

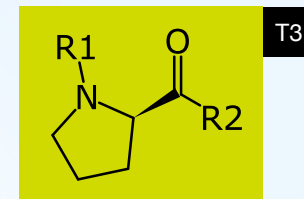
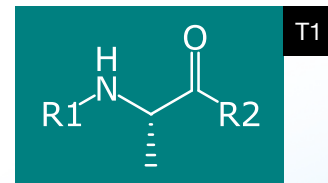
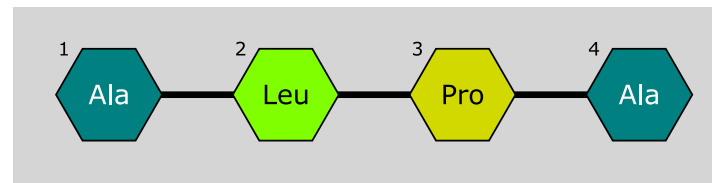
Small molecules,
bigger molecules,
and biologics



Alex M. Clark

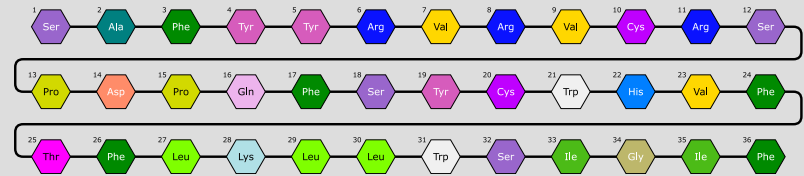
Hiding in the Molfile this whole time

- 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



- Sometimes just sequence is enough

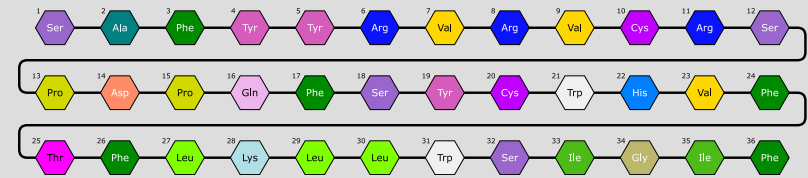
SAFYRVRVCRS
PDPQFSYCVHVF
TFLKLLWSIGIF



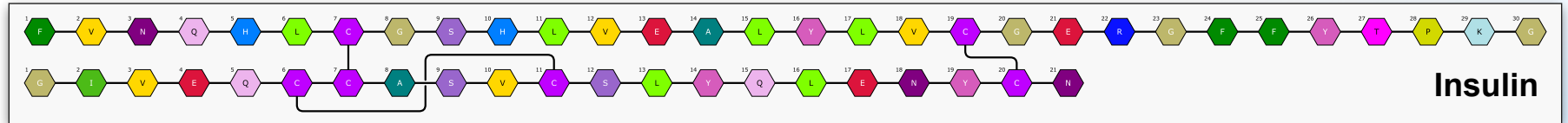
- Exotic oligomers: the sweet spot for **macromolecule** feature

