

# Interfacing to JChem for Excel

- the API-rich solution

*Jan Holst Jensen*  
*CEO, Biochemfusion*

***biochemfusion***  
***- Enabling biochemformatics***

# Proteins – it's all the rage



# Two (mostly separate) worlds

## Macromolecules / sequences

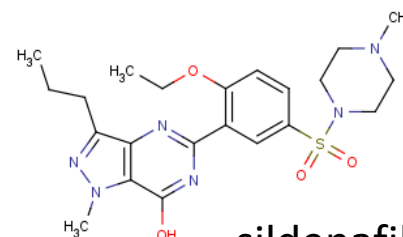
1 HAEGTFTSDVSSYLEGQAAK  
21 EFIAWLVKGR GLP-1

1 GIVEQCCTSI C SLYQLENYC  
21 NFVNQHL C GSHLVEALYLVC  
41 GERGFFYTPKT Human Insulin

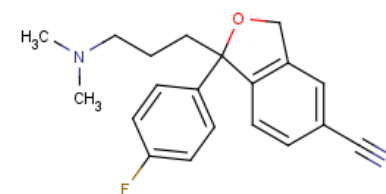
1 GCCSDPRCAWRC  
Alpha-conotoxin Iml

Proteax for Spreadsheets  
*biochemfusion*

## Small molecules / 2D graphs



sildenafil



citalopram

JChem for Excel



Evaluation of pK<sub>a</sub> Estimation Methods on 211 Druglike Compounds

John Manchester\*, Grant Walkup, Olga Rivin and Zhiping You

Infection Discovery, AstraZeneca R&amp;D Boston, 35 Gatehouse Drive, Waltham, Massachusetts 02451

*J. Chem. Inf. Model.*, 2010, 50 (4), pp 565-571

DOI: 10.1021/ci100019p

Publication Date (Web): March 12, 2010

Copyright © 2010 American Chemical Society

\* Corresponding author e-mail: [john.manchester@astrazeneca.com](mailto:john.manchester@astrazeneca.com); tel: (781) 839-4844; fax: (781) 839-4640.

Abstract

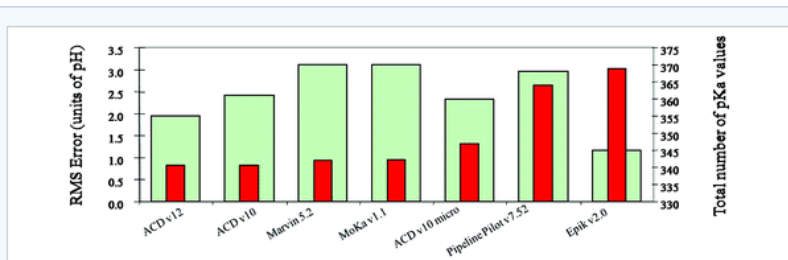
Supporting Info

Full Text HTML

Hi-Res PDF [1610 KB]

PDF w/ Links [268 KB]

## Abstract



The pK<sub>a</sub> values of 211 discovery (druglike) compounds were determined experimentally using capillary electrophoresis coupled with ultraviolet spectroscopy and a novel fitting algorithm. These values were compared to those predicted by five different commercially available pK<sub>a</sub> estimation packages: ACDLabs/pK<sub>a</sub>, Marvin (ChemAxon), MoKa (Molecular Discovery), Epik (Schrodinger), and Pipeline Pilot (Accelrys). Even though the topological method MoKa was noticeably faster than ACD, the accuracy of those two methods and Marvin was statistically indistinguishable, with a root-mean-squared error of about 1 pK<sub>a</sub> unit compared to experiment. Pipeline Pilot and Epik both produced pK<sub>a</sub> estimates in significantly worse agreement with the experiment. Interestingly, on a number of compounds, the predictions due to ACD v12 were in poorer agreement with the experiment than ACD v10. Microscopic and “apparent” pK<sub>a</sub> predictions were also compared using ACD v10. Microscopic pK<sub>a</sub>s gave significantly worse agreement with the experiment than the “apparent” values. In all cases, the errors appeared to be randomly distributed across chemical series.

