows))
  oFx <- cost + (lambda*a/trialCols)
  dir <- c("==", trialRows)
  ulbounds <- list(lower = list(ind = c(1L, 2L), val=c(-Inf, -Inf)))
  onensvm <- Rglpk_solve_LP(oFx, A, dir, b, bounds = ulbounds)
  solution <- list()
  solution$hyperplane <- onensvm$solution[3:(trialCols+2)] -
  onensvm$solution[3 + trialCols:(2 * trialCols + 2)]
  solution$error <-
  onensvm$solution[(2 * trialCols + 3):(2 * trialCols + trialRows + 2)] -
onensvm$solution[(3 + 2 * trialCols + trialRows):]
(2 * trialCols + 2 * trialRows + 2)]
class(solution) <- "1nsvm"
solution
}

# Prediction method for class 1nsvm
predict.1nsvm <- function (object, testSet) {
  hyperplane <- object$hyperplane
  margin <- object$margin
  as.vector(sign(margin + testSet %*% hyperplane))
}

# Internal accuracy calculation for a tuned 1-norm SVM
internalAccuracy <- function (internalTest, internalAssignment) {
  validSet <- vector(mode="numeric")
  for (itSol in rownames(internalTest)) {
    tempSol <- sign(mySlack + internalTest[as.character(itSol),] %*% myPenalty)
    validSet <- append(validSet, tempSol)
  }
  cat(validSet)
  # The following will calculate the percent of correctly assigned activities
  classificationAccuracy <- sum((validSet * internalAssignment) == 1) /
    nrow(internalTest)
}

# cost function for SVM tuning
svmCost <- function(true,predicted) {
  (length(predicted)-sum(diag(table(true,predicted))))/length(predicted)
}

# Matthews Correlation Coefficient for balanced measure of classification
# performance
MCC <- function (TP, TN, FP, FN) {
  ((TP*TN)-(FP*FN))/sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))
}

D.2 1nSVM.R

This is the driver for the 1-norm SVM. It should be run immediately after the SAP and instance-based feature selection protocols.

source(bin/1norm.R)
# #
# 1norm.R
#
# Interactive workflow for tuning 1-norm SVM
# divide the feature space between training and test sets
InternalMatrix <- subset(FPfeatures, subset = uniqueActivities$i_user_model_set == 1)
ExtTesting <- subset(FPfeatures, subset = uniqueActivities$i_user_model_set == 2)

# generate random splits of for 5-fold cross-validation
mySplits <- cvFolds(n = nrow(InternalMatrix), K = 5, R = 15, type = "random")

# perform initial tuning of the SVM
tuning <- list(lambda = c(10^(-12:4)))
# the tuning process may be split among multiple cores on SMP systems
firstTuning <- perryTuning(oneNormSVM, x = InternalMatrix,
y = intTrainClass, tuning = tuning, cost = svmCost, splits = mySplits,
names = c("trialMatrix", "activityAssignment"), ncores = 2)

# examine the tuning parameter
tuning

# perform the second round of tuning
# note that this is an interactive process and the next line must be edited
tuning <- list(lambda = unique(c(seq(from = 1e-7, to = 1e-6, by = (1e-6 - 1e-7)/9),
seq(from = 1e-6, to = 1e-5, by = (1e-5 - 1e-6)/9))))

secondTuning <- perryTuning(oneNormSVM, x = InternalMatrix,
y = intTrainClass, tuning = tuning, cost = svmCost, splits = mySplits,
names = c("trialMatrix", "activityAssignment"), ncores = 2)

# inspect the tuning parameter
secondTuning

# note that the following is interactive and the lambda parameter must be edited
tunedSVM <- oneNormSVM(InternalMatrix, intTrainClass, 7e-07)

mySlack <- tunedSVM$margin
myPenalty <- tunedSVM$hyperplane

tunedAccuracy <- internalAccuracy(InternalMatrix, intTrainClass)
tunedAccuracy

# perform external accuracy calculation
extPrediction <- predict(tunedSVM, ExtTesting)
confusionMatrix(extPrediction, extTestClass, positive = "1")

# calculate Matthews Correlation Coefficient
# note that this is interactive, replace with values from the confusion matrix
MCC(8,2,4,5)

# retrieve the model set compound indices
subsetRange <- seq(from = 1, to = nrow(Acs))
subsetSelect <- subsetRange[modelSets != 0]

# retrieve the prototype conformers
conformerRange <- subsetSelect[myPenalty != 0]

# write to a file for retrieval
write.csv(conformerRange, file="prototypes.csv", quote = FALSE,row.names = FALSE)
APPENDIX E. ADDITIONAL SCRIPTS
E.1 pubmed_cids.py

The following script was developed to retrieve a SMILES file containing known PubChem compounds that have been curated by PubMed IDs. This script takes advantage of the PubChem Power User Gateway.

```
pubmed_cids.py

Retrieves a SMILES file containing known PubChem compounds referenced back to a PubMed ID.

author      David Watson
email       dewatson@icloud.com
copyright    Copyright (c) 2013, David Watson

#!/usr/bin/env python
import sys
import argparse
import urllib
import nltk
import xml.dom.minidom
import time

parser = argparse.ArgumentParser(
    description='Resolves compound lists in SMILES format based upon associated PubMed IDs through PubChem.')
parser.add_argument('-i', metavar='infile', type=argparse.FileType('r'),
    help='PubMed UID file', required=True, dest='pmid')
args = parser.parse_args()


PUG_WAITING_HEAD = """
    <PCT-Data>
        <PCT-Data_input>
            <PCT-InputData>
                <PCT-InputData_request>
                    <PCT-Request>
                    """

PUG_WAITING_TAIL = """
    <PCT-Request_type value="status"/>
"""
```
def processUidFile():
    try:
        for pmid in args.pmid.readlines():
            print "Processing " + pmid
            cids = getCIDsFromUID(pmid.strip())
            if len(cids) > 0:
                processCids(pmid, cids)
    except StopIteration:
        sys.exit("PubMed UID file appears to be empty")
def getCIDsFromUID(UID):
    params = urllib.urlencode({'Db': 'pccompound',
                               'DbFrom': 'pubmed', 'Cmd': 'Link',
                               'LinkName': 'pubmed_pccompound_mesh', 'format': 'text',
                               'report': 'uilist', 'IdsFromResult': UID})
    pubcrawl = urllib.urlopen(PUBMED, params).read()
    raw = nltk.clean_html(pubcrawl)
    tokens = nltk.word_tokenize(raw)
    return tokens

def processCids(smilesName, cidlist):
    querystring = ""
    for cid in cidlist:
        querystring += "<PCT-ID-List_uids_E>
                        + cid + "</PCT-ID-List_uids_E>" + "\n"
    pubquery = PUG_DL_HEAD + querystring + PUG_DL_TAIL
    pub = urllib.urlopen(PUG, pubquery).read()
    pubdom = xml.dom.minidom.parseString(pub)
    handleResponse(smilesName, pubdom)

def getText(nodelist):
    rc = ""
    for node in nodelist:
        if node.nodeType == node.TEXT_NODE:
            rc = rc + node.data
    return rc

def handleResponse(smilesName, dom):
    waitingXml = dom.getElementsByTagName("PCT-Waiting_reqid")
    if waitingXml.length > 0:
        for wait in waitingXml:
            waiturl = getText(wait.childNodes)
            handleWait(smilesName, waiturl)
    else:
        downloadXml = dom.getElementsByTagName("PCT-Download-URL_url")
        print "Received download URL for " + smilesName
        processDownload(smilesName, downloadXml)

def handleWait(smilesName, waitReq):
    print "Waiting on response for job " + smilesName
    waitquery = "<PCT-Request_reqid>" + waitReq + "</PCT-Request_reqid>" + "\n"
    PCT_QUERY = PUG_WAITING_HEAD + waitquery + PUG_WAITING_TAIL
    time.sleep(3)
    newquery = urllib.urlopen(PUG, PCT_QUERY).read()
    newdom = xml.dom.minidom.parseString(newquery)
    handleResponse(smilesName, newdom)
```python
def processDownload(smilesName, downloadXml):
    dlurl = ""
    for node in downloadXml:
        dlurl = getText(node.childNodes)
    smilesFile = smilesName.rstrip() + ".smi"
    urllib.urlretrieve(dlurl, smilesFile)

processUidFile()
```

