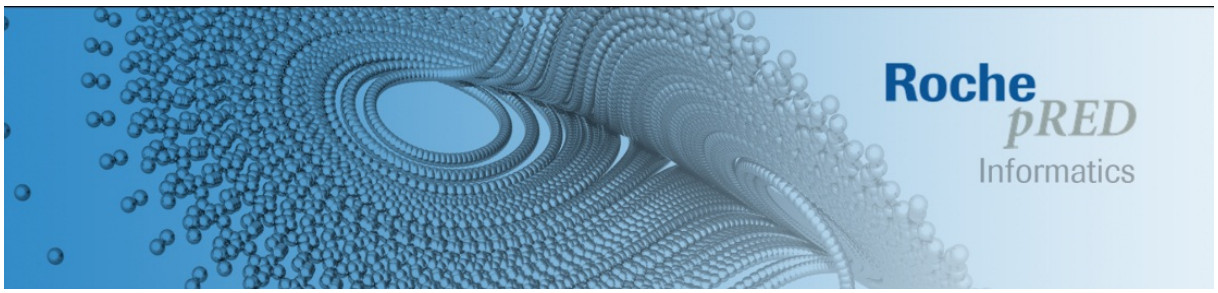
Extending HELM to antibody space to register complex biologics

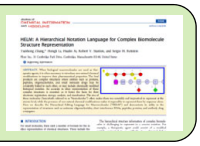
Stefan Klostermann, Roche pREDi



THE STORY IN SHORT

The collaboration

Developed HELM notation and HELM Editor for peptide work



***H**ierarchical **E**diting **L**anguage for **M**acromolecules*

Pistoia Alliance

Phase I

- Release of Pfizer code on GitHub as Open Source repository
- Full technical specification of HELM Notation
- Legal and communication work

Pistoia Alliance

Roche

Phase II

- e.g. XHELM extension
- Extend HELM Editor for antibodies

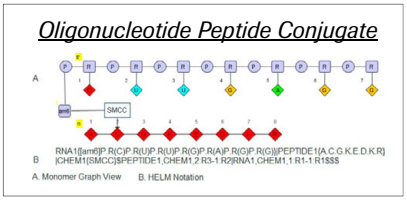
The HELM notation

Structured in 4 hierachical layers

Complexity of HELM hierarchy

Complex polymer

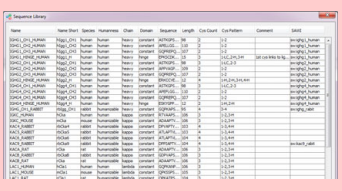
Entire molecule consisting of inter-connected simple polymers



Simple polymer

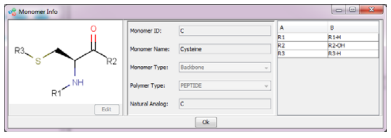
Linear chain of monomers of same type such a peptide or nucleic acid

Roche extension: Domain library for antibody domain recognition



Monomer

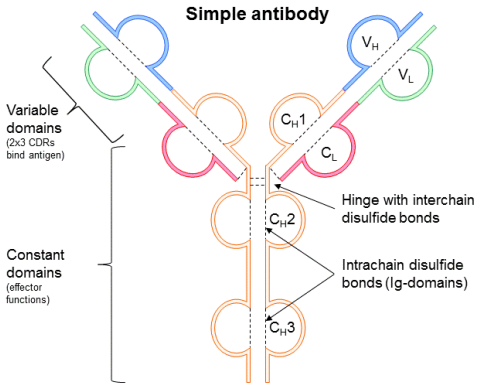
Defined in HELM monomer library in full chemical resolution, i.e. all atoms and bonds



Atom

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

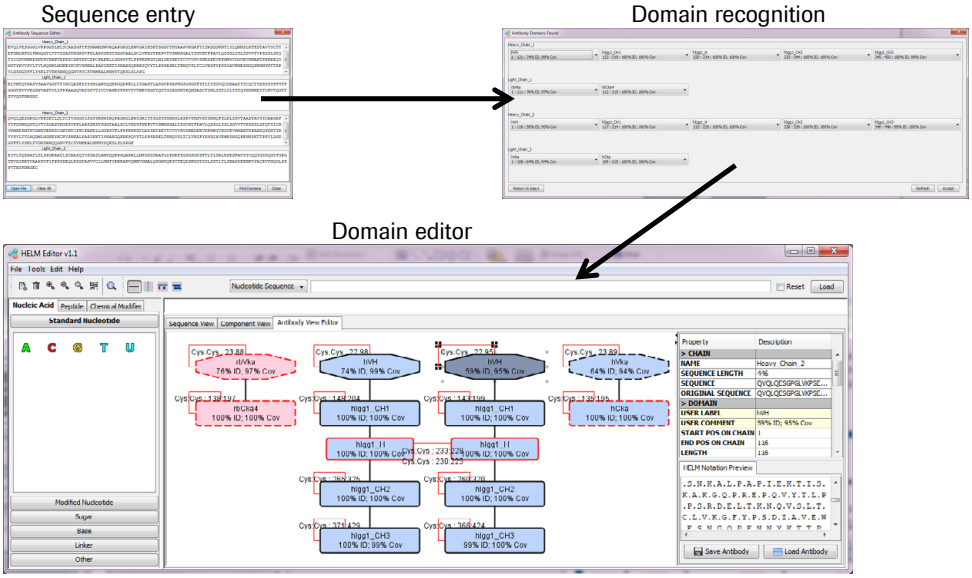
```
>Heavy_Chain_1

EVQLVESGGGLVKPGGSLKLSCAASGFTFSNAWMSWVRQAPGKGLEWVGRIKSKTDGGTTNYAAPVKGRFTISRDDSKNTLYLQMNSLKT
EDTAVYYCTTEFGMLWFGIFWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQS
SGLYSLSSVVTVPSSSLGT                                                    KDTLMISRTPEVTCVVVDV
SHEDPEVKFNWYVDGVEVH                                                    KAKGQPREPQVYTLPPSRD
ELTKNQVSLTCLVKGFYPS                                                    MHEALHNHYTQKSLSLSPG

>Light_Chain_1

ELVMTQTPASVSAAVGGTV                                                    TLTISGVQCDDAATYYCQC
TYDSSSYFFYTFGGGTKVV                                                    IENSKTPQNSADCTYNLSS
TLTLTSTQYNSHKEYTCKV

>Heavy_Chain_2

QVQLQESGPGLVKPSETLS                                                    DTSKNQFSLKLSSVTAADT
AVYYCARGRFTYFDYWGQG                                                    LTSGVHTFPAVLQSSGLYS
LSSVVTVPSSSLGTQTYIC                                                    ISRTPEVTCVVVDVSHEDP
EVKFNWYVDGVEVHNAKTK                                                    PREPQVYTLPPSRDELTKN
QVSLTCLVKGFYPSDIAVE                                                    HNHYSQKSLSLSPGK

>Light_Chain_2

EIVLTQSPATLSLSPGERA                                                    FTLTISRLEPEDFAVYYCQ
QYGSSQGTFGPGTKVDIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTL
TLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC
```