- Sometimes just sequence is enough

SAFYRVRVCRS
PDPQFSYCWVHF
TFLKLLWSIGIF

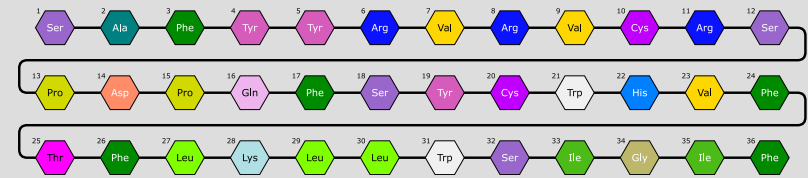


- Exotic oligomers: the sweet spot for **macromolecule** feature

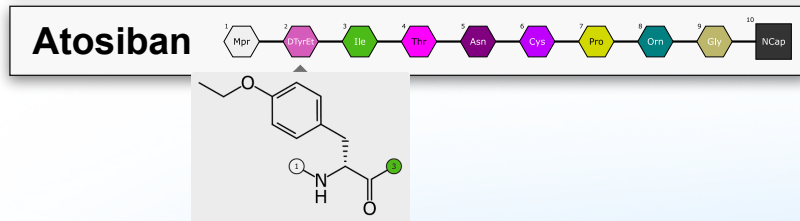
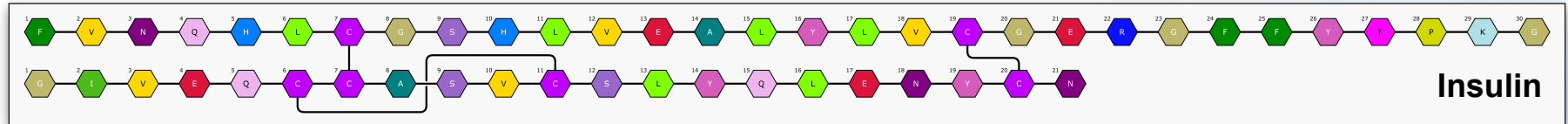


- Sometimes just sequence is enough

SAFYRVRVCRS
PDPQFSYCWVVF
TFLKLLWSIGIF

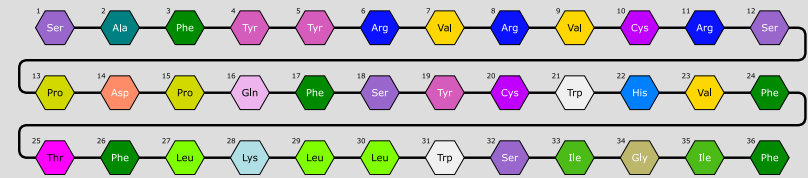


- Exotic oligomers: the sweet spot for **macromolecule** feature

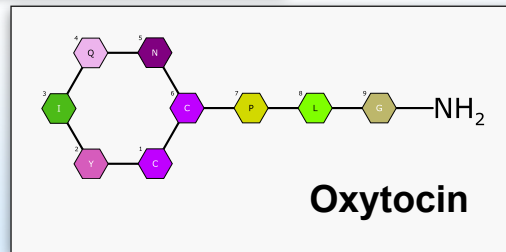
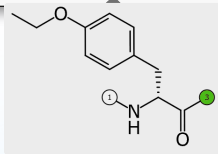
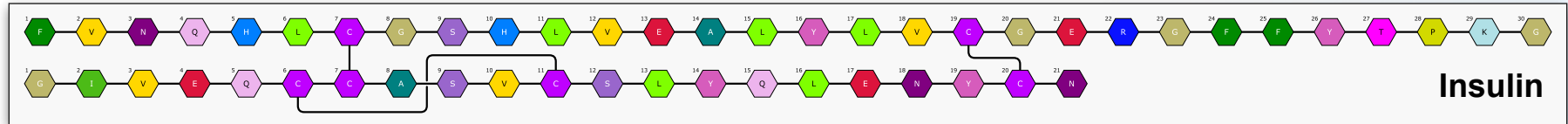


