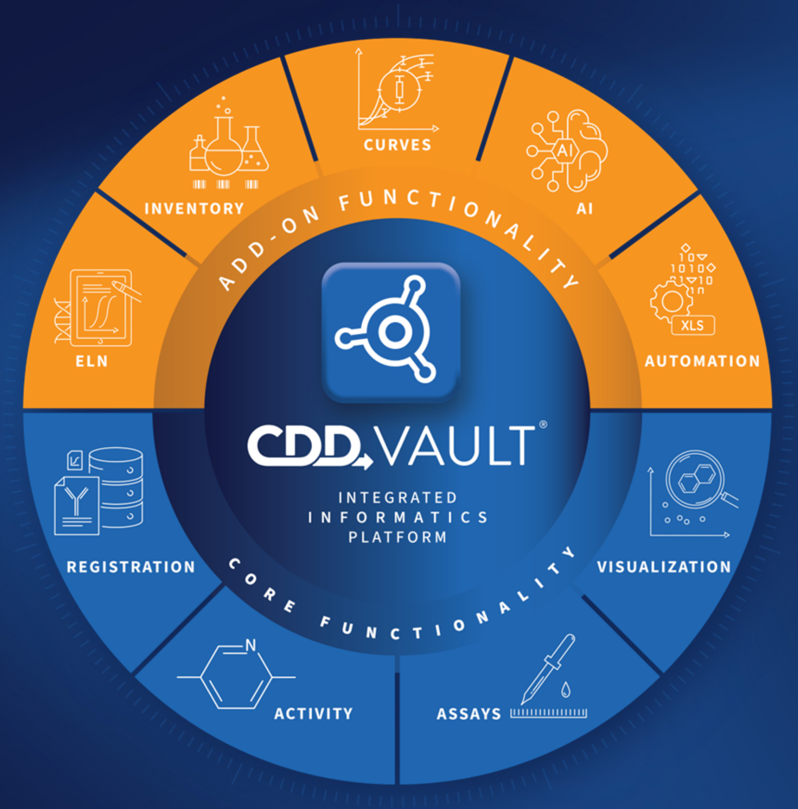
ACS Washington DC Autumn 2025

CDD VAULT®
Complexity Simplified



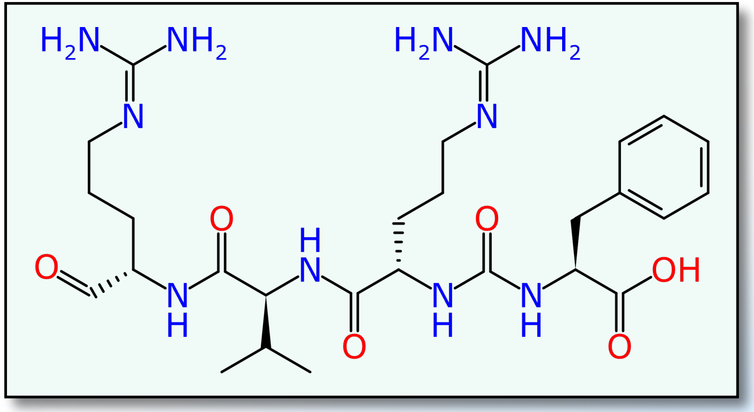
Conjugate biomolecule datastructures for chemistry and biochemistry discovery

Alex M. Clark

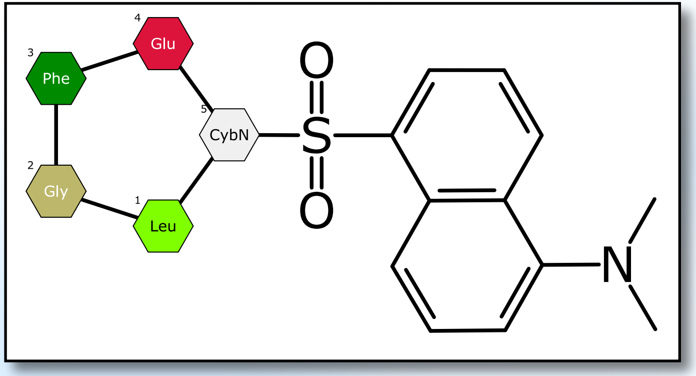
Macromolecules

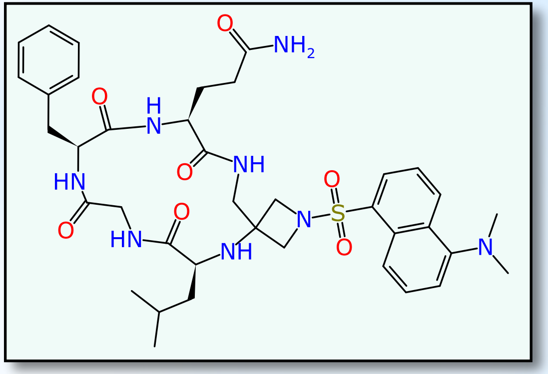
H	R	R	T	S
Q	R	C	E	E
E	R	Y	V	P
V	N	Y	C	C
Q	W	G	M	

