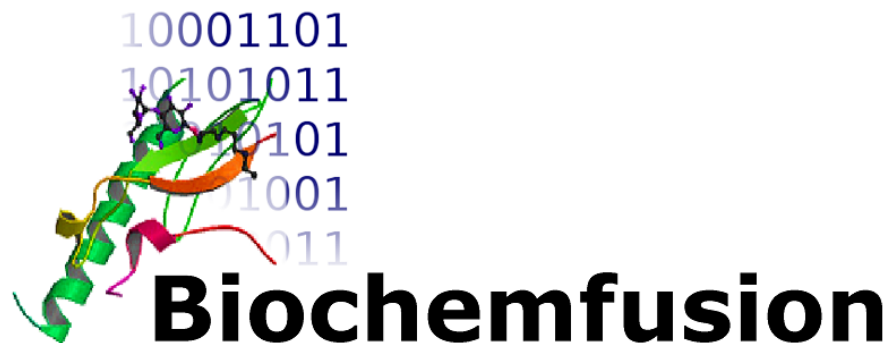


Macromolecule registration and analysis

Jan Holst Jensen
CEO, Biochemfusion ApS



Biochemfusion

- Based in Copenhagen, Denmark
- Developing the Proteax software suite since 2007

- Let's see what it does...

Literature – linear peptide

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3415635/figure/fig01/>

Figure 1

Copyright © 2011 The Author. British Journal of Pharmacology © 2011 The British Pharmacological Society

	7	15	22	36	
GLP-1	HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR				
A7-GLP-1	AAEGTFTSDVSSYLEGQAAKEFIAWLVKGR				
A10-GLP-1	HAEATFTSDVSSYLEGQAAKEFIAWLVKGR				
GLP-1 (9-36)	EGTFTSDVSSYLEGQAAKEFIAWLVKGR				
GLP-1 (15-36)	DVSSYLEGQAAKEFIAWLVKGR				
	1	9	16	30	39
Ex4	HGEGTFTSDLSKQMEEEVARLFI EWLNKGGPSSGAPPPS				
Ex4 (1-30)	HGEGTFTSDLSKQMEEEVARLFI EWLNKGG				
Ex4 (9-39)	DLSKQMEEEVARLFI EWLNKGGPSSGAPPPS				
Ex4 (9-30)	DLSKQMEEEVARLFI EWLNKGG				

A sequence alignment of GLP-1, Ex4 and several analogues mentioned at various stages in the text. All peptides are C-terminally amidated.

PLN (Proteax[®] Protein Line Notation) version :
H-HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-[NH₂]

Literature – cyclic peptide

Cyclosporins are natural undecapeptides derived from cyclosporin A (CsA), cyclo-(-MeBmt¹-Abu²-Sar³-MeLeu⁴-Val⁵-MeLeu⁶-Ala⁷-D-Ala⁸-MeLeu⁹-MeLeu¹⁰-MeVal¹¹-), where MeBmt = (4*R*)-4-[(*E*)-2-butenyl]-4,*N*-dimethyl-L-threonine [(2*S*,3*R*,4*R*,6*E*)-3-hydroxy-4-methyl-2-(*N*-methylamino)-6-octenoic acid)] (Fig. 1), by formal substitution of one or two amino acids. Cyclosporin A is nowadays widely used as immunosuppressant for organ transplantations and treatment of autoimmune diseases (*Consupren*[®], Galena; *Sandimmun*[®], Novartis). Although cyclosporins are structurally almost identical, they strongly differ in their pharmacological activities. Several crystal structures of cyclo-

PLN version :

(cyclo)-[MeBmt]-[Abu]-[Sar]-[MeLeu]-Val-[MeLeu]-Ala-dAla-[MeLeu]-[MeLeu]-[MeVal]-(cyclo)

CRYSTAL STRUCTURE OF CYCLOSPORIN E

Collect. Czech. Chem. Commun. (Vol. 63) (1998) (pp.115 - 120)

Michal HUSAK^{a1}, Bohumil KRATOCHVIL^{a2}, Martin BUCHTA^b, Ladislav CVAK^b
and Alexandr JEGOROV^{c,*}