### E.2  `tanimoto_cluster.py`

The following script clusters compounds based on a user-specified Tanimoto similarity cutoff. A Canvas distance matrix file is required, which should have been generated using Tanimoto similarity measures. The output is a Cytoscape network file that may be useful for visualizing the clusters in the network.

```bash
#!/usr/bin/env python

# `tanimoto_cluster.py`

Cluster compounds based on a specified Tanimoto similarity cutoff

author    David Watson
email     dewatson@icloud.com
copyright Copyright (c) 2013, David Watson

import sys
import argparse

parser = argparse.ArgumentParser(
    description='Convert a Canvas distance matrix to Cytoscape network.
)
parser.add_argument('-i', metavar='infile', type=argparse.FileType('r'),
    help='Canvas distance matrix', required=True, dest='canvas')
parser.add_argument('-o', metavar='outfile', type=argparse.FileType('w'),
    help='Cytoscape network', required=True, dest='cytoscape')
parser.add_argument('-s', metavar='similarity', type=float,
    help='Tanimoto similarity cutoff (default: 0.70)',
    dest='similarity', default=0.70)
args = parser.parse_args()

cytoscape = '%%s %s %s\n'

try:
    compounds = next(args.canvas).rstrip().split(',

except StopIteration:
    sys.exit("Canvas input file does not contain header information")
```

137
if len(compounds) <= 1 or compounds[0] != 'canvas':
sys.exit("The Canvas input file does not appear to be valid")

line_index=0
compound_index=1

while True:
    line_index+=1
    compound_index=line_index+1
    if (line_index == len(compounds)):
        break

try:
    distances = next(args.canvas).rstrip().split(',,')
except StopIteration:
    sys.exit("The input file appears corrupt")

while compound_index < len(compounds):
    if (args.similarity <= float(distances[compound_index])):
        args.cytoscape.write(cytoscape % (compounds[line_index],
                                           compounds[compound_index], distances[compound_index]))
    compound_index+=1
APPENDIX F. BINDING AFFINITY DATA SETS
## APPENDIX F. BINDING AFFINITY DATA SETS

Table F.1: Sigma 1: PTZ guinea pig brain dataset

<table>
<thead>
<tr>
<th>SMILES</th>
<th>Name</th>
<th>pKᵢ</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1cc1c1C[Ce@H]1CCNCCc2cccccc2</td>
<td>140</td>
<td>100</td>
<td>9.444</td>
</tr>
<tr>
<td>c1ccccc1CCNCCc2cccccc2</td>
<td>100</td>
<td>9.938</td>
<td></td>
</tr>
<tr>
<td>c1ccccc1C[Ce@H]1(N+)1CCNCCc2cccccc2</td>
<td>100</td>
<td>9.938</td>
<td></td>
</tr>
<tr>
<td>C1CCN1CCCCc2cccccc2</td>
<td>100</td>
<td>9.938</td>
<td></td>
</tr>
<tr>
<td>C1CCN1CCCC(C2)CCCc3ccccc3</td>
<td>100</td>
<td>9.938</td>
<td></td>
</tr>
<tr>
<td>C1CCN1CCCC(C2)CCCc3ccccc3</td>
<td>100</td>
<td>9.938</td>
<td></td>
</tr>
<tr>
<td>Chemical Structure</td>
<td>Name</td>
<td>MW</td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>-----------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>NC=CN(C)CCCc1cc(c12)Sc3c(cccc3)N2CCCN(C)C</td>
<td>Trifluopromazine</td>
<td>611.0</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C2)[C@H][C@H][C@H]3Cc(c4[@]23C)cccc(c4)O</td>
<td>(+)-alpha-Cyclazocine</td>
<td>772.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C2)[C@H][C@H][C@H]3Cc(c4[@]23C)cccc(c4)O</td>
<td>(-)-alpha-Cyclazocine</td>
<td>5999.4</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C2)[C@H][C@H][C@H]3Cc(c4[@]23C)cccc(c4)O</td>
<td>(+)-beta-Cyclazocine</td>
<td>5131.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C2)[C@H][C@H][C@H]3Cc(c4[@]23C)cccc(c4)O</td>
<td>(-)-beta-Cyclazocine</td>
<td>4544.1</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN(CC1)<a href="N(C)CC2">C@@H</a>Cc(c3[@]12C)ccc(c3)O (+)</td>
<td>(+)-alpha-Pentazocine</td>
<td>8678.1</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN(CC1)<a href="N(C)CC2">C@@H</a>Cc(c3[@]12C)ccc(c3)O (-)</td>
<td>(-)-alpha-Pentazocine</td>
<td>6951.0</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN(CC1)<a href="N(C)CC2">C@@H</a>Cc(c3[@]12C)ccc(c3)O (+)</td>
<td>(+)-beta-Pentazocine</td>
<td>5157.0</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN(CC1)<a href="N(C)CC2">C@@H</a>Cc(c3[@]12C)ccc(c3)O (-)</td>
<td>(-)-beta-Pentazocine</td>
<td>4571.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>Meperidine</td>
<td>5763.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>Morphine</td>
<td>5004.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>Phencyclidine</td>
<td>5857.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>(U-50,488H)</td>
<td>5730.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>Aminocaramiphen</td>
<td>7167.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>Iodocaramiphen</td>
<td>8750.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>(-)-Zenazocine</td>
<td>5484.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>(+)-SKF-10,047</td>
<td>6746.1</td>
<td></td>
</tr>
<tr>
<td>C1CC1CN(C)CCC1(C(=O)OCC)c2ccccc2</td>
<td>GBR-12909</td>
<td>7685.1</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** The table above lists various chemical structures and their corresponding names, along with their molecular weights (MW). Each structure is followed by its name and molecular weight, providing a clear and organized representation of the information.
c1ccccc1[C@@H](C(=O)OC)[C@@H](C2)CN(C3)C[H](C=C@H)4C)(Cc(c5[C@@H]34)Cccc(c5)O