# JChem – High quality chemical predictors

- If pK<sub>a</sub> works well...
  - then pI prediction is probably OK as well
- But does it work for peptides too ?

## Chemical Modification of Conotoxins to Improve Stability and Activity

David J. Craik, and David J. Adams, \*

† Institute for Molecular Bioscience

‡ School of Biomedical Sciences

The University of Queensland, Brisbane 4072, Australia

ACS Chem. Biol., 2007, 2 (7), pp 457-468

DOI: 10.1021/cb700091j

Publication Date (Web): July 20, 2007

Copyright © 2007 American Chemical Society

\* Corresponding author, [dadams@uq.edu.au](mailto:dadams@uq.edu.au).

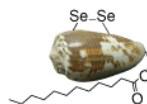
Abstract

Full Text HTML

Hi-Res PDF [780 KB]

PDF w/ Links [542 KB]

### Abstract



Conotoxins are small disulfide-rich peptides from the venom of cone snails. Along with other conopeptides, they target a wide range of membrane receptors, and because of their high potency, they have attracted a great deal of attention for drug development. However, like most peptides, conotoxins have disadvantages of poor absorption, poor stability, and poor activity. Various chemical approaches, including residue substitutions, backbone cyclization, and disulfide-bridge modification, have been reported to increase the stability of conopeptides. These manufactured interventions add to the array of post-translational modifications that occur naturally in conopeptides. They enhance the versatility of these peptides as tools in neuroscience and as drug leads.

<http://pubs.acs.org/doi/abs/10.1021/cb700091j>

<http://www.ncbi.nlm.nih.gov/pubmed/16500898>

Display Settings:  Abstract

Send to:

J Biol Chem. 2006 May 19;281(20):14136-43. Epub 2006 Feb 24.

## Alpha-selenoconotoxins, a new class of potent alpha7 neuronal nicotinic receptor antagonists.

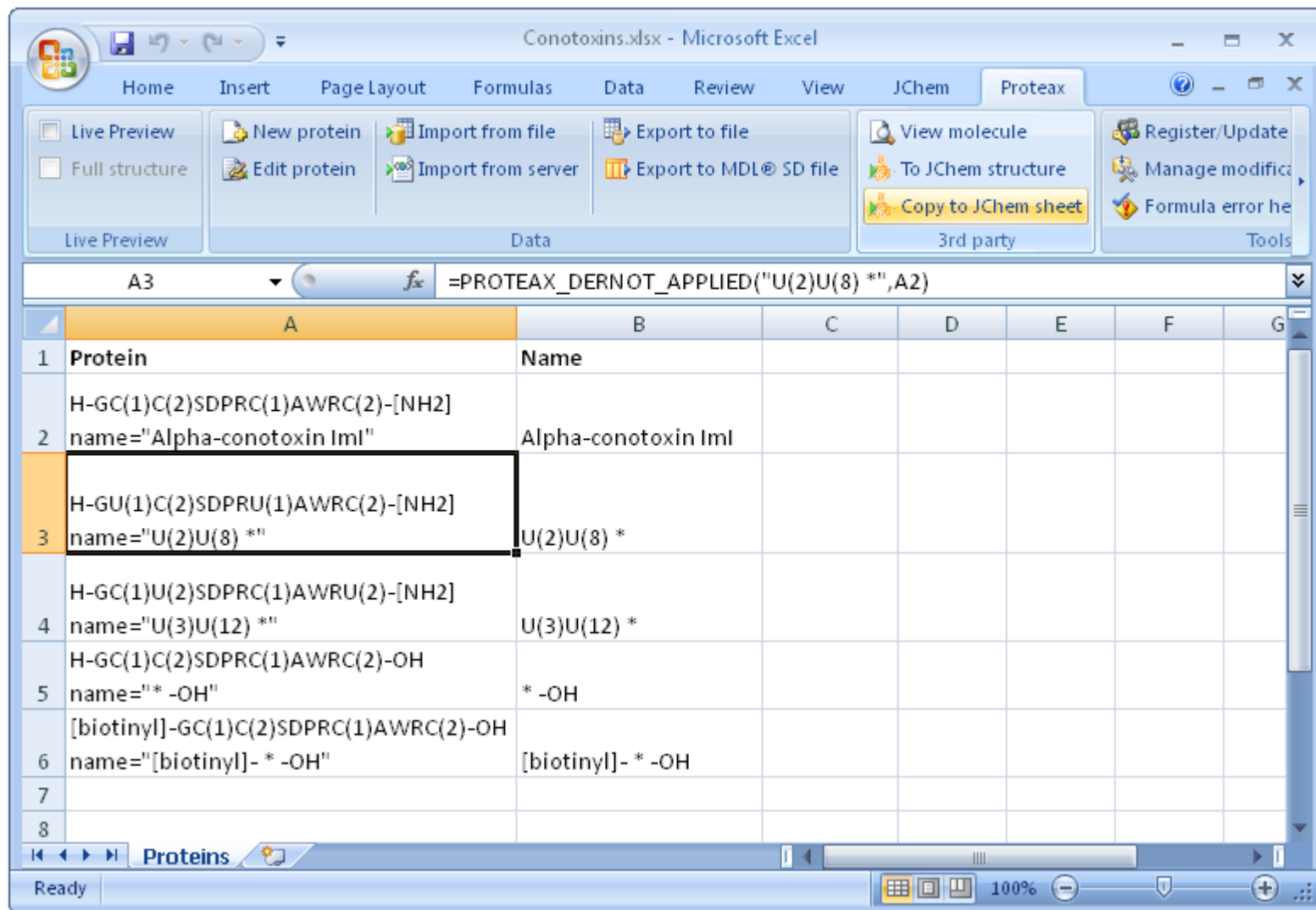
Armishaw CJ, Daly NL, Nevin ST, Adams DJ, Craik DJ, Alewood PF.

Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia.

### Abstract

Disulfide bonds are important structural motifs that play an essential role in maintaining the conformational stability of many bioactive peptides. Of particular importance are the conotoxins, which selectively target a wide range of ion channels that are implicated in numerous disease states. Despite the enormous potential of conotoxins as therapeutics, their multiple disulfide bond frameworks are inherently unstable under reducing conditions. Reduction or scrambling by thiol-containing molecules such as glutathione or serum albumin in intracellular or extracellular environments such as blood plasma can decrease their effectiveness as drugs. To address this issue, we describe a new class of selenoconotoxins where cysteine residues are replaced by selenocysteine to form isosteric and nonreducible diselenide bonds. Three isoforms of alpha-conotoxin Iml were synthesized by t-butoxycarbonyl chemistry with systematic replacement of one ([Sec(2,8)]Iml or [Sec(3,12)]Iml), or both ([Sec(2,3,8,12)]Iml) disulfide bonds with a diselenide bond. Each isoform demonstrated potent and stable antagonism of alpha7 neuronal nicotinic receptors. These results demonstrate that selenoconotoxins can be used as highly stable scaffolds for the design of new drugs.

# *In silico* protein synthesis



The screenshot shows the Microsoft Excel interface with the Proteax ribbon active. The active cell is A3, containing the formula `=PROTEAX_DERNOT_APPLIED("U(2)U(8) **",A2)`. The table below shows the results of this formula, with the third row highlighted.

Protein	Name
H-GC(1)C(2)SDPRC(1)AWRC(2)-[NH2] name="Alpha-conotoxin lml"	Alpha-conotoxin lml
H-GU(1)C(2)SDPRU(1)AWRC(2)-[NH2] name="U(2)U(8) **"	U(2)U(8) *
H-GC(1)U(2)SDPRC(1)AWRU(2)-[NH2] name="U(3)U(12) **"	U(3)U(12) *
H-GC(1)C(2)SDPRC(1)AWRC(2)-OH name="* -OH"	* -OH
[biotinyl]-GC(1)C(2)SDPRC(1)AWRC(2)-OH name="[biotinyl]- * -OH"	[biotinyl]- * -OH

# Transfer to JChem for Excel

The screenshot shows the Microsoft Excel interface with the JChem Proteax ribbon. The Proteax ribbon includes options for Live Preview, Full structure, New protein, Edit protein, Import from file, Import from server, Export to file, Export to MDL@ SD file, View molecule, To JChem structure, Copy to JChem sheet, Register/Update, Manage modifications, and Formula error handling. A dialog box titled "Copy data to JChem for Excel sheet" is open, showing the following settings:

