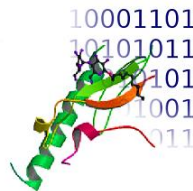


# Hall of mirrors

- Reflections on protein registration

*Jan Holst Jensen*  
*CEO, Biochemfusion ApS*



**biochemfusion**  
*- Enabling biochemformatics*

# Perceiving molecules

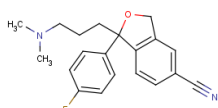
## Cheminformatics

## Bioinformatics

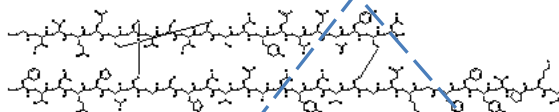
neither-nor / both-and

Molecule graphs

Sequences



```
1  G I V E Q C C T S I C S L Y Q L E N Y C
21 N F V N Q H L C G S H L V E A L Y L V C
41 G E R G F F Y T P K T
```



```
1  M V S Q A L R L L C L L L C L Q C C L A A G G V A K A S G C E T R D M P W K P C
41 P H R V F V T Q E E A H C V L H R R R R A N A F L F L R P G S L E R F C K R E
81 Q C S F E E A R E I F K D A E R T K L F W I S Y S D G D C A S S P C Q N G G S
121 C R D Q L Q S Y I C F C L P A F E C R N C E T H R D D Q L I C V N E N G C C K Q
161 Y C S D H T G T K R S C R C H E G Y S L L A D G V S C T P T V E Y P C K I P I
201 L E K R N A S R P Q G R I V G C K V C P R C E C P W Q V L L L V N G A Q L C C G
241 T L I N T I W V V S A A H C F D K I R N U R N L I A V L G E H D L S E H D G D E
281 Q S R R V A Q V I I P S T Y V P G T I N H D I A L L R L H Q P V V L T D H V V F
321 L C L P E R T F S E R T L A F V F S L V S C W G Q L L D R C A T A L E L M V L
361 N V P R L M T Q D C L Q Q S R V K V C D S P N I T E Y M F C A G Y S D C S K D S C
401 K G D S G C P H A T H Y R G T W Y L T G I V S W G Q G C A T V G H F G V Y T R V
441 S Q Y I E W L Q K L M R S E P R P F C V L L R A P P F
```

100

10k

1M

MW  
Da

# Biologics registration

- Zealand Pharma, Denmark
  - Novel peptide drugs for e.g. diabetes and obesity
  - e.g. lixisenatide, marketed by Sanofi as Lyxumia®
- Joint collaboration with Biochemfusion on a new biologics registration system
  - Project started Feb 1<sup>st</sup>, going live June 30<sup>th</sup>
  - Using Biochemfusion's Proteax toolkit on both server and client side

# Why a joint collaboration ?

## Traditional approach

### Off-the-shelf

- Many product features irrelevant
- 80% of needs covered
- 20% must be custom-coded and added in
- (and what about that protein chemistry ?)

- Premium price for irrelevant features
- Those last 20% may be very hard to do

### Bespoke

- Everything is possible
- (Eventually and with a bit of luck 😊)

- Huge potential risk
- Expect and accept prolonged projects

## BCF + Zealand Approach

### Bespoke app on standard components

- 80% of needs covered from
- 100% relevant features available from client selected standard components

### Remaining 20% done by

- Joint development collaboration
- Maintained in the future from off-the-shelf BCF product

### Build on

- Proteax protein chemistry cartridge
- Standard chemistry database cartridge
- Max. integration with existing tools
  - Standard chemistry sketchers
  - Excel (!)

