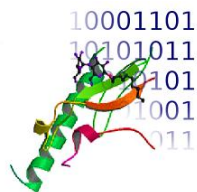


Sequence or structure

- "You say tomaytoe, I say tomahto"

Gershwin, ca. 1937

Jan Holst Jensen
CEO, Biochemfusion



biochemfusion
- Enabling biochemformatics

Dialects in representation

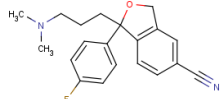
Cheminformatics

Bioinformatics

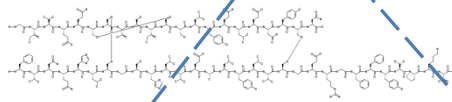
neither-nor / both-and

Molecule graphs

Sequences



```
1 GIVEQCCTSICSLYQLENYC  
21 NFVNQHLCGSHLVEALYLVC  
41 GERGFFYTPKT
```



```
1 MVSQALRLLCLLLCLOGCCLAAGGVAKASGCETRDMPWRPG  
41 PRRVFTQEEAHCVLHRRRANAFLEELRPGSLERICKRE  
81 QCSFEARLIFKDAERTKLFMISYSDGDCASSPCQNGGS  
121 CRDQLQSYICFCLPAFEGRCNETHRDDQLICVNMENGGCRQ  
161 YCSDHTGTRRSRCRCHEGYSLLADGVSCPTVEYPCGRIPY  
201 LEKRNASKPQGRIVGCKVCFKGCPCPQVLLLVNCAQLCGG  
241 LLINTIUVVSAACHCFDKIKNWRNLIAVLGEHDLSEHDCDE  
281 QSRRVAQVIIPSTYVPGTINHDIALLRLHQPVVLTDHVV  
321 LCLPRTFSERTLAFVRFSLVSGWCQLLDRCATALELMV  
361 NVPRLHTQDCLQQSRKWCDSPNITETHFCACYSDGSRDSC  
401 KGDSCGPHATHYRGTWYLTGIVSMGCCCATVGHFGVYTRV  
441 SQYIEWLQKLMRSEPRPCVLLRAPFF
```

100

10k

1M

MW
Da

Two-way translation with Proteax[®]