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

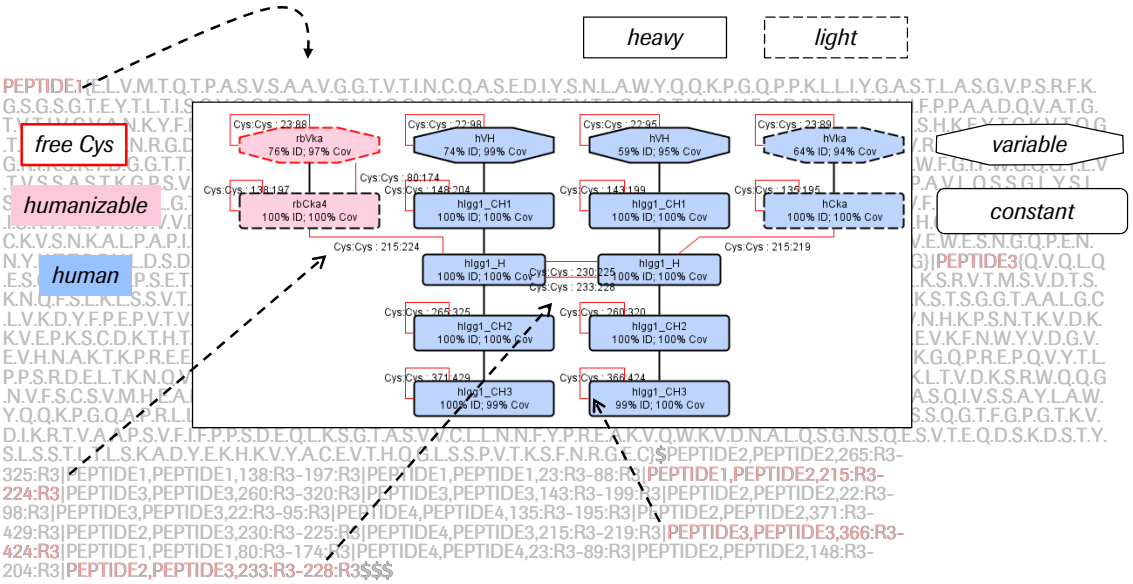
The solution

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



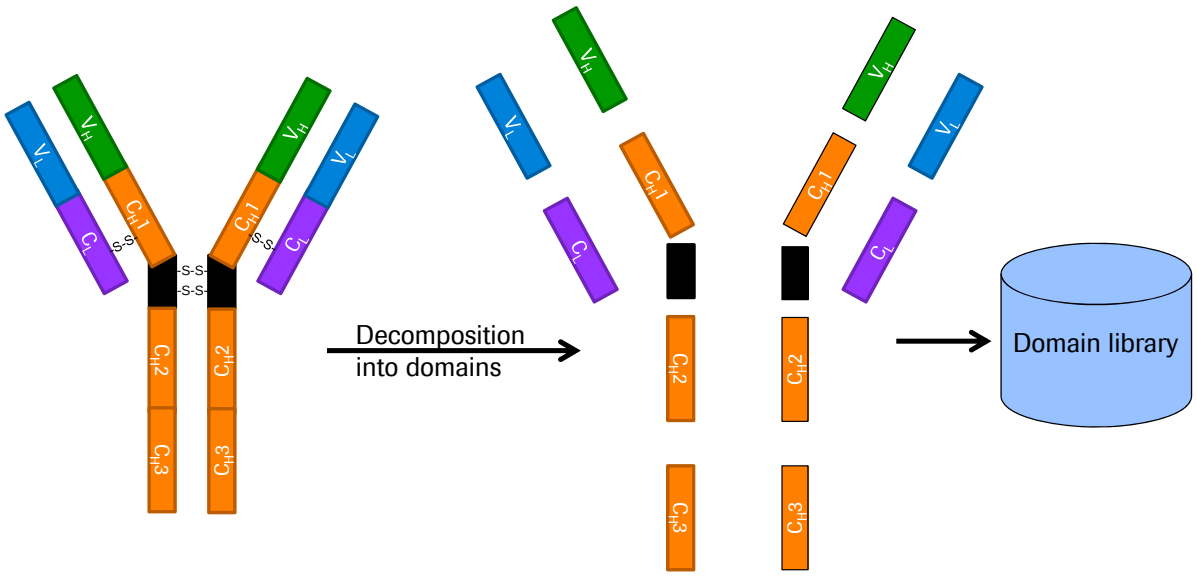
→ **Complete specification on all atoms and bonds of the antibody is created in HELM notation**

