
Extracting Chemical Information within Documents - from Desktop to Enterprise

Wei Deng (David), Daniel Bonniot

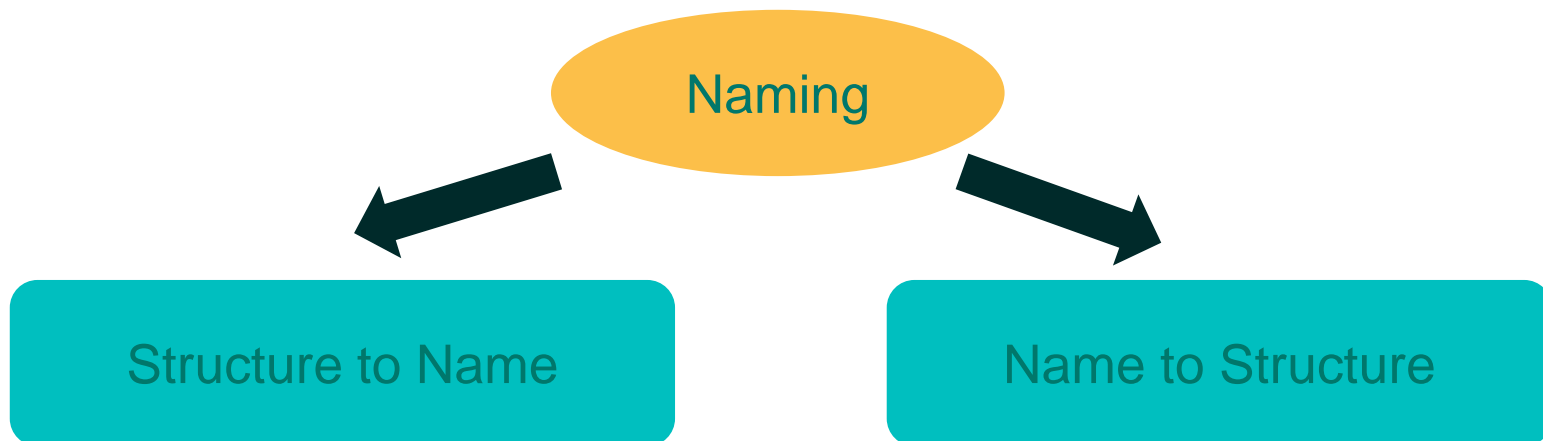
ChemAxon US UGM
Sep 2013



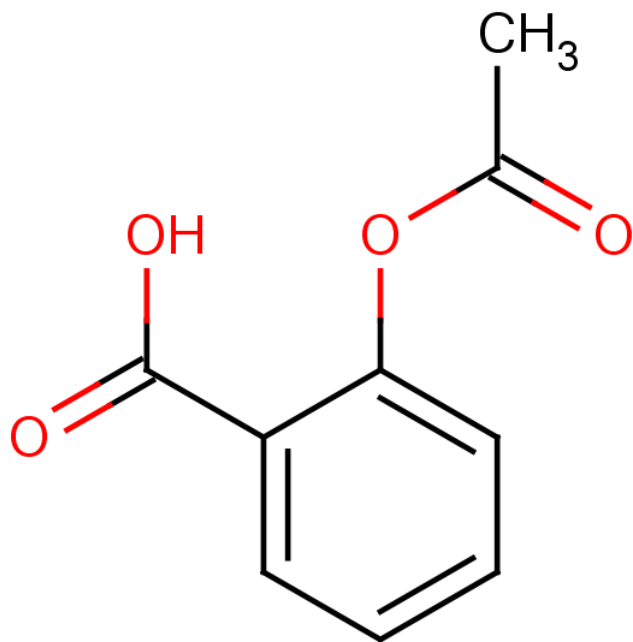
What is *Naming*

- Reliable chemical structure and name conversion
- Available in all ChemAxon products
- Can be used in various ways
- Backbone of ChemAxon's text mining tools

Backbone of Text Mining



Structure to "Name"



S2N

- 2-(acetyloxy) benzoic acid
- Aspirin
- 50-78-2
- 11126-35-5
- 11126-37-7
- 2349-94-2
- 26914-13-6
- 98201-60-6

NEW

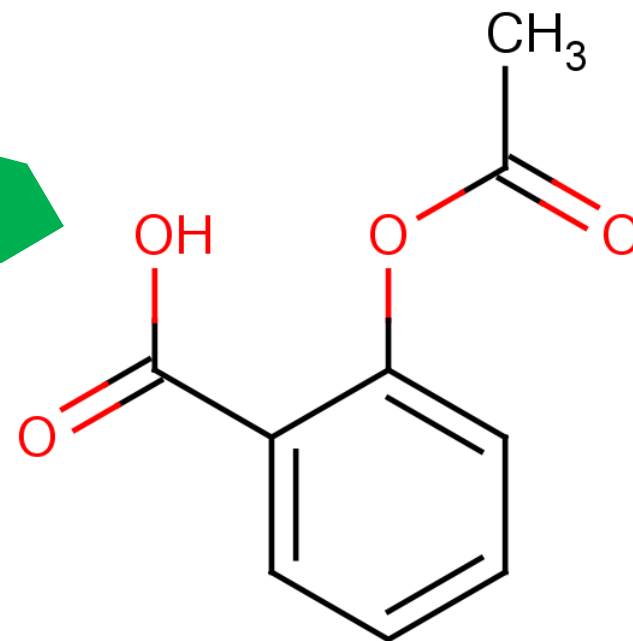
“Name” to Structure



- 2- (乙酰氧基) 苯甲酸
- 阿司匹林

- 2-(acetyloxy) benzoic acid
- Aspirin
- Acetylsalicylate
- Easprin ...