PubChem structures imported from SD file into Proteax for Spreadsheets

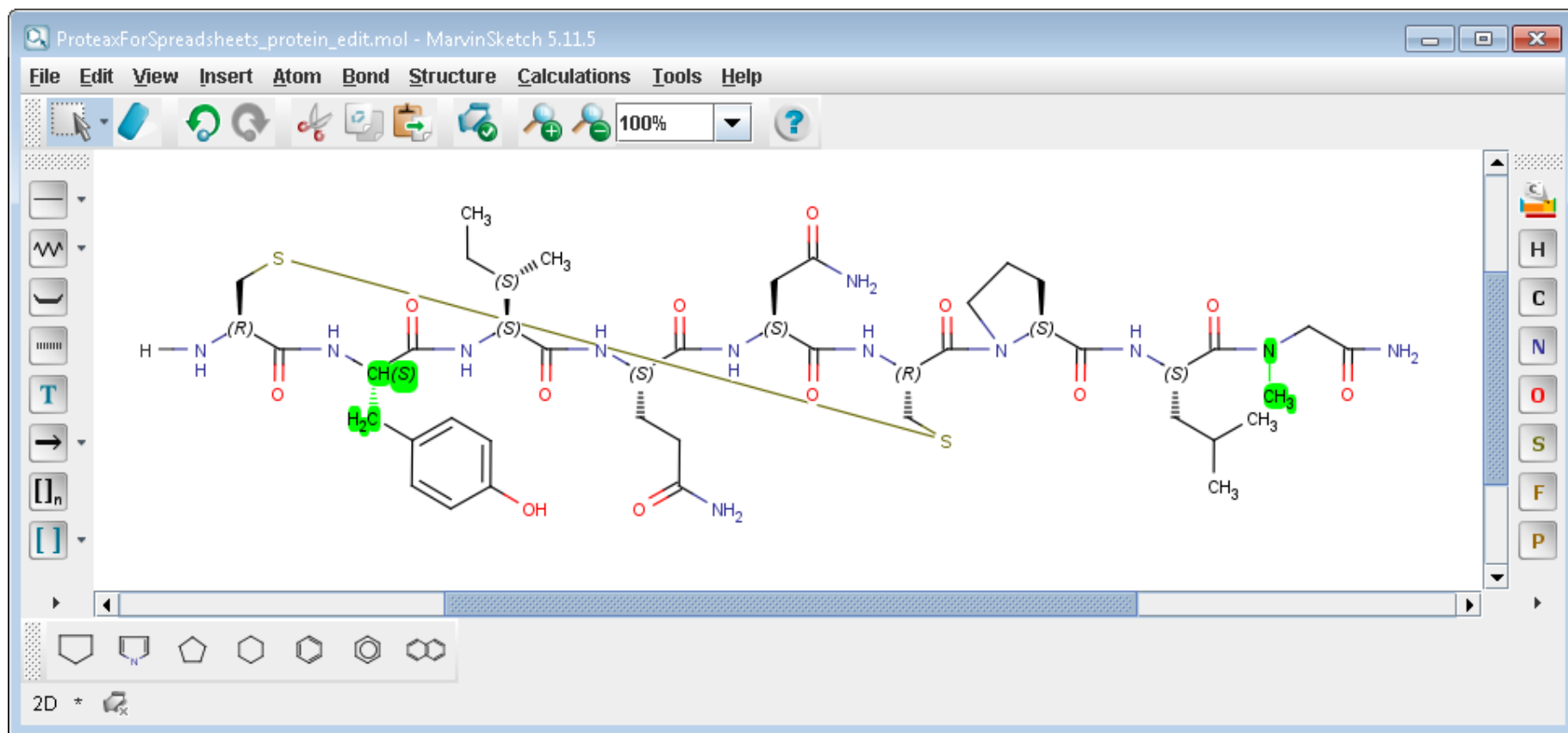
The screenshot shows the Proteax software interface with a spreadsheet of chemical structures. The spreadsheet has columns for structure names and PubChem Compound IDs. An 'Editor options' dialog box is open, allowing the user to choose an editor for the structures. The 'Use chemical structure editor' option is selected, and the path to the executable file is entered.

| Structure | PUBCHEM_COMPOUND_CID |
|---|----------------------|
| H-Cys(1)-dTyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-[NH2] name=638315 | 638315 |
| H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-[NH2] name=439302 | 439302 |
| H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-[NH2] name=194531 | 194531 |
| H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-[NH2] name=167997 | 167997 |
| H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-[NH2] name=63477758 | 63477758 |

Editor options dialog box content:

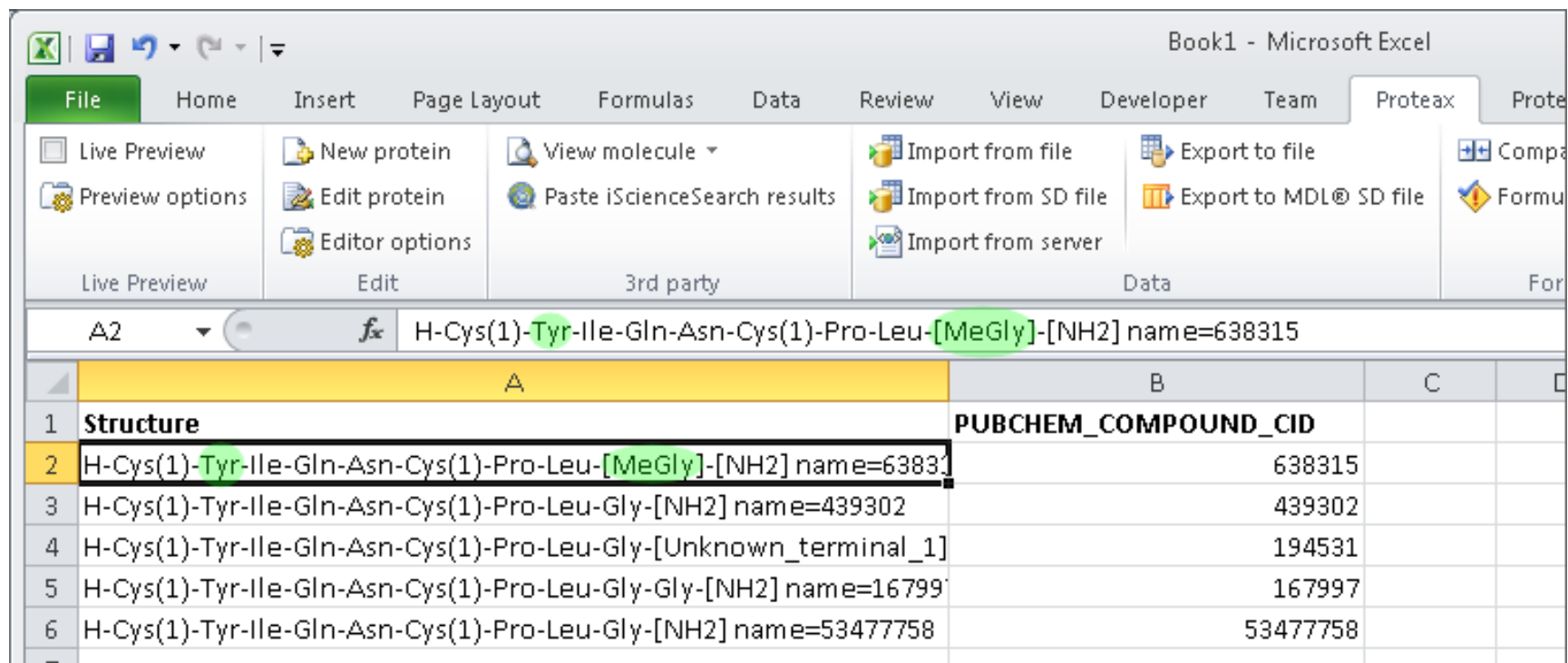
- Editor choice
 - Use Proteax's PLN editor
 - Use GPMW editor
GPMW is a 3rd party product. For more information, visit <http://gpmw.com>
 - Use chemical structure editor
- Editor executable
 -
- Buttons: Use .MOL editor, Use ISIS/Draw helper, OK, Cancel

Edit Proteax-generated structure in MarvinSketch



Change Tyr-2 back into its L-form and methylate the glycine.

File→Save in MarvinSketch detected: Proteax translates back to sequence



Book1 - Microsoft Excel

File Home Insert Page Layout Formulas Data Review View Developer Team Proteax Prote

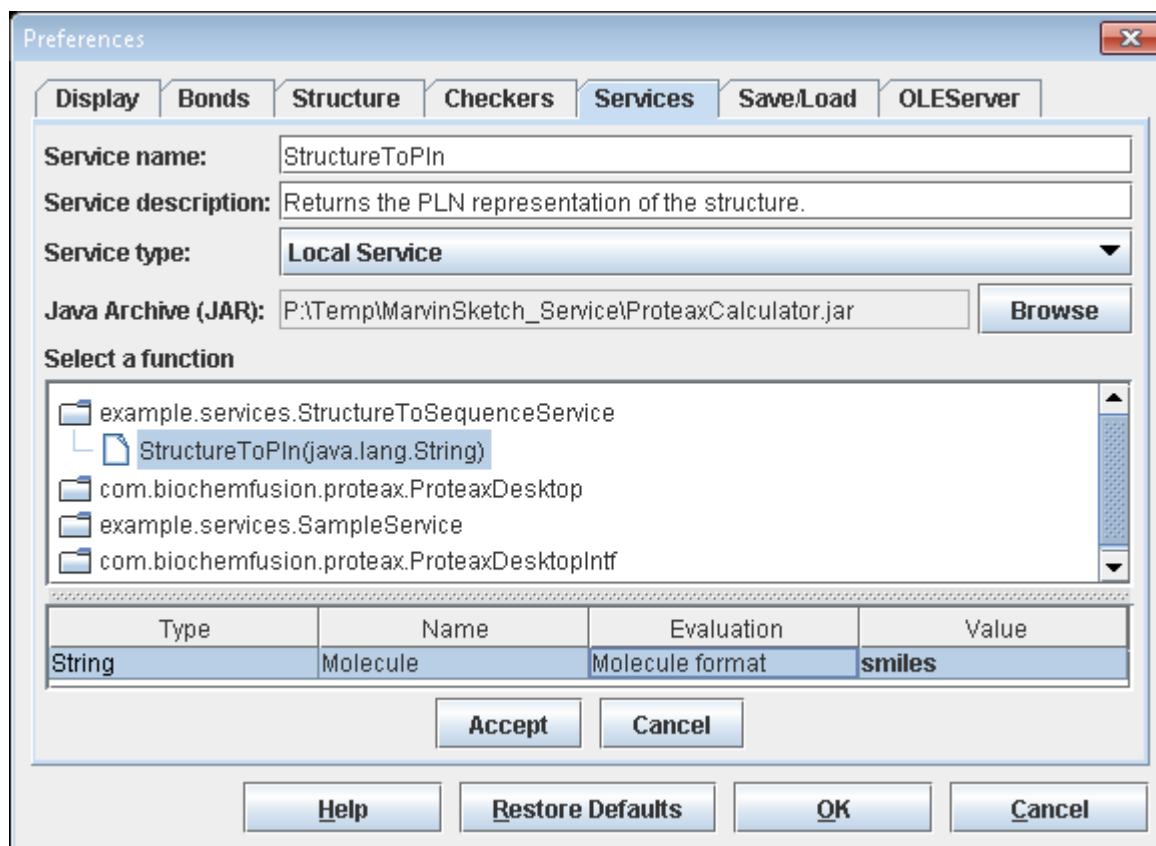
Live Preview Preview options Live Preview New protein Edit protein Editor options Edit 3rd party View molecule Paste iScienceSearch results Import from file Import from SD file Import from server Data Export to file Export to MDL@ SD file Compa Formu