→ **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	Humanness	Chain	Domain	Sequence	Length	Cys Count	Cys Pattern
IGHG1_CH1_HUMAN	hIgg1_CH1	human	human	heavy	constant	ASTKGPS...	98	2	1-2
IGHG1_CH2_HUMAN	hIgg1_CH2	human	human	heavy	constant	APELLGGP...	110	2	1-2
IGHG1_CH3_HUMAN	hIgg1_CH3	human	human	heavy	constant	GQPREPQ...	107	2	1-2
IGHG1_HINGE_HUMAN	hIgg1_H	human	human	heavy	hinge	EPKSCDK...	15	3	1-LC,2-H,3-H
IGHG2_CH1_HUMAN	hIgg2_CH1	human	human	heavy	constant	ASTKGPS...	98	3	1-LC,2-3
IGHG2_CH2_HUMAN	hIgg2_CH2	human	human	heavy	constant	APPVAGP...	109	2	1-2
IGHG2_CH3_HUMAN	hIgg2_CH3	human	human	heavy	constant	GQPREPQ...	107	2	1-2
IGHG2_HINGE_HUMAN	hIgg2_H	human	human	heavy	hinge	ERKCCVE...	12	4	1-H,2-H,3-H,4-H
IGHG4_CH1_HUMAN	hIgg4_CH1	human	human	heavy	constant	ASTKGPS...	98	3	1-LC,2-3
IGHG4_CH2_HUMAN	hIgg4_CH2	human	human	heavy	constant	APEFLGG...	110	2	1-2
IGHG4_CH3_HUMAN	hIgg4_CH3	human	human	heavy	constant	GQPREPQ...	107	2	1-2
IGHG4_HINGE_HUMAN	hIgg4_H	human	human	heavy	hinge	ESKYGPP...	12	2	1-H,2-H
IGHG_CH1_RABBIT	rbIgg_CH1	rabbit	humanizable	heavy	constant	GQPKAPS...	95	4	3-4
IGKC_HUMAN	hCka	human	human	kappa	constant	RTVAAPS...	106	3	1-2,3-H
IGKC_MOUSE	mCka	mouse	humanizable	kappa	constant	ADAAPTV...	106	3	1-2,3-H
KAC4_RABBIT	rbCka4	rabbit	humanizable	kappa	constant	DPVAPTVL...	103	4	1-3,4-H
KAC5_RABBIT	rbCka5	rabbit	humanizable	kappa	constant	ATLAPTVL...	103	4	1-3,4-H
KAC6_RABBIT	rbCka6	rabbit	humanizable	kappa	constant	ATLAPTVL...	104	4	1-3,4-H
KAC9_RABBIT	rbCka9	rabbit	humanizable	kappa	constant	DPPIAPTV...	104	4	1-3,4-H
KACA_RAT	rCka	rat	humanizable	kappa	constant	ADAAPTV...	106	3	1-2,3-H
KACB_RABBIT	rbCka8	rabbit	humanizable	kappa	constant	GDPVAPS...	106	3	1-2,3-H
KACB_RAT	rCka	rat	humanizable	kappa	constant	ADAAPTV...	106	3	1-2,3-H
LAC1_HUMAN	hCla1	human	human	lambda	constant	GQPKANP...	106	3	1-2,3-H
LAC1_MOUSE	mCla1	mouse	humanizable	lambda	constant	QPKSSPS...	105	3	1-2,3-H
LAC1_RAT	rCla1	rat	humanizable	lambda	constant	QPKATPS...	104	3	1-2,3-H
LAC2_HUMAN	hCla2	human	human	lambda	constant	GQPKAAP...	106	3	1-2,3-H
LAC2_MOUSE	mCla2	mouse	humanizable	lambda	constant	QPKSTPTL...	104	3	1-2,3-H
LAC2_RAT	rCla2	rat	humanizable	lambda	constant	QPKSTPTL...	104	3	1-2,3-H
LAC3_HUMAN	hCla3	human	human	lambda	constant	GQPKAAP...	106	3	1-2,3-H
LAC3_MOUSE	mCla3	mouse	humanizable	lambda	constant	QPKSTPTL...	104	3	1-2,3-H
LAC6_HUMAN	hCla6	human	human	lambda	constant	GQPKAAP...	106	3	1-2,3-H
LAC7_HUMAN	hCla7	human	human	lambda	constant	GQPKAAP...	106	3	1-2,3-H
LAC_RABBIT	rbCla	rabbit	humanizable	lambda	constant	QPAVTPS...	105	3	1-2,3-H
VHEAVY-CONS_HUMAN	hVH	human	human	heavy	variable	QVQLVES...	122	2	1-2
VHEAVY-CONS_MOUSE	mVH	mouse	humanizable	heavy	variable	QVQLQQS...	122	2	1-2
VHEAVY-CONS_RABBIT	rbVH	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

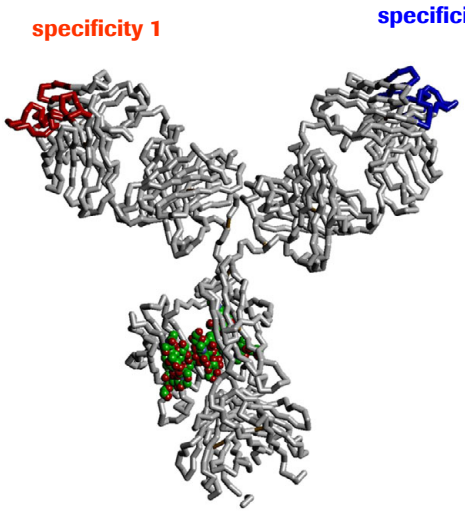
- Review full domain data from recognition module
  - Relabel domains
- Close remaining Cys-Cys bonds
  - Delete erroneous Cys-Cys bonds
- Edit domains using the public domain HELM Editor window
  - Attach peptides, chemical linkers and DNA/RNA to the antibody
  - Shorten / change domains (work in progress)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

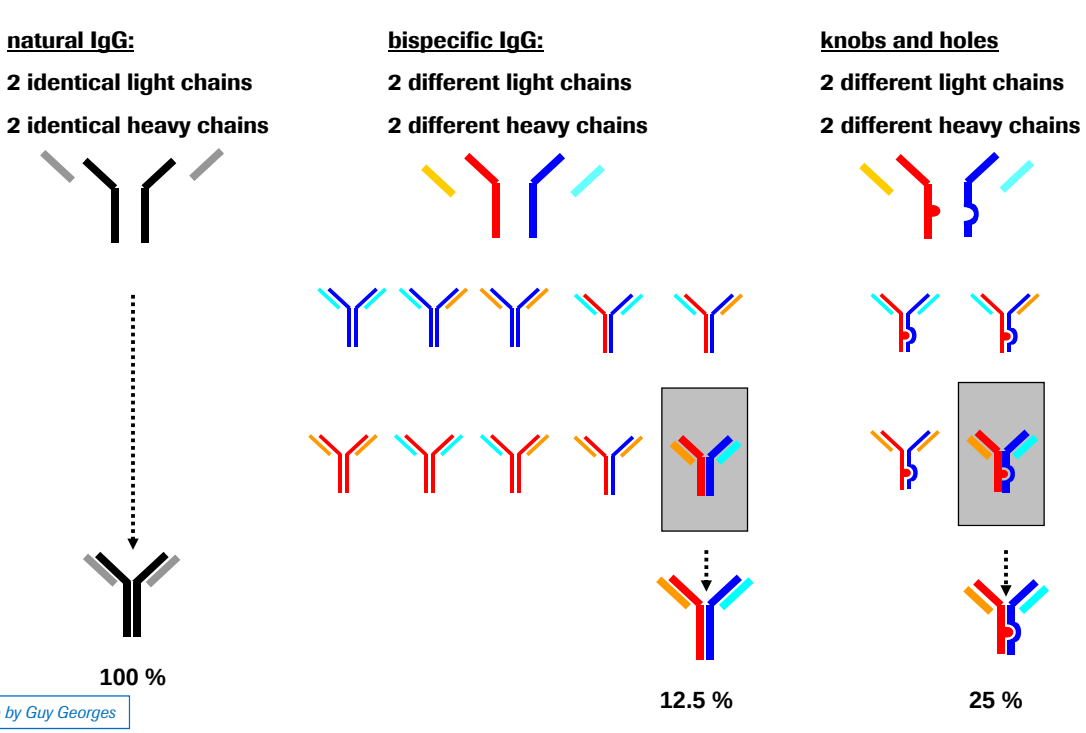
Bispecific antibodies as motivation



Slide by Guy Georges

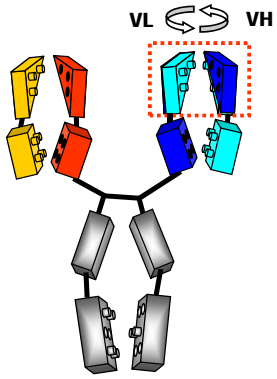
Cross-Mab (includes KiH)