- 50-78-2

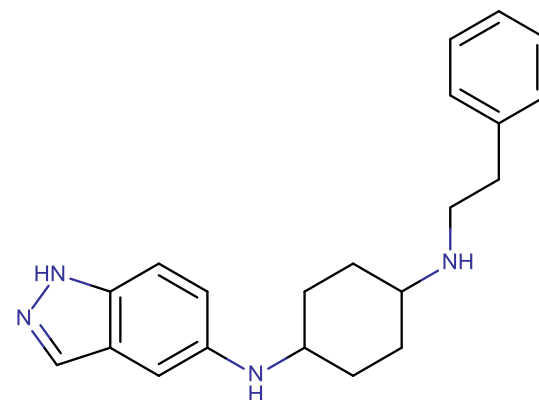


Matured yet still Improving



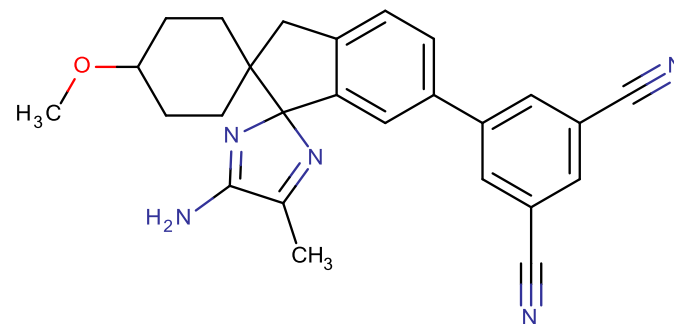
N1-(1H-5-Indazolyl)-N4-phenylethyl-1,4-cyclohexanediamine

6.1



5-(4''-amino-4-methoxy-5''-methyl-3'H-dispiro[cyclohexane-1,2'-indene-1',2''-imidazol]-6'-yl)benzene-1,3-dicarbonitrile

6.2



What About Other Chemical Information?

- Corporate compound ID number
- Unusual common names
- Names in other languages
- Other identifier (e.g. FICTS ID)
- ...

Customized Dictionary

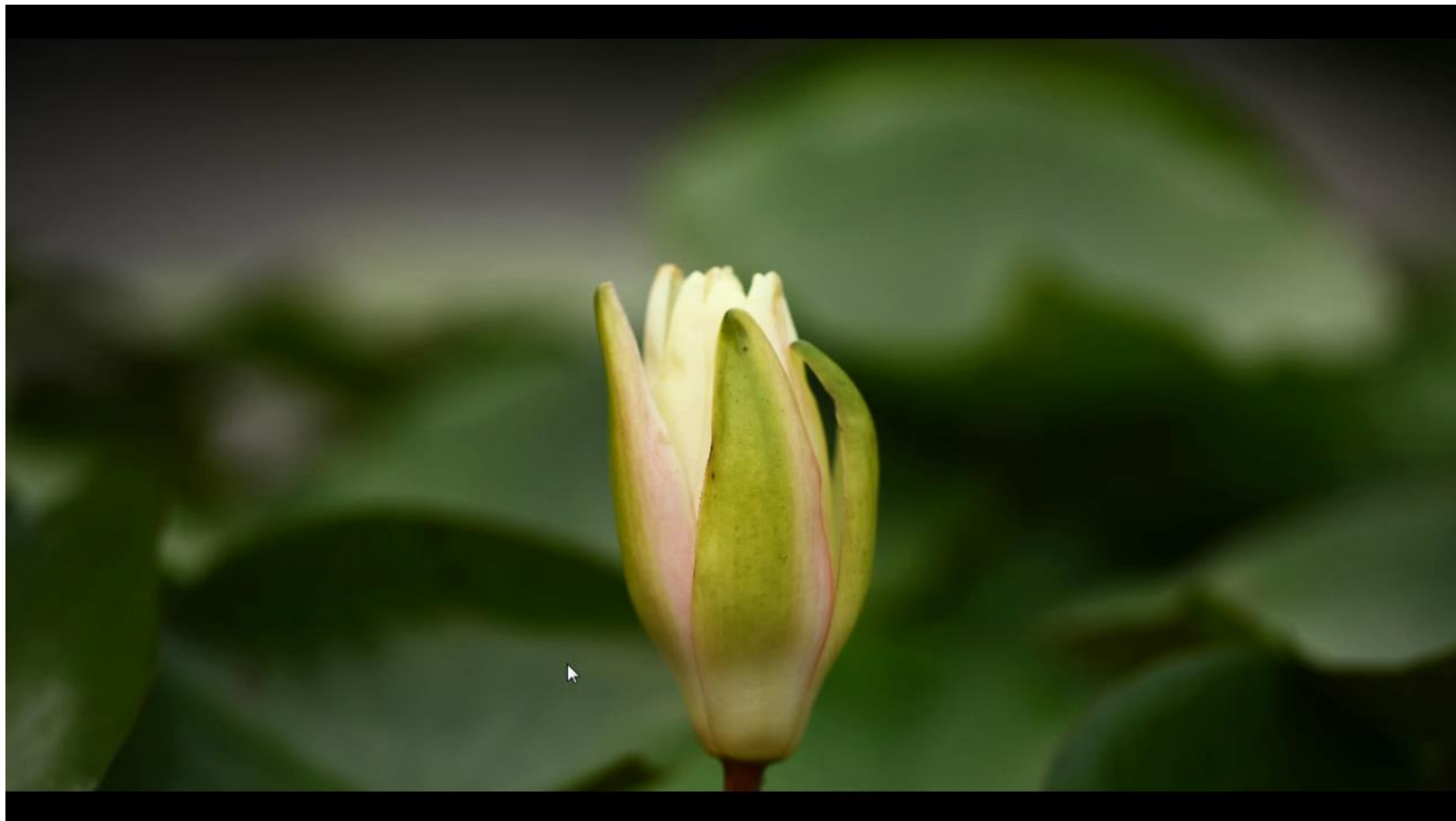
- A SMILES file “custom_names.smi”
- Default location ChemAxon DIR
e.g. in Windows 7 C:\Users\USERNAME\chemaxon\

- Format

c1ccccc1 CXN000001

SMILES Tab ANY text string

Custom Dictionary Demo



Customized Dictionary

- A SMILES file “custom_names.smi”
- Default location ChemAxon DIR
e.g. in Windows 7 C:\Users\USERNAME\chemaxon\

