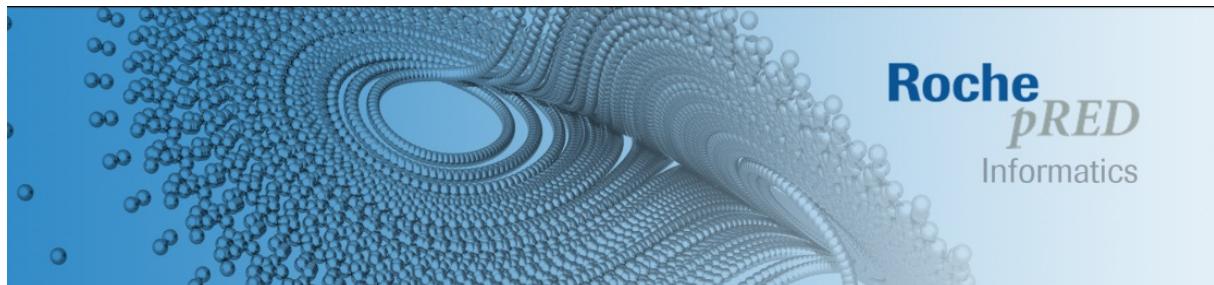


---

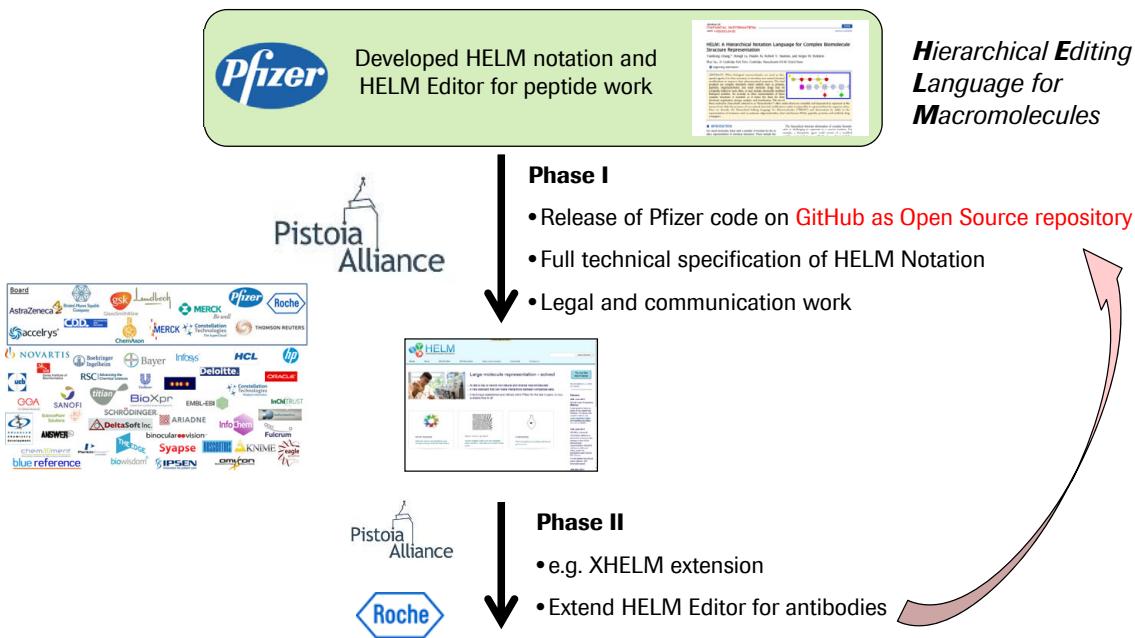
## Extending HELM to antibody space to register complex biologics

*Stefan Klostermann, Roche pREDi*

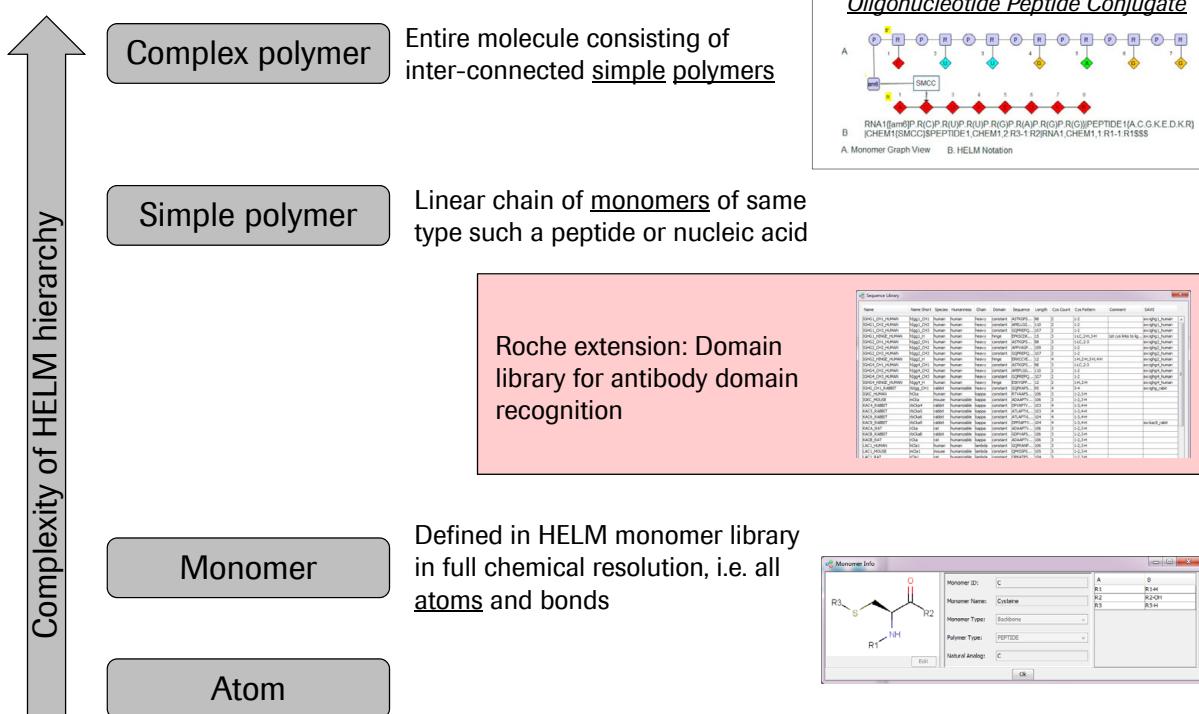


## THE STORY IN SHORT

# The collaboration



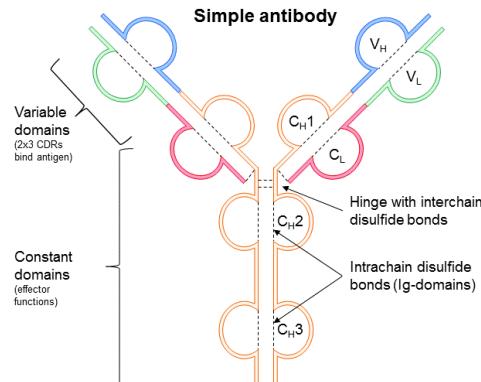
## The HELM notation Structured in 4 hierarchical layers



## The task



- **Register** complex biotherapeutics, e.g. antibodies (incl. bi-specifics, ADCs, ...)
  - to fully exploit the data generated along and across projects, platforms and technologies
- **Prerequisite** is a complete description of all components and their connections
- **Use** has to be as simple as possible



