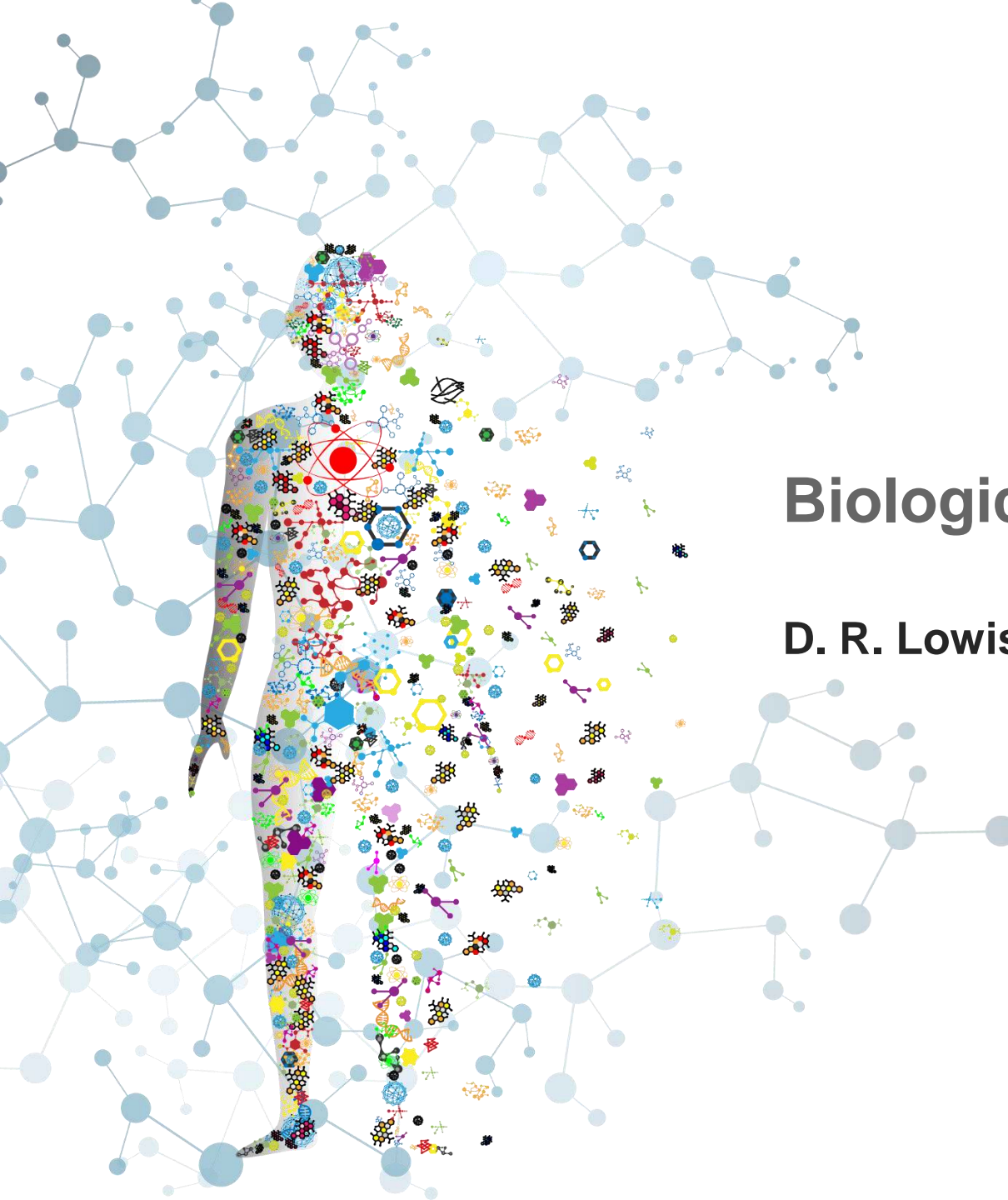
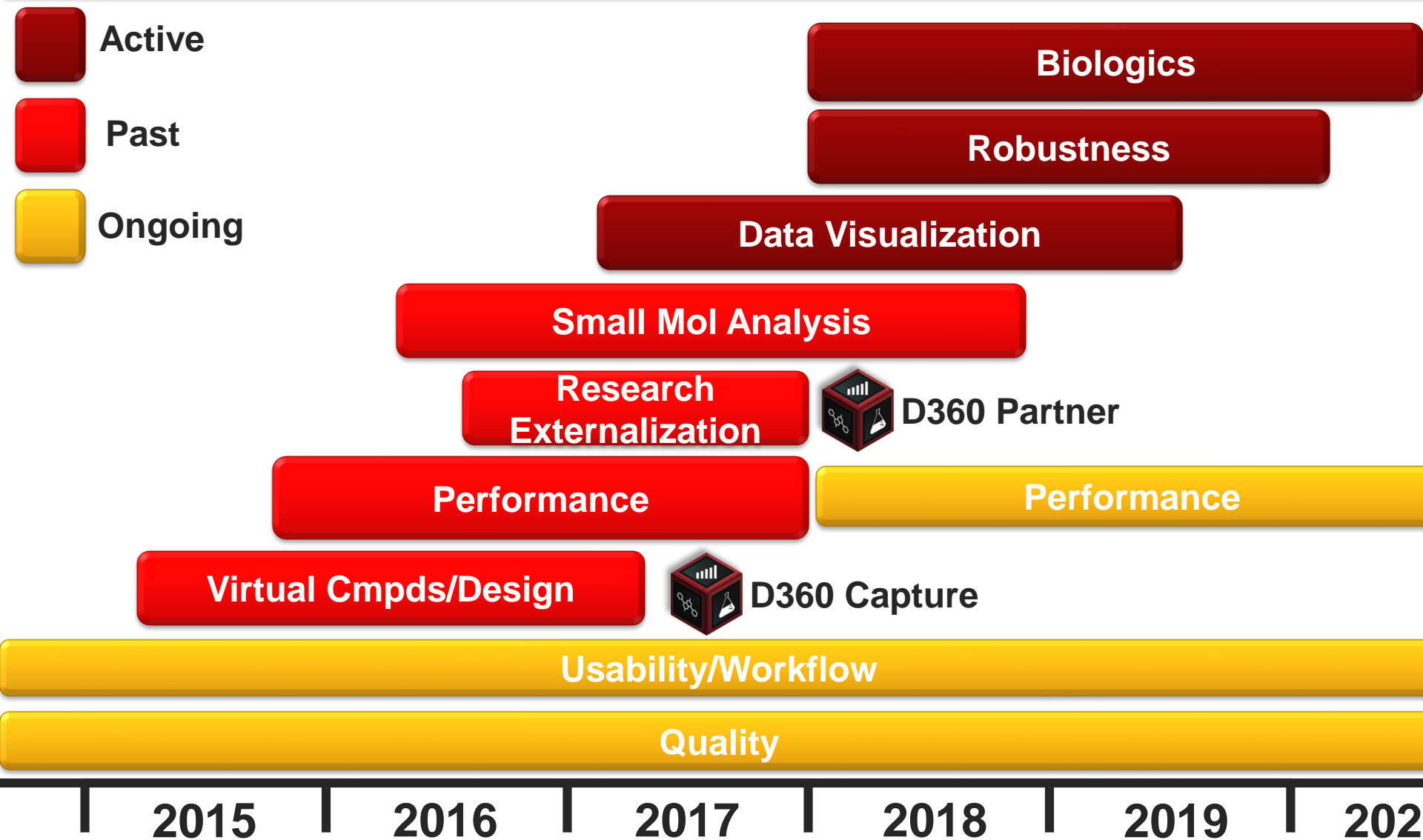