- Format

c1ccccc1 CXN000001

SMILES Tab ANY text string



From Version 6.0, a custom **web service** can also be used



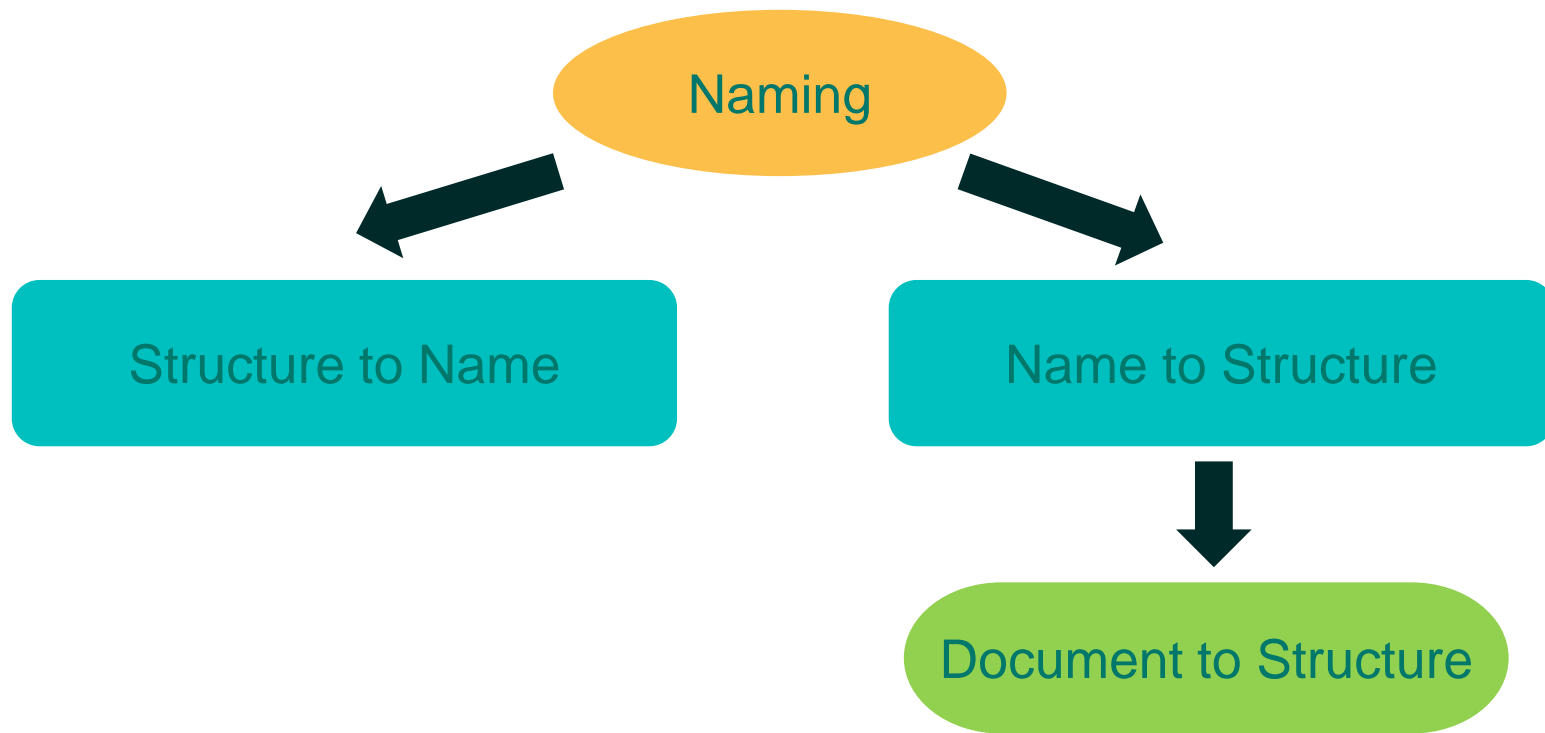
Naming Web Service Demo

```
Untitled - Notepad
File Edit Format View Help
Webservice in general:
http://company.com/ws/n2s?input=

For example:
Cactus service by NIH:
http://cactus.nci.nih.gov/chemical/structure/
[NAME]/sdf

FICTS ID:
F488D82B00A10EC1-FICTS-01-3B
```

Chemical Information in Documents?



Text Chemical Information in Documents

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Discovery of Novel Selective Development of a Protein-I

Sankar Manepalli,[†] Laura M. Geffert,[†] Ch

[†]Department of Chemistry and Biochemistry and C
308 Mellon Hall, Pittsburgh, Pennsylvania 15262, U

[‡]Division of Pharmaceutical Sciences, Mylan School
Pittsburgh, Pennsylvania 15262, United States

ABSTRACT: The serotonin transporter (SERT) sodium symporter (NSS) family, is responsible for synaptic cleft to maintain neurotransmitter homeostasis. SERT is established as an important target in the treatment of anxiety and depression. Because a high-resolution crystal structure is not available, a computational model of SERT was built based upon the X-ray coordinates of the leucine transporter LeuT, a bacterial NSS homologue. The model was used to develop the first SERT structure-based pharmacophore. Virtual screening (VS) of a small molecule structural library using model yielded candidate ligands of diverse scaffolds. Pharmacological analysis of the VS hits identified two SERT-selective compounds, 2 SERT-related medication development.

INTRODUCTION

Signaling between cells in the central nervous system is mediated by the controlled release and reuptake of neurotransmitters in the synaptic cleft. The reuptake of neurotransmitters in the synaptic cleft is mediated by the serotonin transporter (SERT), which terminates the action of these biogenic amines via reuptake into the presynaptic cell. The mechanism of drug action related to the above medical conditions typically the MAT proteins. Tricyclic antidepressants (TCAs) imipramine (Tofranil), developed in the 1950s, alleviate depression by blocking serotonin and norepinephrine transporters (SERT and NET, respectively), thereby extending the lifespan of synaptic serotonin and norepinephrine. Unfortunately, the TCAs also block adrenergic, muscarinic, acetylcholinesterase, and histaminic receptors, responsible for a plethora of adverse effects. Selective serotonin reuptake inhibitors (SSRIs), the first generation of antidepressants led by fluoxetine (Prozac) in the 1980s, carry far fewer adverse effects compared to the TCAs. The SERT, but because the resultant surge of serotonin can activate any of 14 serotonin receptor types, this drug class is not without its own adverse effects.^{1–10}

The driving force for MAT uptake of monoamine substrate is electrogenic, harnessing the inward Na⁺ gradient across the cell membrane.^{11,12} The SERT, NET, and DAT (dopamine transporter) are members of the neurotransmitter sodium symporter (NSS) family as well as members of a larger group of Cl⁻-dependent transporters known as the "solute carrier 6" (SLC6)

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

ARTICLE

pubs.acs.org/jcim

Discovery of Novel Selective Serotonin Reuptake Inhibitors through Development of a Protein-Based Pharmacophore

Sankar Manepalli,[†] Laura M. Geffert,[†] Christopher K. Surratt,[‡] and Jeffrey D. Madura^{*,†}