- Sometimes just sequence is enough

SAFYRVRVCRS
PDPQFSYCWVVF
TFLKLLWSIGIF

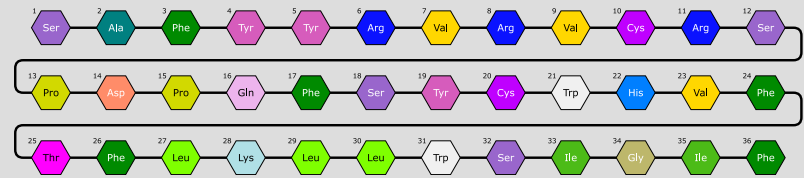


- Exotic oligomers: the sweet spot for **macromolecule** feature

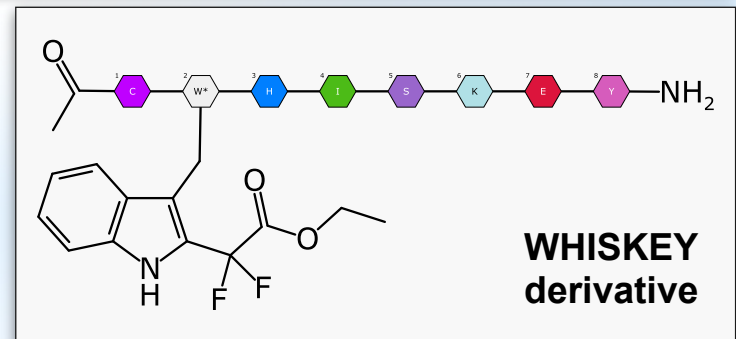
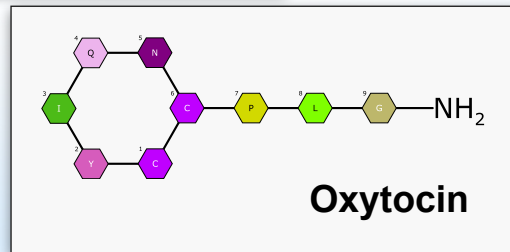
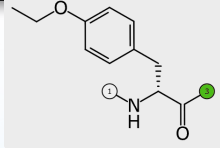
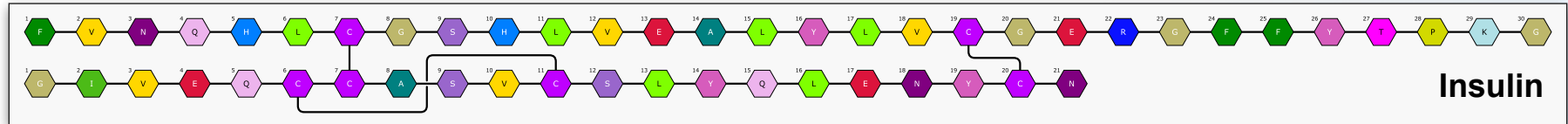