# Biologics and D360

D. R. Lewis D. Phil.

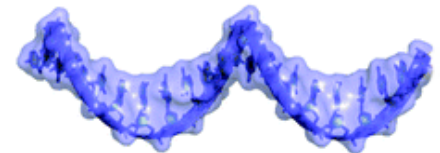
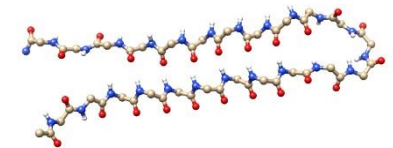
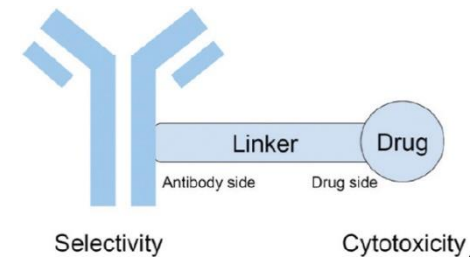
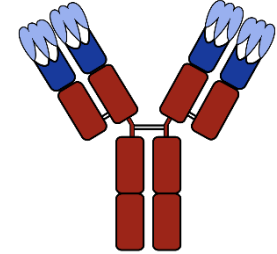


# D360 Product Development Themes



# Multiple Types of Biologic/Research Process

- Antibodies
  - Very different workflow from small molecules
  - Large number of entities, focus jointly on bioprofile and producibility
- ADCs
  - Small number of antibodies, different liners/warheads
  - Reduces to a small-molecule-esque workflow
- Peptides
  - Similar to small molecule workflow – Sequence-Activity relationships
- Oligonucleotides
  - Somewhat similar to small molecule workflow depending on mechanism of action
- Other
  - Vaccines, mixtures of the above concepts, ...



# A Simple Question

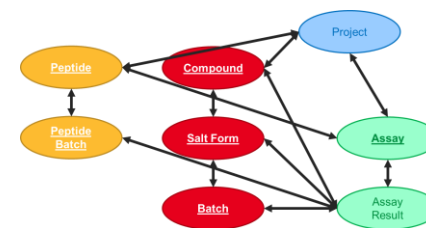
Given that D360 already provides the infrastructure for the data...

What additional tools can be provided that assist data-driven decision making for each new therapeutic modality?