^a Department of Solid State Chemistry, Prague Institute of Chemical Technology, 166 28 Prague 6, Czech Republic; e-mail: ¹ husakm@vscht.cz, ² kratochb@vscht.cz

^b Galena Co., R.&D., 747 70 Opava-Komarov, Czech Republic

^c Galena Co., Research Unit, Branisovska 31, 370 05 Ceske Budejovice, Czech Republic; e-mail: husakm@marvin.jcu.cz

Multi-chain peptide

- Let's "cheat" – import from UniProt

The screenshot shows the Proteax software interface. The ribbon is set to 'Proteax' and the 'Import from server' option is highlighted. A tooltip is displayed over the 'Import from server' button, providing details on how to import protein entries by ID from a network server. The tooltip text includes: 'Retrieves protein entries by ID from a server, e.g. a web server or FTP server on the internet or on your local network.', 'The retrieved protein entries may be in either of the following formats:', a list of formats (PLN, UniProt, GPMWV, FASTA, MDL molfile), and 'Proteax will automatically detect each entry's format and convert as necessary.' At the bottom of the tooltip, it says 'ProteaxMain.xlam Press F1 for add-in help.'

PLN version :

H-FVNQHLC(1)GSHLVEALYLVC(2)GERGFFYTPKT-OH.

H-GIVEQC(3)C(1)TSIC(3)SLYQLENYC(2)N-OH

name="RecName: Full=Insulin;..." id=INS_HUMAN

Registration in Biochemfusion's Proteax-based database BCFReg

CompoundUpload_2016-05-22_Template.xlsx - Microsoft Excel

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

Validate
Upload and register
Upload

C3 fx H-HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-[NH2]

	A	B	C	
1	Status as determined by server.	Server-assigned compound name. You cannot set this yourself.	* Compound in Protein Line Notation.	Select
2	Status	Compound name	PLN	Salt name
3	CMP-2: New compound salt of parent structure "BCF01002".	BCF01002_T	H-HAEGTFTSDVSSYLEGQAAKEFIAWLVKGR-[NH2]	TFA
4	CMP-2: New compound salt of parent structure "BCF00101".	BCF00101_T	(cyclo)-[MeBmt]-[Abu]-[Sar]-[MeLeu]-Val-[MeLeu]-Ala-dAla-[MeLeu]-[MeLeu]	TFA
5	CMP-2: New compound salt of parent structure "BCF00001".	BCF00001_T	H-FVNQHL(1)GSHLVEALYLVC(2)GERGFFYTPKT-OH.H-GIVEQC(3)C(1)TSIC(3)SLY	TFA

- In production at Danish biotech Zealand Pharma since 2015.

BCFReg – registration system

Zealand Pharma - Compound Registration System - V1.0.3.133

File Edit View ShoP Admin tools Help

Query root: compounds

Show fields...

batches\amount >= 3

Query fields: Root = compounds

- molecule
- pln_sequence
- structure_primary_type
- structure_ratio
- salt_ratio
- sum_formula
- mw_avg
- mw_mono
- comments
- created_by
- created_on
- updated_by
- updated_on
- batches
 - name
 - amount
 - nuriv

Add to search panel

Search

Search result / current list: 20 compounds

- BCF00001_N
- BCF00002_N
- BCF00003_N
- BCF00004_N
- BCF00005_N
- BCF00006_N
- BCF00101_N
- BCF00102_N
- BCF00103_N
- BCF00104_N
- BCF00105_N
- BCF00106_N
- BCF01001_N
- BCF01002_N
- BCF01003_N
- BCF01004_N
- BCF01005_N
- BCF01006_N
- BCF11001_N
- BCF11002_N