- Sometimes just sequence is enough

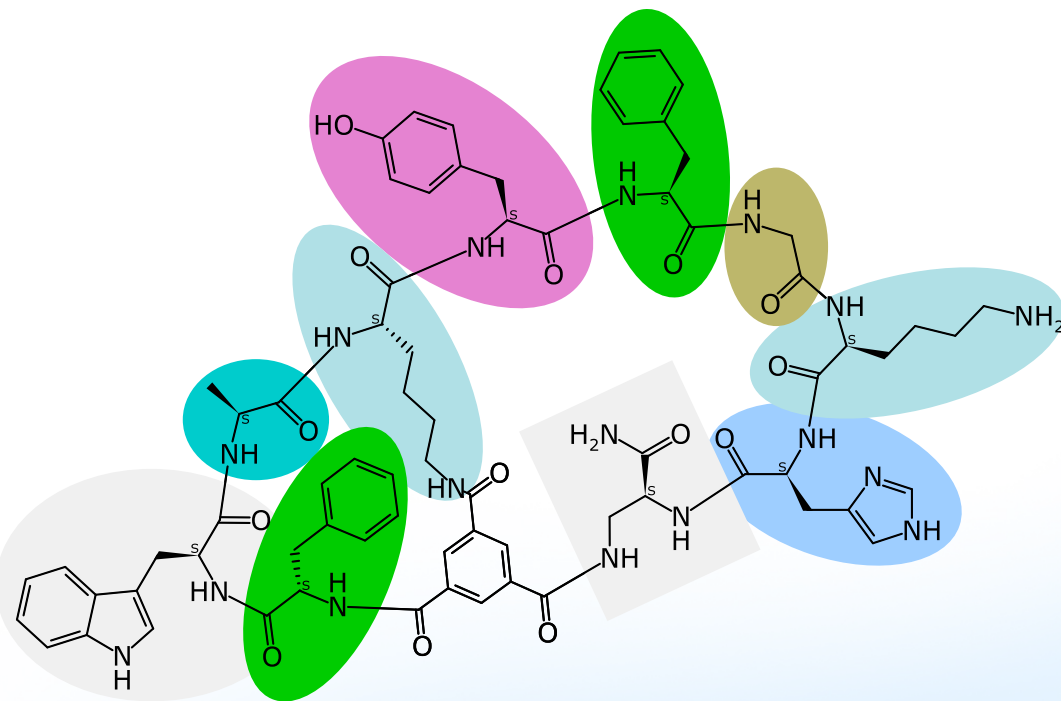
SAFYRVRVCRS
PDPQFSYCWVVF
TFLKLLWSIGIF



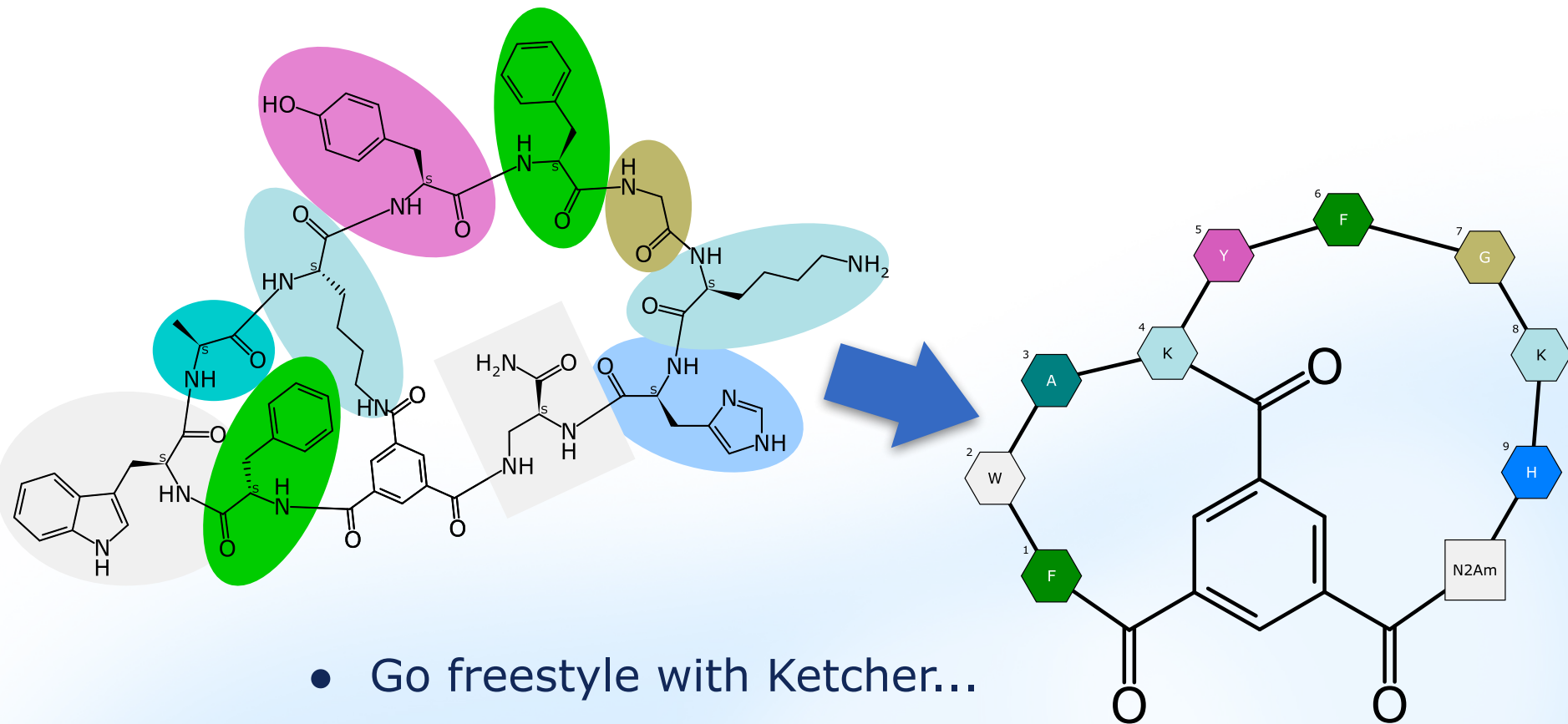
- Exotic oligomers: the sweet spot for **macromolecule** feature