(1S,2R)-6a

C=CCCN(C1)C[H](C=C@H)2C\Cc(c3[C@H](12)ccc(c3)C)O

(+)-SKF10047

1.560 127


c1ccccc1[(N+)[(O-)]=O)c(Cl)cc1CCN[C[C@@H]2(C[C@H](12)ccc(c3)O) (+)-pentazocine

5.860 127

C=CCN1CCN(C[C@H]2(C[C@H]3)CCSC[C[C@H]2(C[C@H]3)CCCC3

(R,S)-22 7.396 129

C=CCN1CCN(C[C@H]2(C[C@H]3)CCSC[C[C@H]2(C[C@H]3)CCCC3

(S,R)-22 7.092 129

COc(c1)ccc1CC(=O)C\Cc(c2ccc(F)cc2)CCCN(C[C@@H]2(C[C@H]3)CCCC3 1

6.582 130

COc(c1)ccc1CC(=O)C\Cc(c2ccc(F)cc2)CCCN(C[C@@H]2(C[C@H]3)CCCC3 2

6.119 130

C1CC1CN(CC2)CCC2O\Cc(c3)[N+]\[O-]=O)j

FCCN(C1)CC(C2)CC2Oc2cc(F)j\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

8\alpha-b 6.897 130

C1CC1CN(CC2)CCC2O\Cc(c3)[N+]\[O-]=O)j

FCCN(C1)CC(C2)CC2Oc2cc(F)j\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

8\beta-b 6.497 130

DTG 7.553 133

N\Cccccc1(CN(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

3 6.838 134

N\Cccccc1(CN(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

2 9.174 136

N\Cccccc1(CN(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

4 8.914 136

N\Cccccc1(CN(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

3 7.614 137

N\Cccccc1(CN(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

3 9.119 137

N\Cccccc1(CN(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

5 8.695 137

N\Cccccc1(CN(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

6 8.034 137

N\Cccccc1(CN(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

8 9.821 137

FCCCN(C1)CC(C2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

8 9.076 137

C1CC1CN(CC2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

10 9.301 137

C1CC1CN(CC2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

11 8.538 137

C1CC1CN(CC2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

12 7.979 137

C1CC1CN(CC2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

13 9.420 137

C1CC1CN(CC2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

14 9.222 137

C1CC1CN(CC2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

15 9.056 137

C1CC1CN(CC2)CC2OC\Cc(c3)Cj\Cc(c2)CCCN(C[C@H]2(C[C@H]3)CCCC3

16 8.770 137
CC(C)(C)Oc(c1)ccc(c1)cnH1(c12)CcH1(4)CcH1)SCG1(C)CCCCN(2CC(2)CC)CCc1Fcc(F)cc3 haloperidol 5.313 163
CC(1)C[ CMS(C)C]C(1)C[ CMS(C)C]CCCCN(2C)CCc1Fcc(F)cc3 10-t-butoxy-ibogamine 6.907 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
Clccc1CCCC2[C@H1]C(CN[C@@H1]2)CCCN(Cccc(F)ccc3)ccc4Ccc(F)cc4
Clccc1Fcc(F)cc(c12)CCCN(C)C[ CMS(C)C]CCCCN(2CC(2)CC)CCc1Fcc(F)cc3
d-tg 7.488 59
[CC@]123c4c5(O)cocc4[CC@H](NCC3)Cc3[CC@H]22CC(=O)[CC@H]1O5 (+)-nordihydrocodeinone 4.777 212
clccc1(C)c1cCc[C@H][C@H]22Cc3c4c5c6c7c8c9ccccc3 1.2 LRI132 8.699 213
c1ccc1(C)c1CNCN2CC2CC 2.0 BD1060 8.523 214
CICCCN1CC(C)=CC2cc(c(C)c1Cc2 2.1 BD1052 8.699 214
c1ccc1(C)c1CNCN[CC(2)C]CC2CCCCC2 2.2 BD1067 8.699 214
clccc1(C)c1CCNCN2CC2CC 2.3 UMB98 7.602 215
CICCCN1CC(C)=CC2CCc(C)c1Cc2 2.4 UMB99 7.796 215
C0c(c1)cc1CNC[CC(2)C]CC2CCCCC2 2.5 UMB100 7.620 215
clccc1(C)c1CCNCN2CC2CC 2.6 UMB101 7.495 215
clccc1(C)c1CCNCN2CC2CC 2.7 UMB102 7.602 215
C0c(c1)cc1CNC[CC(2)C]CC2CCCCC2 2.8 UMB103 7.678 215
clccc1(C)c1CCNCN2CC2CC 2.9 CM777 6.447 219

158
c1c(Br)cc(OC)c1C(=O)NCCCCN(C2)CCc(c3)c2cc(c3)OCO4
2. 7.085 253

c1c(Br)cc(OC)c1C(=O)NCCCCN(C2)c(c3)c3cc4c(c3)OCO4
3. 6.471 253

c1c(Br)cc(OC)c1C(=O)NCCCN(C2)Cc(c3)c3cc4c(c3)OCO4
4. 5.845 253

c1c(Br)cc(OC)c1C(=O)NCCCCNCC2ccc(CC)cc6
5. 6.056 253

c1c(OC)cc6NCC2CCN(c3)ccc4c(c3)CCO4
2a. 5.570 254

c1c(OC)cc6NCC2CCN(C3)ccc4c(c3)CCO4
3a. 6.431 254

c1c(OC)cc6NCC3CCN(C3)ccc4c(c3)CCO4
4a. 6.578 254

Haloperidol
8.036 255

CCCc(cc1)cc(c12)sc(=O)n2CCN3CCCCCC3
9. 8.432 255

CC(C)(C)OC(=O)NCCC(=O)N1CCN(CC1)CCCc2ccccc2
10. 7.066 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
11. 8.387 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
12. 6.830 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
13. 8.699 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
14. 8.387 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
15. 8.409 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
16. 8.854 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
17. 8.886 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
18. 8.584 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
19. 8.534 255

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
20. 5.834 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
21. 6.839 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
22. 8.276 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
23. 8.086 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
24. 7.066 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
25. 9.495 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
26. 8.387 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
27. 8.830 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
28. 6.695 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
29. 8.585 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
30. 6.742 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
31. 6.725 257

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
32. 7.542 258

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
33. 7.860 258

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
34. 8.939 258

NCCCC(=O)N1CCN(CC1)CCCc2ccccc2
35. 6.644 258
<table>
<thead>
<tr>
<th>Molecular Structure</th>
<th>Energy (kcal/mol)</th>
<th>Formation (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1CCCN1CCCN(C)CCN(C)Cc2cc(Cl)c(Cl)cc2</td>
<td>8.097</td>
<td>258</td>
</tr>
<tr>
<td>C1CCCN1CCN(C)CCN(C)CCN2CCCC2</td>
<td>7.111</td>
<td>258</td>
</tr>
<tr>
<td>C1CCCN1CCCN(C)CCN(C)CCN2CCCC2</td>
<td>6.893</td>
<td>258</td>
</tr>
<tr>
<td>C1CCCN1CCCN(C)CCN(C)CCN2CCCC2</td>
<td>5.950</td>
<td>258</td>
</tr>
<tr>
<td>C1CCCN1CCCN(C)CCN(C)CCN2CCCC2</td>
<td>7.390</td>
<td>258</td>
</tr>
<tr>
<td>C1CCCN1CCCN(C)CCN(C)CCN2CCCC2</td>
<td>6.409</td>
<td>258</td>
</tr>
<tr>
<td>C1CCCN1CCCN(C)CCN(C)CCN2CCCC2</td>
<td>6.614</td>
<td>258</td>
</tr>
<tr>
<td>C1CCCN1CCCN(C)CCN(C)CCN2CCCC2</td>
<td>7.268</td>
<td>258</td>
</tr>
<tr>
<td>SMILES</td>
<td>Name</td>
<td>pK_a</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------------</td>
<td>------</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-8</td>
<td>7.279</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-9</td>
<td>7.620</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-10</td>
<td>7.284</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-4</td>
<td>7.666</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-6</td>
<td>7.418</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-7</td>
<td>7.650</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-5</td>
<td>8.194</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-4</td>
<td>7.090</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-5</td>
<td>7.577</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-6</td>
<td>7.334</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-7</td>
<td>6.910</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-1</td>
<td>4.602</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-2d</td>
<td>6.812</td>
</tr>
<tr>
<td>CB-64D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CB-182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-10c</td>
<td>6.069</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (+)-10a</td>
<td>6.025</td>
<td></td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-2c</td>
<td>6.217</td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-10b</td>
<td>6.175</td>
<td></td>
</tr>
<tr>
<td>PR0084</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(-)-1d</td>
<td>6.827</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(-)-2a</td>
<td>6.883</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(-)-2b</td>
<td>6.939</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(-)-2c</td>
<td>6.321</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(-)-2c</td>
<td>6.876</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(-)-2c</td>
<td>6.025</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-10a</td>
<td>6.217</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-10b</td>
<td>6.120</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-10b</td>
<td>6.071</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-10c</td>
<td>6.069</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)-10c</td>
<td>6.105</td>
</tr>
<tr>
<td>C(C)=O[Nc1ccccccc1C=C(<a href="C">C@@H</a>23c4cc(OC)ccc4]</td>
<td>(+)-3a</td>
<td>3.851</td>
</tr>
<tr>
<td>c1ccccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3a</td>
<td>3.587</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3e</td>
<td>5.658</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3c</td>
<td>6.588</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3g</td>
<td>6.300</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3b</td>
<td>6.080</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3f</td>
<td>6.368</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3f</td>
<td>6.767</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3d</td>
<td>5.721</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+2d</td>
<td>6.312</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3h</td>
<td>5.691</td>
</tr>
<tr>
<td>c1ccccc1C=C(<a href="C2">C@@H</a>N(C)(CC3)C)=O[C@H]23c4cc(OC)ccc4</td>
<td>(+)+3e</td>
<td>7.367</td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-2c</td>
<td>6.530</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-2b</td>
<td>6.530</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-2a</td>
<td>7.286</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-1f</td>
<td>6.717</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-1f</td>
<td>6.419</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-2d</td>
<td>6.368</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-2c</td>
<td>6.249</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-2b</td>
<td>6.161</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-2a</td>
<td>5.697</td>
<td></td>
</tr>
<tr>
<td>CC(C)=CCN([C@H]3C)C)c(c3C@H12C)c3ccccc3 (-)-3a</td>
<td>6.461</td>
<td></td>
</tr>
</tbody>
</table>
The image contains a page of a document with chemical structures and their corresponding annotations. The text appears to be part of a scientific or medicinal context, likely discussing chemical compounds and their properties. The structures are complex and include various functional groups and modifications, such as bromine (Br) and iodine (I) substitutions. The annotations include descriptions like "haloperidol" and "(+)-7a", indicating specific compounds or modifications. The page seems to be part of a larger document discussing chemical substances, possibly in the context of pharmacology or organic chemistry.
c1ccccc1CCCN2CCN(CC2)CCc3ccc(N)cc3 14 7.854 257