# D360 Support for Biologics - Strategy

- Implement General Concepts

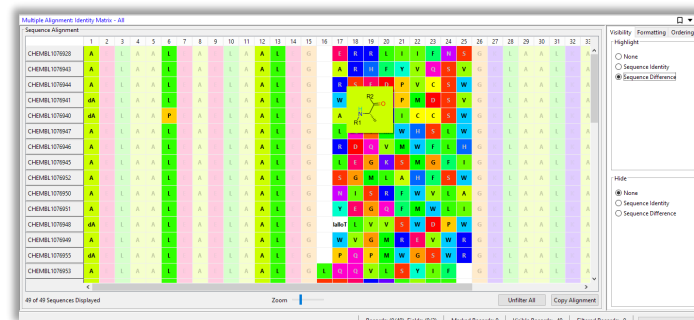
- Multiple new entities – a potential data catalog explosion and data configuration
- General representation of entities



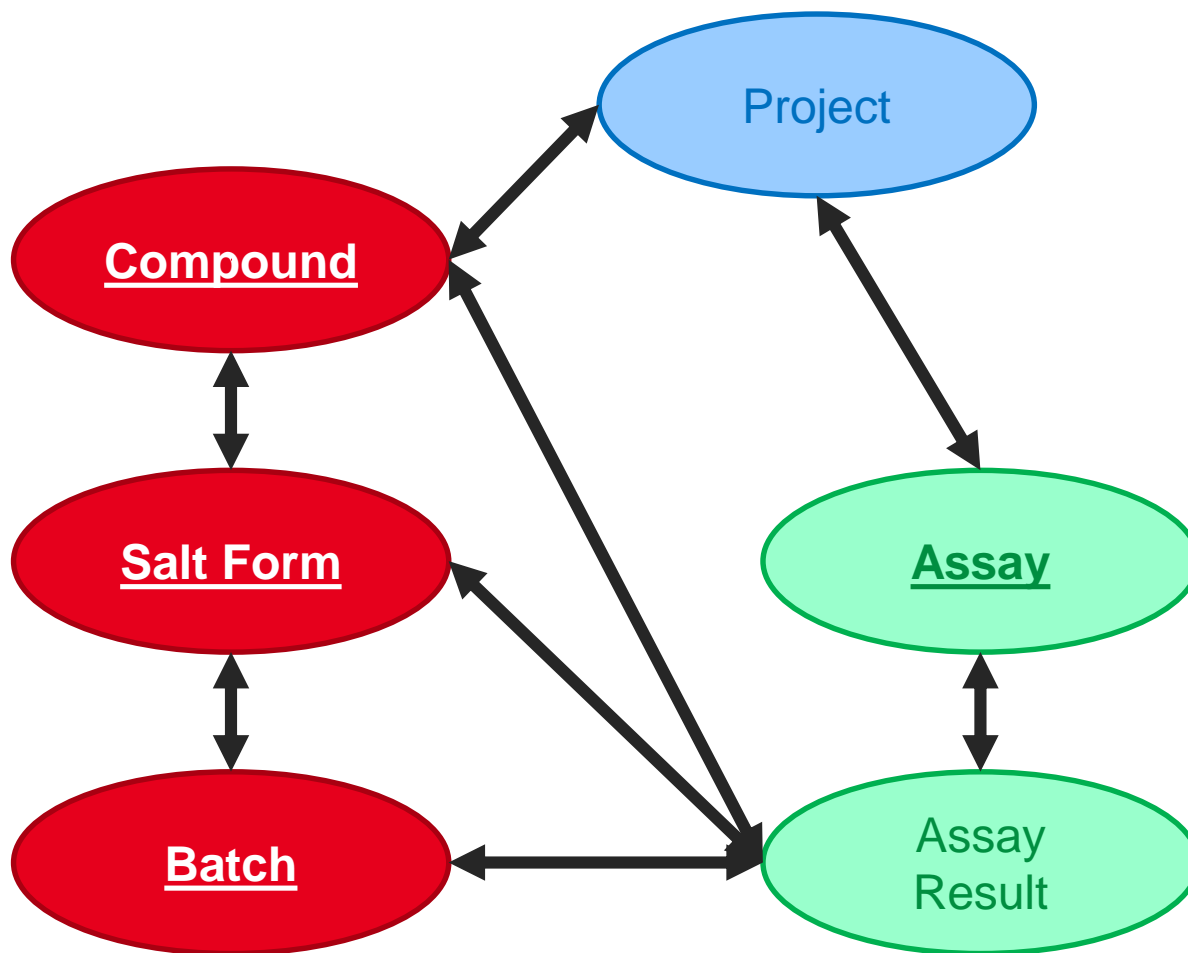
- Implement Modality-specific Concepts

1. Specific entity representation
  - Sequence (including non-natural monomers)
  - Search, formatting, find, filtering, ...
2. Basic analysis
  - Determining measures of similarity
  - Comparison methods
    - Alignment
    - Clustering
3. Advanced Analysis
  - Extraction of therapeutic-relevant knowledge
  - “Structure”-Activity relationships

CHEMEL ID	Name Sequence	Assay ID (PKA)
1	CHEMEL1076928	
2	CHEMEL1076932	
3	CHEMEL1076934	
4	CHEMEL1076941	
5	CHEMEL1076942	
6	CHEMEL1076947	