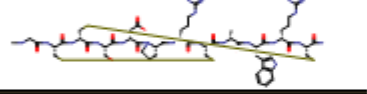
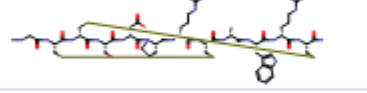
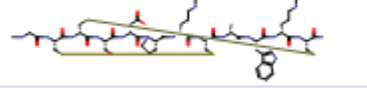
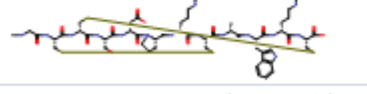
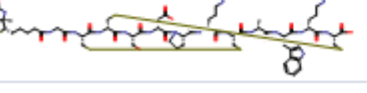
- SD fields:  First row defines field names
- Mol Field name: Protein (checked), Name (checked)
- Output format:  Full structures,  Condensed structures
- Place in:  New JChem sheet,  Existing sheet (Sheet1)

The Excel spreadsheet contains the following data:

	A	B
1	Protein	Name
2	H-GC(1)C(2)SDPRC(1)AWRC(2)-[NH2] name="Alpha-conotoxin lml"	Alpha-conotoxin lml
3	H-GU(1)C(2)SDPRU(1)AWRC(2)-[NH2] name="U(2)U(8) **"	U(2)U(8) *
4	H-GC(1)U(2)SDPRC(1)AWRU(2)-[NH2] name="U(3)U(12) **"	U(3)U(12) *
5	H-GC(1)C(2)SDPRC(1)AWRC(2)-OH name="* -OH"	* -OH
6	[biotinyl]-GC(1)C(2)SDPRC(1)AWRC(2)-OH name="[biotinyl]- * -OH"	[biotinyl]- * -OH
7		
8		

# And we have live structures...

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H
1	<b>Structure</b>	<b>Name</b>						
2		Alpha-conotoxin lml						
3		U(2)U(8) *						
4		U(3)U(12) *						
5		*-OH						
6		[biotinyl]- *-OH						
7								

The formula bar shows: `=JCSYSstructure("4A0DE978B01F348A38365E242B3C4246")`

# Adding pI predictions

The screenshot shows the Microsoft Excel interface with the JChem ribbon active. The formula bar displays `=JCIsoelectricPoint(A2)`. The spreadsheet has columns A through H and rows 1 through 7. Row 1 contains headers: A1: Structure, B1: Name, C1: pI, D1: pI delta. Row 2 shows a chemical structure in A2, 'Alpha-conotoxin lml' in B2, and the formula `=int(A2)` in C2. A dialog box titled 'Function Arguments' is open, showing the function `JCIsoelectricPoint` with the argument 'Molecule' set to 'A2'. The result is displayed as `= 11.81656906`. The dialog also includes a 'Help on this function' link and 'OK' and 'Cancel' buttons.

	A	B	C	D	E	F	G	H
1	Structure	Name	pI	pI delta				
2		Alpha-conotoxin lml	<code>=int(A2)</code>					
3		U(2)U						
4		U(3)U						
5		*-OH						
6		[biotin]						
7								

# Results

Conotoxins.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View JChem Proteax

Import from Database Add/Edit From SMILES From Image To SMILES To Image Filter R-group Decomposition Options Help

D3  $=C3-\$C\$2$

	A	B	C	D	E	F	G	H
1	Structure	Name	pI	pI delta				
2		Alpha-conotoxin lml	11.81657					
3		U(2)U(8) *	11.75302	-0.06355				
4		U(3)U(12) *	11.75598	-0.06059				
5		*-OH	9.629033	-2.18754				
6		[biotinyl]-*-OH	7.685836	-4.13073				
7								