# Registration in action

The screenshot displays the 'Zealand Pharma - Compound Registration System - V0.1.0.23' application window. The interface includes a menu bar (File, Edit, View, Help) and a 'Query root' dropdown menu currently set to 'compounds'. A 'Show fields...' button is visible below the dropdown. The main area is divided into a 'Search result' panel (currently empty) and a 'Result details' panel. A 'Query fields' dialog box is open in the foreground, listing fields for the 'Root = compounds' query: 'name', 'molecule' (highlighted with a magnifying glass icon), 'sequence', 'structure\_primary\_type', 'structure\_ratio', 'salt\_ratio', 'created\_by', 'created\_on', and 'batches'. An 'Add to search panel' button is located at the bottom of the dialog. The Zealand Pharma logo, 'REVOLUTIONARY HEALTH SOLUTIONS', is visible in the bottom right of the application window. The status bar at the bottom of the window shows '3.1.10 [Viewer]'.

# Registration in action

The screenshot displays the 'Zealand Pharma - Compound Registration System - V0.1.0.23' interface. The 'Query root' is set to 'compounds'. The 'Search results (6)' list includes BCF00101\_N, BCF00102\_N, BCF00103\_N, BCF00104\_N, BCF00105\_N, and BCF00106\_N. The 'Compound name' field shows 'BCF00101\_N'. The 'molecule' section displays a chemical structure of a peptide derivative. The 'Sequence' tab is active, showing the sequence: 1 Thr Ala Gly Leu Val Leu Ala Ala Leu Leu Val. The 'Full structure' tab is also visible. The bottom status bar shows 'Test zp\$jhje@192.168.1.10 [Viewer]'. The Zealand Pharma logo is present at the bottom of the interface.

# Registration in action

The screenshot displays the Zealand Pharma Compound Registration System (V0.1.0.23) interface. The window title is "Zealand Pharma - Compound Registration System - V0.1.0.23". The menu bar includes "File", "Edit", "View", and "Help".

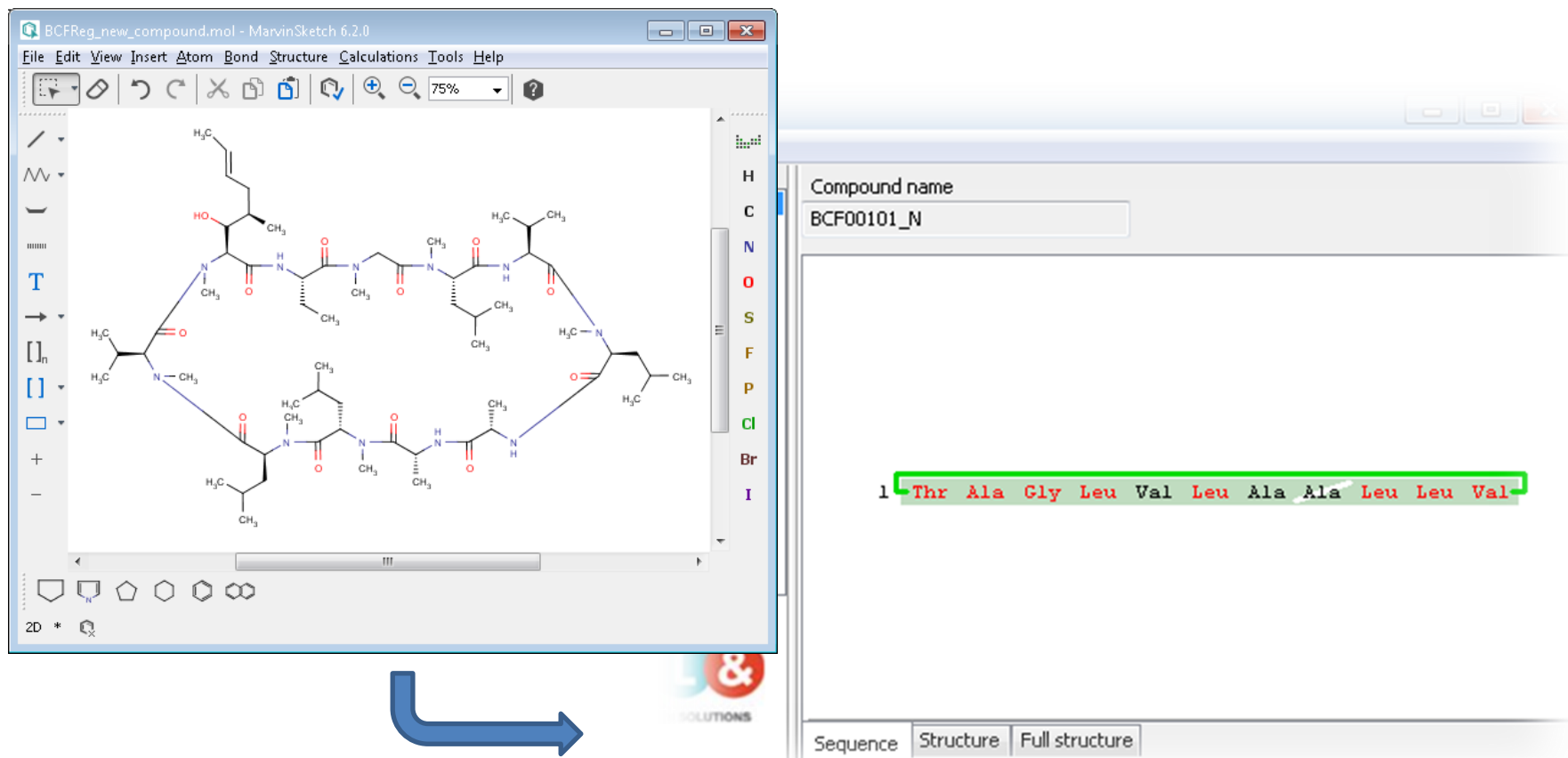
The interface is divided into several sections:

- Query root:** A dropdown menu set to "compounds" and a "Show fields..." button.
- molecule:** A chemical structure viewer showing a fragment of a molecule with a nitrogen atom, an oxygen atom, and a terminal alkene group.
- Search results (6):** A list of search results: BCF00101\_N (highlighted), BCF00102\_N, BCF00103\_N, BCF00104\_N, BCF00105\_N, and BCF00106\_N.
- Compound name:** A text field containing "BCF00101\_N".
- Chemical structure:** A large, detailed chemical structure of a cyclic peptide, specifically a cyclic heptapeptide with a long side chain containing an alkene. The structure is shown in a 2D representation with stereochemistry.
- Navigation:** A set of tabs at the bottom right: "Sequence", "Structure" (selected), and "Full structure".
- Status bar:** A green bar at the bottom left showing "Test zp\$hje@192.168.1.10 [Viewer]".

The Zealand Pharma logo is visible at the bottom center of the interface, featuring the letters "ZEAL&" in a stylized font with the tagline "REVOLUTIONARY HEALTH SOLUTIONS" below it.

# And of course...

- *Just Works* with MarvinSketch

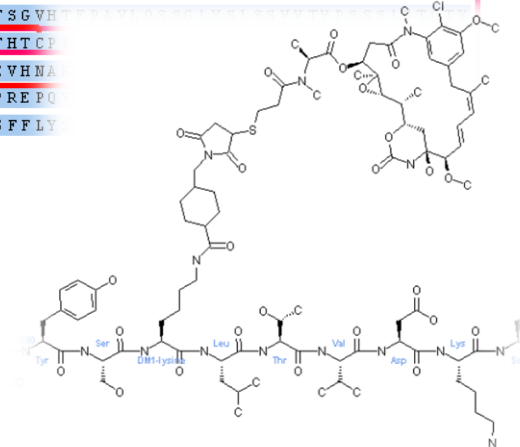




# “What about the big guys ?”

- Proteax composite structure key
- Provides a unique identifier for proteins from 0 – 1 MDa
- Calculated in milliseconds for antibody-drug conjugates
- ICCS 2014, Noordwijkerhout: Presenting the poster “A computationally efficient structure key for large proteins” in collaboration with Gerd Blanke, StructurePendium Technologies, Germany

```
1 DIQMTQSPSSLSASVGDRTVITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPS
61 RFGSGRSRSGTDFTLTISLQPEDFATYYCQQHYTTPPTFGQGTKEIKRTVAAPSVFIFPP
121 SDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSSTYLSSTLT
181 LSKADYEKHAVYACEVTHQGLSSPVTKSFNRGECDEVQLVE SGGGLVQPGGSLRLSCAASG
241 FNIADTYIHWRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTTISADTSKNTAYLQMNSL
301 RAEDTAVYYCSRUGGDFYAMDYWGQGTLVTVSSASTKGPSVFLAPSSKSTSGGTAALG
361 CLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVY
421 HKPSNTKVDKKEVPPKSCDKTHTCPPCPAPELGGPSVFLFPPKPKDTLMISRTPEVTCV
481 VVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCK
541 VSNKALPAPIEKTI SKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWE
601 SNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFS CSMHEALHNHYTQKLSL
661 LSPGKDIQMTQSPSSLSASVGDRTVITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLY
721 SGVPSRFSGRSGTDFTLTISLQPEDFATYYCQQHYTTPPTFGQGTKEIKRTVAAPSV