- Bioinformatics vs. Cheminformatics
- Need both…



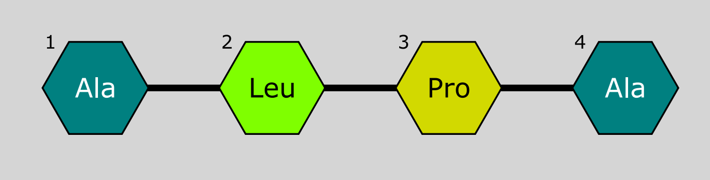
L,G,F,E,CybN-SO2NphNMe2,*

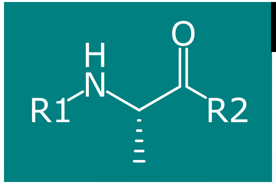




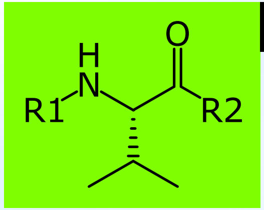
Datastructure

- V3000 molfile feature called **SCSR**
- **Layout** information as 2D sketch
- Monomers as **templates**: scalable
- Self contained, no external **dictionary**
- Full **atomic** structure is implied
- Contains the **sequence** data for bioinformatics

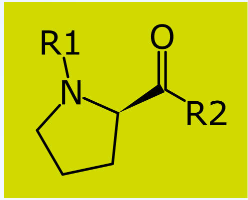




T1



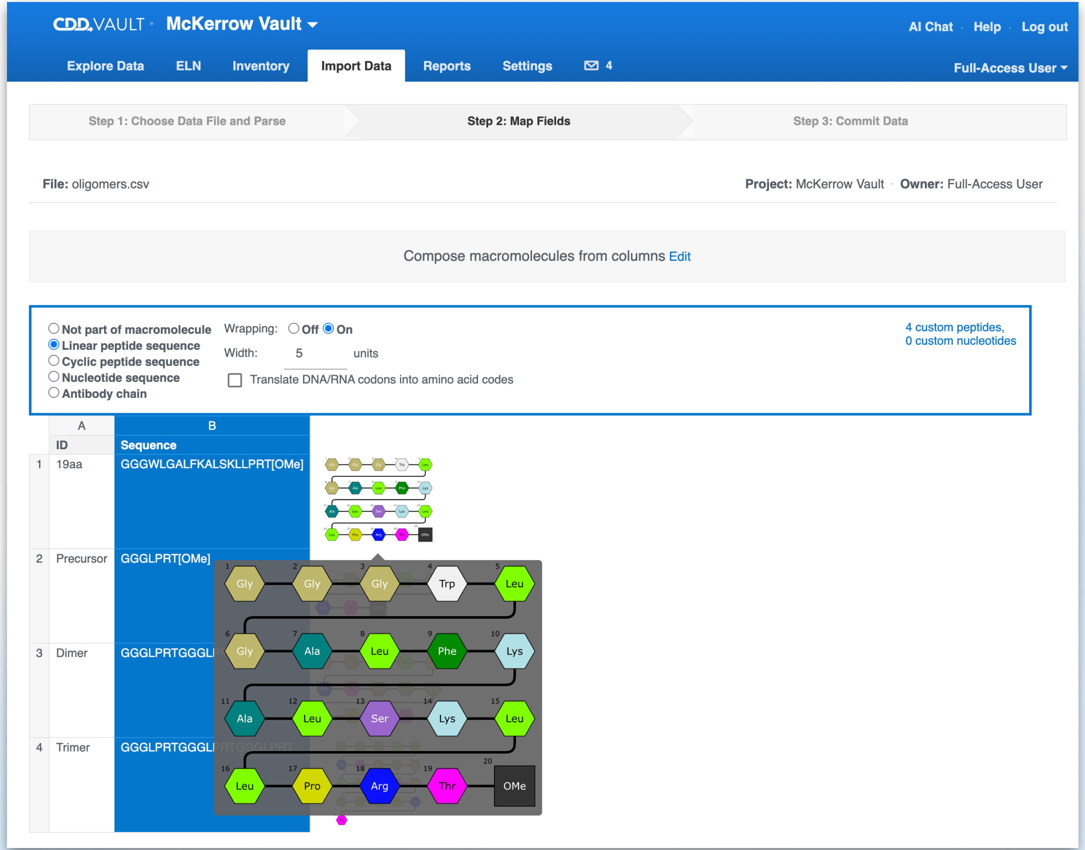
T2



T3

Content creation

- Draw anything with Ketcher
- Bulk import: markup sequences
- Custom monomers optional



File: oligomers.csv

Project: McKerrow Vault · **Owner:** Full-Access User

Compose macromolecules from columns Edit

○ Not part of macromolecule
● Linear peptide sequence
○ Cyclic peptide sequence
○ Nucleotide sequence
○ Antibody chain