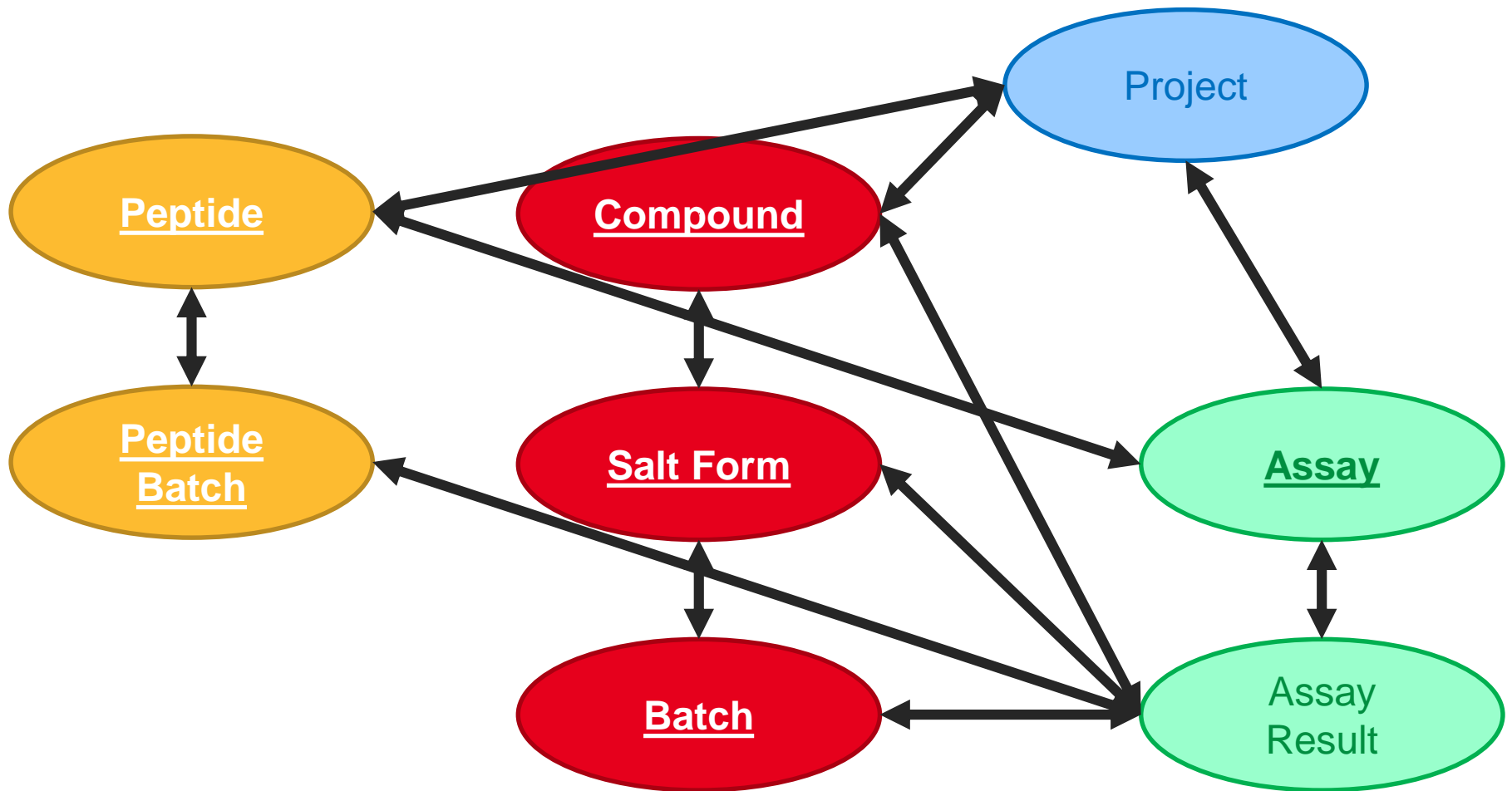


# Multiple Entities/Data Configuration



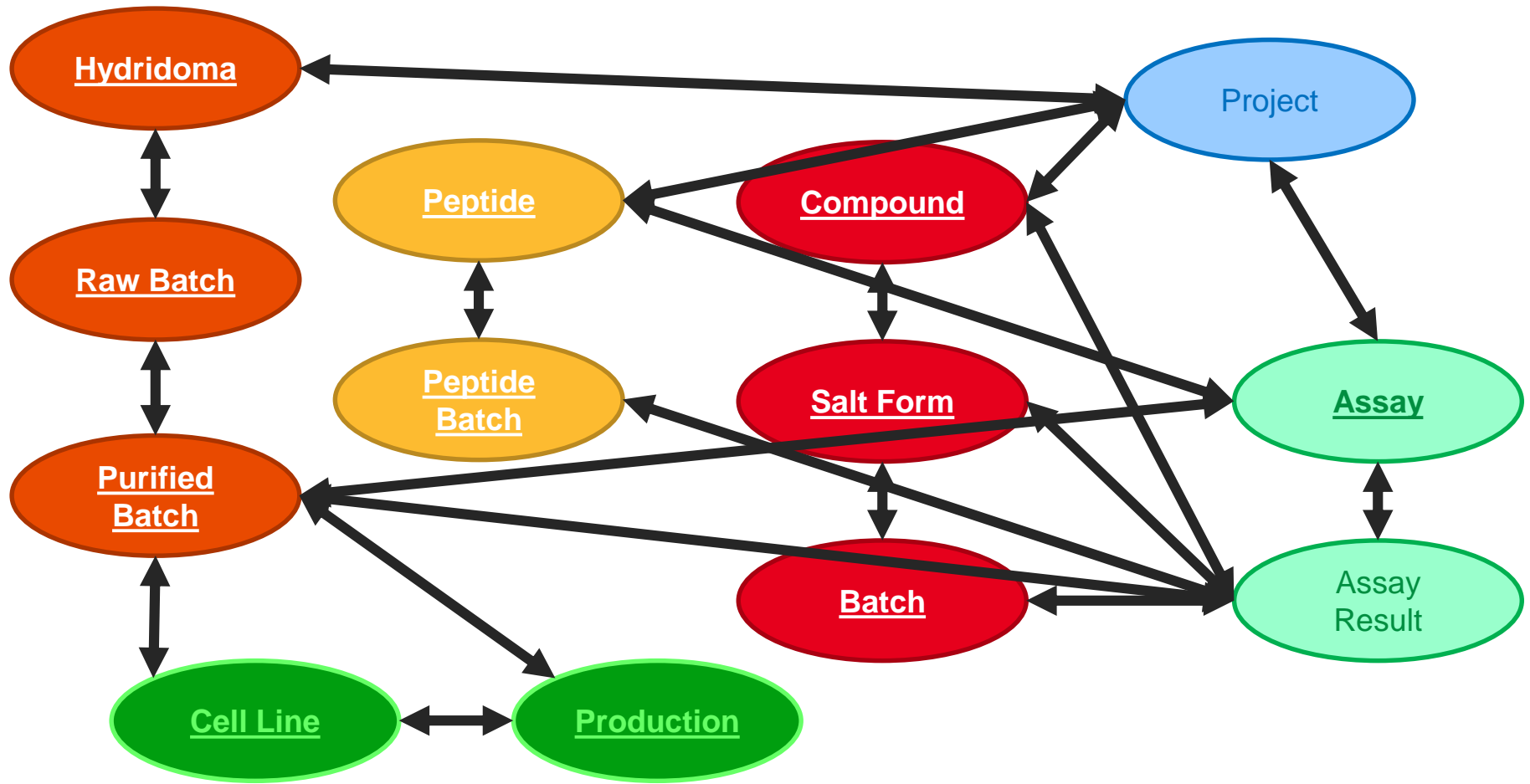
Example Small Molecule Data Catalog

# Multiple Entities/Data Configuration



...adding Peptides

# Multiple Entities/Data Configuration

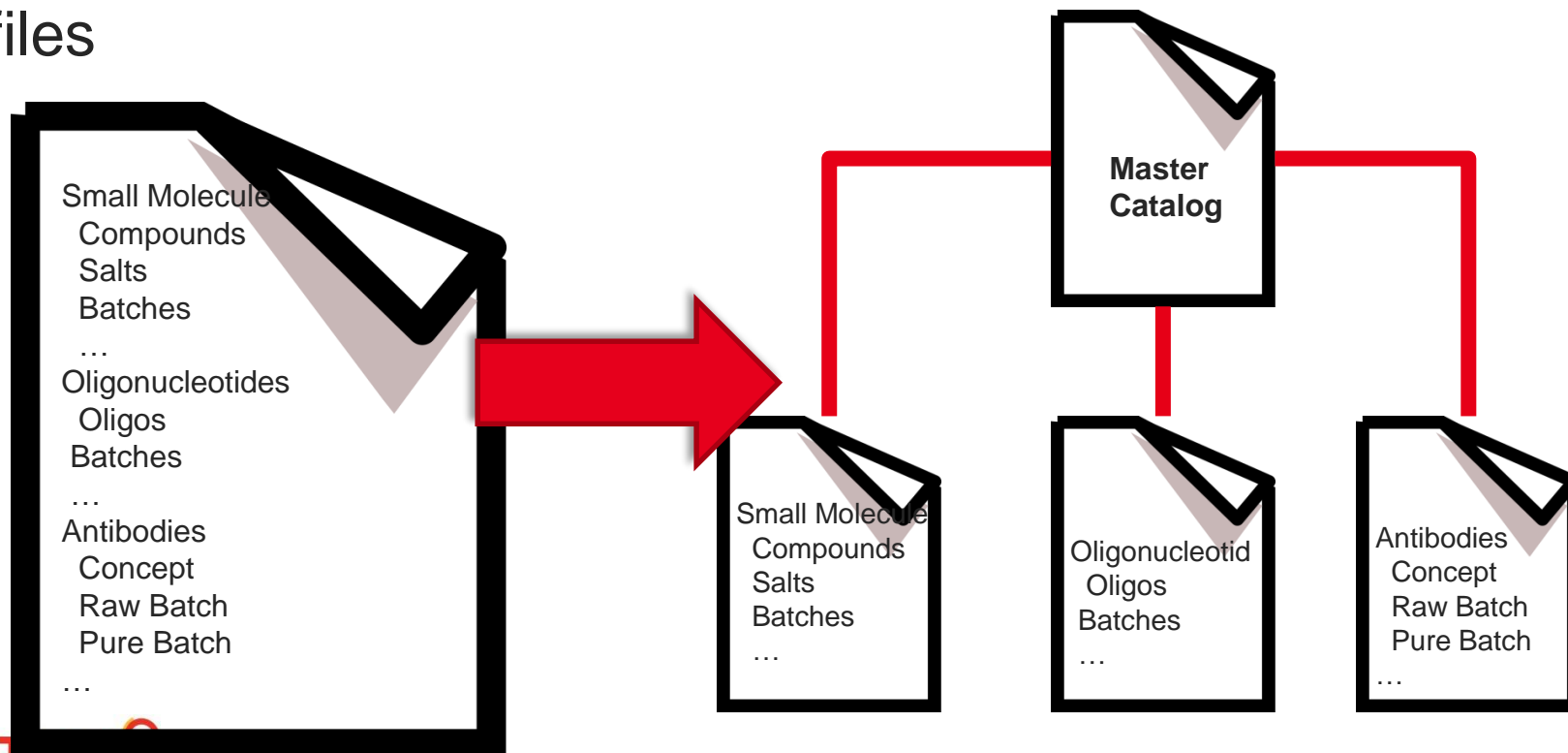


...adding Antibodies



# Assisting the Configuration

- General Configuration
  - Not an issue for D360 – at its core it is data agnostic
  - But.... More entities, more data fields, more relationships can make configuration management more complex
- D360 Updated to allow data catalog to be split over multiple files



# Assisting the End User: Data Category Selection

Different Data Domains  
on Different Tabs  
- rather than one giant list

The image displays two overlapping screenshots of a software dialog box titled "Choose Data Category".