Bispecific antibodies as motivation

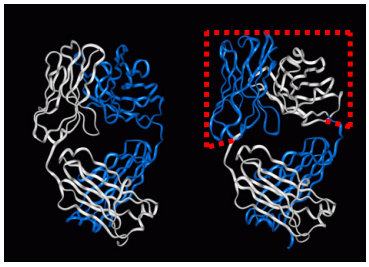


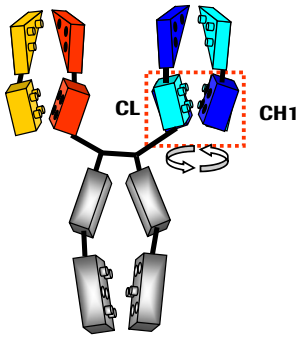
Slide by Guy Georges

Cross-Mab (includes KiH)

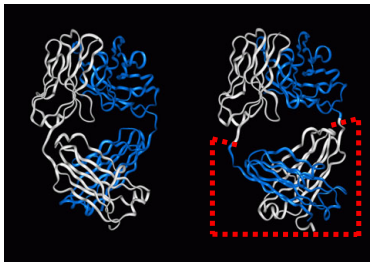


VH-VL-CrossMAb



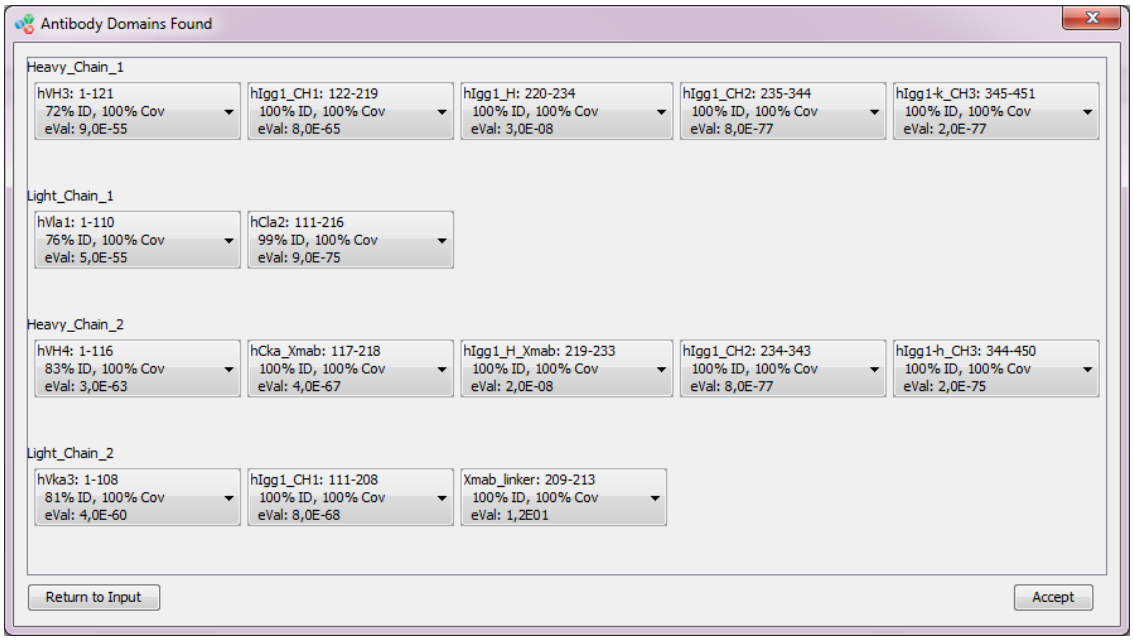


CH1-CL-CrossMAb

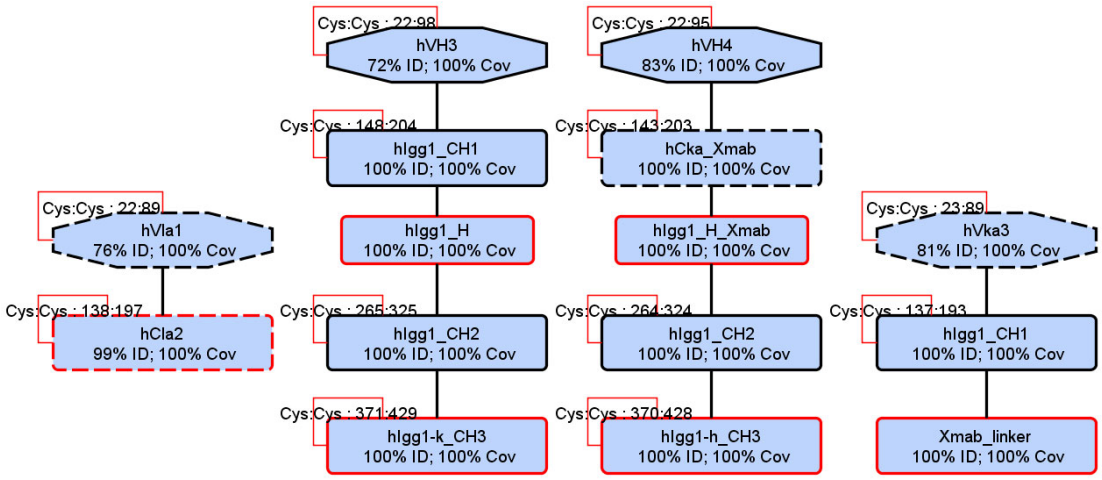


Slide by Guy Georges

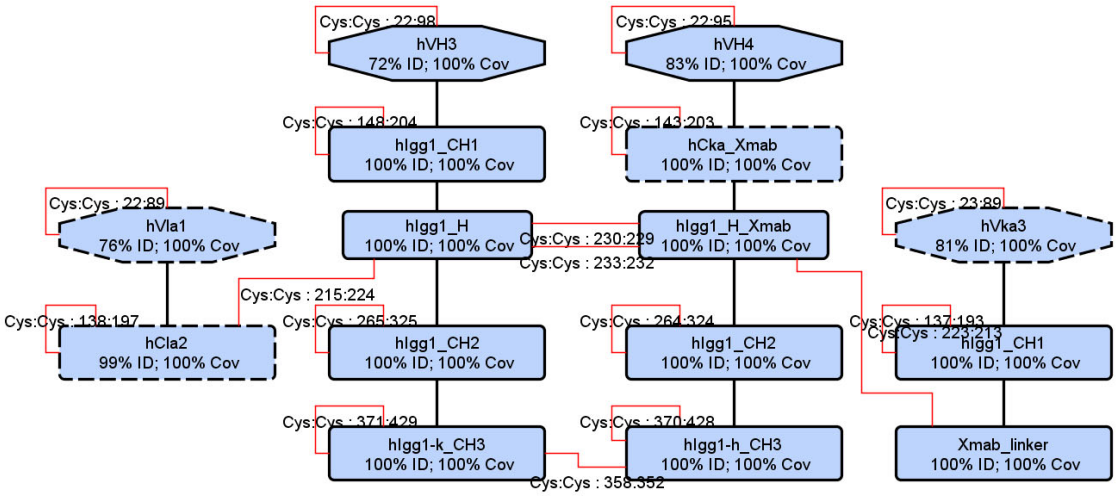
Cross-Mab (includes KiH)