Wrapping: ○ Off ● On

Width: 5 units

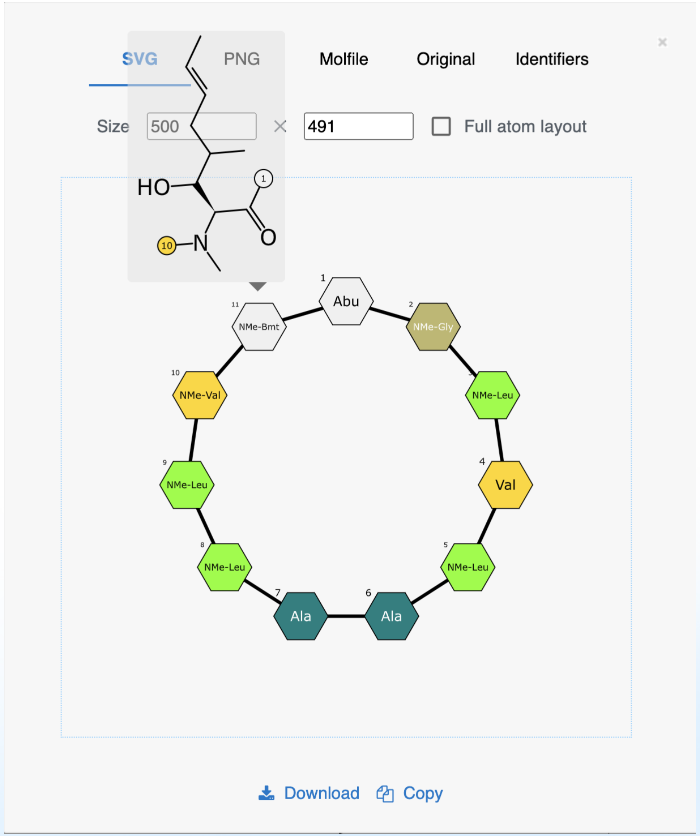
☐ Translate DNA/RNA codons into amino acid codes

4 custom peptides,
0 custom nucleotides

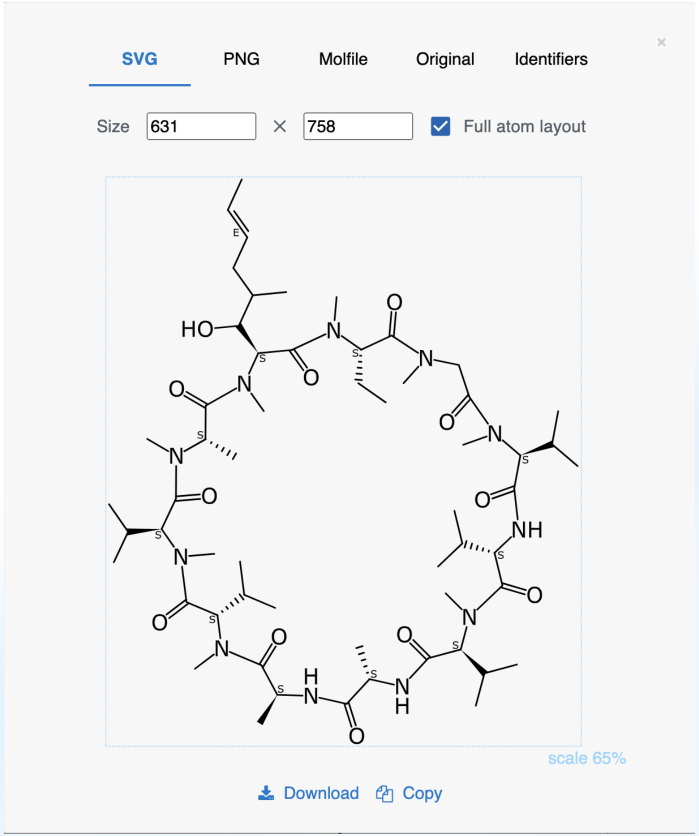
	A	B
	ID	Sequence
1	19aa	GGGWLGALFKALSKLLPRT[OMe]
2	Precursor	GGGLPRT[OMe]
3	Dimer	GGGLPRTGGGL
4	Trimer	GGGLPRTGGGLPRTGGGLPRT