Perception is everything



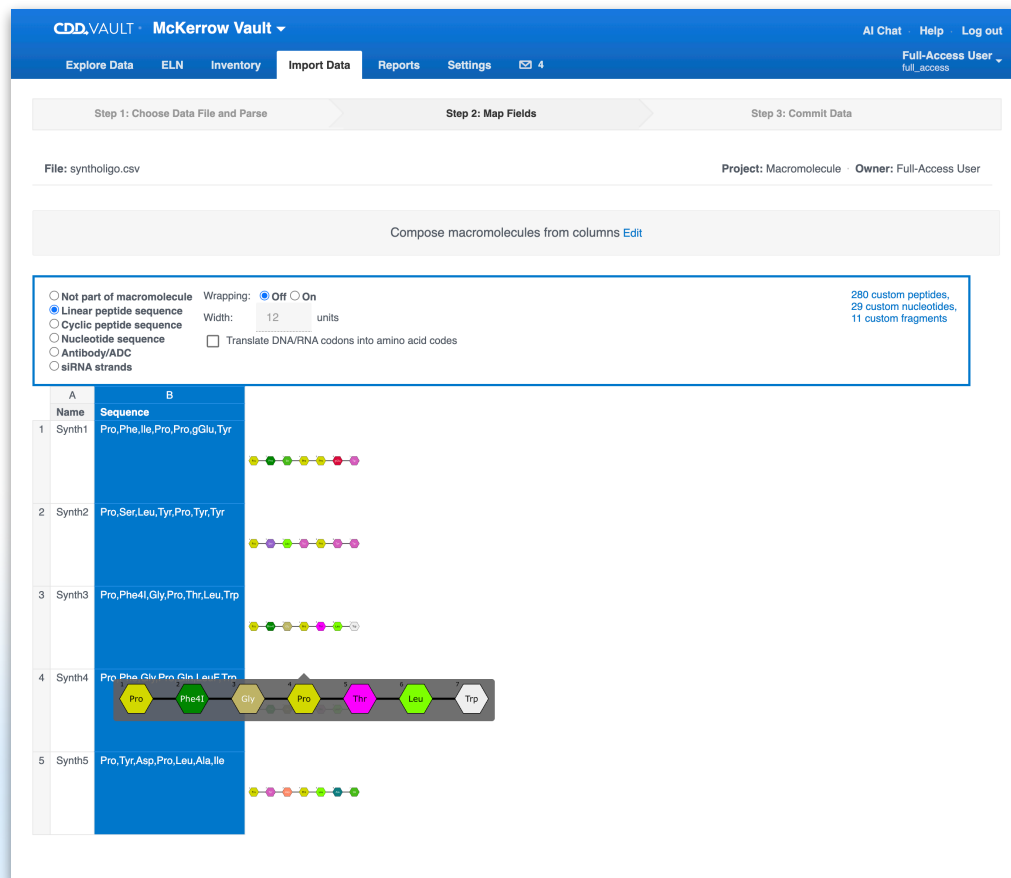
Perception is everything



- Incoming data:
 - sequences
 - metadata
- Markup to macromolecules
- Custom types
- Custom monomers

Bulk import with Vault

- Incoming data:
 - sequences
 - metadata
- Markup to macromolecules
- Custom types
- Custom monomers



CDD.VAULT - McKerrow Vault

AI Chat Help Log out

Full-Access User
full_access

Explore Data ELN Inventory **Import Data** Reports Settings 4

Step 1: Choose Data File and Parse Step 2: **Map Fields** Step 3: Commit Data

File: syntholigo.csv Project: Macromolecule Owner: Full-Access User

Compose macromolecules from columns Edit

Not part of macromolecule Wrapping: Off On
 Linear peptide sequence Width: 12 units
 Cyclic peptide sequence
 Nucleotide sequence Translate DNA/RNA codons into amino acid codes
 Antibody/ADC
 siRNA strands