CCc3ccc(OC)cccc3 15 8.125 257

c1ccccc1CCCN2CCN(CC2)CCc3cc(OC)ccc3 16 7.991 257

c1ccccc1CCCN2CCN(CC2)CCc3ccc(cc3)OC 17 7.544 257

COc(cc1)ccc1CCN2CCN(CC2)C(=O)Cc3cc(OC)ccc3 19 7.928 257

Clc3ccc1CCCN2CCN(CC2)CCc3cc(OC)ccc3 20 6.128 257

Clc3ccc1CCCN2CCN(CC2)CCc3cc(OC)ccc3 21 7.770 257

Clc3ccc1CCCN2CCN(CC2)CCc3cc(OC)ccc3 22 7.538 257

Cc3ccc1CCCN2CCN(CC2)CCc3ccc(OCCccc3 23 7.928 257

CC(C)(C)OC(=O)NCC(=O)N1CCN(CC1)CCc2ccccc2 24 6.757 257

CC(C)(C)OC(=O)NCC(=O)N1CCN(CC1)CCc2ccccc2 25 5.341 257

CC(C)(C)OC(=O)NCC(=O)N1CCN(CC1)CCc2ccccc2 26 6.900 257

NCC(=O)N1CCN(CC1)CCc2ccccc2 27 6.519 257

NCC(=O)N1CCN(CC1)CCc2ccccc2 28 5.749 257

NCC(=O)N1CCN(CC1)CCc2ccccc2 29 5.416 257

c1ccccc1CCCN2CCN(CC2)CCN 30 6.726 257

N1CCN(CC1)CCc2ccccc2 31 5.699 257

N1CCN(CC1)CCc2ccccc2 32 6.000 257

c1cc(Cl)c(Cl)c1CCN2CCC2 3 7.111 258

c1cc(Cl)c(Cl)c1CCN(C)CCN(C)CCN2CC2 4 7.842 258

c1cc(Cl)c(Cl)c1CCN(C)CCN(C)CCN2CC2 5 7.271 258

c1cc(Cl)c(Cl)c1CCN(C)CCN(C)CCN(C)CCN2CC2 6 7.690 258

C1CCCN1CCN(C)CCN(C)CCN(C)CCN2CC2 7 7.738 258

C1CCCN1CCN(C)CCN(C)CCN(C)CCN2CC2 8 7.738 258

C1CCCN1CCN(C)CCN(C)CCN(C)CCN2CC2 9 7.279 258

C1CCCN1CCN(C)CCN(C)CCN(C)CCN2CC2 10 7.917 258

C1cc(Cl)c(Cl)c1CCN(C)CCN(C)CCN2CC2 11 7.827 258

C1cc(Cl)c(Cl)c1CCN(C)CCN(C)CCN2CC2 12 6.818 258

C1cc(Cl)c(Cl)c1CCN(C)CCN(C)CCN2CC2 13 6.939 258

C1cc(Cl)c(Cl)c1CCN(C)CCN2CC2 14 6.870 258

NCCNCCc1cc(Cl)c(Cl)c1 15 5.626 258

C1cc(Cl)c(Cl)c1CCNCCCN 16 6.270 258

C1cc(Cl)c(Cl)c1CCNCCCN2CC2 17 6.955 258
Table F.3: Sigma 2: DTG/PTZ rat brain dataset

<table>
<thead>
<tr>
<th>SMILES</th>
<th>Name</th>
<th>pKᵢ</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>C(=O)c1ccccc1CCN(CC)CCN2CCCCC2</td>
<td>Fluphenazine</td>
<td>6.682</td>
<td>265</td>
</tr>
<tr>
<td>C(=O)c1ccccc1CCN(CC)CCN2CCCCC2</td>
<td>Fluphenazine</td>
<td>6.702</td>
<td>265</td>
</tr>
<tr>
<td>C(=O)c1ccccc1CCN(CC)CCN2CCCCC2</td>
<td>Fluphenazine</td>
<td>6.585</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>5.935</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>5.987</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.086</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.237</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.260</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.398</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.367</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.509</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.824</td>
<td>265</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>5.887</td>
<td>130</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.055</td>
<td>130</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>6.730</td>
<td>131</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>6.808</td>
<td>131</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>6.833</td>
<td>131</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>6.632</td>
<td>131</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>6.859</td>
<td>131</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>7.104</td>
<td>131</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>5.516</td>
<td>132</td>
</tr>
<tr>
<td>c1ccc(c1)CCN2CCOC</td>
<td>Trifluperidol</td>
<td>6.569</td>
<td>132</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>5.276</td>
<td>132</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>1.625</td>
<td>266</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>2.068</td>
<td>266</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>3.959</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>4.853</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>5.841</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>9.353</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>2.953</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>7.968</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>8.925</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>9.816</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>9.816</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>10.915</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>11.602</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>12.481</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>13.114</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>1.777</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>2.718</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>3.793</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>4.768</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>4.786</td>
<td>267</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>5.623</td>
<td>268</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.658</td>
<td>268</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.365</td>
<td>268</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.461</td>
<td>268</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>5.194</td>
<td>268</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.580</td>
<td>269</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.580</td>
<td>269</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>7.252</td>
<td>269</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>8.566</td>
<td>269</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.496</td>
<td>6</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.572</td>
<td>6</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>9.535</td>
<td>6</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>5.593</td>
<td>6</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>5.269</td>
<td>6</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>9.737</td>
<td>6</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.917</td>
<td>6</td>
</tr>
<tr>
<td>C1CCN(c1)C=C=O</td>
<td>Risperidone</td>
<td>6.682</td>
<td>6</td>
</tr>
</tbody>
</table>
Table F.4: Sigma 2: DTG/PTZ rat liver dataset