[†]Department of Chemistry and Biochemistry and Center for Computational Science, Duquesne University, 600 Forbes Avenue, 308 Mellon Hall, Pittsburgh, Pennsylvania 15262, United States

[‡]Division of Pharmaceutical Sciences, Mylan School of Pharmacy, Duquesne University, 600 Forbes Avenue, 401 Mellon Hall, Pittsburgh, Pennsylvania 15262, United States

ABSTRACT: The serotonin transporter (SERT), a member of the neurotransmitter sodium symporter (NSS) family, is responsible for the reuptake of serotonin from the synaptic cleft to maintain neurotransmitter homeostasis. SERT is established as an important target in the treatment of anxiety and depression. Because a high-resolution crystal structure is not available, a computational model of SERT was built based upon the X-ray coordinates of the leucine transporter LeuT, a bacterial NSS homologue. The model was used to develop the first SERT structure-based pharmacophore. Virtual screening (VS) of a small molecule structural library using the generated SERT computational model yielded candidate ligands of diverse scaffolds. Pharmacological analysis of the VS hits identified two SERT-selective compounds, potential lead compounds for further SERT-related medication development.

INTRODUCTION

Signaling between cells in the central nervous system is mediated by the controlled release and reuptake of neurotransmitters in the synaptic cleft. An excess or deficit of the monoamines serotonin, dopamine, or norepinephrine in the synaptic cleft has been associated with various psychiatric and neurological disorders including depression, anxiety, compulsivity, attention deficit hyperactivity disorder (ADHD), substance abuse, Parkinson's disease, and schizophrenia.^{1–7} Control of synaptic monoamine levels is affected by the plasma membrane monoamine transporters (MATs), which terminate the action of these biogenic amines via reuptake into the presynaptic cell. The mechanisms of action of drugs related to the above medical conditions typically involve the MAT proteins. Tricyclic antidepressants (TCAs) such as imipramine (Tofranil), developed in the 1950s, alleviate depression by blocking serotonin and norepinephrine transporters (SERT and NET, respectively), thereby extending the lifespan of synaptic serotonin and norepinephrine. Unfortunately, the TCAs also block adrenergic, muscarinic, acetylcholinesterase, and histaminic receptors, responsible for a plethora of adverse effects.^{8,9} Selective serotonin reuptake inhibitors (SSRIs), the first generation of antidepressants led by fluoxetine (Prozac) in the 1980s, carry far fewer adverse effects compared to the TCAs. The SSRIs block SERT, but because the resultant surge of serotonin can activate any of 14 serotonin receptor types, this drug class is not without its own adverse effects.^{10–12}

The driving force for MAT uptake of monoamine substrate is electrogenic, harnessing the inward Na⁺ gradient across the cell membrane.^{13,14} The SERT, NET, and DAT (dopamine transporter) are members of the neurotransmitter sodium symporter (NSS) family as well as members of a larger group of Na⁺ and Cl⁻-dependent transporters known as the "solute carrier 6" (SLC6)

family.¹⁵ The lack of a high-resolution 3-D MAT structure had hindered structure–function and therapeutic development efforts until a breakthrough was achieved in the form of crystallization of the bacterial NSS homologue LeuT, a leucine transporter.¹⁶ The LeuT X-ray structure has provided a template to build credible MAT computational models.^{17–21} Although model quality increases with sequence identity with the template, structural similarity also plays a significant role.

In the absence of high-resolution 3D structures for SERT, development of a ligand-based pharmacophore^{22–24} or QSAR^{25–28} are feasible alternatives to obtain structural information about the binding pocket. Ligand-based approaches analyze a set of ligands and generate possible protein–ligand interaction patterns without knowledge of the protein structure. A limitation of ligand-based approaches is that flexible alignment using dissimilar scaffolds is less reliable; knowledge of the bioactive conformation of at least one active molecule significantly improves alignment accuracy.²⁹ These limitations can be overcome by using structure-based approaches, in which diverse scaffolds can be used to capture ligand interactions and the binding pocket environment in general. Docking, a structure-based technique, is capable of reliably predicting the bioactive conformation for a noncrystallized ligand.³⁰ In the absence of an experimental 3D structure, however, a reasonable 3D model can be constructed for a receptor using crystallographic or NMR data from genetically and functionally related proteins.^{31,32}

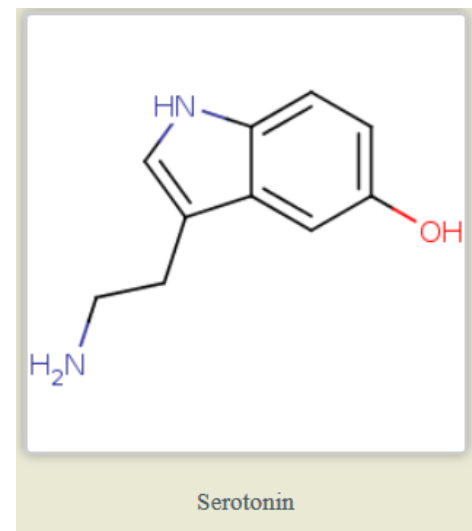
Comparative modeling correctly predicts the 3D fold of a protein in most cases and has often provided insight into the

Received: June 17, 2011

Published: August 12, 2011

2017

dx.doi.org/10.1021/acs.jcim.1b00048 | J. Chem. Inf. Model. 2011, 41, 2017–2024



Structure Images in Documents

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Discovery of Novel Selective Development of a Protein-I

Sankar Manepalli,[†] Laura M. Geffert,[†] Ch

[†]Department of Chemistry and Biochemistry and G
308 McLean Hall, Pittsburgh, Pennsylvania 15262, U

[‡]Division of Pharmaceutical Sciences, Mylan School
Pittsburgh, Pennsylvania 15262, United States

ABSTRACT: The serotonin transporter (SERT) sodium symporter (NSS) family, is responsible for synaptic cleft to maintain neurotransmitter homeostasis. SERT is established as an important target in the treatment of anxiety and depression. Because a high-resolution crystal structure is not available, a computational model of SERT was built based upon the X-ray coordinates of the leucine transporter LeuT, a bacterial NSS homologue. The model was used to develop the first SERT structure-based pharmacophore. Virtual screening (VS) of a small molecule structural library using the generated SERT computational model yielded candidate ligands of diverse scaffolds. Pharmacological analysis of the VS hits identified two SERT-selective compounds, p SERT-related medication development.