## The input



>Heavy\_Chain\_1

```
EVQLVESGGGVVKPGGSLKLSCAAAGFTFSNAWMWVRQAPGKGLEWVGRIKS KTDGGTTNYAAPVKGRFTI SRRDDSKNTLYLQMNSLKT  
EDTAVYVYCTTEFGMLWFGIFWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAA LGCLVKDYFPEPVTVWSN GALTSGVHTFP AVLQS  
SGLYSLSVVTVPSSSLGTC  
SHEDPEVKFNWYVDGVEVH  
ELTKNQVSLTCLVKGFYFPG
```

>Light\_Chain\_1

```
ELVMTQTTPASVSAAVGGTV  
TYDSSSYFFYTFGGGTKV  
TLTLTSTQYNSHKEYTCKV
```

>Heavy\_Chain\_2

```
QVQLQESGPGLVKPSETLS  
AVYYCARGRFTYFDYWGQ  
LSSVVTVPSSSLGTQTYIC  
EVFKFNWYVDGVEVHNAKTI  
QVSLTCLVKGFYFPSDIAVE
```

>Light\_Chain\_2

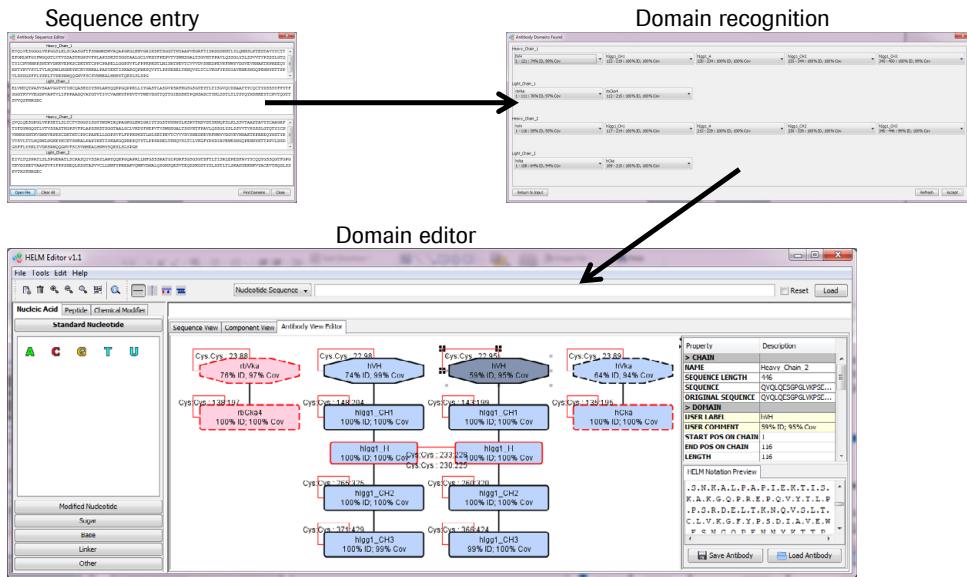
```
EIVLTQSPATLSLSPGERA  
QYGSQQGTFGPGTKVDIKRTVAAPSVFIFPPSDEQLKSGTASVVCLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYLSSTL  
TLSKADYEKHKVYACEVTHQGLSPVTKSFRGEC
```

### The challenge:

- Detect and annotate **domains**
- Auto-connect known **Cys-Cys bonds**
- Enable easy specification of remaining Cys-Cys bonds and connection of chains based on **intuitive overview**
- Alert for non-human domains and remaining free Cys residues

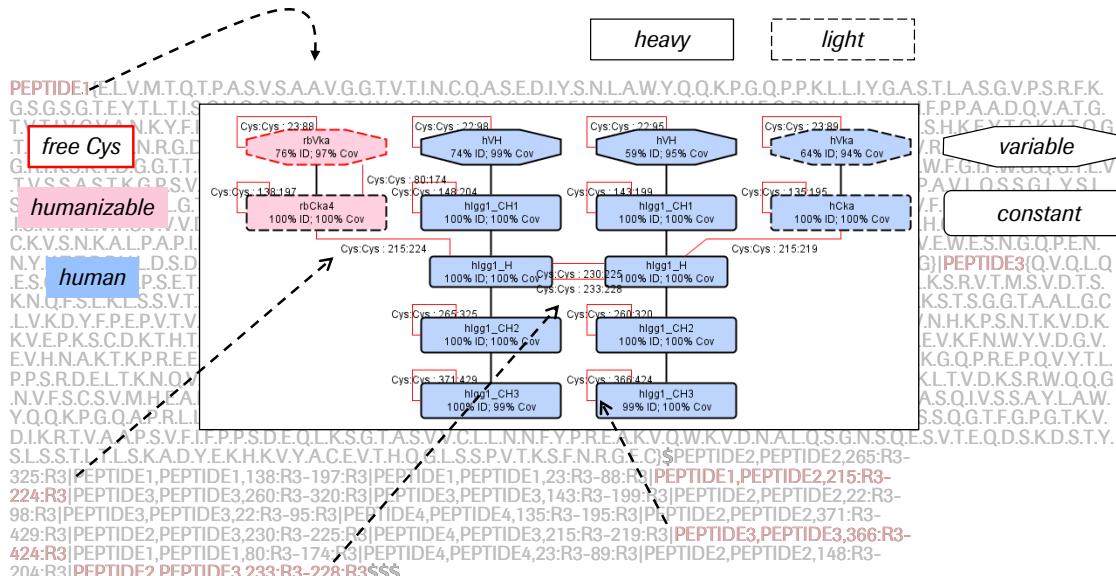
# The solution

In-house extension of public domain HELM Editor



→ Single antibody sequences converted into fully annotated antibody model including almost all Cys-Cys bonds

# The output

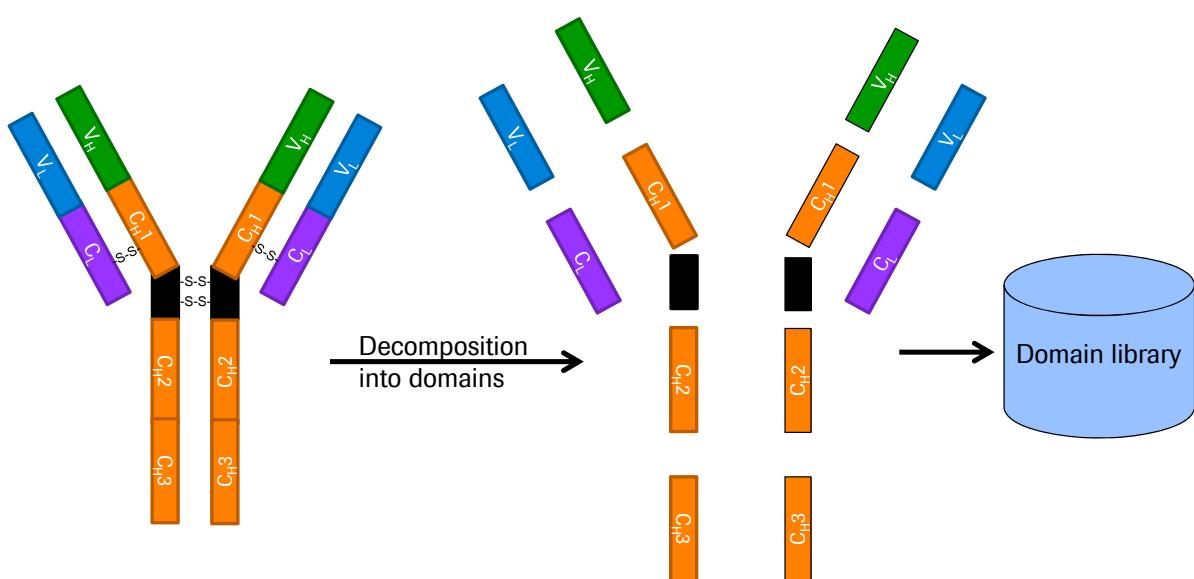


→ Complete specification on all atoms and bonds of the antibody is created in HELM notation

→ Registration of the antibody and all individual components (domains, linker, conjugates, ...)