<table>
<thead>
<tr>
<th>SMILES</th>
<th>Name</th>
<th>pk_{i}</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>9</td>
<td>5.991</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>10</td>
<td>6.152</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>11</td>
<td>6.083</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>12</td>
<td>6.432</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>13</td>
<td>6.371</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>14</td>
<td>6.602</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>15</td>
<td>6.697</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>23</td>
<td>7.022</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>24</td>
<td>6.879</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>25</td>
<td>7.046</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>26</td>
<td>7.194</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>27</td>
<td>7.268</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>28</td>
<td>7.398</td>
<td>103</td>
</tr>
<tr>
<td>c1cccc1C@H)<a href="C3">C@H</a>C(C@H)[C@H][C@H]C4[C@H]C[CC]24O</td>
<td>29</td>
<td>7.469</td>
<td>103</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>R-4</td>
<td>7.708</td>
<td>70</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>S-4</td>
<td>7.747</td>
<td>70</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>R-10</td>
<td>7.592</td>
<td>70</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>S-10</td>
<td>7.730</td>
<td>70</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>R-11</td>
<td>8.036</td>
<td>70</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>S-11</td>
<td>7.801</td>
<td>70</td>
</tr>
<tr>
<td>CCN1C@H]2(C2)CCCC2</td>
<td>13</td>
<td>6.457</td>
<td>70</td>
</tr>
<tr>
<td>C1CCCCC1C@H](C2)(C3)C4CCCCC3</td>
<td>12</td>
<td>8.328</td>
<td>70</td>
</tr>
<tr>
<td>C1CCCCC1C@H](C2)(C3)C4CCCCC3</td>
<td>24</td>
<td>7.939</td>
<td>70</td>
</tr>
<tr>
<td>C1CCCCC1C@H](C2)(C3)C4CCCCC3</td>
<td>26</td>
<td>6.697</td>
<td>70</td>
</tr>
<tr>
<td>c1cccc1NC=O]%N]c(C2)cccc2</td>
<td>dtg</td>
<td>7.790</td>
<td>143</td>
</tr>
<tr>
<td>c1cccc1NC=O]%N]c(C2)cccc2</td>
<td>cis-7</td>
<td>8.437</td>
<td>143</td>
</tr>
<tr>
<td>c1cccc1NC=O]%N]c(C2)cccc2</td>
<td>trans-7</td>
<td>8.321</td>
<td>143</td>
</tr>
<tr>
<td>C0c1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>cis-8</td>
<td>8.799</td>
<td>143</td>
</tr>
<tr>
<td>C0c1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>trans-8</td>
<td>8.917</td>
<td>143</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>cis-9</td>
<td>8.688</td>
<td>143</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>trans-9</td>
<td>8.783</td>
<td>143</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>cis-10</td>
<td>8.162</td>
<td>143</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>cis-11</td>
<td>8.101</td>
<td>143</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>trans-11</td>
<td>9.678</td>
<td>143</td>
</tr>
<tr>
<td>Fc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>cis-12</td>
<td>7.851</td>
<td>143</td>
</tr>
<tr>
<td>Fc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>trans-12</td>
<td>8.714</td>
<td>143</td>
</tr>
<tr>
<td>c1cccc1Cl=O]%N]c(C2)cccc2</td>
<td>cis-13</td>
<td>7.676</td>
<td>143</td>
</tr>
<tr>
<td>c1cccc1Cl=O]%N]c(C2)cccc2</td>
<td>trans-13</td>
<td>8.305</td>
<td>143</td>
</tr>
<tr>
<td>c1cccc1Cl=O]%N]c(C2)cccc2</td>
<td>cis-14</td>
<td>8.780</td>
<td>143</td>
</tr>
<tr>
<td>c1cccc1Cl=O]%N]c(C2)cccc2</td>
<td>trans-14</td>
<td>8.783</td>
<td>143</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>R-9</td>
<td>7.996</td>
<td>144</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>S-9</td>
<td>8.228</td>
<td>144</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>R-11</td>
<td>7.842</td>
<td>144</td>
</tr>
<tr>
<td>COc1cccc1C@H]<a href="C3">C@H</a>C(NCCN(C3)C4CCCCC4</td>
<td>S-11</td>
<td>8.070</td>
<td>144</td>
</tr>
<tr>
<td>C1CCCN1(C2)C2NCCO<a href="CC3">C@H</a>c3cccc2</td>
<td>R-14</td>
<td>7.955</td>
<td>144</td>
</tr>
<tr>
<td>C1CCCN1(C2)C2NCCO<a href="CC3">C@H</a>c3cccc2</td>
<td>S-14</td>
<td>8.327</td>
<td>144</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>20</td>
<td>7.635</td>
<td>144</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>21</td>
<td>7.733</td>
<td>144</td>
</tr>
<tr>
<td>C1CCCN1(C2)C2NCCO[3c]ccc4</td>
<td>22</td>
<td>8.620</td>
<td>144</td>
</tr>
<tr>
<td>C1CCCN1(C2)C2NCCO[3c]ccc4</td>
<td>2</td>
<td>8.061</td>
<td>145</td>
</tr>
<tr>
<td>C1CCCN1(C2)C2NCCO[3c]ccc4</td>
<td>11a</td>
<td>8.580</td>
<td>145</td>
</tr>
<tr>
<td>C1CCCN1(C2)C2NCCO[3c]ccc4</td>
<td>11b</td>
<td>7.583</td>
<td>145</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>12a</td>
<td>7.889</td>
<td>145</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>12b</td>
<td>7.674</td>
<td>145</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>13a</td>
<td>7.293</td>
<td>145</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>13b</td>
<td>6.331</td>
<td>145</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>15a</td>
<td>8.315</td>
<td>145</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>15b</td>
<td>7.575</td>
<td>145</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>dtg</td>
<td>7.580</td>
<td>145</td>
</tr>
<tr>
<td>c1cccc1OC=O]%N]c(C2)cccc2</td>
<td>dtg</td>
<td>7.609</td>
<td>146</td>
</tr>
<tr>
<td>COCN1(C@H)(10)C2CN1C2cccc2</td>
<td>3b</td>
<td>6.290</td>
<td>149</td>
</tr>
<tr>
<td>COCN1(C@H)(10)C2CN1C2cccc2</td>
<td>3c</td>
<td>6.752</td>
<td>149</td>
</tr>
<tr>
<td>COCN1(C@H)(10)C2CN1C2cccc2</td>
<td>3d</td>
<td>7.153</td>
<td>149</td>
</tr>
<tr>
<td>COCN1(C@H)(10)C2CN1C2cccc2</td>
<td>3e</td>
<td>6.879</td>
<td>149</td>
</tr>
<tr>
<td>COCN1(C@H)(10)C2CN1C2cccc2</td>
<td>3f</td>
<td>6.967</td>
<td>149</td>
</tr>
<tr>
<td>COCN1(C@H)(10)C2CN1C2cccc2</td>
<td>4a</td>
<td>6.427</td>
<td>151</td>
</tr>
</tbody>
</table>
[Data and information from the image]
<table>
<thead>
<tr>
<th>SMILES</th>
<th>Name</th>
<th>$pK_i$</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMILES</td>
<td>Name</td>
<td>$pK_i$</td>
<td></td>
</tr>
<tr>
<td>c1cccc1CCCN(C2)CC2Cc3cccc3</td>
<td>5</td>
<td>8.509</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c1cccc1CCCN(C2)CC2Cc3cccc3</td>
<td>7</td>
<td>8.553</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>8</td>
<td>7.102</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>9</td>
<td>8.155</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>10</td>
<td>7.013</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>11</td>
<td>7.301</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>12</td>
<td>4.264</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>13</td>
<td>6.081</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>14</td>
<td>6.256</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>15</td>
<td>7.201</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>16</td>
<td>7.155</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>17</td>
<td>7.051</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>18</td>
<td>5.101</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>19</td>
<td>6.015</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>20</td>
<td>5.014</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>21</td>
<td>6.502</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>22</td>
<td>5.048</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>23</td>
<td>5.235</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>24</td>
<td>5.672</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>25</td>
<td>5.667</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>26</td>
<td>5.841</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>101</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>27</td>
<td>7.215</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>28</td>
<td>7.420</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>29</td>
<td>7.319</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>30</td>
<td>7.444</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>31</td>
<td>7.000</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>32</td>
<td>7.721</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>33</td>
<td>7.194</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>34</td>
<td>6.620</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>35</td>
<td>7.398</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>36</td>
<td>6.710</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>37</td>
<td>6.456</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>38</td>
<td>7.854</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>39</td>
<td>8.444</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>40</td>
<td>8.886</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>41</td>
<td>6.125</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>42</td>
<td>8.509</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>43</td>
<td>7.481</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>44</td>
<td>8.796</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>45</td>
<td>7.602</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>46</td>
<td>9.000</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>42</td>
<td>8.678</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>43</td>
<td>7.959</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>44</td>
<td>6.629</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>45</td>
<td>7.959</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>46</td>
<td>8.086</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>47</td>
<td>7.015</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>48</td>
<td>8.786</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>49</td>
<td>6.538</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>50</td>
<td>7.921</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>51</td>
<td>6.827</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C1CN(C)CC2Cc3cccc3</td>
<td>52</td>
<td>7.699</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c1cccc1CCCCN(CC2)CC2Cc3cccc3</td>
<td>53</td>
<td>6.205</td>
<td></td>
</tr>
<tr>
<td>53</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c1cccc1CCCCN(CC2)CC2Cc3cccc3</td>
<td>54</td>
<td>8.073</td>
<td></td>
</tr>
<tr>
<td>54</td>
<td>298</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c1cccc1CCCCN(CC2)CC2Cc3cccc3</td>
<td>55</td>
<td>8.084</td>
<td></td>
</tr>
<tr>
<td>55</td>
<td>298</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX G. SIGMA-2 TRAINING/TEST SETS
title: Abate:2011p4733:11a
pKi: 8.58
model set: 2

title: Xu:2005p8:Ifenprodil
pKi: 8.569
model set: 1

title: Wirt:2007p462:(1S)-15c
pKi: 8.553
model set: 1

---

title: Mach:2003p380:4
pKi: 8.538
model set: 1

---

title: MouithysMickalad:2002p1149:10
pKi: 8.523
model set: 1

---

title: Mach:2001p339:3n
pKi: 8.509
model set: 2

---

title: Efange:1997p3905:24b
pKi: 8.481
model set: 1

---

title: Wirt:2007p462:(1R)-14b
pKi: 8.444
model set: 1
Vangveravong:2010p5291:7
pKi: 5.487
model set: 1