INTRODUCTION

Signaling between cells in the central nervous system is mediated by the controlled release and reuptake of neurotransmitters in the synaptic cleft. The reuptake of neurotransmitters, such as serotonin, dopamine, and norepinephrine, is associated with various psychiatric and neurological disorders including depression, anxiety, compulsivity, attention deficit hyperactivity disorder (ADHD), substance abuse, Parkinson's disease, and schizophrenia.¹⁻⁷ Control of synaptic monoamine levels is affected by the plasma membrane monoamine transporter (MAT), which terminates the action of these biogenic amines via reuptake into the presynaptic cell. The mechanism of drug inhibition to the above medical conditions typically targets the MAT proteins. Tricyclic antidepressants (TCAs) imipramine (Tofranil), developed in the 1950s, alleviate depression by blocking serotonin and norepinephrine transporters (SERT and NET, respectively), thereby extending the lifespan of synaptic serotonin and norepinephrine. Unfortunately, the TCAs also block adrenergic, muscarinic, acetylcholinergic, and histamine receptors, responsible for a plethora of adverse effects.⁸⁻¹⁰ Selective serotonin reuptake inhibitors (SSRIs), the most common of antidepressants led by fluoxetine (Prozac) in the 1990s, carry far fewer adverse effects compared to the TCAs. The SSRIs block SERT, but because the resultant surge of serotonin can activate any of 14 serotonin receptor types, this drug class is not without its own adverse effects.¹¹⁻¹³

The driving force for MAT uptake of monoamine substrate is electrogenic, harnessing the inward Na⁺ gradient across the cell membrane.^{14,15} The SERT, NET, and DAT (dopamine transporter) are members of the neurotransmitter sodium symporter (NSS) family as well as members of a larger group of Na⁺ and Cl⁻ dependent transporters known as the "solute carrier 6" (SLC6)

JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

ARTICLE

pubs.acs.org/jcim

Discovery of Novel Selective Serotonin Reuptake Inhibitors through Development of a Protein-Based Pharmacophore

Sankar Manepalli,[†] Laura M. Geffert,[†] Christopher K. Surratt,[‡] and Jeffrey D. Madura^{*,†}

[†]Department of Chemistry and Biochemistry and Center for Computational Science, Duquesne University, 600 Forbes Avenue, 308 Mellon Hall, Pittsburgh, Pennsylvania 15262, United States

[‡]Division of Pharmaceutical Sciences, Mylan School of Pharmacy, Duquesne University, 600 Forbes Avenue, 411 Mellon Hall, Pittsburgh, Pennsylvania 15262, United States

ABSTRACT: The serotonin transporter (SERT), a member of the neurotransmitter sodium symporter (NSS) family, is responsible for the reuptake of serotonin from the synaptic cleft to maintain neurotransmitter homeostasis. SERT is established as an important target in the treatment of anxiety and depression. Because a high-resolution crystal structure is not available, a computational model of SERT was built based upon the X-ray coordinates of the leucine transporter LeuT, a bacterial NSS homologue. The model was used to develop the first SERT structure-based pharmacophore. Virtual screening (VS) of a small molecule structural library using the generated SERT computational model yielded candidate ligands of diverse scaffolds. Pharmacological analysis of the VS hits identified two SERT-selective compounds, potential lead compounds for further SERT-related medication development.

INTRODUCTION

Signaling between cells in the central nervous system is mediated by the controlled release and reuptake of neurotransmitters in the synaptic cleft. An excess or deficit of the monoamines serotonin, dopamine, or norepinephrine in the synaptic cleft has been associated with various psychiatric and neurological disorders including depression, anxiety, compulsivity, attention deficit hyperactivity disorder (ADHD), substance abuse, Parkinson's disease, and schizophrenia.¹⁻⁷ Control of synaptic monoamine levels is affected by the plasma membrane monoamine transporters (MATs), which terminate the action of these biogenic amines via reuptake into the presynaptic cell. The mechanism of action of drug inhibition to the above medical conditions typically involves the MAT proteins. Tricyclic antidepressants (TCAs) such as imipramine (Tofranil), developed in the 1950s, alleviate depression by blocking serotonin and norepinephrine transporters (SERT and NET, respectively), thereby extending the lifespan of synaptic serotonin and norepinephrine. Unfortunately, the TCAs also block adrenergic, muscarinic, acetylcholinergic, and histamine receptors, responsible for a plethora of adverse effects.^{8,9} Selective serotonin reuptake inhibitors (SSRIs), the most common of antidepressants led by fluoxetine (Prozac) in the 1990s, carry far fewer adverse effects compared to the TCAs. The SSRIs block SERT, but because the resultant surge of serotonin can activate any of 14 serotonin receptor types, this drug class is not without its own adverse effects.¹⁰⁻¹²

The driving force for MAT uptake of monoamine substrate is electrogenic, harnessing the inward Na⁺ gradient across the cell membrane.^{13,14} The SERT, NET, and DAT (dopamine transporter) are members of the neurotransmitter sodium symporter (NSS) family as well as members of a larger group of Na⁺ and Cl⁻ dependent transporters known as the "solute carrier 6" (SLC6)

family.¹⁵ The lack of a high-resolution 3-D MAT structure had hindered structure-function and therapeutic development efforts until a breakthrough was achieved in the form of crystallization of the bacterial NSS homologue LeuT, a leucine transporter.¹⁶ The LeuT X-ray structure has provided a template to build credible MAT computational models.¹⁷⁻²¹ Although model quality increases with sequence identity with the template, structural similarity also plays a significant role.