Cyclic peptides

CDD VAULT®
Complexity Simplified



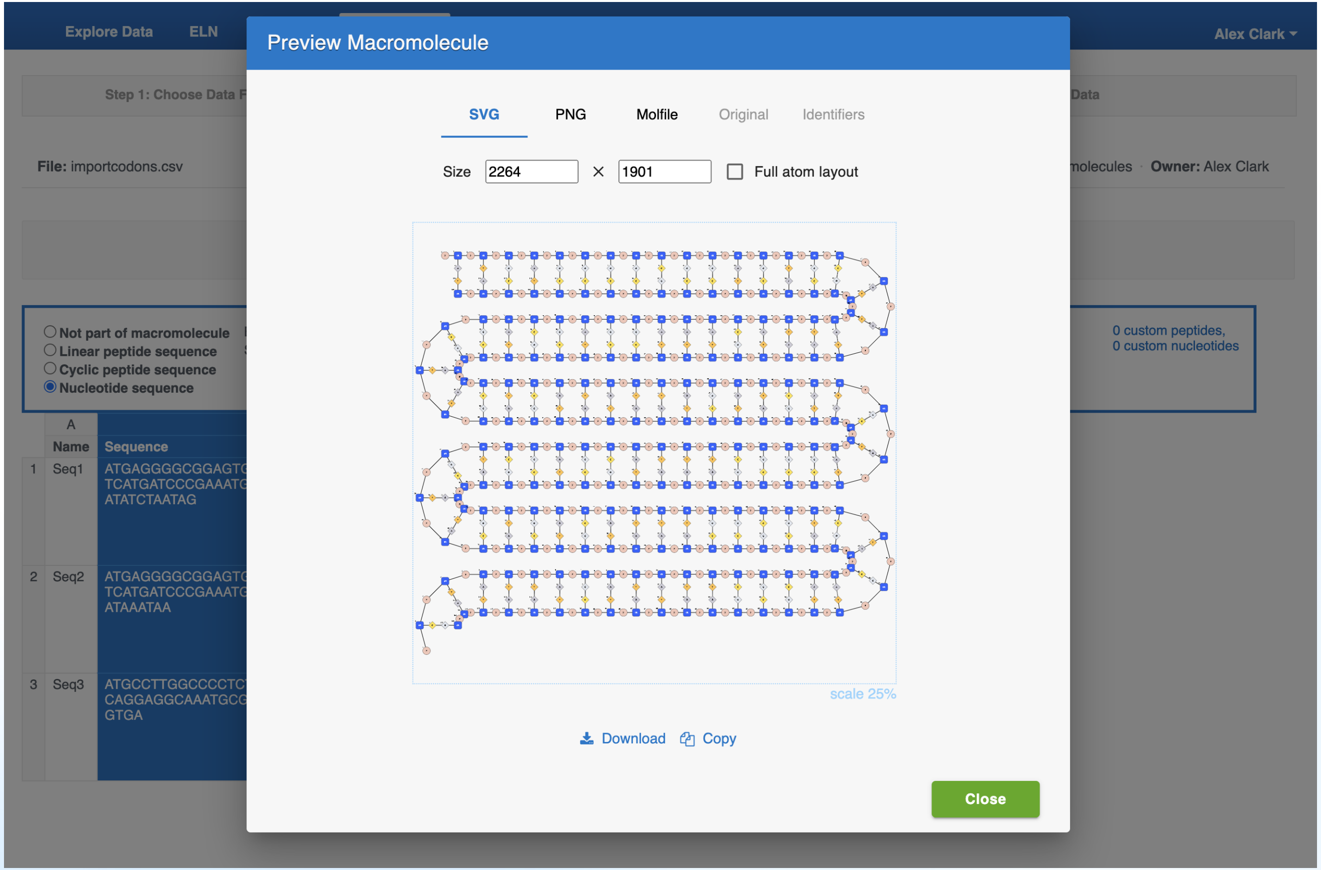
Size 500 × 491 ☐ Full atom layout



Size 631 × 758 ☑ Full atom layout

Nucleotides

CDD VAULT®
Complexity Simplified



File: importcodons.csv

molecules · Owner: Alex Clark

○ Not part of macromolecule
○ Linear peptide sequence
○ Cyclic peptide sequence
● Nucleotide sequence

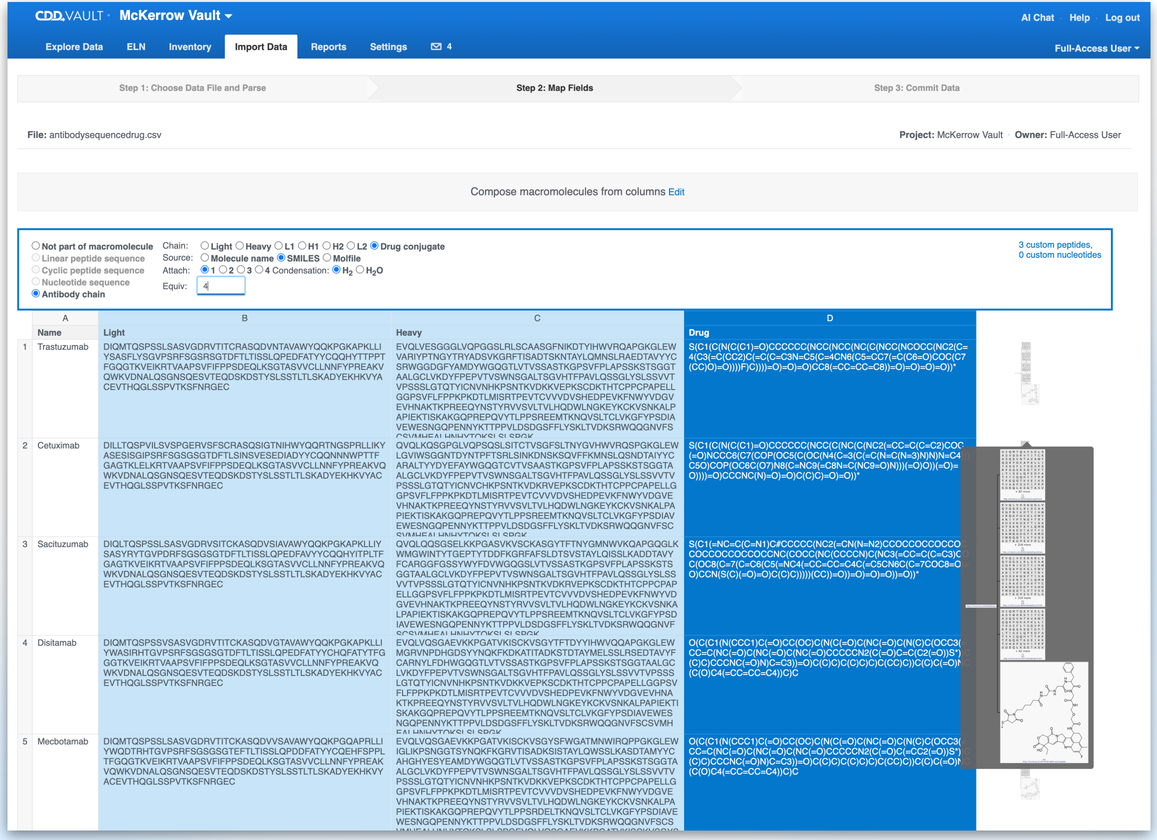
0 custom peptides,
0 custom nucleotides

	A	
	Name	Sequence
1	Seq1	ATGAGGGGCGGAGTG TCATGATCCCGAAATG ATATCTAATAG
2	Seq2	ATGAGGGGCGGAGTG TCATGATCCCGAAATG ATAAATAA
3	Seq3	ATGCCTTGGCCCCTC CAGGAGGCAAATGCG GTGA