Holl:2009p1445:ent-14
pKi: 5.444
model set: 2

Holl:2009p1445:16
pKi: 5.423
model set: 1

Xu:2005p8:3-
Tropanylindole-3-c...
pKi: 5.413
model set: 2

Ferorelli:2007p4648:26
pKi: 5.349
model set: 1

Fontanilla:2009p934:tryptamine
pKi: 5.309
model set: 1

Geiger:2010p4212:ent-20
pKi: 5.292
model set: 2

Tu:2011p1555:20e
pKi: 5.225
model set: 1

Wiedemeyer:2006p2321:29alpha
pKi: 5.213
model set: 1

Vangveravong:2011p3502:10
pKi: 5.194
model set: 1

Chu:2005p77:13d
pKi: 5.151
model set: 1

Chu:2005p77:13g
pKi: 5.142
model set: 1
title: Holl:2009p2126:ent-23a
pKi: 4.243
model set: 1

Hajipour:2011p7435:5i
pKi: 4.169
model set: 2

Hajipour:2011p7435:5g
pKi: 4.127
model set: 1

Ren:2009p1692:F-FBZA
pKi: 3.921
model set: 1

Hajipour:2010p4397:4
pKi: 3.898
model set: 2

Hajipour:2010p4397:16
pKi: 3.875
model set: 1

Hajipour:2010p4397:6
pKi: 3.627
model set: 2

Fontanilla:2009p934:tyramine
pKi: 3.217
model set: 1

Fontanilla:2008p7205:5
pKi: 2.272
model set: 1
APPENDIX H. SIGMA-1 TRAINING/TEST SETS
APPENDIX H. SIGMA-1 TRAINING/TEST SETS
title: Berardi:2005p8237:31
pKi: 10.523
model set: 1

pKi: 10.377
model set: 1

pKi: 10.347
model set: 2

Title: Hudkins:1994p1964:15
pKi: 10.301
model set: 1

Title: Ablordeppey:2000p2105:9
pKi: 10.155
model set: 1

Title: Akunne:1997p51:PD144418
pKi: 10.097
model set: 2

Title: Abate:2011p73:trans-9
pKi: 10.0
model set: 2

Title: Glennon:2004p2217:8
pKi: 9.886
model set: 2

Title: Berardi:2009p205:S-3
pKi: 9.886
model set: 1

Title: Glennon:2004p2217:9
pKi: 10.046
model set: 1

Title: Ablordeppey:1998p625:57
pKi: 10.05
model set: 1
I
SO
ON

pKi: 9.854
model set: 1

pKi: 9.745
model set: 1

pKi: 9.699
model set: 2

pKi: 9.602
model set: 2

pKi: 9.585
model set: 1

pKi: 9.569
model set: 1

pKi: 9.553
model set: 2

pKi: 9.456
model set: 2
Ablordeppey: 2002p2759:10b
pKi: 9.444
model set: 1

Nahas: 2008p755:3i
pKi: 9.432
model set: 1

Waterhouse: 1997p45:1
pKi: 9.42
model set: 2

Dence: 1997p333:2b
pKi: 9.42
model set: 1

Foster: 2003p749:4
pKi: 9.409
model set: 1

Schlager: 2011p6704:28d
pKi: 9.367
model set: 2

Vangveravong: 2006p6988:3b
pKi: 9.337
model set: 1

Foster: 2003p749:7
pKi: 9.337
model set: 1

Glennon: 2004p2217:16a
pKi: 9.319
model set: 2

Glennon: 1994p1214:34
pKi: 9.319
model set: 1

May: 1998p311:1a
pKi: 9.31
model set: 1

Waterhouse: 1997p1657:10
pKi: 9.301
model set: 2

206
Hanner:1996p8072: Corticosterone
pKi: 4.553
model set: 1

DeHavenHudkins:1992p371: (-)-be...
pKi: 4.544
model set: 1

Bertha:1994p3163: (-)-(1R,5R)-4
pKi: 4.498
model set: 2

Chu:2011p7568: 4a
pKi: 4.463
model set: 1

Linders:1993p2499: (+)-3
pKi: 4.228
model set: 1
APPENDIX I. COPYRIGHT PERMISSIONS RELEASES
APPENDIX I. COPYRIGHT PERMISSIONS RELEASES
May 15, 2012

David Watson
Medicinal Chemistry
University of Mississippi
425 Faser Hall
University, MS  38677

Email: dewatson@go.olemiss.edu

Dear Mr. Watson:

This is to grant you permission to reproduce the following figure in your thesis entitled “Quasi-comprehensive scaffold perception, pharmacophore development, and structure-affinity relationships of SIGMA site ligands” for the University of Mississippi:

Figure 12 from D T Manallack, M G Wong, M Costa, P R Andrews, and P M Beart, Receptor site topographies for phencyclidine-like and sigma drugs: predictions from quantitative conformational, electrostatic potential, and radioreceptor analyses, *Mol Pharmacol* December 1988 34:863-879

Permission to reproduce the figure is granted for worldwide use in all languages, translations, and editions, and in any format or medium including print and electronic. The authors and the source of the materials must be cited in full, including the article title, journal title, volume, year, and page numbers.

Sincerely yours,

Richard Dodenhoff
Journals Director
## European journal of pharmacology. Molecular pharmacology section

**Order detail ID:** 62467368  
**Article Title:** Heterocyclic amino alcohols related to ifenprodil as \( \sigma \) receptor ligands: binding and conformational analyses  
**Author(s):** Beart, Philip M. ; et al  
**DOI:** 10.1016/0922-4106(94)90086-8  
**Date:** Oct 01, 1994  
**ISSN:** 0922-4106  
**Publication Type:** Journal  
**Volume:** 269  
**Issue:** 2  
**Start page:** 193  
**Publisher:** ELSEVIER SCIENCE PUBLISHERS, 

**Permission Status:** Granted  
**Permission type:** Republish or display content  
**Type of use:** reuse in a thesis/dissertation  
**Order License Id:** 2910400866100  

### Order Details

<table>
<thead>
<tr>
<th><strong>Permission Status</strong></th>
<th><strong>Permission type</strong></th>
<th><strong>Type of use</strong></th>
<th><strong>Order License Id</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Granted</td>
<td>Republish or display content</td>
<td>reuse in a thesis/dissertation</td>
<td>2910400866100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Number of pages</strong></th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Portion</strong></td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td><strong>Number of figures/tables/illustrations</strong></td>
<td>1</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>both print and electronic</td>
</tr>
<tr>
<td><strong>Are you the author of this Elsevier article?</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Will you be translating?</strong></td>
<td>No</td>
</tr>
<tr>
<td><strong>Order reference number</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Title of your thesis/dissertation:** Quasi-comprehensive scaffold perception, pharmacophore development, and structure–affinity relationships of sigma site ligands  
**Expected completion date:** Jul 2012  
**Estimated size (number of pages):** 200  
**Elsevier VAT number:** GB 494 6272 12  
**Permissions price:** 0.00 USD  
**VAT/Local Sales Tax:** 0.00 USD / 0.0 GBP

**Note:** This item was invoiced separately through our RightsLink service. More info
## Customer Information
- **Customer:** David Watson  
- **Account Number:** 3000533869  
- **Organization:** David Watson  
- **Phone:** +1 (662)4738003  
- **Payment Method:** Invoice

