
Proteax Cartridge

- protein chemistry for Oracle® databases

biochemfusion



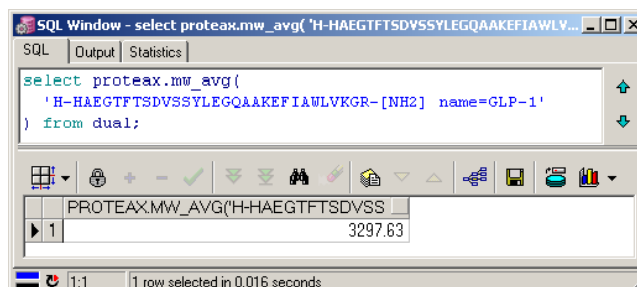
Biochemfusion Aps
Copenhagen, Denmark

Mail: info@biochemfusion.com
Phone: +45 3048 0050
Web: <http://www.biochemfusion.com>

Proteax Cartridge - protein chemistry for Oracle® databases

Proteax® Cartridge lets you work with chemically or post-translationally modified protein sequences in Oracle® databases via normal SQL statements.

- Directly supports modified residues, terminals, disulfide bridges and lactam cyclizations.
- Average and mono-isotopic molecular weights based on the actual molecular structure of the protein entries.
- Protein comparison, manipulation, and SAR table reports via DerNot (**Derivatives Notation**) expressions.
- Sequence and structural identity checks.
- On-the-fly generation of 2D chemical structures for visualization or chemical indexing.



A simple SQL statement that calculates the average molecular weight of the GLP-1 molecule.

Industry standard formats

Proteax Cartridge accepts protein entries in the following standard protein file formats:

- UniProt - the UniProt Consortium
<http://www.uniprot.org>
- GPMW - by Lighthouse Data
<http://www.gpmw.com>
- FASTA - (unmodified sequences only)
<http://www.ncbi.nlm.nih.gov/blast/fasta.shtml>

```
select protein_text, proteax.mw_mono(protein_text)
from mixed_compounds;
```

PROTEIN_TEXT	PROTEAX.MW_MONO(PROTEIN_TEXT)
1 ID 14331_MAIZE AC P49106; DT 01-FEB-1996	29643.9
2 H-GIVEQC(1)C(2)TSIC(1)SLYOLENYC(3)N-OH H-FVNGHLC(2)GSHLVEALYLVC(3)GERGFFYPKT-OH name="Human insulin"	5803.64
3 Exendin-4_NCBIS YCT 39 H2N1 NH2 HGEFTFTSDLSKQMEEEAVRLFIIEWLKNKGGPSSGAPPPS*	4184.03
4 >Human insulin GIVEQCCTSIICSLYOLENYCNFVNOHLSHSLVEALYLVCGER GFFYPYTKT	5791.67

Any supported protein input format is accepted.

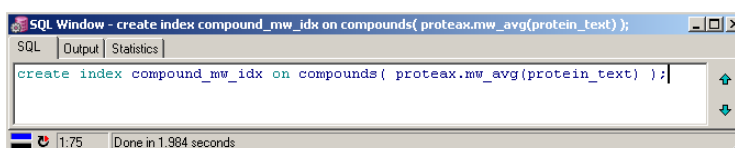
In addition, Biochemfusion's own PLN format lets you work with complex protein structures in an easily read format. The PLN notation was designed with specific attention to trouble-free e-mail and spreadsheet exchange of protein entries.

Protein entries can be freely converted to any of the formats PLN, UniProt, and GPMW without loss of chemical information.

Functional indexing

Proteax Cartridge provides a number of functions that can be indexed by standard Oracle commands.

Since all Proteax indexes build on Oracle's normal support for functional indexes the Oracle optimizer will be used correctly.