In the absence of high-resolution 3D structures for SERT, development of a ligand-based pharmacophore²²⁻²⁴ or QSAR²⁵⁻²⁸ are feasible alternatives to obtain structural information about the binding pocket. Ligand-based approaches analyze a set of ligands and generate possible protein-ligand interaction patterns without knowledge of the protein structure. A limitation of ligand-based approaches is that flexible alignment using dissimilar scaffolds is less reliable; knowledge of the bioactive conformation of at least one active molecule significantly improves alignment accuracy.²⁹ These limitations can be overcome by using structure-based approaches, in which diverse scaffolds can be used to capture ligand interactions and the binding pocket environment in general. Docking, a structure-based technique, is capable of reliably predicting the bioactive conformation for a crystallized ligand.³⁰ In the absence of an experimental 3D structure, however, a reasonable 3D model can be constructed for a receptor using crystallographic or NMR data from genetically and functionally related proteins.^{31,32}

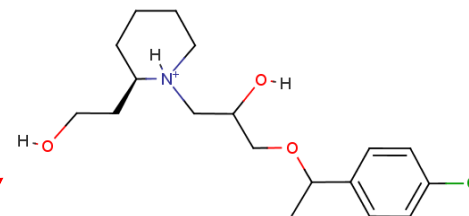
Comparative modeling correctly predicts the 3D fold of a protein in most cases and has often provided insight into the

Received: June 17, 2011

Published: August 12, 2011

2017

doi:10.1021/acs.jcim.1b00044



D2S


OSRA (NIH)


6.1 Supports **NEW**

CLiDE (Keymodule)

Imago (GGA)

Non-Text Document

(19)  **Europäisches Patentamt**
European Patent Office
Office européen des brevets

(11)  **EP 2 377 850 A1**

(12) **EUROPEAN PATENT APPLICATION**

(43) Date of publication: **19.10.2011 Bulletin 2011/42**

(21) Application number: **10158292.2**

(22) Date of filing: **30.03.2010**

(51) Int Cl.:
C07D 209/32 (2006.01) C07D 235/26 (2006.01)
C07D 263/58 (2006.01) C07D 265/38 (2006.01)
C07D 401/12 (2006.01) C07D 403/12 (2006.01)
C07D 407/12 (2006.01) A61K 31/404 (2006.01)
A61K 31/4184 (2006.01) A61K 31/423 (2006.01)
A61K 31/538 (2006.01) A61P 29/00 (2006.01)
A61P 11/00 (2006.01) A61P 13/00 (2006.01)
A61P 17/00 (2006.01)

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO SE SI SK SM TR
Designated Extension States:
AL BA ME RS

(71) Applicant: **Pharmeste S.r.l.**
44100 Ferrara (IT)

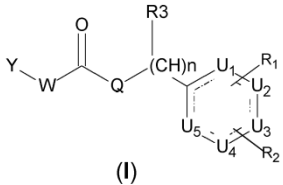
(72) Inventors:
• **NAPOLETANO, Mauro**
20127, MILANO (IT)

(74) Representative: **Minoja, Fabrizio**
Bianchetti Bracco Minoja S.r.l.
Via Plinio, 63
20129 Milano (IT)

(84) Designated Contracting States:
• **TREVISANI, Marcello**
44017, GAVELLO DI BONDENO (FE) (IT)
• **PAVANI, Maria Giovanna**
44049, VIGARANO MAINARDA (FE) (IT)
• **FRUTTAROLO, Francesca**
44100, FERRARA (IT)

(54) **TRPV1 vanilloid receptor antagonists with a bicyclic portion**

(57) The invention discloses compounds of formula I



(I)

wherein Y is selected from a group of formula

EP 2 377 850 A1

Printed by Jouve, 73001 PARIS (FR)

(Cont. next page)



Full Text

Automatic OCR Error Correction

(2R)-2-**rn**ethylsulfany**y**1-3-hydr**0**xybutanedi**0**ate



(2R)-2-methylsulfanyl-3-hydroxybutanedioate

Ar-benzyl-**Ar**-[3-(**1**H-tetrazol-5-yl)phenyl]propanamide



N-benzyl-*N*-[3-(1H-tetrazol-5-yl)phenyl]propanamide



我们目前**止**在研究开**友**中文化**字**名称的**OCR**自动纠错**工**力能



我们目前正在研究开**发**中文化**学**名称的**OCR**自动纠错功能

Document to Structure Demo

EP2377850A1.pdf - Adobe Reader

File Edit View Window Help

1 / 50 95.6%

Tools Sign Comment

Sign In

Export PDF

Create PDF

Adobe CreatePDF


Convert files to PDF and easily combine them with other file types with a paid subscription.


Select File to Convert to PDF:

Select File

Send Files

Store Files

(19)  Europäisches Patentamt
European Patent Office
Office européen des brevets

(11)  EP 2 377 850 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication: 19.10.2011 Bulletin 2011/42

(21) Application number: 10158292.2

(22) Date of filing: 30.03.2010

(51) Int Cl.:
C07D 209/32^(2006.01) C07D 235/26^(2006.01)
C07D 263/58^(2006.01) C07D 265/36^(2006.01)
C07D 401/12^(2006.01) C07D 403/12^(2006.01)
C07D 407/12^(2006.01) A61K 31/404^(2006.01)
A61K 31/4184^(2006.01) A61K 31/423^(2006.01)
A61K 31/538^(2006.01) A61P 29/00^(2006.01)
A61P 11/00^(2006.01) A61P 13/00^(2006.01)
A61P 17/00^(2006.01)

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL
PT RO SE SI SK SM TR
Designated Extension States:
AL BA ME RS

(71) Applicant: Pharmeste S.r.l.
44100 Ferrara (IT)

(72) Inventors:
• NAPOLETANO, Mauro
20127, MILANO (IT)