The left screenshot shows the "Small molecules" tab selected under the "Scientific Data Domain". The categories listed are:

- Compounds
- Salt Forms
- Batches

The "Advanced Categories" dropdown is set to "Compounds".

The right screenshot shows the "Oligonucleotides" tab selected. The categories listed are:

- Oligonucleotides
- Oligonucleotide Batches

The "Advanced Categories" dropdown is set to "Oligo Concepts".

A red callout box with white text points to the tabs, stating: "Different Data Domains on Different Tabs - rather than one giant list".

# Moving to Representation - General and Specific



# General Representation

- Support for HELM Notification

- Allows HELM to be presented
- Integration with HELM viewer/editor



The screenshot displays the HELM software interface. On the left, a vertical list of panels shows chemical structures for various ChEMBL IDs (e.g., CHEMBL2370397, CHEMBL2370407, CHEMBL2110209, CHEMBL2370393, CHEMBL2370404) with their corresponding species and inhibition data. The central 'Grid Viewer' shows a grid of these structures. A 'Spreadsheet' window is overlaid on the grid, listing ChEMBL IDs and their corresponding HELM notations. On the right, the 'Web HELM Editor' is open, showing a 'Detailed Sequence' view of a peptide sequence: n1-ac-2-R-3-G-4-D-5-V-6-am. The sequence is represented by colored boxes (ac, R, G, D, V, am) with numbers 1-6 above them. Below the sequence, there are tabs for 'Sequence', 'HELM', 'Properties', and 'Structure View', along with 'Format' and 'Apply' buttons.

- Product support for BioVia Renderer

- Allows Biovia notation presented



# Oligonucleotide Representations

- Multiple, sequence-based, oligonucleotide representations
  - HELM, Biovia, single sequence custom format, multi-sequence custom format

The image shows a screenshot of a spreadsheet application. Two blue callout boxes are overlaid on the spreadsheet. The left callout box points to the 'Base Sequence' column and contains the text: 'Oligonucleotide Base Sequence Formatted by color and capitalization the backbone sequence and Sugar sequence data'. The right callout box points to the 'Backbone Sequence' column and contains the text: 'Oligonucleotide Backbone Sequence Formatted by the color based on sequence composition'. The spreadsheet has a header row with columns: 'Oligo ID', 'Base Sequence', and 'Backbone Sequence'. Below the header are several rows of data, each representing an oligonucleotide with its ID, base sequence, and backbone sequence.

	Oligo ID	Base Sequence	Backbone Sequence
515:	<input type="checkbox"/> Oligo882	GUNUUECUUTctccTcAAGCTacucUCUn	RPSXPXXRPPXSMESXSREXMOMXOSXR
516:	<input type="checkbox"/> Oligo208	NeaActGANeUNEUEAnAGUUtCnTAnG	SMSMMOMPRRROMPESRSRXORPMRXSXS
517:	<input type="checkbox"/> Oligo159	tEaEEGNANEGaTEgECUECCEaETGTnU	EOERPOPMPRPSEXXMXXPMEXROOXEO
518:	<input type="checkbox"/> Oligo557	tEGNNUTuUecNUenuUTUGaaCAnGTET	RPSOOPXESEXPRMXESOXRREROMMXX
519:	<input type="checkbox"/> Oligo541	tEuCcgUEnNUNccTCNtGETGTCgecNc	PEXXSPMPSMEPPPXRORPSRMMOSMMO
520:	<input type="checkbox"/> Oligo168	tteGATCuNNGaNtEgTtuUGCnGNtGUg	OPPMEEMEPEROSRXSERROOPOPSSPEM

# Oligonucleotide Support: Filtering and Equations

## Filtering

Filter x  
Filters  
Add Logic Delete All Clear All Criteria  
Disable Filters  
Find:  
Base Sequence  
GEAAGCATCCTGT <= Min Edit Distance:

Filter by text or Similarity

## Sequence Similarity Equation

Spreadsheet

Concept UID	HELM Sequence	Base Sequence
1: RTR24383	[Sequence]	GEAAGCATCCTGT
2: RTR56971	[Sequence]	GEAAGCATCCTGT
3: RTR34441	[Sequence]	GEAAGCATCTG
4: RTR34442	[Sequence]	GEAAGCATCCTG
5: RTR45759	[Sequence]	GEAAGCATCCTG
6: RTR49513	[Sequence]	GEAAGCATCCTG