# DOMAIN LIBRARY & DETECTION

## The domain library *Concept / Creation*



# The domain library

## Content

- Name of domains (long / short)
  - Species, humaness (humanizable)
  - Chain, domain
  - Sequence / length
  - Cysteins: number, binding pattern
- on**
- Antibody domains (h, m, r, rb)
  - Linkers, peptides
  - Conjugate proteins

Name	Name Short	Species	Humaness	Chain	Domain	Sequence	Length	Cys Count	Cys Pattern	
IGHG1_CH1_HUMAN	hIgg1_CH1	human	heavy	constant	ASTKGPS...	98	2	1-2		
IGHG1_CH2_HUMAN	hIgg1_CH2	human	heavy	constant	APELGGP...	110	2	1-2		
IGHG1_CH3_HUMAN	hIgg1_CH3	human	heavy	constant	GQPRREPO...	107	2	1-2		
IGHG1_HINGE_HUMAN	hIgg1_H	human	heavy hinge	hinge	EPKSCDK...	15	3	1+L,2+H,3+H		
IGHG2_CH1_HUMAN	hIgg2_CH1	human	heavy	constant	ASTKGPS...	98	3	1+L,C,2-3		
IGHG2_CH2_HUMAN	hIgg2_CH2	human	heavy	constant	APPVAGP...	109	2	1-2		
IGHG2_CH3_HUMAN	hIgg2_CH3	human	heavy	constant	GQPRREPO...	107	2	1-2		
IGHG2_HINGE_HUMAN	hIgg2_H	human	heavy hinge	hinge	ERKCCVE...	12	4	1+H,2+H,3+H,4H		
IGHG4_CH1_HUMAN	hIgg4_CH1	human	heavy	constant	ASTKGPS...	98	3	1+L,C,2-3		
IGHG4_CH2_HUMAN	hIgg4_CH2	human	heavy	constant	APELGG...	110	2	1-2		
IGHG4_HINGE_HUMAN	hIgg4_CH3	human	heavy	constant	GQPRREPO...	107	2	1-2		
IGKC_HINGE_HUMAN	hIgk_H	human	heavy	constant	GQKGPSP...	12	2	1+H,2+H		
IGHG_CH1_RABBIT	rIgg_CH1	rabbit	humanizable	heavy	constant	GQPNAPS...	95	4	3+L	
IGKC_HUMAN	hIgk_H	human	kappa	constant	RTVIAAPS...	106	3	1-2,3+H		
IGKC_MOUSE	mIgk_H	mouse	humanizable	kappa	constant	ADAAAPTV...	106	3	1-2,3+H	
KAC4_RABBIT	rbCk4	rabbit	humanizable	kappa	constant	DPIVAPTM...	103	4	1-3+H	
KAC5_RABBIT	rbCk5	rabbit	humanizable	kappa	constant	ATLAPTM...	103	4	1-3+H	
KAC6_RABBIT	rbCk6	rabbit	humanizable	kappa	constant	ATLAPTM...	104	4	1-3+H	
KAC9_RABBIT	rbCk9	rabbit	humanizable	kappa	constant	DPIIAPTV...	104	4	1-3+H	
KACA_RAT	rCka	rat	humanizable	kappa	constant	ADAAAPTV...	106	3	1-2,3+H	
KACB_RABBIT	rbCk8	rabbit	humanizable	kappa	constant	GDPVAPS...	106	3	1-2,3+H	
KACB_RAT	rCkb	rat	humanizable	kappa	constant	ADAAAPTV...	106	3	1-2,3+H	
LAC1_HUMAN	hCl1	human	lambda	lambda	constant	GQPKANP...	106	3	1-2,3+H	
LAC1_MOUSE	mCl1	mouse	humanizable	lambda	constant	QPKSPSP...	105	3	1-2,3+H	
LAC1_RAT	rCl1	rat	humanizable	lambda	constant	QPKATPS...	104	3	1-2,3+H	
LAC2_HUMAN	hCl2	human	lambda	lambda	constant	GQPKAAP...	106	3	1-2,3+H	
LAC2_MOUSE	mCl2	mouse	humanizable	lambda	constant	QPKSTPL...	104	3	1-2,3+H	
LAC2_RAT	rCl2	rat	humanizable	lambda	constant	QPKSTPL...	104	3	1-2,3+H	
LAC3_HUMAN	hCl3	human	lambda	lambda	constant	GQPKAAP...	106	3	1-2,3+H	
LAC3_MOUSE	mCl3	mouse	humanizable	lambda	constant	QPKSTPL...	104	3	1-2,3+H	
LAC6_HUMAN	hCl6	human	lambda	lambda	constant	GQPKAAP...	106	3	1-2,3+H	
LAC6_RABBIT	rbCk7	rabbit	humanizable	lambda	constant	GQPKAAP...	106	3	1-2,3+H	
VHEAVY-CONS_HUMAN	mH	human	heavy	variable	QVQLVER...	125	2	1-2		
VHEAVY-CONS_MOUSE	mH	mouse	humanizable	heavy	variable	QVQLQQS...	122	2	1-2	
VHEAVY-CONS_RABBIT	ybH	rabbit	humanizable	heavy	variable	QESLEES...	123	4	1-4,2-3	

*Can be extended also to any protein class with defineable domains and peptides used as building blocks*

# The algorithm of domain detection

- Run **BLAST** of entered sequence(s) against domain library
  - Retrieve hits with name, E-value, % coverage, % identity
- Apply **filters** (adjustable, >50% coverage, >50% identity)
- Apply **sorting** rules according to E-value thresholds
  - %coverage \* % identity > E-value only > % coverage
- Present top hit for each region with options to change (**GUI**)
- Match **Cys-Cys** binding pattern and auto-connect
  - **intra-domain** Cys-Cys bonds
  - **inter-chain** Cys-Cys bonds (e.g. hinge)

## The functions of domain editing

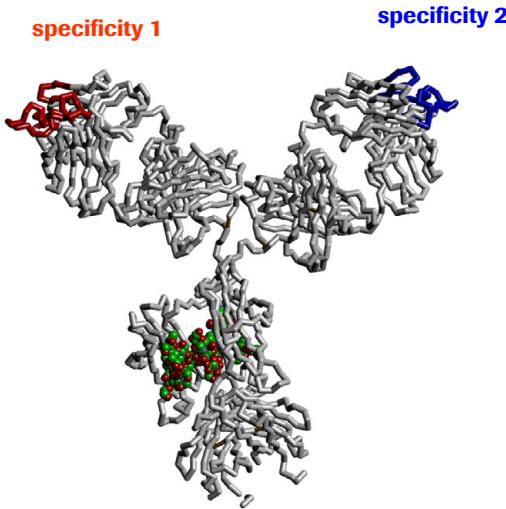
- Review full **domain data** from recognition module
  - Relabel domains
- Close remaining **Cys-Cys** bonds
  - Delete erroneous Cys-Cys bonds
- Edit domains using the **public domain HELM Editor** window
  - Attach peptides, chemical linkers and DNA/RNA to the antibody
  - Shorten / change domains (work in progress)
- **Add domains** from domain library
  - Re-annotate domains
- Up-to-date **HELM notation**

Examples will be part of a live demonstration

## EXAMPLES

# Cross-Mab (includes KiH)

Bispecific antibodies as motivation



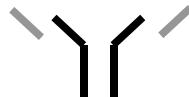
Slide by Guy Georges

# Cross-Mab (includes KiH)

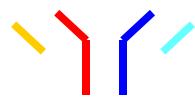
Bispecific antibodies as motivation



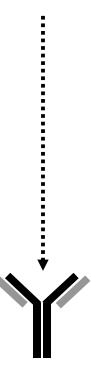
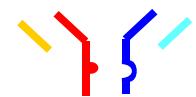
natural IgG:  
2 identical light chains  
2 identical heavy chains