781 FIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSYSL
841 SSTLTLSKADYEKHAVYACEVTHQGLSSPVTKSFNRGECDEVQLVESGGGLVQPGGSLRLS
901 CAASGFNIADTYIHWRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTTISADTSKNTAYL
961 QMNSLR AEDTAVYYCSRUGGDFYAMDYWGQGTLVTVSSASTKGPSVFLAPSSKSTSGG
1021 TAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYI
1081 ICNVNHKPSNTKVDKKEVPPKSCDKTHTCPPCPAPELGGPSVFLFPPKPKDTLMISRTPEVTCV
1141 EVTCVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCK
1201 EYKCKVSNKALPAPIEKTI SKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWE
1261 AVEWESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNVFS CSMHEALHNHYTQKLSL
1321 QKLSLSLSPGK
```

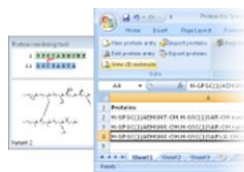


# Proteax<sup>®</sup> – easily integrated

## Proteax for Spreadsheets

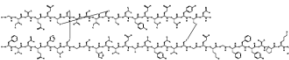
Spreadsheet functions,  
macro programming

MS Excel<sup>®</sup>  
OpenOffice.org<sup>®</sup>



Sequence	Molecular weights	Da
1 NEVYDQVETGK	Average	5687.51922
21 NEVYDQVETGK	Standard Unifrac	5795.64874
41 NEVYDQVETGK		

3D molecular



Java<sup>®</sup> applet, Flash<sup>®</sup>-based,  
JavaScript<sup>™</sup>, .NET, and Pascal  
visualizers

Structure and  
sequence visualization

## Proteax for KNIME

KNIME nodes for graphical  
processing of protein data

KNIME<sup>™</sup>  
graphical workflows



## Proteax Desktop

.NET, Java<sup>®</sup>, Python,  
C/C++, Pascal

Scripting,  
programming

Proteax toolkit

Proteax Cartridge  
Database integration  
registration and querying

Oracle<sup>®</sup>

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

PostgreSQL<sup>®</sup>

# Integration partners

- iScienceSearch, by AKos GmbH
  - <http://isciencesearch.com/iss/default.aspx>
  - Proteax enables sequence searching across multiple chemical structure databases on the web.
- Sysment Notebook and Reaction Tool
  - <http://www.sysment.hu/>
  - Sysment is using Proteax Desktop so their product line can support both small and large molecules
- Proteax4Morphit, The Edge
  - [http://www.edge-ka.com/products/morphit/morphit\\_plugins](http://www.edge-ka.com/products/morphit/morphit_plugins)
  - Proteax plugin for Morphit provides sequence and structure visualization plus protein comparison and analysis

# Thanks

- Zealand Pharma
  - Jakob Anker Hansen
  - Jakob Lind Tolborg
  - Lars Bo Hansen
- StructurePendium
  - Gerd Blanke
- RDKit
  - Greg Landrum