Indexes are created by normal SQL commands.

```
select proteax.name (protein_text)
from compounds
where proteax.mw_avg(protein_text) > 20000;
```

Description	Object owner	Object name	Cost	Cardinality	Bytes
SELECT STATEMENT, GOAL = ALL_ROWS			7	46	4002
TABLE ACCESS BY INDEX ROWID	BCF_DEMO_...	COMPOUNDS	7	46	4002
INDEX RANGE SCAN	BCF_DEMO_...	COMPOUND_MW_IDX	2	8	

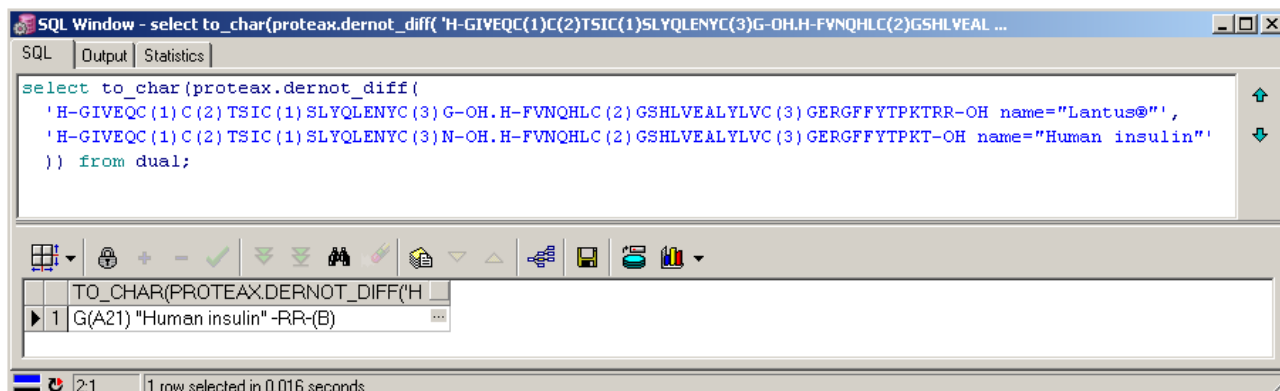
Indexed Proteax functions are utilized as expected by the Oracle cost-based optimizer.

Try the online demo at <http://www.biochemfusion.com/demo>

Contact info@biochemfusion.com or call +45 3048 0050 to obtain a free trial license.

Creating and naming protein derivatives

DerNot (**Derivatives Notation**) expressions are strings that resemble IUPAC trivial naming of peptides. Proteax Cartridge can calculate the DerNot expression that describes the difference between two protein entries.



```
SQL Window - select to_char(proteax.dernot_diff('H-GIVEQC(1)C(2)TSIC(1)SLYQLENYC(3)G-OH.H-FVNQHL(2)GSHLVEAL ...
SQL | Output | Statistics
select to_char(proteax.dernot_diff(
  'H-GIVEQC(1)C(2)TSIC(1)SLYQLENYC(3)G-OH.H-FVNQHL(2)GSHLVEALYLVC(3)GERGFFYTPKTRR-OH name="Lantus@"',
  'H-GIVEQC(1)C(2)TSIC(1)SLYQLENYC(3)N-OH.H-FVNQHL(2)GSHLVEALYLVC(3)GERGFFYTPKT-OH name="Human insulin"'
)) from dual;
```

TO_CHAR(PROTEAX.DERNOT_DIFF('H-GIVEQC(1)C(2)TSIC(1)SLYQLENYC(3)G-OH.H-FVNQHL(2)GSHLVEALYLVC(3)GERGFFYTPKTRR-OH name="Lantus@"', 'H-GIVEQC(1)C(2)TSIC(1)SLYQLENYC(3)N-OH.H-FVNQHL(2)GSHLVEALYLVC(3)GERGFFYTPKT-OH name="Human insulin"'))

1 | G(A21) "Human insulin" -RR-(B)

2:1 | 1 row selected in 0.016 seconds

You may also create derived proteins from existing entries via DerNot expressions.