Cross-Mab (includes KiH)



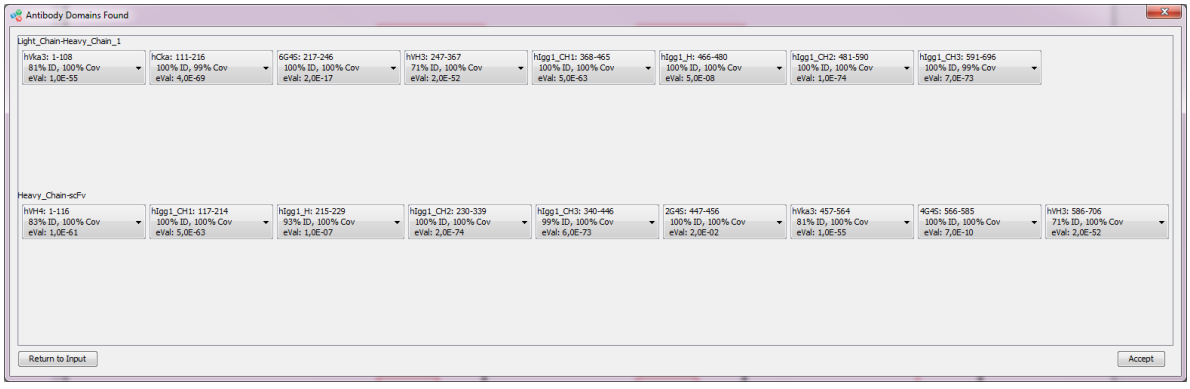
Cross-Mab (includes KiH)