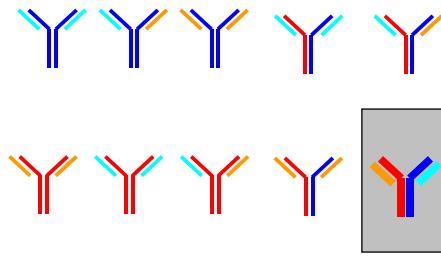
bispecific IgG:  
2 different light chains  
2 different heavy chains



knobs and holes  
2 different light chains  
2 different heavy chains



100 %



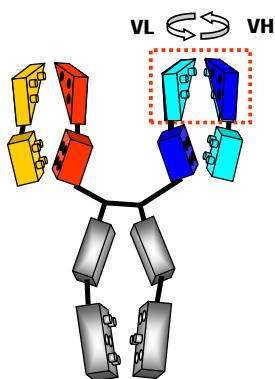
12.5 %



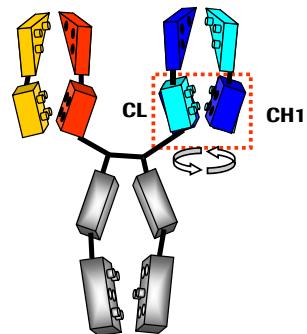
25 %

Slide by Guy Georges

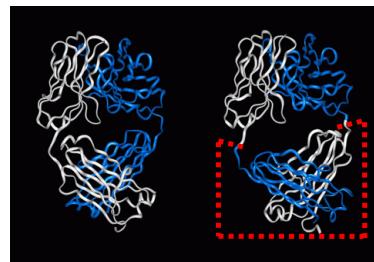
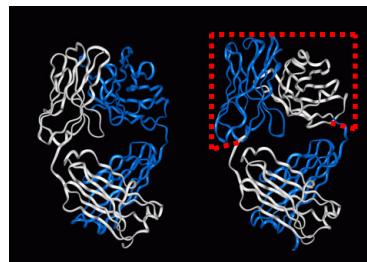
## Cross-Mab (includes KiH)



VH-VL-CrossMAb



CH1-CL-CrossMAb



Slide by Guy Georges

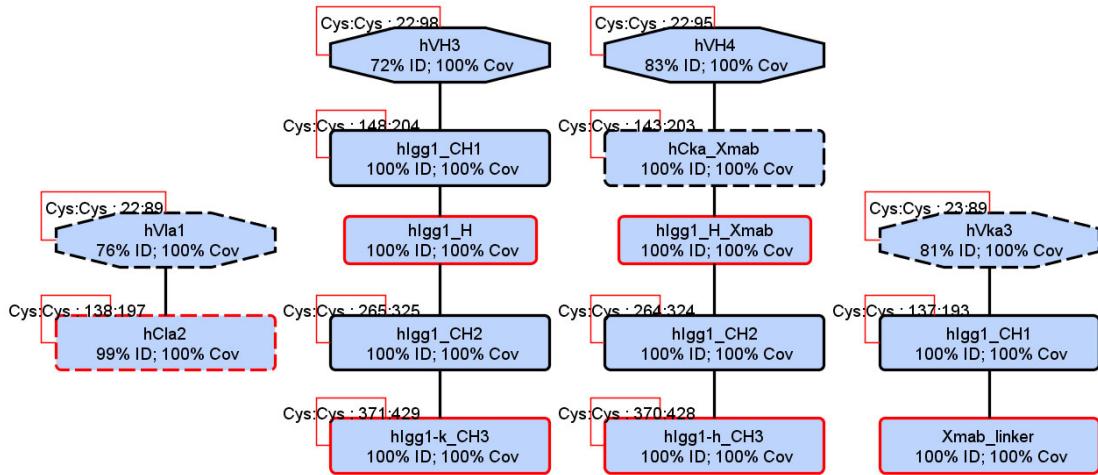
## Cross-Mab (includes KiH)



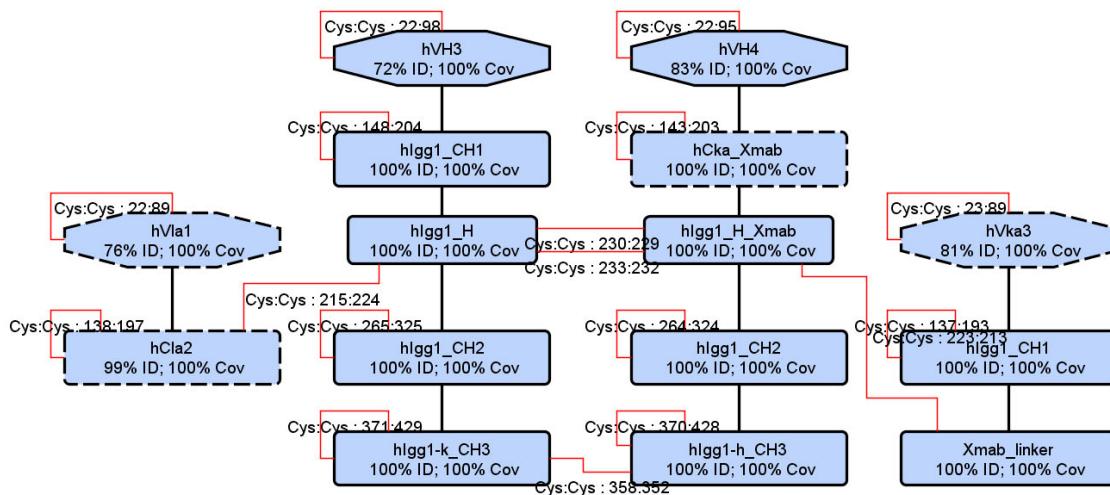
Antibody Domains Found

<b>Heavy_Chain_1</b> hVH3: 1-121 72% ID, 100% Cov eVal: 9,0E-55	hIgg1_CH1: 122-219 100% ID, 100% Cov eVal: 8,0E-65	hIgg1_H: 220-234 100% ID, 100% Cov eVal: 3,0E-08	hIgg1_CH2: 235-344 100% ID, 100% Cov eVal: 8,0E-77	hIgg1-k_CH3: 345-451 100% ID, 100% Cov eVal: 2,0E-77
<b>Light_Chain_1</b> hVLa1: 1-110 76% ID, 100% Cov eVal: 9,0E-55				
<b>Heavy_Chain_2</b> hVH4: 1-116 83% ID, 100% Cov eVal: 3,0E-63				
<b>Light_Chain_2</b> hVka3: 1-108 81% ID, 100% Cov eVal: 4,0E-60				
<b>IgG1_H_Xmab</b> : 117-218 100% ID, 100% Cov eVal: 4,0E-67				hIgg1_H_Xmab: 219-233 100% ID, 100% Cov eVal: 2,0E-08
<b>IgG1_CH2</b> : 234-343 100% ID, 100% Cov eVal: 8,0E-77				hIgg1-h_CH3: 344-450 100% ID, 100% Cov eVal: 2,0E-75
<b>Xmab_linker</b> : 209-213 100% ID, 100% Cov eVal: 1,2E01				

## Cross-Mab (includes KiH)



## Cross-Mab (includes KiH)



## Single chain variations



- Example 1: LC and HC coded by 1 protein chain
- Example 2: C-terminally added scFv