Size 2264 × 1901 ☐ Full atom layout

Antibodies: Basic

- Just have sequences and drug conjugate structures?
- No problem: collection of objects



File: antibodysequencedrug.csv

Project: McKerrow Vault · **Owner:** Full-Access User

Compose macromolecules from columns Edit

○ Not part of macromolecule
○ Linear peptide sequence
○ Cyclic peptide sequence
○ Nucleotide sequence
● Antibody chain

Chain: ○ Light ○ Heavy ○ L1 ○ H1 ○ H2 ○ L2 ● Drug conjugate
Source: ○ Molecule name ● SMILES ○ Molfile
Attach: ● 1 ○ 2 ○ 3 ○ 4 Condensation: ● H_2 ○ H_2O
Equiv: 4

3 custom peptides,
0 custom nucleotides

	A	B	C	D
	Name	Light	Heavy	Drug
1	Trastuzumab	DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSRSGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC	EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFS	S(C1(C(N(C(C1)=O)CCCCCC(NCC(NCC(NC(C(NCC(NCOCC(NC2(C=4(C3(=C(CC2)C(=C(C=C3N=C5(C=4CN6(C5=CC7(=C(C6=O)COC(C7(CC)O)=O))))F)C))))=O)=O)=O)CC8(=CC=CC=C8))=O)=O)=O)=O))*
2	Cetuximab	DILLTQSPVILSVSPGERVSFSCRASQSIGTNIHWYQQRTNGSPRLLIKYASESISGIPSRFSGSGSGTDFTLSINSVESEDIADYYCQQNNNWPTTFGAGTKLELKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC	QVQLKQSGPGLVQPSQSLSITCTVSGFSLTNYGVHWVRQSPGKGLEWLGVIWSGGNTDYNTPFTSRLSINKDNSKSQVFFKMNSLQSNDTAIYYCARALTYYDYEFAYWGQGTCVTVSAASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVCHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSC	S(C1(C(N(C(C1)=O)CCCCCC(NCC(C(NC(C(NC2(=CC=C(C=C2)COC(=O)NCCC6(C7(COP(OC5(C(OC(N4(C=3(C(=C(N=C(N=3)N)N)N=C4)))C5O)COP(OC6C(O7)N8(C=NC9(=C8N=C(NC9=O)N)))(=O)O))(=O)=O))))=O)CCCNC(N)=O)=O)C(C)C)=O)=O))*
3	Sacituzumab	DIQLTQSPSSLSASVGDRVSITCKASQDVSIAVAWYQQKPGKAPKLLIYSASYRYTGVPDRFSGSGSGTDFTLTISSLQPEDFAVYYCQQHYITPLTFGAGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC	QVQLQQSGSELKKPGASVKVSCKASGYTFTNYGMNWVKQAPGQGLKWMGWINTYTGEPTYTDDFKGRFAFSLDTSVSTAYLQISSLKADDTAVYFCARGGFGSSYWYFDVWGQGSLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKRVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVF	S(C1(=NC=C(C=N1)C#CCCCC(NC2(=CN(N=N2)CCOCCOCCOCCOCCOCCOCCOCCOCCNC(COCC(NC(CCCCN)C(NC3(=CC=C(C=C3)COC(OC8(C=7(C=C6(C5(=NC4(=CC=CC=C4C(=C5CN6C(C=7COC8=O)=O)CCN(S(C)(=O)=O)C(C)C))))))(CC))=O))=O)=O)=O))=O))*
4	Disitamab	DIQMTQSPSSVSASVGDRVTITCKASQDVGTAVAWYQQKPGKAPKLLIYWASIRHTGVPSRFSGSGSGTDFTLTISSLQPEDFATYYCHQFATYTFGGGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC	EVQLVQSGAEVKKPGATVKISCKVSGYTFTDYYIHWVQQAPGKGLEWMGRVNPDHGDSYYNQKFKDKATITADKSTDTAYMELSSLRSEDTAVYFCARNYLFDHWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMH	O(C(C1(N(CCC1)C(=O)CC(OC)C(N(C(=O)C(NC(=O)C(N(C)C(OCC3(=CC=C(NC(=O)C(NC(=O)C(NC(=O)CCCCCN2(C(=O)C=C(C2(=O))S*))C(C)C)CCCNC(=O)N)C=C3))=O)C(C)C)C(C)C)C)C(CC)C))C(C)C(=O)NC(C(O)C4(=CC=CC=C4))C)C
5	Mecbotamab	DIQMTQSPSSLSASVGDRVTITCKASQDVVSAVAWYQQKPGQAPRLLIYWQDTRHTGVPSRFSGSGSGTEFTLTISSLQPDDFATYYCQEHFSPPLTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC	EVQLVQSGAEVKKPGATVKISCKVSGYSFWGATMNWIRQPPGKGLEWIGLIKPSNGGTSYNQKFKGRVTISADKSISTAYLQWSSLKASDTAMYYCAHGHYESYEAMDYWGQGTLVTVSSASTKGPSVFPLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCS	O(C(C1(N(CCC1)C(=O)CC(OC)C(N(C(=O)C(NC(=O)C(N(C)C(OCC3(=CC=C(NC(=O)C(NC(=O)C(NC(=O)CCCCCN2(C(=O)C(=CC2(=O))S*))C(C)C)CCCNC(=O)N)C=C3))=O)C(C)C)C(C)C)C)C(CC)C))C(C)C(=O)NC(C(O)C4(=CC=CC=C4))C)C