Spreadsheet

Oligo ID	Base Sequence	MED to Ref
1: Oligo508	AaUcGUeccc	23
2: Oligo521	aCAAuNUTnA	24
3: Oligo778	AeUNINraeg	23
4: Oligo349	CACcNuIGcU	23
5: Oligo193	CnGENUaUEn	
6: Oligo708	CTuaNETAcI	
7: Oligo605	FeTNgAGtAc	
Oligo924	EnneUUGGsu	
Oligo902	EUUucNetuA	
Oligo651	gCCAGUGtuE	
Oligo593	GCnUtUATcT	
Oligo466	GEIqUaCUcU	
Oligo154	gGCaCNaATu	
Oligo672	NACNENtUtN	
Oligo865	naEgTIGCC	
Oligo908	NANaUaGcC	
Oligo810	neEAETqgNU	
Oligo499	NeEeAgotNA	
Oligo91	tcEeATaaA	
Oligo174	teTuCancun	
Oligo787	tgTcnNEEAn	
Oligo267	TUnAACNetG	
Oligo849	UCcNAngTuA	
Oligo817	UTcgnuINC	
Oligo645	UtgUACTIONC	
Oligo307	uTNCnTCuNu	
Oligo380	aaTAsAEINcC	
Oligo304	aEGTCAGUEIE	
Oligo762	aNEAEgUCAuT	
Oligo98	AtcCanUqeEE	

Edit Equation Column

Available Data: [Dropdown]

Equation Functions: min

- Column Statistics
  - min\_col
- General Math
  - min
- Text
  - mineditdistance

Equation: mineditdistance(C4,\*GUNUUEUCUUTCTCTCAAGTACUCUCUN\*)

Column Name: MED to Ref

Sequence Similarity Equation Function

# Peptide/Protein Representations

- Multiple, sequence-based, oligonucleotide representations
  - Peptides: HELM, Biovia, single sequence custom format
  - Proteins: Standard sequence representations
  - Added complication of non-natural monomers
    - Splits concepts of proteins vs. peptides

	ChEMBL Id	Helm Sequence	AssayX GMean IC50 (nM)
1:	<input type="checkbox"/> CHEMBL1076928	A-E-L-A-A-L-E-A-E-L-A-A-L-E-G-E-R-R-L-I-I-F-N-S-G-K-L-A-A-L-K-A-K-L-A-A-L-K-A	61
2:	<input type="checkbox"/> CHEMBL1076943	A-E-L-A-A-L-E-A-E-L-A-A-L-E-G-A-R-H-F-Y-V-Q-S-V-G-K-L-A-A-L-K-A-K-L-A-A-L-K-A	75
3:	<input type="checkbox"/> CHEMBL1076944	A-E-L-A-A-L-E-A-E-L-A-A-L-E-G-R-S-E-D-P-V-C-S-W-G-K-L-A-A-L-K-A-K-L-A-A-L-K-A	96
4:	<input type="checkbox"/> CHEMBL1076941	dA-E-L-A-A-L-E-A-E-L-A-A-L-E-G-W-G-S-Q-P-M-D-S-V-G-K-L-A-A-L-K-A-K-L-A-A-L-K-A	98
5:	<input type="checkbox"/> CHEMBL1076940	dA-E-L-A-A-P-E-A-E-L-A-A-L-E-G-A-S-R-R-I-C-G-S-W-G-K-L-A-A-L-K-A-K-L-A-A-L-K-A	140
6:	<input type="checkbox"/> CHEMBL1076947	A-E-L-A-A-L-E-A-E-L-A-A-L-E-G-L-Q-T-M-W-H-S-L-W-G-K-L-A-A-L-K-A-K-L-A-A-L-K-A	164
		A-E-L-A-A-L-E-A-F-I-A-A-I-F-G-I-F-G-K-S-M-G-F-I-G-K-I-A-A-I-K-A-K-I-A-A-I-K-A	



# Moving to Basic Analysis - Sequence Alignment





# Medicinally Relevant Peptide Sequence Alignment

- Bioinformatics studies protein (and DNA/RNA) sequences via sequence alignment

```
R Y D S R T T I F S P . . E G R L Y Q V E Y A M E A I G N A . G S A I G I L S
R Y D S R T T I F S P L R E G R L Y Q V E Y A M E A I S H A . G T C L G I L S
R Y D S R T T I F S P . . E G R L Y Q V E Y A Q E A I S N A . G T A I G I L S
R Y D S R T T I F S P . . E G R L Y Q V E Y A M E A I S H A . G T C L G I L A
R Y D S R T T I F S P . . E G R L Y Q V E Y A M E A I G H A . G T C L G I L A
R Y D S R T T I F S P . . E G R L Y Q V E Y A M E A I G N A . G S A L G V L A
R Y D S R T T I F S P . . E G R L Y Q V E Y A L E A I N N A . S I T I G L I T
S Y D S R T T I F S P . . E G R L Y Q V E Y A L E A I N H A . G V A L G I V A
```