## Single chain variations



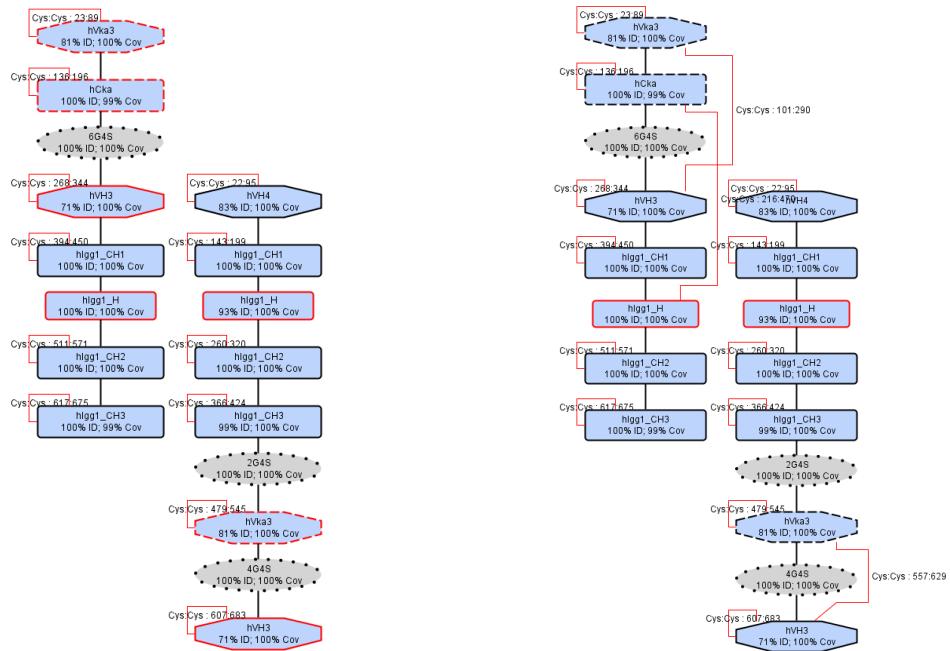
Antibody Domains Found

Light_Chain-Heavy_Chain_1							
hvKa3: 1-108 81% ID, 100% Cov eVal: 1,0E-55	hcKa: 111-216 100% ID, 99% Cov eVal: 4,0E-69	6G4S: 217-246 100% ID, 100% Cov eVal: 2,0E-17	hVH3: 247-367 71% ID, 100% Cov eVal: 2,0E-52	hIgG1_CH1: 368-465 100% ID, 100% Cov eVal: 5,0E-63	hIgG1_H: 466-480 100% ID, 100% Cov eVal: 5,0E-08	hIgG1_CH2: 481-590 100% ID, 100% Cov eVal: 1,0E-74	hIgG1_CH3: 591-696 100% ID, 99% Cov eVal: 7,0E-73

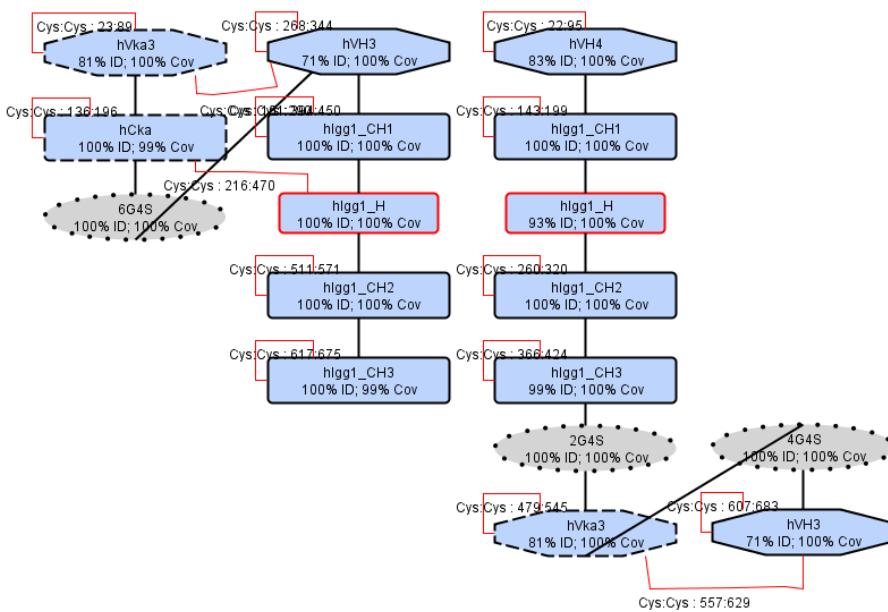
  

Heavy_Chain-scFv								
hvH4: 1-116 83% ID, 100% Cov eVal: 1,0E-61	hIgG1_CH1: 117-214 100% ID, 100% Cov eVal: 5,0E-03	hIgG1_H: 215-229 93% ID, 100% Cov eVal: 1,0E-07	hIgG1_CH2: 230-339 100% ID, 100% Cov eVal: 2,0E-74	hIgG1_CH3: 340-446 99% ID, 100% Cov eVal: 6,0E-73	2G4S: 447-456 100% ID, 100% Cov eVal: 2,0E-02	hvKa3: 457-564 81% ID, 100% Cov eVal: 1,0E-55	4G4S: 566-585 100% ID, 100% Cov eVal: 7,0E-10	hVH3: 586-706 71% ID, 100% Cov eVal: 2,0E-52