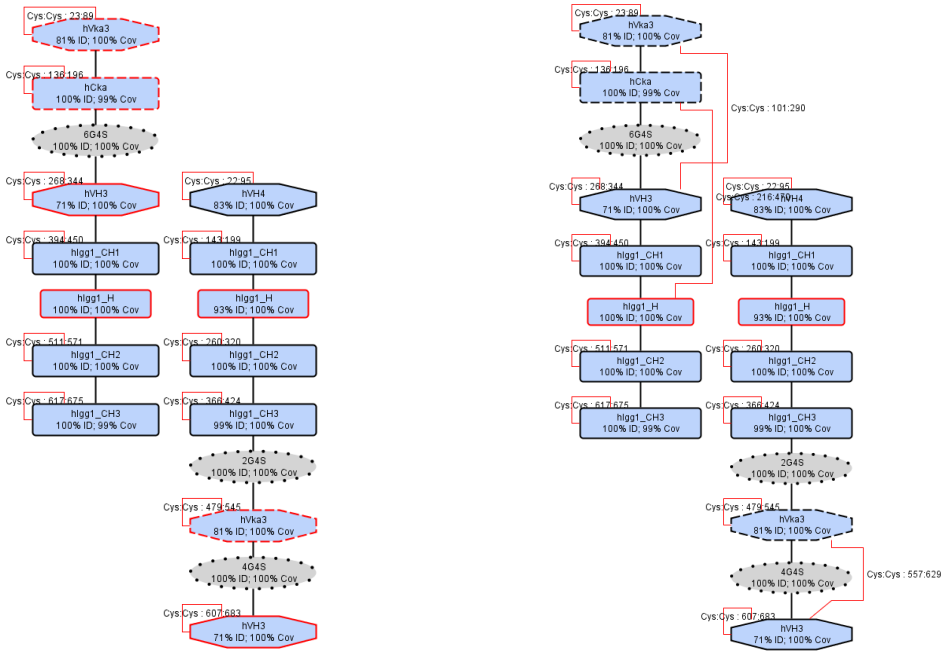
Single chain variations

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

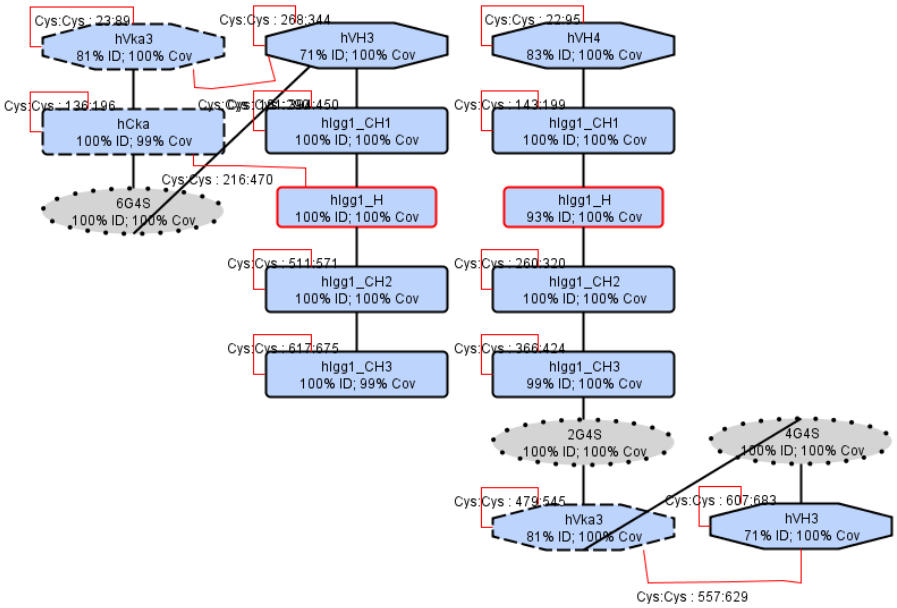
Single chain variations



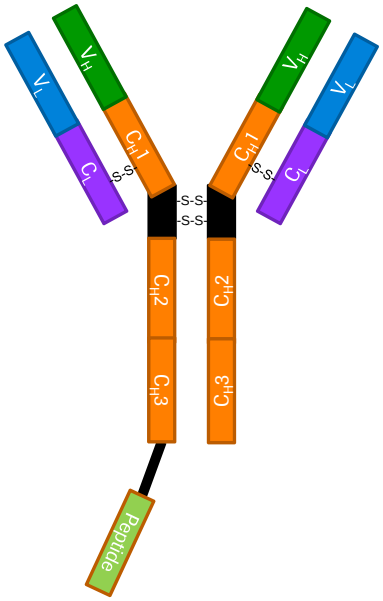
Single chain variations



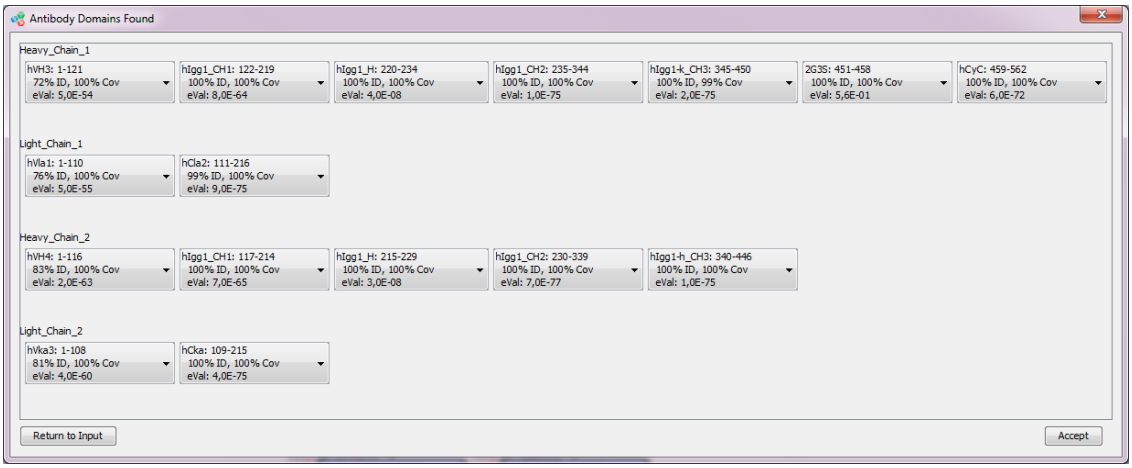
Single chain variations



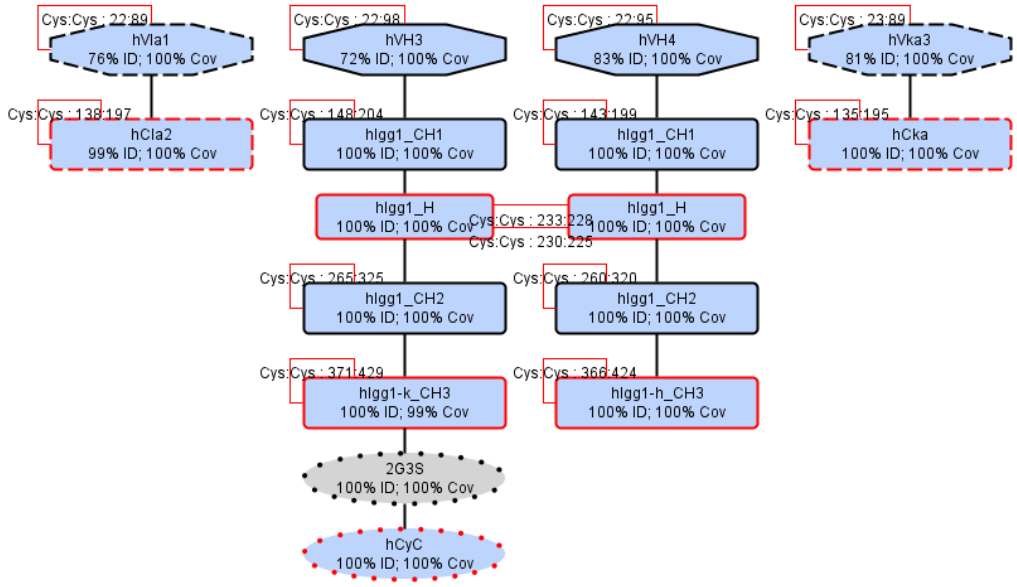
Antibody conjugates



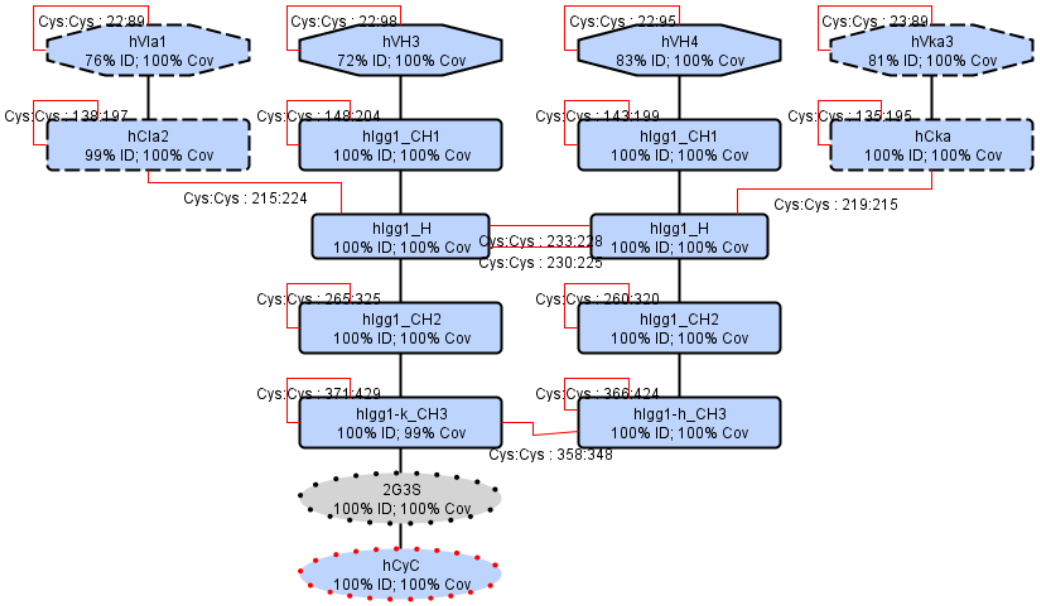
Antibody conjugates



Antibody conjugates



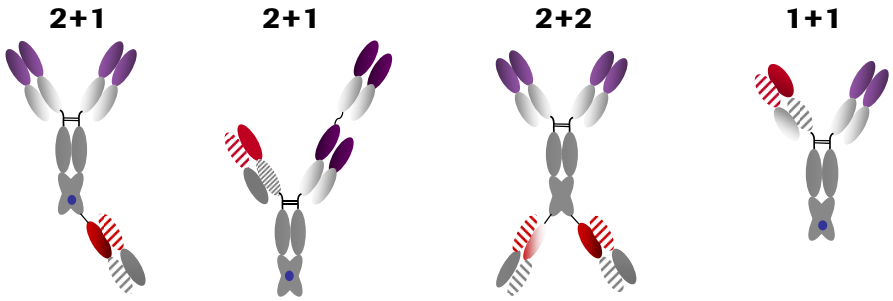
Antibody conjugates



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

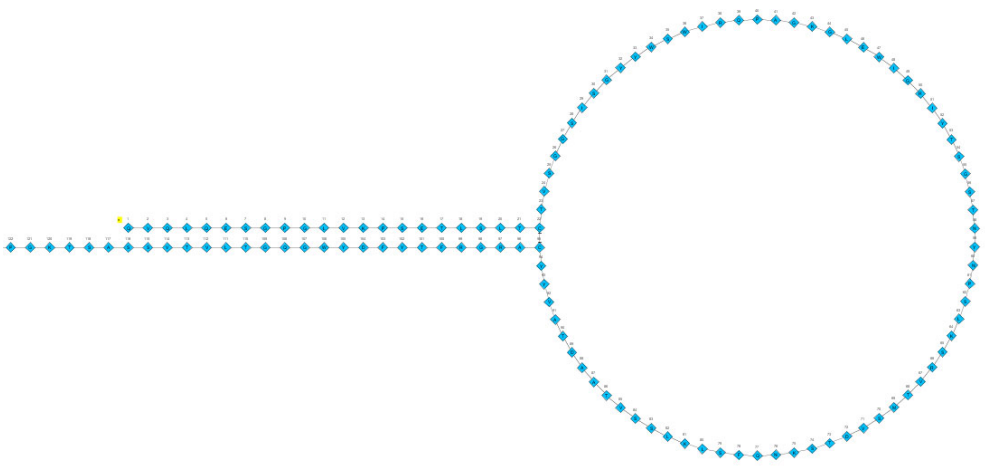
QVQLQESGPGLVKPSETLSLTCTVSGGSISGYYWSWIRQPAGKGLEWIGRIYTSGST