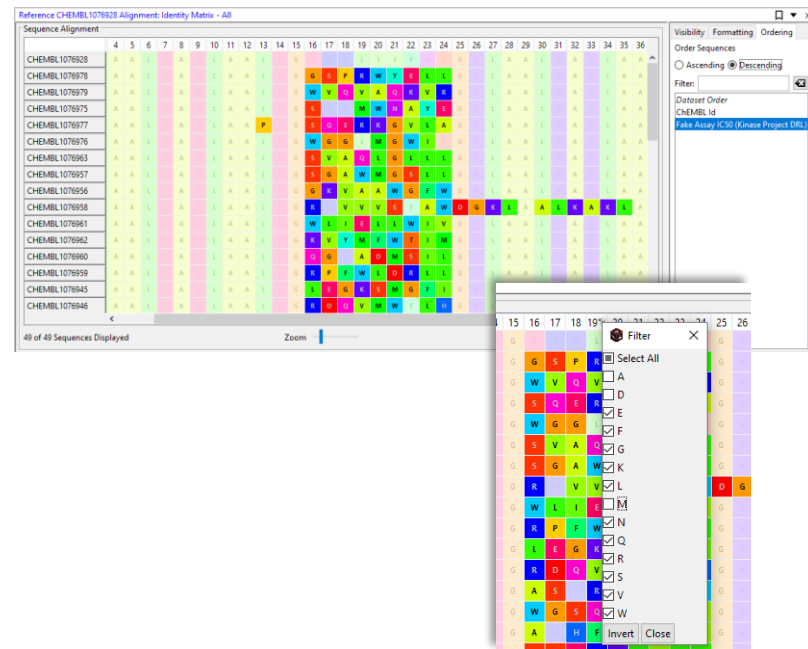
- BUT:
  - Bioinformatics looks for evolutionary relationships between sequences
  - Bioinformatics deals only with naturally occurring monomers
- What we want:
  - Must deal with unnatural monomers
  - We care about structural and chemical similarity NOT evolutionary distance
- So:
  - We utilize standard bioinformatics alignment algorithms but apply chemical similarity matrices to score alignments
  - Bioinformatics: X matched Y scores 0.7 since there is a 70% chance of an evolutionary mutation
  - D360: X matched Y scores 0.6 since X and Y are 60% chemically similar
  - **This makes D360 sequence alignments medicinally relevant**

# What can we do with the Sequence Alignment?

- Alignment options
  - Align to reference sequence
    - Comparison of sequences with a reference
  - Multiple alignment
    - View overall similarity/differences across a set of sequences
  - For both natural and unnatural monomers

- Format and Highlight
  - Color by monomer properties
  - Highlight differences at aligned positions
  - Order sequences by assay data

- Filter sequences
  - By monomer X at position Y



**To allow users to understand what sequence changes give improved bio-properties**

# Basic Analysis: Peptide Sequence Alignment Viewer

## Aligned Peptide Sequences

- color coded by monomer chemical properties

An example of an unnatural monomer

Chemical structures are displayed with tooltip or my zooming in

Tools to Highlight/Show/Hide differences and similarities



Sequence Alignment shows similarities and differences between sequences which can be compared with changes in biological activity

More Formatting and sequence ordering tools

# What about oligo and protein sequence alignment?

- Protein sequences
  - Just need to add in standard similarity matrices
- Oligo sequence alignment
  - Need to consider the alignment against the gene

Oligonucleotide Alignment: CM\_ATXN2\_11

Show Alignments for ID:

Species: Cynomolgus monkey

Gene Filter:

Gene Symbol	# Alignment Blocks	# Sequences in largest block
ATXN2	217	314
BCL7B	1	4
GABRB1	1	2

Alignment Blocks: Show Blocks with > 2 Sequences

Alignment Block	# Sequences	Block Start
CM_ATXN2_9	24	91110
CM_ATXN2_10	22	93729
CM_ATXN2_11	21	4216
CM_ATXN2_12	20	99715
CM_ATXN2_13	18	87740

151 of 217 Alignment Blocks shown

Sequence Alignment

Display Sequence: Base Sequence Text size: 12 Copy

```
RTR77817      ECATATATTTACCTTTEE
RTR77819      TECATATATTTACCTTTEE
RTR77818      TECATATATTTACCTTTE
RTR77820      TTECATATATTTACCTTTE
RTR77821      TTTCATATATTTACETT
RTR77822      TTTCATATATTTAEETT
RTR77823      ETTCCATATATTTACETT
RTR77825      ETTCCATATATTTACET
RTR77827      AETTTCCATATATTTACET
RTR77824      ETTCCATATATTTAE
RTR77826      AETTTCCATATATTTAE
RTR77829      AAETTTCCATATATTTAE
RTR77828      AAETTTCCATATATTTAE
RTR77831      TAAETTTCCATATATTTAE
RTR77830      TAAETTTCCATATATTTAE
RTR77832      TAAETTTCCATATATTTA
RTR77832      TATTTTAACTTTCCATAT
RTR77834      ATATTTTAACTTTCCATAT
RTR77833      ATATTTTAACTTTCEATA
RTR77836      AATATTTTAACTTTCEATA
RTR77835      AATATTTTAACTTTTEEAT
```

21 of 21 Sequences Displayed Block Start Position: 4216

Settings

Highlight Sequences

Identity  Difference

None

Base Sequence

Sugar Sequence

Stereo Sequence

Chirality Sequence

Sequence Format

None

Sequence Order

Filter:

Oligo Concept UID

Base Sequence

Assay X IC50 GMean (nM)

Gene Relative Start

Ascending  Descending

# Where Next...

	Oligos	Peptides	Antibodies
General Representation	Complete for presentation.  Sequence parsers created ad-hoc.	Complete for presentation.  Sequence parsers created ad-hoc.	Complete for presentation.  Sequence parsers from standard formats.
Basic Analysis	Alignment In Design	Alignment complete.	Alignment to be based on Peptide
Advanced Analysis		In Progress: Extraction of sub-sequence/chemical structure	

**Questions?**