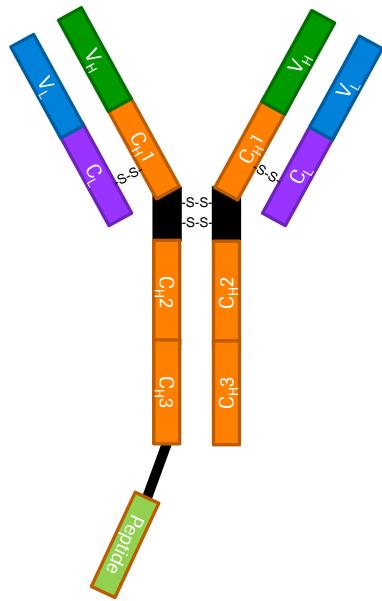
# Single chain variations



# Single chain variations



# Antibody conjugates



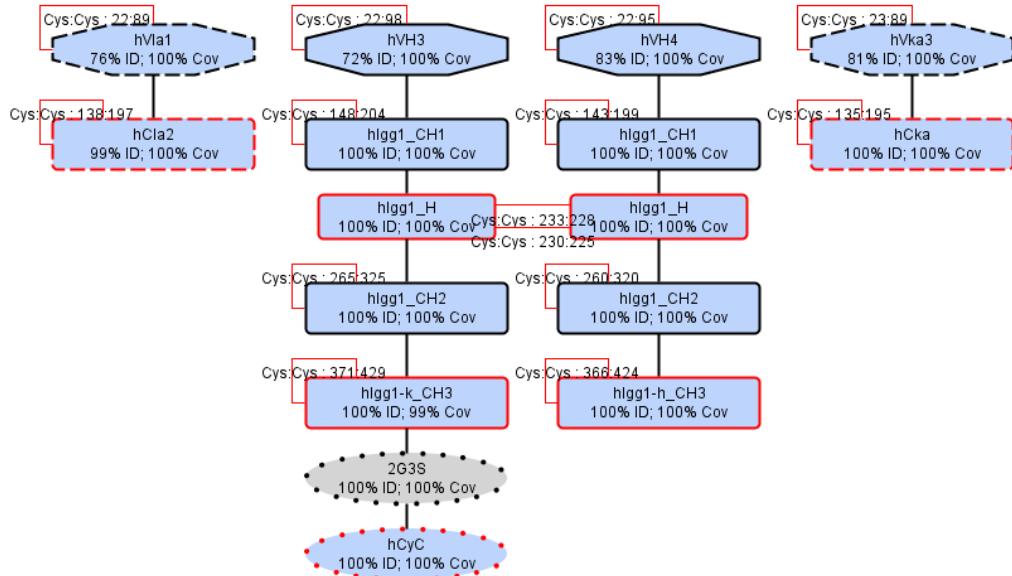
# Antibody conjugates



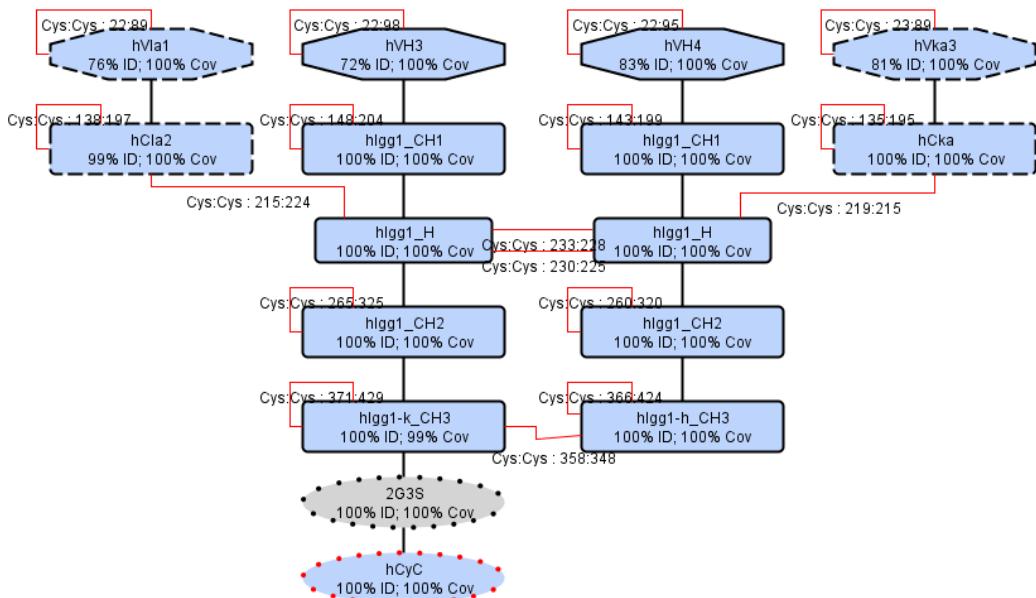
Antibody Domains Found

Heavy_Chain_1	hVH3: 1-121 72% ID, 100% Cov eVal: 5,0E-54	hIgG1_CH1: 122-219 100% ID, 100% Cov eVal: 8,0E-64	hIgG1_H: 220-234 100% ID, 100% Cov eVal: 4,0E-03	hIgG1_CH2: 235-344 100% ID, 100% Cov eVal: 1,0E-75	hIgG1_L_CH3: 345-450 100% ID, 99% Cov eVal: 2,0E-75	2G8: 451-458 100% ID, 100% Cov eVal: 5,0E-01	hCvC: 459-562 100% ID, 100% Cov eVal: 6,0E-72
Light_Chain_1	hVLa1: 1-110 76% ID, 100% Cov eVal: 5,0E-55	hClq2: 111-216 99% ID, 100% Cov eVal: 9,0E-75					
Heavy_Chain_2	hVH4: 1-116 83% ID, 100% Cov eVal: 2,0E-63	hIgG1_CH1: 117-214 100% ID, 100% Cov eVal: 7,0E-65	hIgG1_H: 215-229 100% ID, 100% Cov eVal: 3,0E-08	hIgG1_CH2: 230-339 100% ID, 100% Cov eVal: 7,0E-77	hIgG1_L_CH3: 340-446 100% ID, 100% Cov eVal: 1,0E-75		
Light_Chain_2	hVLa3: 1-108 81% ID, 100% Cov eVal: 4,0E-60	hClq3: 109-215 100% ID, 100% Cov eVal: 4,0E-75					

# Antibody conjugates



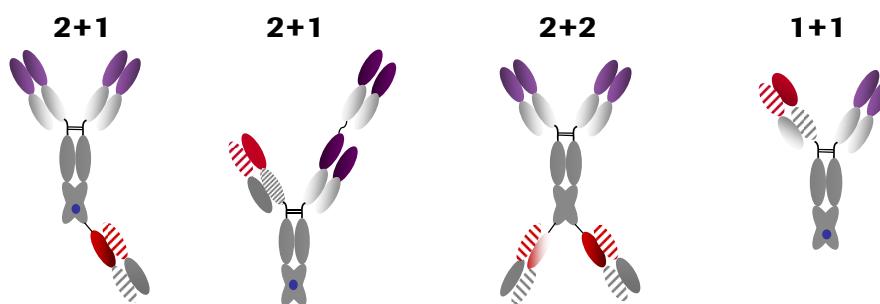
# Antibody conjugates



# WHY DEVELOP SOMETHING NEW?

## The motivation

'Standard' formats (cross-Fab based)



And glycovariants ...

... and ADCs ...

**How to store them without an added PPT?**

## Example antibody sequences



>HC heavy chain

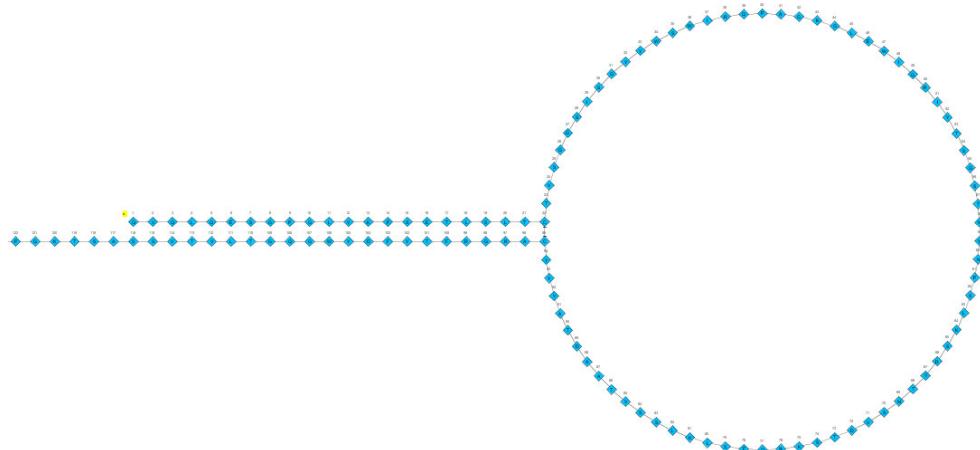
QVQLQESGPGLVKPSETLSLTCTVSGGSISGYYWSWIRQPAGKGLEWIGRIYTSGST  
NYNPSLKSRTMSVDTSKNQFSKLSSVTAADTAVYYCARGRFTYFDYWGQGTLVTV  
SSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVWSNSGALTSGVHTFPA  
VLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPC  
PAPELLGGPSVFLFPPKPDKTLMISRTPETCVVVDVSHEDEPEVKFNWYVDGVEVHN  
AKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQP  
REPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLD  
DGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK

and

>LC light chain

EIVLTQSPATLSLSPGERATLSCRASQIVSSAYILAWYQQKPGQAPRLLLFGSSSRAT  
GIPDRFSGSGSGTDFTLTISRLPEDFAVYYCQQYGSSQGTFGPGTKVDIKRTVAAP  
SVFIFFPSDEQLKSGTASVVCLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKD  
STYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRC

## Load HC, draw 1. C-C bridge manually



## **Load full antibody (2x HC, 2x LC)**



The figure shows a screenshot of the BioEdit software interface. At the top, there is a menu bar with File, Tools, Edit, Help, and several icons for file operations like Open, Save, Print, and Find. Below the menu is a toolbar with icons for Undo, Redo, Cut, Copy, Paste, and other functions. The main workspace is divided into two main sections. On the left, there is a panel for "Chemical Modifier" which includes tabs for "Nucleic Acid" (selected), "Peptide", and "Standard Nucleotide". Below this are four colored buttons: A (red), C (green), G (yellow), T/U (blue). At the bottom of this panel is a section for "Modified Nucleotide" with tabs for "Sugar", "Base", "Linker", and "Other". On the right, the main workspace displays a peptide sequence: LNNFYPREAKYQWKV/DNALQSGHNSQESVTEQDSKDSLYSLSLTLSKADYEKIKY/ACEVTHQGLSPVTKSFNRGEC. There are two horizontal scroll bars below the sequence. At the bottom of the workspace, there are two tabs: "Sequence View" (selected) and "Component View". At the very bottom, there is a legend for "Number of Modifications" with four entries: 0 (black square), 1 (green square), 2 (magenta square), and Modified Phosphate (dark blue square).