## Order Details

### European journal of medicinal chemistry

<table>
<thead>
<tr>
<th>Order detail ID:</th>
<th>62467429</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Article Title:</strong></td>
<td>5D-QSAR for spirocyclic σ1 receptor ligands by Quasar receptor surface modeling</td>
</tr>
<tr>
<td><strong>Author(s):</strong></td>
<td>Oberdorf, Christoph ; Schmidt, Thomas J ; Wünsch, Bernhard</td>
</tr>
<tr>
<td><strong>DOI:</strong></td>
<td>10.1016/j.ejmech.2010.03.048</td>
</tr>
<tr>
<td><strong>Date:</strong></td>
<td>Jan 01, 2010</td>
</tr>
<tr>
<td><strong>ISSN:</strong></td>
<td>1768-3254</td>
</tr>
<tr>
<td><strong>Publication Type:</strong></td>
<td>e-Journal</td>
</tr>
<tr>
<td><strong>Volume:</strong></td>
<td>45</td>
</tr>
<tr>
<td><strong>Issue:</strong></td>
<td>7</td>
</tr>
<tr>
<td><strong>Start page:</strong></td>
<td>3116</td>
</tr>
<tr>
<td><strong>Publisher:</strong></td>
<td>Elsevier Science</td>
</tr>
</tbody>
</table>

**Permission Status:** Grant

**Permission type:** Republish or display content  
**Type of use:** reuse in a thesis/dissertation  
**Order License Id:** 2910431098505

| **Number of pages** | 9 |
| **Portion** | figures/tables/illustrations |
| **Number of figures/tables/illustrations** | 1 |
| **Format** | both print and electronic |

**Note:** This item was invoiced separately through our RightsLink service. More info

### Order Information
- **Total order items:** 1  
- **Order Total:** $0.00
Confirmation Number: 10998015  
Order Date: 05/15/2012

Customer Information
Customer: David Watson  
Account Number: 3000533869  
Organization: David Watson  
Email: dawatson@go.olemiss.edu  
Phone: +1 (662)4738003  
Payment Method: Invoice

Order Details

<table>
<thead>
<tr>
<th>Bioorganic &amp; medicinal chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Order detail ID:</strong></td>
</tr>
<tr>
<td><strong>Article Title:</strong></td>
</tr>
<tr>
<td><strong>Author(s):</strong></td>
</tr>
<tr>
<td><strong>DOI:</strong></td>
</tr>
<tr>
<td><strong>Date:</strong></td>
</tr>
<tr>
<td><strong>ISSN:</strong></td>
</tr>
<tr>
<td><strong>Publication Type:</strong></td>
</tr>
<tr>
<td><strong>Volume:</strong></td>
</tr>
<tr>
<td><strong>Issue:</strong></td>
</tr>
<tr>
<td><strong>Start page:</strong></td>
</tr>
<tr>
<td><strong>Publisher:</strong></td>
</tr>
<tr>
<td><strong>Permission Status:</strong></td>
</tr>
<tr>
<td><strong>Permission type:</strong></td>
</tr>
<tr>
<td><strong>Type of use:</strong></td>
</tr>
<tr>
<td><strong>Order License Id:</strong></td>
</tr>
<tr>
<td><strong>Number of pages</strong></td>
</tr>
<tr>
<td><strong>Portion</strong></td>
</tr>
<tr>
<td><strong>Number of figures/tables/illustrations</strong></td>
</tr>
<tr>
<td><strong>Format</strong></td>
</tr>
<tr>
<td><strong>Are you the author of this Elsevier article?</strong></td>
</tr>
<tr>
<td><strong>Will you be translating?</strong></td>
</tr>
<tr>
<td><strong>Order reference number</strong></td>
</tr>
<tr>
<td><strong>Expected completion date</strong></td>
</tr>
<tr>
<td><strong>Estimated size (number of pages)</strong></td>
</tr>
<tr>
<td><strong>Elsevier VAT number</strong></td>
</tr>
<tr>
<td><strong>Permissions price</strong></td>
</tr>
<tr>
<td><strong>VAT/Local Sales Tax</strong></td>
</tr>
</tbody>
</table>

**Note:** This item was invoiced separately through our RightsLink service. More info

Total order items: 1  
Order Total: $0.00
Confirmation Number: 10998021
Order Date: 05/15/2012

Customer Information

- Customer: David Watson
- Account Number: 3000533869
- Organization: David Watson
- Email: dewatson@go.olemiss.edu
- Phone: +1 (662)4738003
- Payment Method: Invoice

Order Details

**Journal of computer-aided molecular design**

- Order detail ID: 62467456
- Article Title: GRIND-derived pharmacophore model for a series of o-tropanyl derivative ligands of the sigma-2 receptor*
- Author(s): Cratteri, Paola ; et al
- DOI: 10.1023/B:JCAM.0000047815.22931.31
- Date: May 01, 2004
- ISSN: 1573-4951
- Publication Type: e-Journal
- Volume: 18
- Issue: 5
- Start page: 361
- Publisher: KLUWER ACADEMIC PUBLISHERS (DORDRECHT)

Permission Status: ✓ Granted
- Permission type: Republish or display content
- Type of use: use in a thesis/dissertation
- Order License ID: 2910450173219

- Portion: Figures
- Author of this Springer article: No
- Title of your thesis / dissertation: Quasi-comprehensive scaffold perception, pharmacophore development, and structure--affinity relationships of sigma site ligands
- Expected completion date: Jul 2012
- Estimated size(pages): 200

Note: This item was invoiced separately through our RightsLink service. More info

Total order items: 1
Order Total: $0.00

Billing Status: N/A

Get Permission | License Your Content | Products & Solutions | Partners | Education | About CCC
Privacy Policy | Terms & Conditions
Copyright 2012 Copyright Clearance Center
Confirmation Number: 10998029  
Order Date: 05/15/2012

Customer Information
Customer: David Watson  
Account Number: 3000533869  
Organization: David Watson  
Phone: +1 (662)4738003  
Payment Method: Invoice

Order Details

<table>
<thead>
<tr>
<th>Bioorganic and medicinal chemistry letters</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Order detail ID:</strong> 62467470</td>
</tr>
<tr>
<td><strong>Article Title:</strong> A 3D-pharmacophore model for α2 receptors based on a series of substituted benzo[d]oxazol-2(3H)-one derivatives</td>
</tr>
<tr>
<td><strong>Author(s):</strong> Laurini, Erik ; et al</td>
</tr>
<tr>
<td><strong>DOI:</strong> 10.1016/J.BMCL.2010.03.009</td>
</tr>
<tr>
<td><strong>Date:</strong> Jan 01, 2010</td>
</tr>
<tr>
<td><strong>ISSN:</strong> 1464-3405</td>
</tr>
<tr>
<td><strong>Publication Type:</strong> e-Journal</td>
</tr>
<tr>
<td><strong>Volume:</strong> 20</td>
</tr>
<tr>
<td><strong>Issue:</strong> 9</td>
</tr>
<tr>
<td><strong>Start page:</strong> 2954</td>
</tr>
<tr>
<td><strong>Publisher:</strong> PERGAMON</td>
</tr>
</tbody>
</table>

|Permission Status: ▶️ Granted |
|**Permission type:** Republish or display content  |
|**Type of use:** reuse in a thesis/dissertation  |
|**Order License Id:** 2910461171361|
|**Number of pages:** 4  |
|**Portion:** figures/tables/illustrations  |
|**Number of figures/tables/illustrations:** 1  |
|**Format:** both print and electronic  |
|**Are you the author of this Elsevier article?** No  |
|**Will you be translating?** No  |
|**Order reference number**  |
|**Title of your thesis/dissertation**  |
|Quasi-comprehensive scaffold perception, pharmacophore development, and structure--affinity relationships of sigma site ligands  |
|**Expected completion date**  |
|Jul 2012  |
|**Estimated size (number of pages)**  |
|200  |
|**Elsevier VAT number**  |
|GB 494 6272 12  |
|**Permissions price**  |
|0.00 USD  |
|**VAT/Local Sales Tax**  |
|0.00 USD / 0.0 GBP  |

Note: This item was invoiced separately through our RightsLink service. More info