Sheet1 Proteins

Ready 100%

# pl-s from the real world

<http://www.jbc.org/content/99/3/741.full.pdf+html>

The screenshot shows the top of a web browser window. The browser's address bar contains the URL <http://www.jbc.org/content/99/3/741.full.pdf+html>. The page header features the logo for 'jbc THE JOURNAL OF BIOLOGICAL CHEMISTRY'. Below the logo is a search bar with fields for 'Author:', 'Keyword:', 'Year:', 'Vol:', 'Page:', and a 'Go' button. A navigation menu includes links for 'Home', 'Current issue', 'Archive', 'Papers in Press', 'Minireviews', 'Classics', 'Reflections', and 'Paper'. The browser's address bar shows '1 / 13' and a search box with the word 'Find'. The main content of the page is a scientific article titled 'THE ISOELECTRIC POINT OF INSULIN' with the subtitle 'ELECTRICAL PROPERTIES OF ADSORBED AND CRYSTALLINE INSULIN\*'. The author is listed as 'BY OSKAR WINTERSTEINER AND HAROLD A. ABRAMSON'. The article's origin is cited as '(From the Department of Biological Chemistry, College of Physicians and Surgeons, Columbia University, New York)'. The date of receipt for publication is '(Received for publication, December 6, 1932)'. The first paragraph of the article discusses the history of insulin purification and the determination of its isoelectric point. A red circle highlights the sentence: 'The pH of a solution from which insulin crystallized in satisfactory yields was 5.60 to 5.65. These solutions contained ammonia, pyridine, brucine, and acetic acid in high concentration. Since such a medium is highly instrumental in holding insulin in'. On the right side of the page, there is a vertical text string: 'Downloaded from www.jbc.org by guest on May 6, 2010'.

<http://www.lantus.com/hcp/works.aspx>

“LANTUS® long-acting (basal) insulin differs from human insulin in that the physio/chemical properties have been modified. The amino acid asparagine at position A21 is replaced by glycine, and two arginines are added to the C-terminus of the B-chain. The effect of these changes is to shift the isoelectric point, producing a solution that is completely soluble at pH 4.1.”

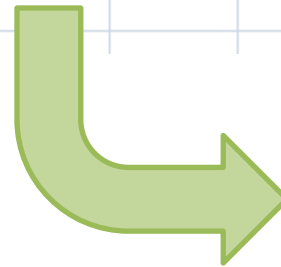
[http://en.wikipedia.org/wiki/Insulin\\_analog](http://en.wikipedia.org/wiki/Insulin_analog)

## “Glargine insulin

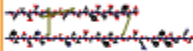
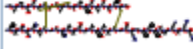
Sanofi-Aventis developed glargine as a longer lasting insulin analogue, and markets it under the trade name Lantus. It was created by modifying three amino acids. Two positively charged arginine molecules were added to the C-terminus of the B-chain, and they shift the isoelectric point from 5.4 to 6.7, making glargine more soluble at a slightly acidic pH and less soluble at a physiological pH.”

# Just checking the walls...

	A	B	C	D	E	F
1	Protein	Name				
2	H-GIVEQC(1)C(2)TSI	Human insulin				
3	H-GIVEQC(1)C(2)TSI	G(A21) * -RR-(B)				



Spot-on!

	A	B	C	D
1	Structure	Name	pI	
2		Human insulin	5.642626	
3		G(A21) * -RR-(B)	7.207238	
4				

[http://expasy.org/cgi-bin/pi\\_tool](http://expasy.org/cgi-bin/pi_tool)

10 20 30 40 50  
GIVEQCCTSI CSLYQLENYC NQVFNQHLGCS HLVEALYLVC GERGFYTPK T

Theoretical pI/Mw: 5.39 / 5795.65

10 20 30 40 50  
GIVEQCCTSI CSLYQLENYC GFVNQHLGCS HLVEALYLVC GERGFYTPK TRR

Theoretical pI/Mw: 6.88 / 6050.97

# Talking to JChem – what's it like ?

- JChem for Excel public API
  - Well-documented, plenty of examples
- Industry standard data formats
  - Data exchange via MDL SDF files
- Bottom line
  - A walk in the park
  - You should be familiar with VBA

# Go forth...

- Grab JChem for Excel
  - <http://www.chemaxon.com/download/jchem-for-excel/>
  - and add some calculators for good fun
  
- Grab Proteax for Spreadsheets
  - <http://www.biochemfusion.com/downloads/>
  - "1.2 2010-05" with the JChem for Excel interface