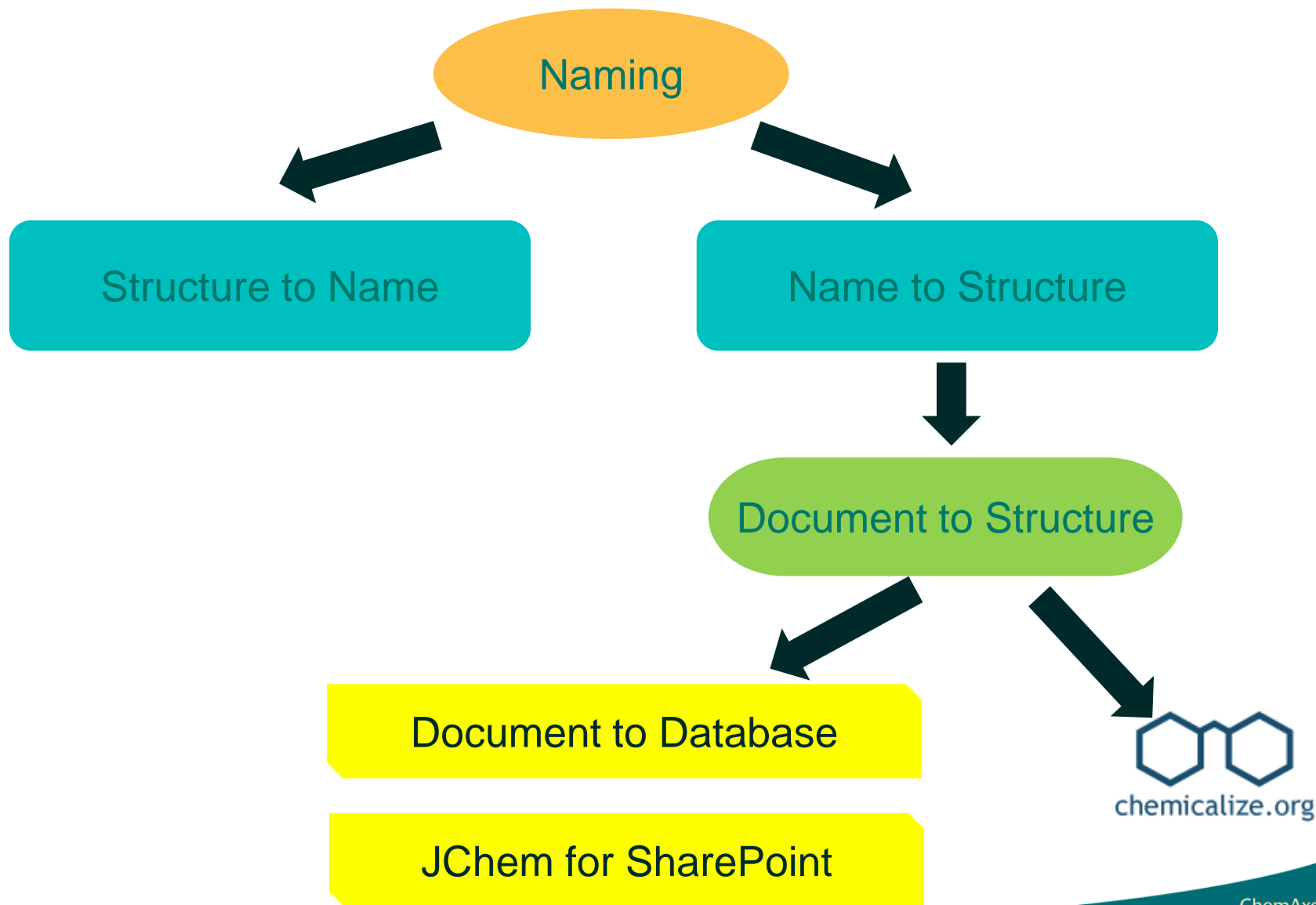
(74) Representative: Minoja, Fabrizio
Bianchetti Bracco Minoja S.r.l.
Via Plinio, 63
20129 Milano (IT)

(54) TRPV1 vanilloid receptor antagonists with a bicyclic portion

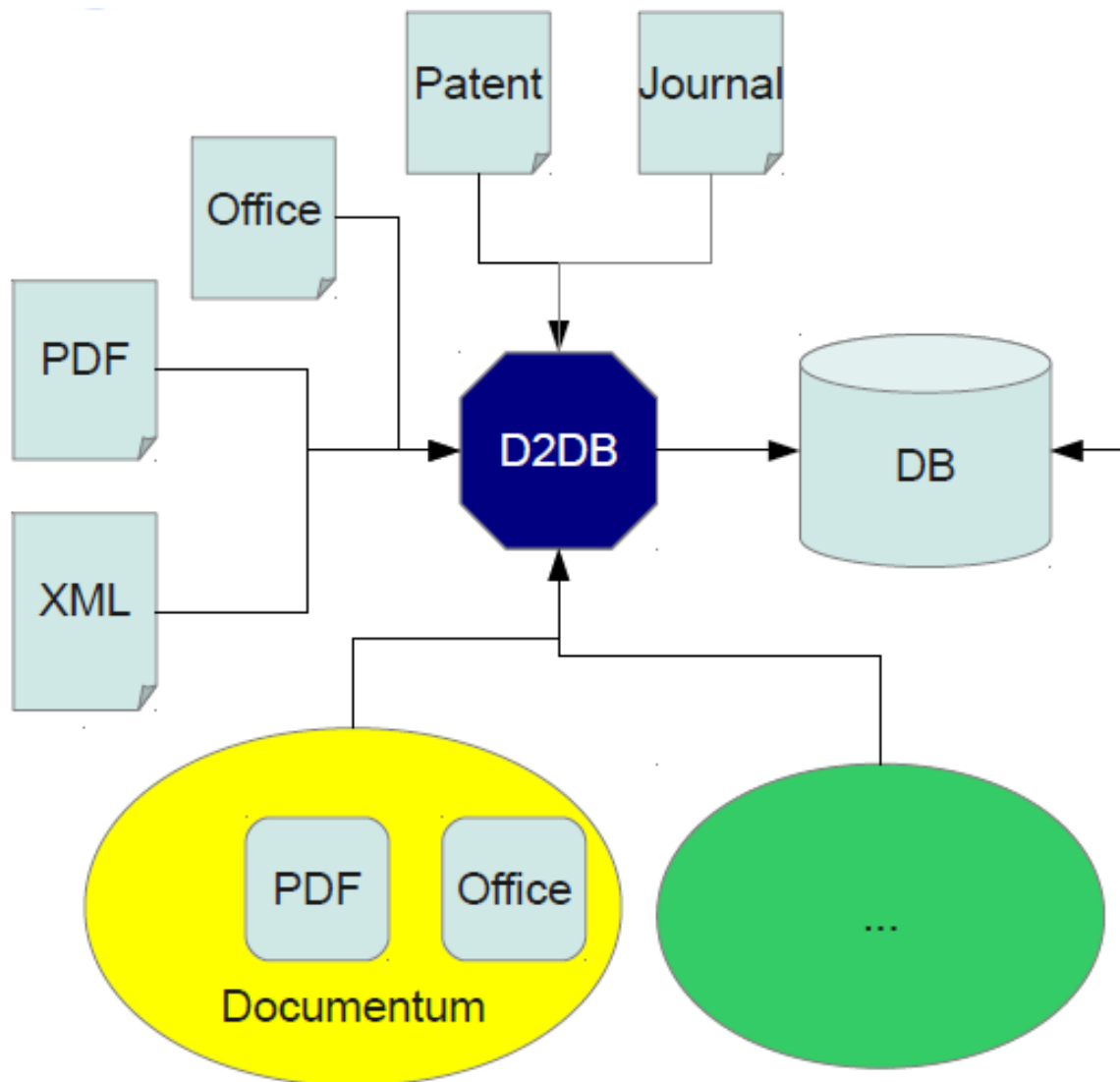
Bookmarks

- Bibliography
- Claims
- Description
 - Description 1
 - Description 2
- Abstract
- Search-Report

Too Many Documents to Manage?