A2 fx H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-[MeGly]-[NH2] name=638315

| | A | B | C | D |
|---|---|-----------------------------|---|---|
| 1 | Structure | PUBCHEM_COMPOUND_CID | | |
| 2 | H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-[MeGly]-[NH2] name=638315 | 638315 | | |
| 3 | H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-[NH2] name=439302 | 439302 | | |
| 4 | H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-[Unknown_terminal_1] | 194531 | | |
| 5 | H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-Gly-[NH2] name=167997 | 167997 | | |
| 6 | H-Cys(1)-Tyr-Ile-Gln-Asn-Cys(1)-Pro-Leu-Gly-[NH2] name=53477758 | 53477758 | | |

Proteax calculator service example for MarvinSketch

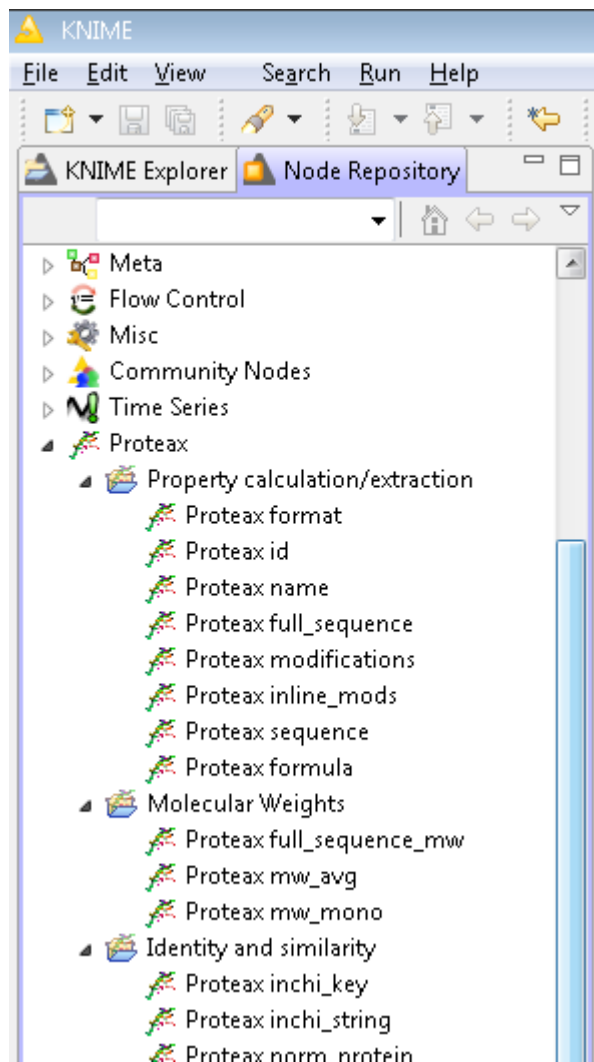
- Just add protein chemistry intelligence



Proteax calculator service in action

The image displays two software windows. The left window, titled 'pubchem_20225183.sdf - MarvinSketch 5.11.5', shows a chemical structure of a peptide chain with a long aliphatic chain and a terminal carboxylic acid group. The right window, titled 'StructureToPln', shows the same structure in a smaller view, with an 'Update' button below it. Below the structure view, the 'Arguments' section contains a 'Molecule' field with the value 'smiles'. The 'Results' section displays the output: '[Unknown_terminal_1]-Gly-Gly-OH inline-mod=N-terminal,[U'. At the bottom of the right window, there is a 'Calculate automatically' checkbox and a 'Calculate' button.