# Antibody HELM converter – POC in Excel

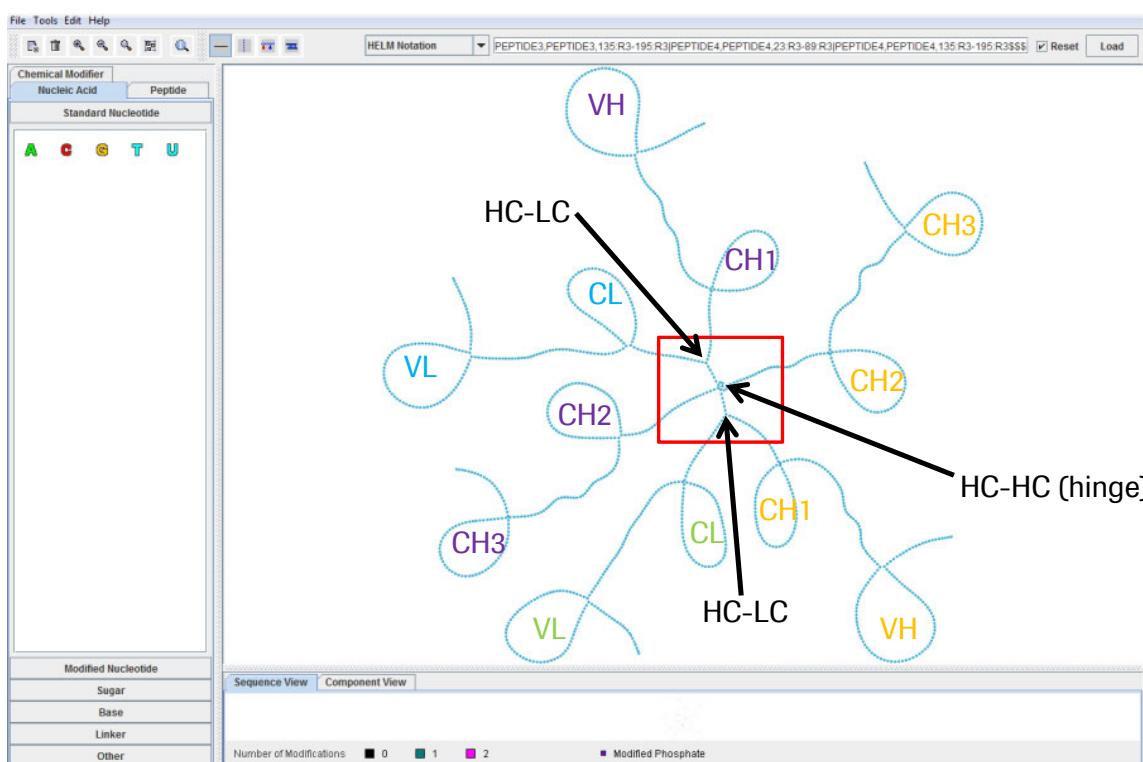


# Input to HELM Editor

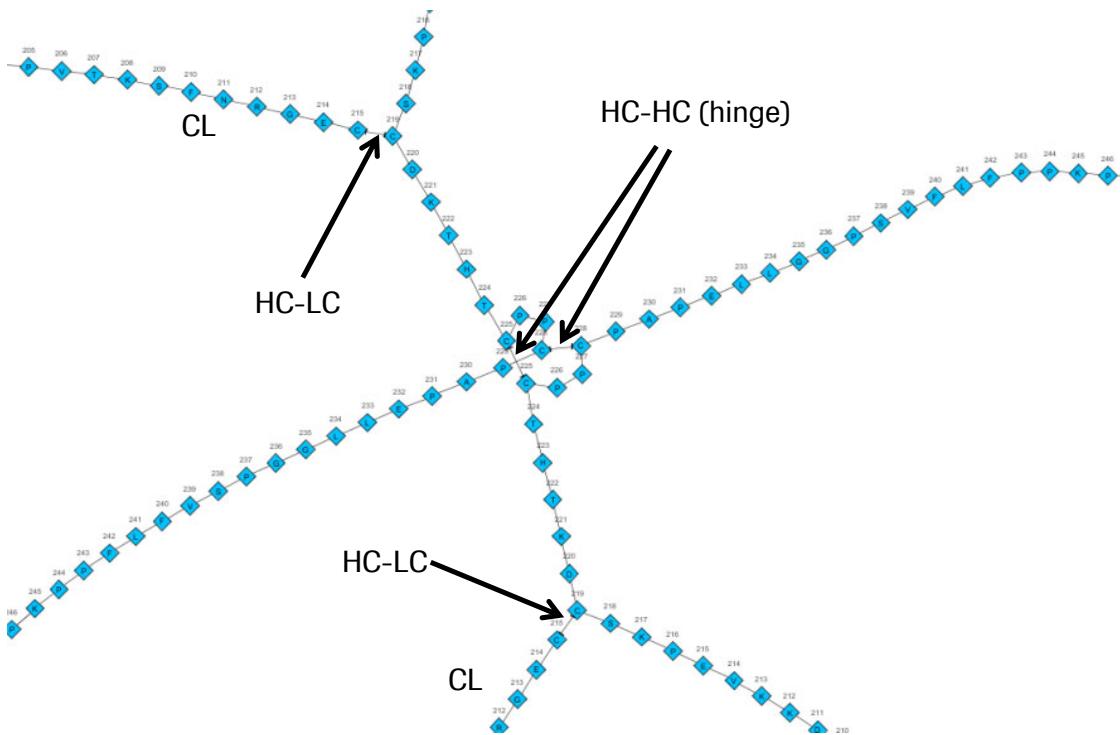


• PEPTIDE1{Q.V.Q.L.Q.E.S.G.P.G.L.V.K.K.P.S.E.T.L.S.L.T.C.T.V.S.G.G.S.I.S.G.Y.Y.W.S.W.I.R.Q.P.A.G.K.G.L.E.W.I.G.R.I.Y.T.S.G.S.T.N.Y.N.P.S.L.K.S.R.V.T.M.S.V.D.T.S.K.N.Q.F.S.L.K.L.S.S.V.T.A.A.D.T.A.V.Y.Y.C.A.R.G.R.F.T.Y.F.D.Y.W.G.Q.G.T.L.V.T.V.S.S.A.S.T.K.G.P.S.V.F.P.L.A.P.S.S.K.S.T.S.G.G.T.A.A.L.G.C.L.V.K.D.Y.F.P.E.P.V.T.V.S.W.N.S.G.A.L.T.S.G.V.H.T.F.P.A.V.L.Q.S.S.G.L.Y.S.L.S.S.V.V.T.V.P.S.S.S.S.L.G.T.Q.T.Y.I.C.N.V.N.H.K.P.S.N.T.K.V.D.K.K.V.E.P.K.S.C.D.K.T.H.T.C.P.P.E.T.C.C.V.V.D.V.S.H.E.D.P.E.V.K.F.N.W.Y.V.D.G.V.E.V.H.N.A.K.T.K.P.R.E.E.Q.Y.N.S.T.Y.R.V.V.S.V.L.T.V.L.H.Q.D.W.L.N.G.K.E.Y.K.C.K.V.S.N.K.A.L.P.A.P.I.E.L.T.C.L.V.K.G.F.Y.P.S.D.I.A.V.E.W.E.S.N.G.Q.P.E.N.N.Y.K.T.T.P.P.V.L.D.S.D.G.S.F.F.L.Y.S.K.L.T.V.D.K.S.R.W.Q.Q.G.N.V.F.S.C.S.V.M.H.E.A.L.H.N.H.Y.T.Q.K.S.L.S.P.G.K}|PEPTIDE2{Q.V.Q.L.Q.E.S.G.P.G.L.V.K.P.S.E.T.L.S.L.T.C.T.V.S.G.G.S.I.S.G.Y.Y.W.S.W.I.R.Q.P.A.G.K.G.L.E.W.I.G.R.I.Y.T.S.G.S.T.N.Y.N.P.S.L.K.S.R.V.T.M.S.V.D.T.S.K.N.Q.F.S.L.K.L.S.S.V.T.A.A.D.T.A.V.Y.Y.C.A.R.G.R.F.T.Y.W.G.Q.G.T.L.V.T.V.S.S.A.S.T.K.G.P.S.V.F.P.L.A.P.S.S.K.S.T.S.G.G.T.A.A.L.G.C.L.V.K.D.Y.F.P.E.P.V.T.V.S.W.N.S.G.A.L.T.S.G.V.H.T.F.P.A.V.L.Q.S.S.G.L.Y.S.L.S.S.V.V.T.V.P.S.S.S.S.L.G.T.Q.T.Y.I.C.N.V.N.H.K.P.S.N.T.K.V.D.K.K.V.E.P.K.S.C.D.K.T.H.T.C.P.P.E.C.P.A.P.E.L.L.G.G.P.S.V.F.L.F.P.P.K.P.K.D.T.L.M.I.S.R.T.P.E.V.T.C.V.V.D.V.S.H.E.D.P.E.V.K.F.N.W.Y.V.D.G.V.E.V.H.N.A.K.T.K.P.R.E.E.Q.Y.N.S.T.Y.R.V.V.S.V.L.T.V.L.H.Q.D.W.L.N.G.K.E.Y.K.C.K.V.S.N.K.A.L.P.A.P.I.E.K.T.I.S.K.A.K.G.Q.P.R.E.P.Q.V.Y.T.L.P.P.S.R.D.E.L.T.K.N.Q.V.S.L.T.C.L.V.K.G.F.Y.P.S.D.I.A.V.E.W.E.S.N.G.Q.P.E.N.N.Y.K.T.T.P.P.V.L.D.S.D.G.S.F.F.L.Y.S.K.L.T.V.D.K.K.V.E.P.K.S.C.D.K.T.H.T.C.P.P.E.K.S.L.S.P.G.K}|PEPTIDE3{E.I.V.L.T.Q.S.P.A.T.L.S.L.P.G.E.R.A.T.L.S.C.R.A.S.Q.I.V.S.S.A.Y.L.A.W.Y.Q.Q.K.P.G.Q.A.P.R.L.L.M.F.G.S.S.S.R.A.T.G.I.P.D.R.F.S.G.S.G.S.G.T.D.F.T.L.T.I.S.R.L.E.P.E.D.F.A.V.Y.Y.C.Q.Q.Y.G.S.S.Q.G.T.F.G.P.G.T.K.V.D.I.K.R.T.V.A.A.P.S.V.F.I.F.P.P.S.D.E.Q.L.K.S.G.T.A.S.V.V.C.L.L.N.N.F.Y.P.R.E.A.K.V.Q.W.K.V.D.N.A.L.Q.S.G.N.S.Q.E.S.V.T.E.Q.D.S.K.D.S.T.Y.S.L.S.S.T.L.T.L.S.K.A.D.Y.E.K.H.K.V.Y.A.C.E.V.T.H.Q.G.L.S.S.P.V.T.K.S.F.N.R.G.E.C}|PEPTIDE4{E.I.V.L.T.Q.S.P.A.T.L.S.L.S.P.G.E.R.A.T.L.S.C.R.A.S.Q.I.V.S.A.Y.L.A.W.Y.Q.Q.K.S.L.S.L.S.P.G.K}|PEPTIDE1, PEPTIDE2, 225:R3-95:R3|PEPTIDE1, PEPTIDE3, 219:R3-215:R3|PEPTIDE1, PEPTIDE2, 225:R3-225:R3|PEPTIDE1, PEPTIDE2, 228:R3-228:R3|PEPTIDE1, PEPTIDE1, 260:R3-320:R3|PEPTIDE1, PEPTIDE1, 366:R3-424:R3|PEPTIDE2, PEPTIDE2, 143:R3-199:R3|PEPTIDE2, PEPTIDE2, 143:R3-199:R3|PEPTIDE2, PEPTIDE4, 219:R3-215:R3|PEPTIDE2, PEPTIDE2, 260:R3-320:R3|PEPTIDE2, PEPTIDE2, 366:R3-424:R3|PEPTIDE3, PEPTIDE3, 23:R3-89:R3|PEPTIDE3, PEPTIDE3, 135:R3-195:R3|PEPTIDE4, PEPTIDE4, 23:R3-89:R3|PEPTIDE4, PEPTIDE4, 135:R3-195:R3\$\$\$