NYNPSLKSRVTMSVDTSKNQFSLKLSSVTAADTAVYYCARGRFTYFDYWGQGTLVTV
SSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPA
VLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPC
PAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHN
AKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQP
REPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDS
DGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK

and

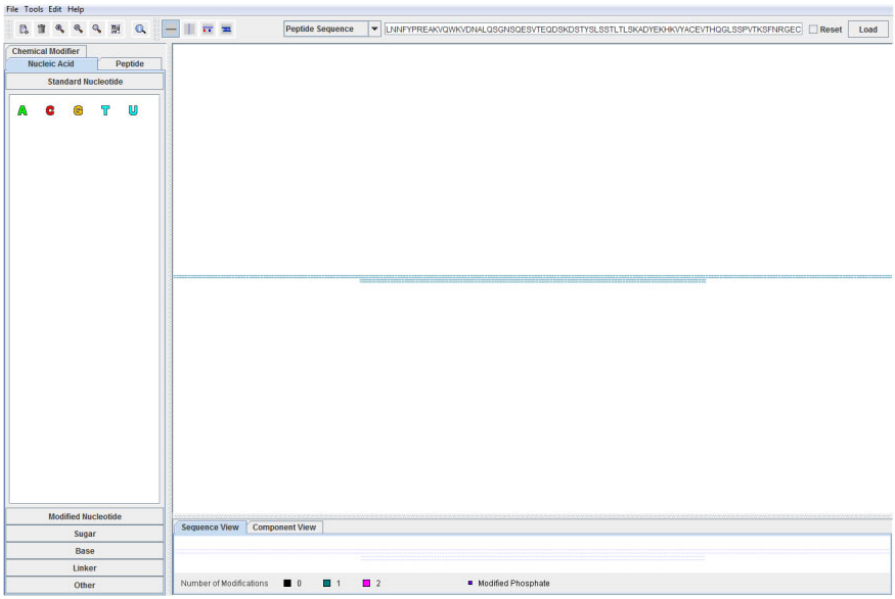
>LC light chain

EIVLTQSPATLSLSPGERATLSCRASQIVSSAYLAWYQQKPGQAPRLLMFGSSSRAT
GIPDRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSQGTFGPGTKVDIKRTVAAP
SVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKD
STYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC

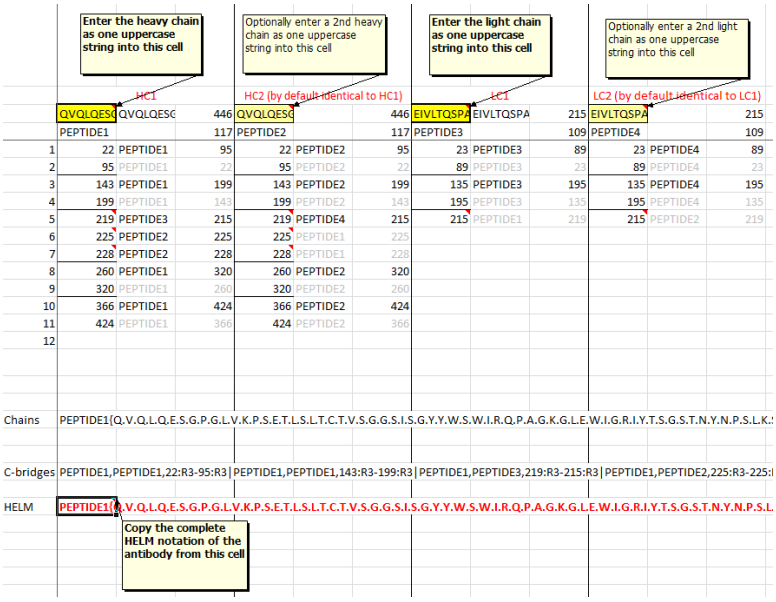
Load HC, draw 1. C-C bridge manually



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



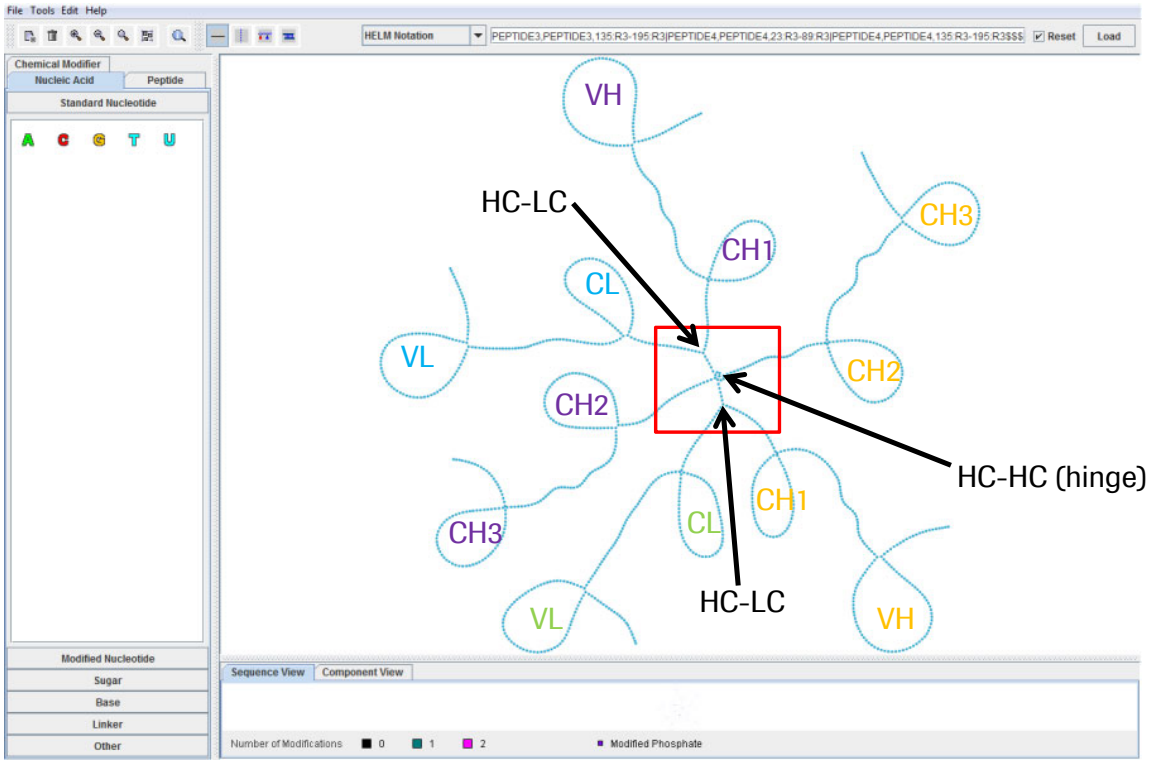
Antibody HELM converter – POC in Excel



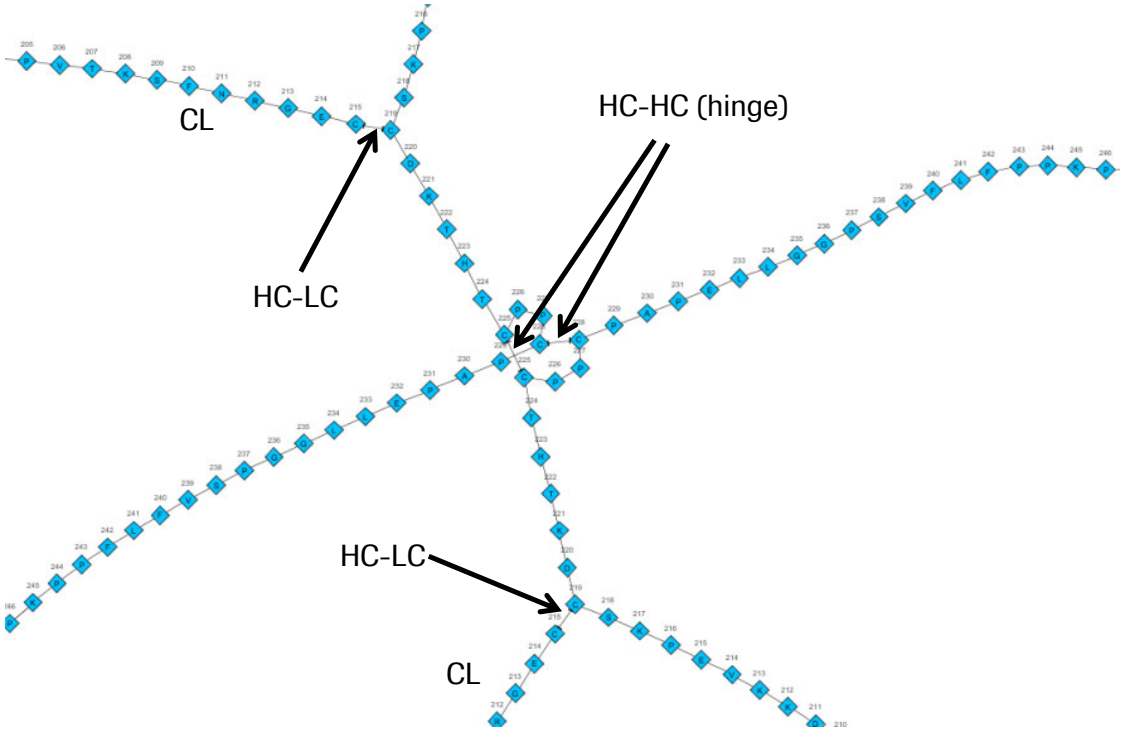
Input to HELM Editor

- PEPTIDE1{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.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.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.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.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.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.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.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.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.
  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.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.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.L.S.P.G.K}|PEPTIDE3{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.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.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}$PEPTIDE1,PEPTIDE1,22:R3-
  95:R3|PEPTIDE1,PEPTIDE1,143:R3-199: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,PEPTIDE2,22:R3-95: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$$$

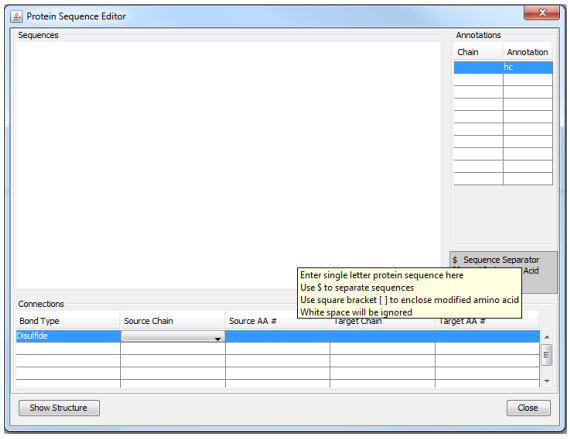
HELM Editor Output



Zoom reveals Cys-Cys bridges

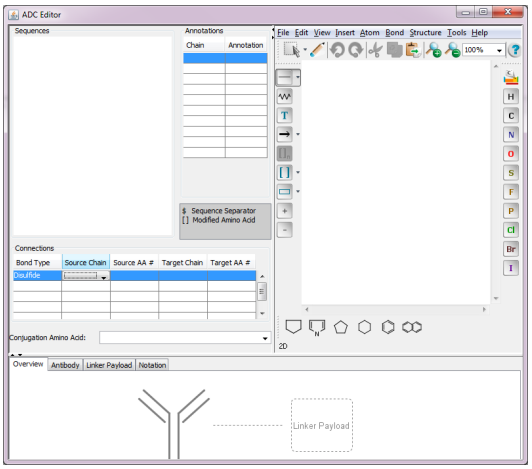


HELM Editor build-in alternatives



Protein Sequence Editor

ADC Editor



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.

Roche

Doing now what patients need next