Antibodies: Metadata

Light

DIQMTQSPSSLSASVGDRVTITCRASQDVNTAVAWYQQKPGKAPKLLIYSASFLYSGVPSRFSGSR
SGTDFTLTISSLQPEDFATYYCQQHYTTPPTFGQGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVV
CLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLTLSKADYEKHKVYACEVTH
QGLSSPVTKSFNRGEC**{VL:1-107}{CL:108-214}{@214:1}**

Heavy

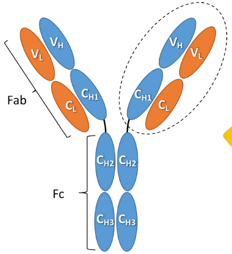
EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVK
GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLVTVSSASTKGPSVF
PLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSS
LGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPE
VTCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKC
KVSNKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQP
ENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFSCSVMHEALHNHYTQKSLSLSPGK
{VH:1-120}{CH1:121-218}{CH2:234-343}{CH3:344-450}{H:219-233}{@223:1}{@229:2}
{@232:3}{$264:1}{$324:1}{$370:1}{$428:1}

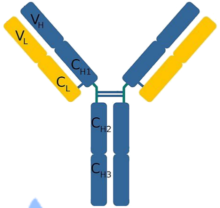
🟦	Domain Region	🟨	Disulfide Bridge
🟪	Hinge Region	🟩	Drug Conjugate Site