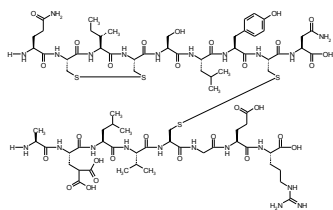
Protein identity checks

Proteax Cartridge contains powerful normalization functions that can be used to determine the identity of two protein entries even in the presence of cyclic chains. Normalization can be done to check sequence identity or to obtain a good measure of structural identity.

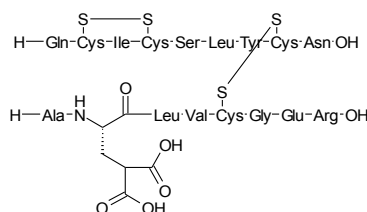
The normalization functions can be indexed to ensure excellent performance of identity checks.

Chemical indexing

Proteax Cartridge can effectively extend the usable molecular size range of existing chemistry cartridges. Chemistry databases will readily index small peptides but will normally not scale well to large proteins. Proteax Cartridge can produce *condensed representation* molecules - a format that enhances registration and search performance by orders of magnitude. This will allow your chemistry database to store and chemically search protein structures with thousands of residues.



Full-structure representation
H-QC(1)IC(1)SLYC(3)N-OH.H-A[G1a]LVC(3)GER-OH



Equivalent condensed representation. All unmodified residues are represented by single "pseudo atoms". Graph complexity is reduced considerably and chemical indexing is therefore much faster.

Proteax Cartridge generates condensed chemical structures that adhere to the rules defined in the article "Building a BioChemformatics Database", Jan H. Jensen, Thomas Hoeg-Jensen and Søren B. Padkjær, *J. Chem. Inf. Model.*, November 2008, <http://pubs.acs.org/doi/abs/10.1021/ci800128b>.

Open Source visualization tools

Sequence and molecule visualization tools can be downloaded for free from the Biochemfusion web site. The tools are released with source code under a BSD-style license which lets you adapt and translate the tools to your environment of choice. Note that the tools are only hundreds of lines of code, not thousands.


The Proteax demo site shows how these tools can be used to visualize protein sequences and molecules.

Rendering		Molecular weights	Da
1	GIVEQCCTSTICSLYQLENYC	Average	5807.57022
21	NFVNQHLCGSHLVEALYLVC		
41	GERGFFYTPKT	Standard UniProt	5795.64974

Part of the Proteax Live demo web page showing sequence graphics created by Biochemfusion's Flash rendering tool.

System requirements

Proteax Cartridge installs on servers running a 32- or 64-bit version of Linux® or Microsoft® Windows®. In order to install Proteax Cartridge you will also need to have an existing Oracle 10g or 11g database running on your server. The following Oracle configurations are supported.

	Oracle 10gR2	Oracle 11gR1+R2	Oracle XE (10g)
Windows	32-bit ✓ 64-bit ✓	32-bit ✓ 64-bit ✓	32-bit ✓
Linux 	32-bit ✓ 64-bit ✓	32-bit ✓ 64-bit ✓	32-bit ✓

Licensing

Commercial licensing details are available by contacting sales@biochemfusion.com. Educational and non-profit research institutions may apply for free academic licenses by contacting info@biochemfusion.com.

Proteax® is a registered trademark of Biochemfusion, <http://www.biochemfusion.com>
Oracle® is a registered trademark of Oracle corp. and its affiliates, <http://www.oracle.com>
GPMaw is a 3rd party product by Lighthouse Data, <http://www.gpmaw.com>
Microsoft® and Windows® are registered trademarks of Microsoft corp. and its affiliates, <http://www.microsoft.com>
Linux® is a registered trademark of Linus Torvalds in the U.S. and other countries
The Tux penguin logo is the creation of Larry Ewing, lewing@isc.tamu.edu
Lantus® is a registered trademark of sanofi-aventis U.S. LLC

Try the online demo at <http://www.biochemfusion.com/demo>
Contact info@biochemfusion.com or call +45 3048 0050 to obtain a free trial license.

May 2011