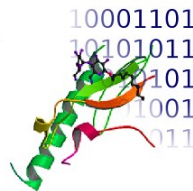


Hall of mirrors

- Reflections on protein registration

Jan Holst Jensen
CEO, Biochemfusion ApS



biochemfusion
- Enabling biochemformatics

Perceiving molecules

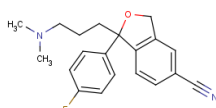
Cheminformatics

Bioinformatics

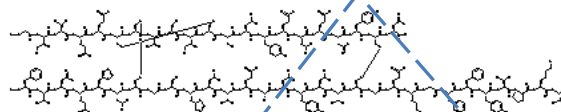
neither-nor / both-and

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 G 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 H 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 1st, going live June 30th
 - 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 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 complex organic molecule with a nitrogen atom, an oxygen atom, and a terminal aldehyde 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".
- Sequence:** A sequence viewer showing "1 Thr Ala Gly Leu Val Leu Ala Ala Leu Leu Val" with a green highlight around the sequence.
- Navigation:** Buttons for "Sequence", "Structure", and "Full structure" at the bottom of the sequence viewer.
- Footer:** A status bar showing "Test zp\$jhje@192.168.1.10 [Viewer]" and the Zealand Pharma logo with the tagline "REVOLUTIONARY HEALTH SOLUTIONS".

Registration in action

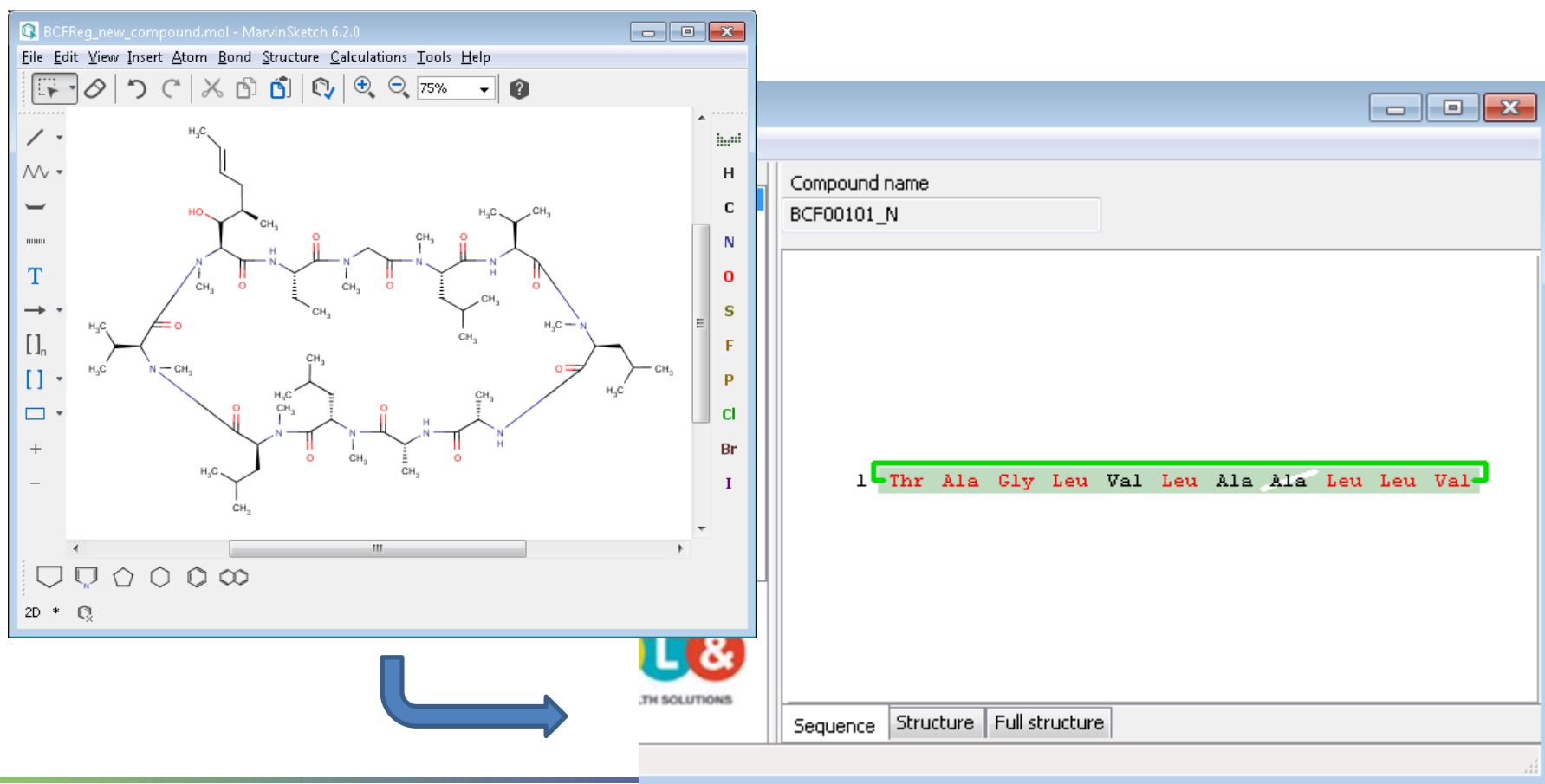
The screenshot displays the 'Zealand Pharma - Compound Registration System - V0.1.0.23' window. The interface is divided into several sections:

- Query root:** A dropdown menu set to 'compounds' with a 'Show fields...' button below it.
- molecule:** A chemical structure of a small molecule, possibly a peptide fragment, is shown. Below it is a dropdown menu set to 'SSS' and a 'Search' button.
- Search results (6):** A list of identifiers: BCF00101_N (highlighted), BCF00102_N, BCF00103_N, BCF00104_N, BCF00105_N, and BCF00106_N.
- Compound name:** A text box containing 'BCF00101_N'.
- Chemical structure:** A large, complex cyclic peptide structure is displayed in the main area.
- Navigation:** At the bottom of the structure area are tabs for 'Sequence', 'Structure', and 'Full structure'.
- Status bar:** At the bottom left, it shows 'Test zp\$hje@192.168.1.10 [Viewer]'.

The Zealand Pharma logo, 'ZEAL& REVOLUTIONARY HEALTH SOLUTIONS', is visible in the bottom center of the interface.

And of course...

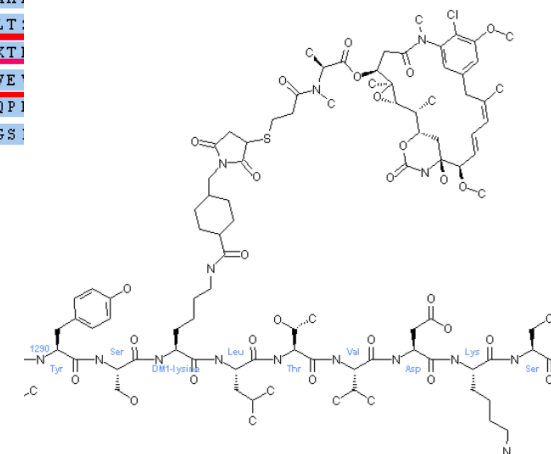
- *Just Works* with MarvinSketch



“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 DIQMTQSPSSLSASVGRVITITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPS
61 RFGSRSRSGTDFTLTISLQPEDFATYYCQQHYTTPPTFGQGTKEIKRTVAAPSVFIFPP
121 SDEQLKSGTASVVCLLNMFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSSTYLSSTLT
181 LSKADYEKHAVYACEVTHQGLSSPVTKSFNRGECDEVQLVE SGGGLVQPGGSLRLSCAASG
241 FNIADTYIHWRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSL
301 RAEDTAVYYCSRUGGDFYAMDYWGQGTLVVSSASTKGPSVFPFLAPSSKSTSGGTAALG
361 CLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYICNVY
421 HKPSNTKVDKKEVPPKSCDKTHTCPPCPAPELGGPSVFLFPPKPKDTLMISRTPEVTCV
481 VVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCK
541 VSNKALPAPIEKTI SKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWE
601 SNGQPENNYKTTTPVLDSDGSFFLYSLKLVTKSRWQQGNVFS CVMHEALHNHYTQKLSL
661 LSPGKDIQMTQSPSSLSASVGRVITITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLY
721 SGVPSRFSGRSGTDFTLTISLQPEDFATYYCQQHYTTPPTFGQGTKEIKRTVAAPSV
781 FIFPPSDEQLKSGTASVVCLLNMFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSYSL
841 SSTLTLSKADYEKHAVYACEVTHQGLSSPVTKSFNRGECDEVQLVESGGGLVQPGGSLRLS
901 CAASGFNIADTYIHWRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYL
961 QMNSLRAEDTAVYYCSRUGGDFYAMDYWGQGTLVVSSASTKGPSVFPFLAPSSKSTSGG
1021 TAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVPSSSLGTQTYI
1081 ICNVNHKPSNTKVDKKEVPPKSCDKTHTCPPCPAPELGGPSVFLFPPKPKDTLMISRTPEV
1141 EVTCVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEY
1201 EYKCKVSNKALPAPIEKTI SKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWE
1261 AVEWESNGQPENNYKTTTPVLDSDGSFFLYSLKLVTKSRWQQGNVFS CVMHEALHNHYTQKLSL
1321 QKLSLSLSPGK
```

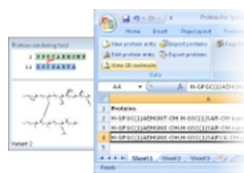


Proteax[®] – easily integrated

Proteax for Spreadsheets

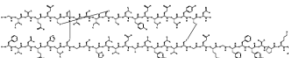
Spreadsheet functions,
macro programming

MS Excel[®]
OpenOffice.org[®]



| Sequence | Molecular weights | Da |
|------------------|-------------------|------------|
| 1 NEVVDGCTGKGGK | Average | 5887.51922 |
| 21 NEVVDGCTGKGGK | Standard Unifrac | 5795.64874 |
| 41 NEVVDGCTGKGGK | | |

3D molecular



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[®]

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
 - 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

... and of course!