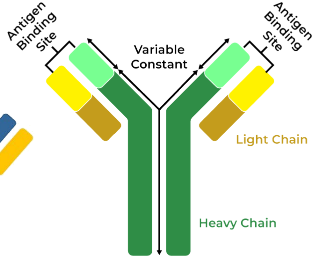
Scale

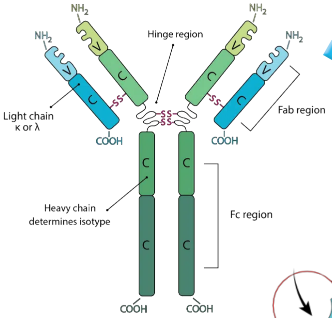
Display

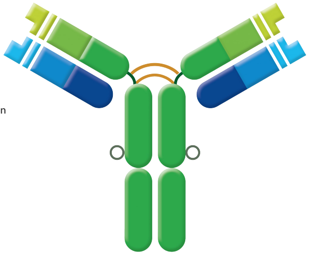
CDD VAULT®
Complexity Simplified

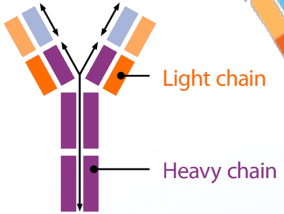


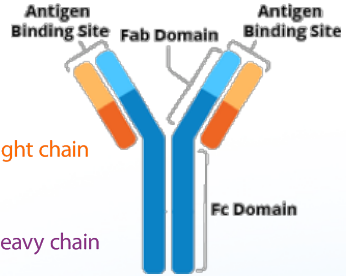


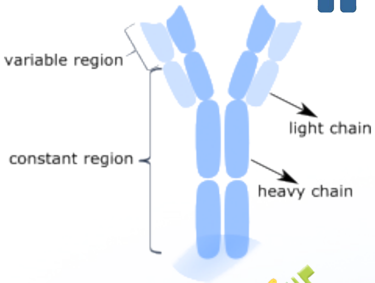


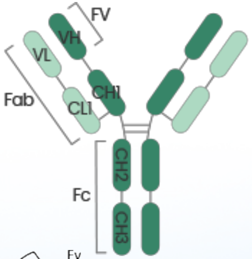


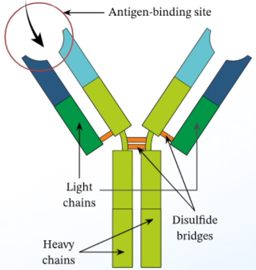


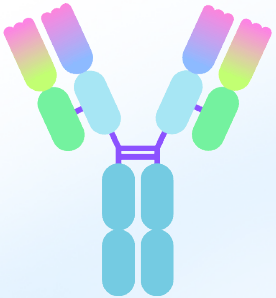


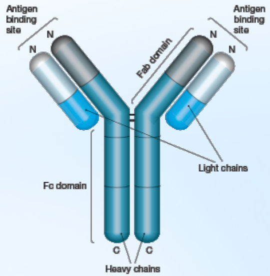


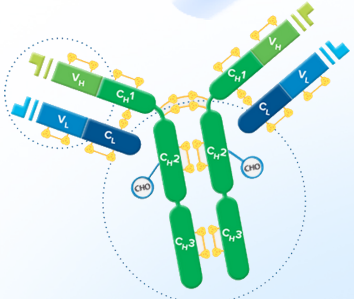


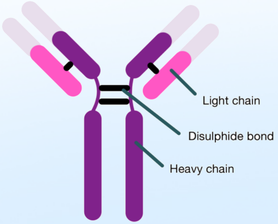


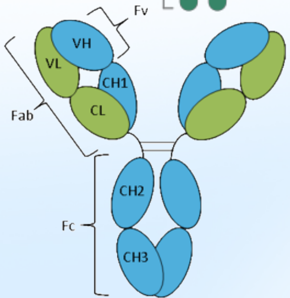


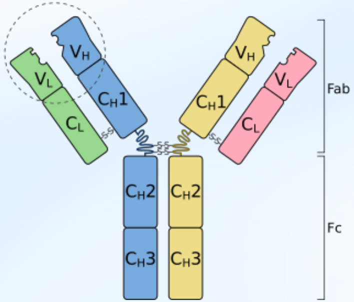


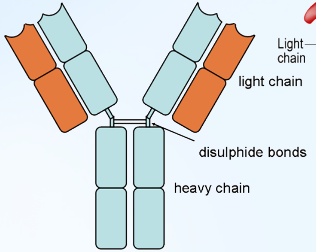