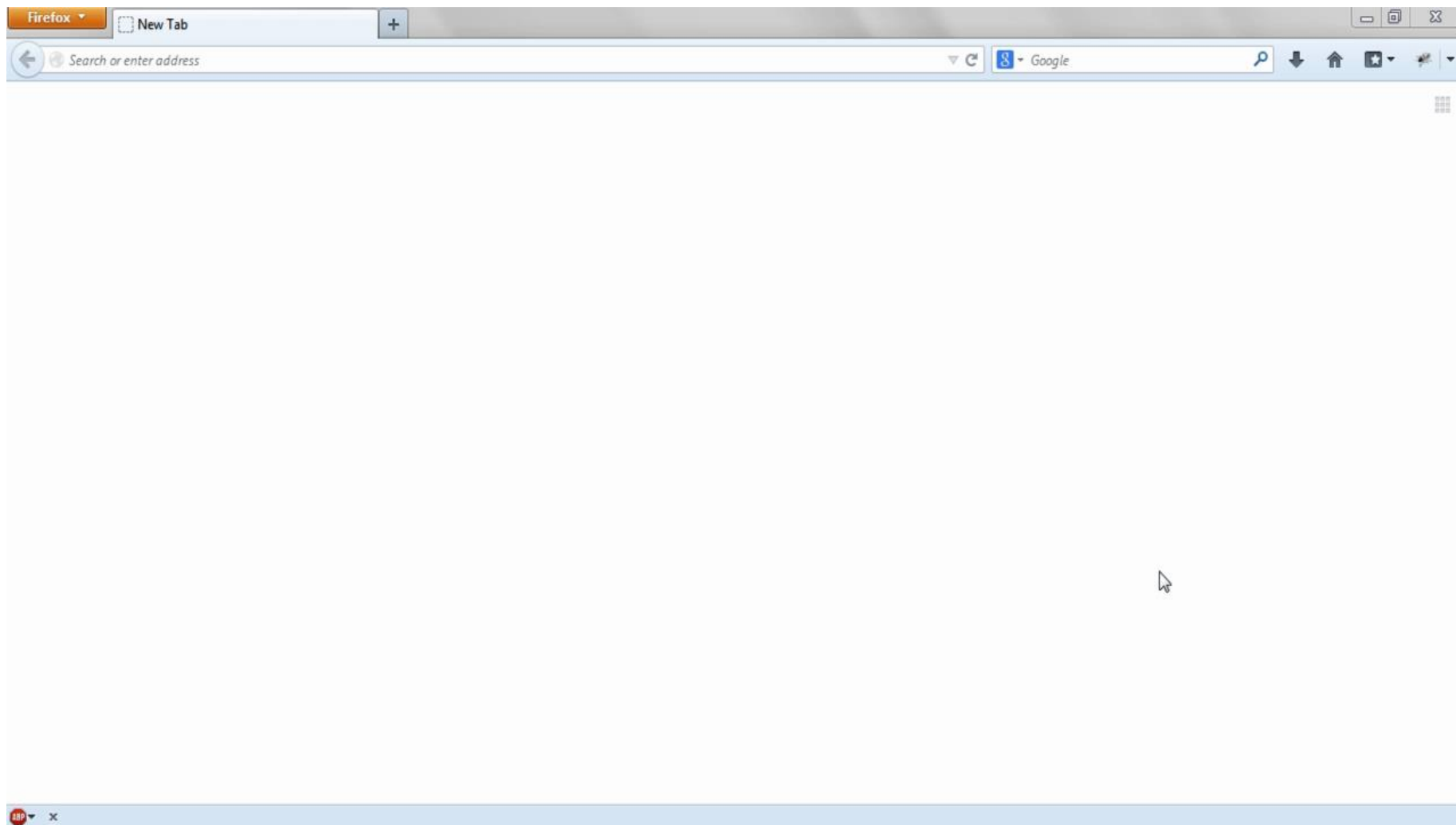
Document to Database



- Chemical information
- Document metadata
- Searchable
- Customizable
- Document annotation



Document to Database Demo



Document Annotation

The screenshot shows a Windows Explorer window with the following details:

- Address Bar:** David > Dropbox > ChemAxon_backup > My_Demos > Each_product > Naming_D2S > annotated_documents
- Search:** Search annotated_documents
- Menu Bar:** Organize, Open with Adobe Reader XI, Share with, Print, New folder
- Table of Files:**

Name	Date modified	Type	Size
US08088500-20120103_files	8/26/2013 4:50 PM	File folder	
200710037874.html	8/19/2013 12:23 PM	Firefox HTML Doc...	2,557 KB
200710037874NEW.XML	1/6/2009 11:24 AM	XML Document	177 KB
ci200280m.pdf	10/12/2011 10:38 ...	Adobe Acrobat D...	3,660 KB
ci200280m_annotated.html	7/2/2013 7:04 AM	Firefox HTML Doc...	4,088 KB
US08088500-20120103.html	8/26/2013 4:50 PM	Firefox HTML Doc...	5,765 KB
US08088500-20120103.xml	8/29/2013 2:08 PM	XML Document	322 KB

Taskbar: ci200280m.pdf, Adobe Acrobat Document, Date modified: 10/12/2011 10:38 PM, Date created: 8/26/2013 4:48 PM, Size: 3.57 MB

In Summary, *Naming* Becomes

- Faster and more accurate
- Multilingual
- More versatile
- Backbone of Text Mining

What's Coming for *Naming*?

- Other language support in N2S and D2S
- D2S: MultiCore OCR/OSR
- D2DB: OpenText, Google Drive

Acknowledgements

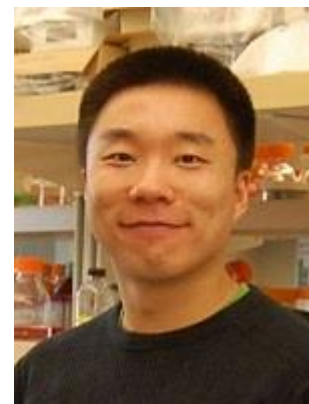
Daniel Bonniot

N2S



邓巍

CN2S



Gustavo Santucho
Mario Burdman