Proteax in KNIME



- Proteax KNIME nodes
 - Released with Proteax Desktop 2.0
- Free download for Proteax users

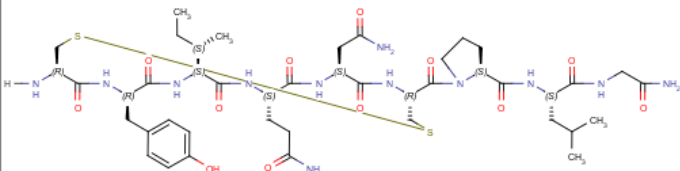
http://biochemfusion.com/downloads/proteax_knime_nodes/

Easily combined with Marvin nodes

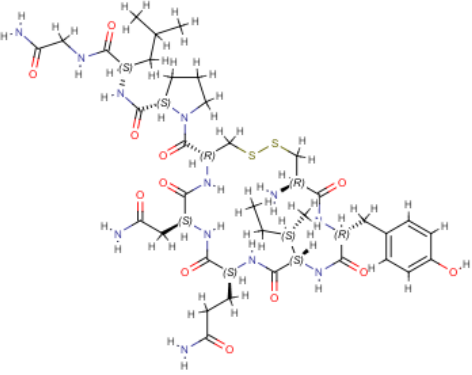
The screenshot shows a KNIME workflow titled "2: Clean peptide structures". The workflow consists of the following nodes:

- SDF Reader**: Read PubChem peptide structures
- Proteax as_pln**: To sequence
- Proteax as_molfile**: To molecule
- Molecule Type Cast**: Molfile type conversion
- MarvinView**: Structures cleaned by Proteax
- MarvinView**: Structures as shown in PubChem

Two MarvinView windows are shown, displaying chemical structures. The top window, titled "MarvinView - 2:5 - MarvinView(Structures as)", shows a table with the following data:

| # | structure | \$MolName | Molec. |
|------|--|-----------|---------|
| Row0 |  | 638315 | [H]OC1- |
| Row1 | | 439302 | [H]OC1- |

The bottom window, titled "MarvinView - 2:2 - MarvinView(Structures as)", shows a table with the following data:

| # | structure | \$Mol |
|------|--|-------|
| Row0 |  | 6383 |

Cleaning peptide structures with Proteax and presenting them in MarvinView

iScienceSearch – by AKos GmbH

Firefox | iSS Querypage | Biochemfusion - Proteax - protein edi... | +

isciencesearch.com/iss/proteax.aspx

Include similar structures 90%

Protein text - PLN format

H-Cys (1) -Tyr-Ile-Gln-Asn-Cys (1) -Pro-Leu-Gly-[NH2]

N-terminal C-terminal

Ala Arg Asn Asp Cys Glu Gln Gly His Ile Leu Lys Met

Phe Pro Ser Thr Trp Tyr Val Sec Pyl Xaa

Sequence Molecule Sum formula: C43 H66 N12 O12 S2 Avg. MW: 1007,18734

| | |
|---|-----------|
| 1 | CYIQNCPLG |
|---|-----------|

1:1

- Searches structure databases by queries created from sequence by Proteax
- See also the Biochemfusion Partner Presentation 2012
 - iScienceSearch was formerly known as GlobalSearch

Links...

- Proteax for Spreadsheets, Proteax Desktop, Proteax KNIME nodes
 - <http://biochemfusion.com/downloads/>
- iScienceSearch, by AKos GmbH
 - <http://isciencesearch.com/iss/default.aspx>
- Sysment Notebook and Reaction Planner
 - <http://www.sysment.hu/>
 - Sysment is using Proteax Desktop so their product line can support both small and large molecules