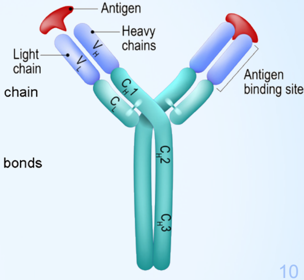






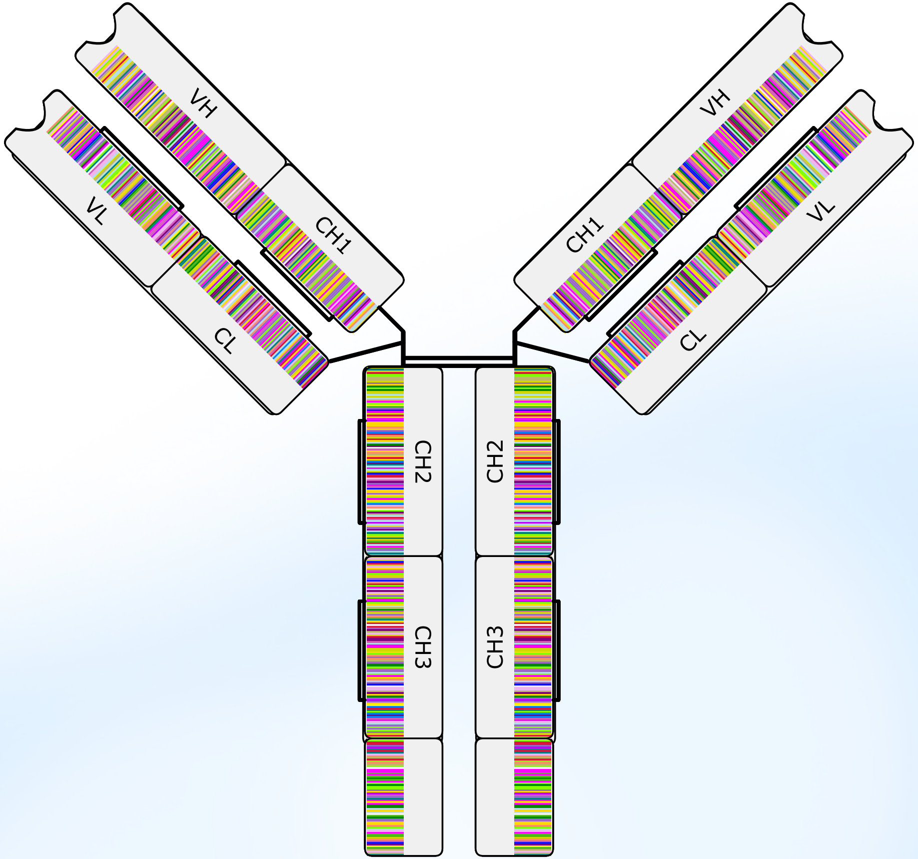






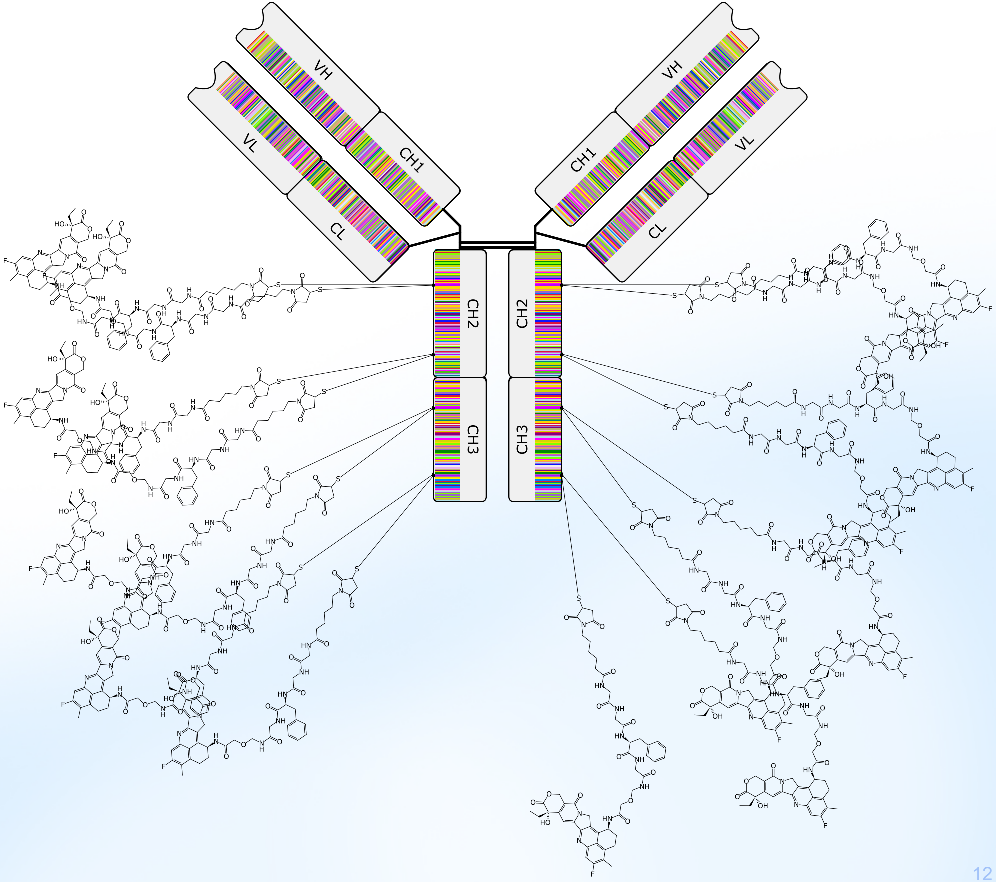
Cartoon and content

- Draw basic Y-shape outline
- Segment domain regions
- Inter-chain disulfide links
- Intra-chain disulfide links
- Sequence barcoding

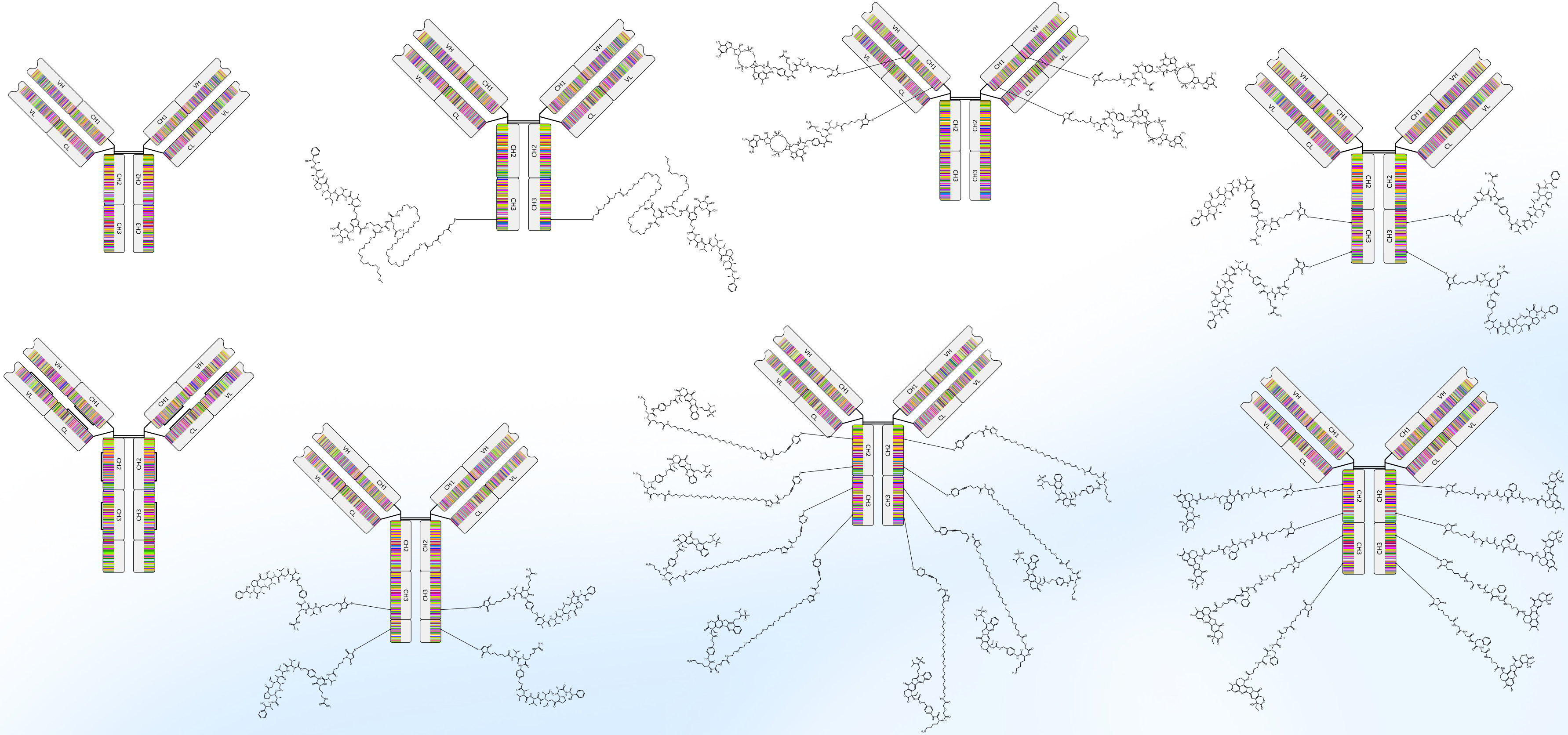


Drug conjugates

- Tethered initial placements
- Rigid body position optimization
- Anchor point is indicated



Gallery



Conclusion

- **Diagrams** for biologists
- **Sequences** for bioinformaticians
- **Atoms** & **bonds** for chemists
- Standard formats and open source tools where possible
- Provide all information: get all the functionality
- Complete informatics is fundamental to everything