All compounds in search result

Drag a column header here to group by that column

compound_name	salt_name	pln
BCF01002_N	None	H-HAEGTFTSDVSSYLEGQAAKEFIAWLKGR-[NH2] name="GLP-1 sequence (7-36) -
BCF01003_N	None	H-HAEGTFTSDVSSYLEGQAA[N6-(gamma-Glu(N1-hexadecanoyl))-lysine]EFIAWLVRGRG-OH
BCF01004_N	None	H-HAEGTFTSDVSSYLEGQAA[N6-(gamma-Glu(N1-hexadecanoyl))-lysine]EFIAWLVRGRG-OH

Compound: BCF01003_N

Salt: None

Comments:

Sum formula: C172 H265 N43 O51

Avg. MW: 3751.202

Mono MW: 3748.9465

Created by: BCF_REG

Created on: 2015-05-17 13:09:58

Parent sum formula: C172 H265 N43 O51

Parent avg. MW: 3751.202

Parent mono MW: 3748.9465

Salt ratio: 1:1

Sequence Condensed structure Full structure

H-HAEGTFTSDVSSYLEGQAA[N6-(gamma-Glu(N1-hexadecanoyl))-lysine]EFIAWLVRGRG-OH
name="Liraglutide according to <http://www.ama-assn.org/ama1/pub/upload/mm/365/liraglutide.pdf>"

Batches Compound synonyms Project names

Drag a column header here to group by that column

name	amount	comments	created_on	updated_on
B23	6840		2015-05-17 13:10:06	

Prod zp\$jhje@192.168.0.120 [Admin]

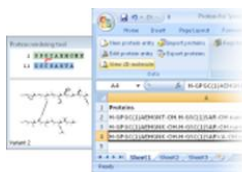
ZEALD
REVOLUTIONARY HEALTH SOLUTIONS

Proteax – broad tooling support

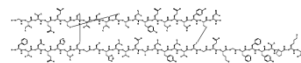
Proteax for Spreadsheets

Spreadsheet functions,
macro programming

MS Excel®
OpenOffice.org®



Sequence	Molecular weights	Da
1 NEVYDQGLR	Average	5887.51922
21 NEVYDQGLR	Standard Unifrac	5795.64874
41 NEVYDQGLR		



Java® applet, Flash®-based,
JavaScript™, .NET, and Pascal
visualizers

Structure and
sequence visualization

Proteax for KNIME

KNIME nodes for graphical
processing of protein data

KNIME™
graphical workflows



Proteax Desktop

.NET, Java®, Python,
C/C++, Pascal

Scripting,
programming

Proteax toolkit

Proteax Cartridge
Database integration
registration and querying

Oracle®

```
select * from compounds  
where proteax.mw_avg(protein_text) > 5000;
```

PostgreSQL®

Does not replace
yourchemistry
cartridge. Works
together with.e.g.
**JChem
Cartridge.**

BCFReg – Excel integration

The screenshot shows an Excel spreadsheet with the following data:

parent_name	pln	Diff
BCF00001	1 GIVEQCCTSIICSLYQLENYC 21 NFNQHLGSHLVEALYLVC 41 GERGFYYTKT	*
BCF00002	1 GIVEQCCTSIICSLYQLENYC 21 NFNQHLGSHLVEALYLVC 41 GERGFYYTKT	endo-P(B29) des-P(B28) *
BCF00003	1 GIVEQCCTSIICSLYQLENYC 21 NFNQHLGSHLVEALYLVC 41 GERGFYYTKT	D(B28) *
BCF00004	1 GIVEQCCTSIICSLYQLENYC 21 NFNQHLGSHLVEALYLVC 41 GERGFYYTKT	K(B3)E(B29) *
BCF00005	1 GIVEQCCTSIICSLYQLENYC 21 NFNQHLGSHLVEALYLVC 41 GERGFYYTKT	G(A21) * -RR-(B)
BCF00006	1 GIVEQCCTSIICSLYQLENYC 21 NFNQHLGSHLVEALYLVC 41 GERGFYYTKT	des-T(B30) [N6-C14fattyacid-lysine](B29) * inline-mod=K-residue,[N6-C14fattyacid-

- Export to peptide SAR tables.
- **Peptides can be directly converted to JChem for Excel structures**

Thank you for your attention

- Please check our booth and see BCFReg in action.