# HELM Editor Output



## Zoom reveals Cys-Cys bridges



## HELM Editor build-in alternatives



**Protein Sequence Editor**

This screenshot shows the Protein Sequence Editor interface. It includes a sequences table, annotations for the HC chain, and a text input field for protein sequences. A connections table allows users to define bonds between amino acids across different chains.

## ADC Editor

**ADC Editor**

This screenshot shows the ADC Editor interface, which is designed for antibody-drug conjugate (ADC) construction. It features a sequences table, annotations, and a detailed connections table. The interface also includes a toolbar for common operations and a molecular structure viewer at the bottom.

## Release to public domain



- Roche Extension of HELM Editor for complex antibody formats expected in Q3 2014

## Acknowledgements

*To the developers*



**Domain editor and framework** (Pharma Research and Early Development Informatics, Roche Diagnostics GmbH, Penzberg, Germany)

- Pandu Raharja
- Stefan Zilch

**Domain recognition** (quattro research GmbH, Martinsried, Germany)

- Anne Mund
- Marco Lanig

## **Associated HELM talk on the Bio-IT 2014**

*Please visit!*



HARNESSING DATA & STANDARDS

**WEDNESDAY, APRIL 30**

### **1:55 HELM: An Open Standard for the Representation of Complex Biomolecules**

*Sergio Rotstein, Ph.D., Director, Research Business Technology, Pfizer, Inc.*

The steady increase in therapeutic research involving complex biologic entities has exposed a gap in the ability of traditional informatics tools to deal gracefully with these types of molecules. HELM is a new open standard that enables the representation of a diverse set of complex macromolecules such as oligonucleotides, proteins, antibodies, antibody-drug conjugates, etc., including those containing unnatural and chemically-modified components. This presentation will describe the HELM standard and associated toolkit, its origins and use at Pfizer, and the Pistoia Alliance HELM project that has transitioned the technology into the open source, making it available freely and openly to the biopharmaceutical industry at large.



***Doing now what patients need next***