Total order items: 1  
Order Total: $0.00
<table>
<thead>
<tr>
<th>Journal of medicinal chemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISSN: 0022-2623</td>
</tr>
<tr>
<td>Publication year(s): 1963 - present</td>
</tr>
<tr>
<td>Author/Editor: AMERICAN CHEMICAL SOCIETY</td>
</tr>
<tr>
<td>Publication type: Journal</td>
</tr>
<tr>
<td>Publisher: AMERICAN CHEMICAL SOCIETY</td>
</tr>
<tr>
<td>Rightsholder: AMERICAN CHEMICAL SOCIETY</td>
</tr>
</tbody>
</table>

**Permission type selected:** Republish or display content

**Type of use selected:** reuse in a Thesis/Dissertation

---

**PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE**

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

---

**Article Title:** Novel piperidine \( \sigma \) receptor ligands as potential antipsychotic drugs

**Author(s):** Gilligan, Paul J.; et al

**DOI:** 10.1021/JM00101A012

**Date:** Nov 1, 1992

**Volume:** 35

**Issue:** 23

---

Get Permission | License Your Content | Products & Solutions | Partners | Education | About CCC
Privacy Policy | Terms & Conditions
Copyright 2012 Copyright Clearance Center
Structural Features Important for σ1 Receptor Binding

Glennon, Richard A. ; et al

DOI: 10.1021/JM00034A020

Date: Apr 1, 1994

Volume: 37

Issue: 8

Permission type selected: Republish or display content

Type of use selected: reuse in a Thesis/Dissertation

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.
Synthesis and Structure−Activity Relationships of N-(1-Benzylpiperidin-4-yl)arylacetamide Analogues as Potent σ1 Receptor Ligands

Huang, Yunsheng; et al

DOI: 10.1021/JM010384J

Date: Dec 1, 2001
Volume: 44
Issue: 25

Permission is granted for your request in both print and electronic formats, and translations.

If figures and/or tables were requested, they may be adapted or used in part.

Please print this page for your records and send a copy of it to your publisher/graduate school.

Appropriate credit for the requested material should be given as follows: “Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society.” Insert appropriate information in place of the capitalized words.

One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.
Novel (4-Phenylpiperidinyl)- and (4-Phenylpiperazinyl)alkyl-Spaced Esters of 1-Phenylcyclopentanecarboxylic Acids as Potent \( \sigma \)-Selective Compounds

Author(s): Hudkins, Robert L.; Mailman, Richard B.; DeHaven-Hudkins, Diane L.

DOI: 10.1021/JM00039A008

Volume: 37

Issue: 13

Date: Jun 1, 1994
Article title: Discovery of High-Affinity Ligands of σ1Receptor, ERG2, and Emopamil Binding Protein by Pharmacophore Modeling and Virtual Screening

Author(s): Laggner, Christian; et al

DOI: 10.1021/JM049073+

Date: Jul 1, 2005

Volume: 48

Issue: 15

Permission type selected: Republish or display content

Type of use selected: reuse in a Thesis/Dissertation

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.
Synthesis, Biological Evaluation, and Three-Dimensional in Silico Pharmacophore Model for \( \sigma \) Receptor Ligands Based on a Series of Substituted Benzo[\textit{d}]oxazol-2(3H)-one Derivatives

Zampieri, Daniele; et al

DOI: 10.1021/JM900366Z

Date: Sep 10, 2009

Volume: 52

Issue: 17

Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society. Insert appropriate information in place of the capitalized words.
Hi David,

Of course, you can use the figures for your dissertation. At the moment, I don't have easy access to the original figures, so you will probably need to scan the figures from the books. If I find them, I will forward them to you.

Ken Domino
Thank you all very much for clarifying the matter for me, and for providing the permissions to use the selected figures in my dissertation.

On Jun 5, 2012, at 1:56 AM, Straalen van, Berendina, Springer SBM NL wrote:

Dear Dr. Watson,
The permission is also granted for UMI dissertations.

From: Honour, Carolyn, Springer US
Sent: maandag 4 juni 2012 19:19
To: Straalen van, Berendina, Springer SBM NL
Subject: Fw: MCRE - Permissions request...

Dear Berendina
Please see below. Are electronic dissertations permissible in this regard?

Best
Carolyn

From: Stephen Cutler [mailto:cutler@olemiss.edu]
Sent: Monday, June 04, 2012 11:23 AM
To: Honour, Carolyn, Springer US
Subject: FW: MCRE - Permissions request...

Hi Carolyn,

We have a graduate student, David Watson, who is interested in using some graphs/tables from previous Med Chem Res publications. Some of these articles were published more than 10 years ago. About 2 years ago the University of Mississippi Graduate School converted from a paper dissertation to an electronic version. It appears that the use of materials from Springer might be in conflict with the trend among Graduate Schools in the US. Can you confirm if dissertations are exempt from the electronic clause of copyright permissions by Springer?

Thanks,
S.
VITA

David E. Watson

Department of Medicinal Chemistry
School of Pharmacy
425 Faser Hall
University, MS 38677
Date of Birth: April 11, 1974
Citizenship: United States

Phone: (662) 915-7101
Cell: (662) 473-8003
Fax: (662) 915-5638
Email: dewatson@go.olemiss.edu

Education

Ph.D. Pharmaceutical Sciences, University of Mississippi, School of Pharmacy, Department of Medicinal Chemistry, expected 2013.
Dissertation Advisor: Christopher R. McCurdy; Dissertation Title: “Quasi-comprehensive scaffold perception, pharmacophore development, and structure–affinity relationships of sigma site ligands”

B.S. Chemistry, University of Mississippi, School of Liberal Arts, Department of Chemistry, 1999. Minor: German.

Scientific Appointments/Experience

Graduate Research Assistant, Department of Medicinal Chemistry, School of Pharmacy, University of Mississippi, University, MS. Advisor: Christopher McCurdy, January 2008–Present.
Certifying Scientist and Laboratory Technician, ElSohly Laboratories, Oxford, MS. Fall 2000–Present.

Student and Technician Supervision

Emily Carrell, 05/2012–08/2012
Andrew Mullen, 06/2010–08/2011
Scientific Associations

American Chemical Society, Member, 2008–2013
  Medicinal Chemistry Division, Member, 2008–2013
  Computers in Chemistry Division, Member, 2008–2013

American Association of Pharmaceutical Scientists UM Student Chapter, Member, 2010–2012

University of Mississippi Medicinal Chemistry Journal Club, Member, 2009-2013

Honors and Awards

Rho Chi Society, University of Mississippi, 2009

Phi Eta Sigma, University of Mississippi, 1993

Service as Scientific Referee

Invited National Institutes of Health Referee (ad hoc)
  NIH-NCRR COBRE CORE-NPN Predoctoral Fellowship ad hoc Study Section, 2010.
  NIH-NCRR COBRE CORE-NPN Predoctoral Fellowship ad hoc Study Section, 2009.

Invited Journal Referee
  Journal of Natural Products, 2011–Present

Professional Service

American Association of Pharmaceutical Scientists Student Chapter, University of Mississippi.
  Chair, 2011
  Chair Elect, 2010

University Committees, and Service

University of Mississippi, School of Pharmacy
  Information, Resources and Computing Committee, Graduate Student at Large, 2009-2010.

Presentations

Regional Presentations
  Watson, D.E. Comparative modeling of prolylcarboxypeptidase to elucidate factors responsible for the selective hydrolysis of kinins. 38th Annual MALTO Medicinal Chemistry and Pharmacognosy meeting, Houston, TX, May 23, 2011
Local Presentations

Watson, D.E. Voltage Gated Sodium Channel Ligands: Opportunities and Challenges. Department of Medicinal Chemistry, School of Pharmacy, University, MS, November 3, 2009

Watson, D.E. Hot or not? Selective TRPV1 antagonists. Department of Medicinal Chemistry, School of Pharmacy, University, MS, September 30, 2008

Watson, D.E. Using BibTeX to Manage References for Scientific Publications. Mississippi Center for Supercomputing Research, University, MS, May 6, 2008

Watson, D.E. Typesetting with LaTeX. Mississippi Center for Supercomputing Research, University, MS, April 10, 2008

Bibliography

Peer-reviewed scientific research


Editor