280 custom peptides,
29 custom nucleotides,
11 custom fragments

A	B
Name	Sequence
1 Synth1	Pro,Phe,Ile,Pro,Pro,Glu,Tyr
2 Synth2	Pro,Ser,Leu,Tyr,Pro,Tyr,Tyr
3 Synth3	Pro,Phe4I,Gly,Pro,Thr,Leu,Trp
4 Synth4	Pro,Phe,Glu,Pro,Glu,Ile,Tyr
5 Synth5	Pro,Tyr,Asp,Pro,Leu,Ala,Ile

Bulk import with Vault

- Incoming data:
 - sequences
 - metadata
- Markup to macromolecules
- Custom types
- Custom monomers

CDD.VAULT - McKerrow Vault

AI Chat Help Log out

Full-Access User
full_access

Explore Data ELN Inventory **Import Data** Reports Settings 4





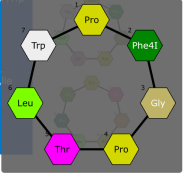
Step 1: Choose Data File and Parse Step 2: **Map Fields** Step 3: Commit Data

File: syntholigo.csv Project: Macromolecule Owner: Full-Access User

Compose macromolecules from columns [Edit](#)

Not part of macromolecule Orientation: North East South West
 Linear peptide sequence Direction: Clockwise Anti-clockwise
 Cyclic peptide sequence
 Nucleotide sequence
 Antibody/ADC
 siRNA strands

280 custom peptides,
29 custom nucleotides,
11 custom fragments

	A	B	
	Name	Sequence	
1	Synth1	Pro,Phe,Ile,Pro,Pro,Glu,Tyr	
2	Synth2	Pro,Ser,Leu,Tyr,Pro,Tyr,Tyr	
3	Synth3	Pro,Phe4I,Gly,Pro,Thr,Leu,Trp	
4	Synth4	Pro,Phe,Gly,Pro,Glu,Leu,F,Trp	
5	Synth5	Pro,Tyr,Asp,Pro,Leu,Ala,	

Bulk import with Vault

- Incoming data:
 - sequences
 - metadata
- Markup to macromolecules
- Custom types
- Custom monomers

CDD.VAULT - McKerrow Vault

AI Chat Help Log out

Full-Access User
full_access

Explore Data ELN Inventory **Import Data** Reports Settings 4

Step 1: Choose Data File and Parse Step 2: **Map Fields** Step 3: Commit Data

File: syntholigo.csv Project: Macromolecule Owner: Full-Access User

Preview Macromolecule

SVG PNG Molfile Original Identifiers

Size 172 × 166 Full atom layout

Not part of macromolecule
 Linear peptide sequence
 Cyclic peptide sequence
 Nucleotide sequence
 Antibody/ADC
 siRNA strands

A	B
Name	Sequence
1 Synth1	Pro,Phe,Ile,Pro,Pro,...
2 Synth2	Pro,Ser,Leu,Tyr,Pro,T...
3 Synth3	Pro,Phe4,Gly,Pro,Thr...
4 Synth4	Pro,Phe,Gly,Pro,Gln,S...
5 Synth5	Pro,Tyr,Asp,Pro,Leu,...

280 custom peptides,
29 custom nucleotides,
11 custom fragments

Download Copy

Close

Bulk import with Vault

- Incoming data:
 - sequences
 - metadata
- Markup to macromolecules
- Custom types
- Custom monomers

CDD.VAULT - McKerrow Vault

AI Chat Help Log out

Full-Access User
full_access

Explore Data ELN Inventory Import Data Reports Settings 4

Step 1: Choose Data File and Parse Step 2: Map Fields Step 3: Commit Data

File: syntholigo.csv Project: Macromolecule Owner: Full-Access User

Preview Macromolecule

SVG PNG Molfile Original Identifiers

Size 615 x 518 Full atom layout

280 custom peptides,
29 custom nucleotides,
11 custom fragments

Not part of macromolecule
 Linear peptide sequence
 Cyclic peptide sequence
 Nucleotide sequence
 Antibody/ADC
 siRNA strands

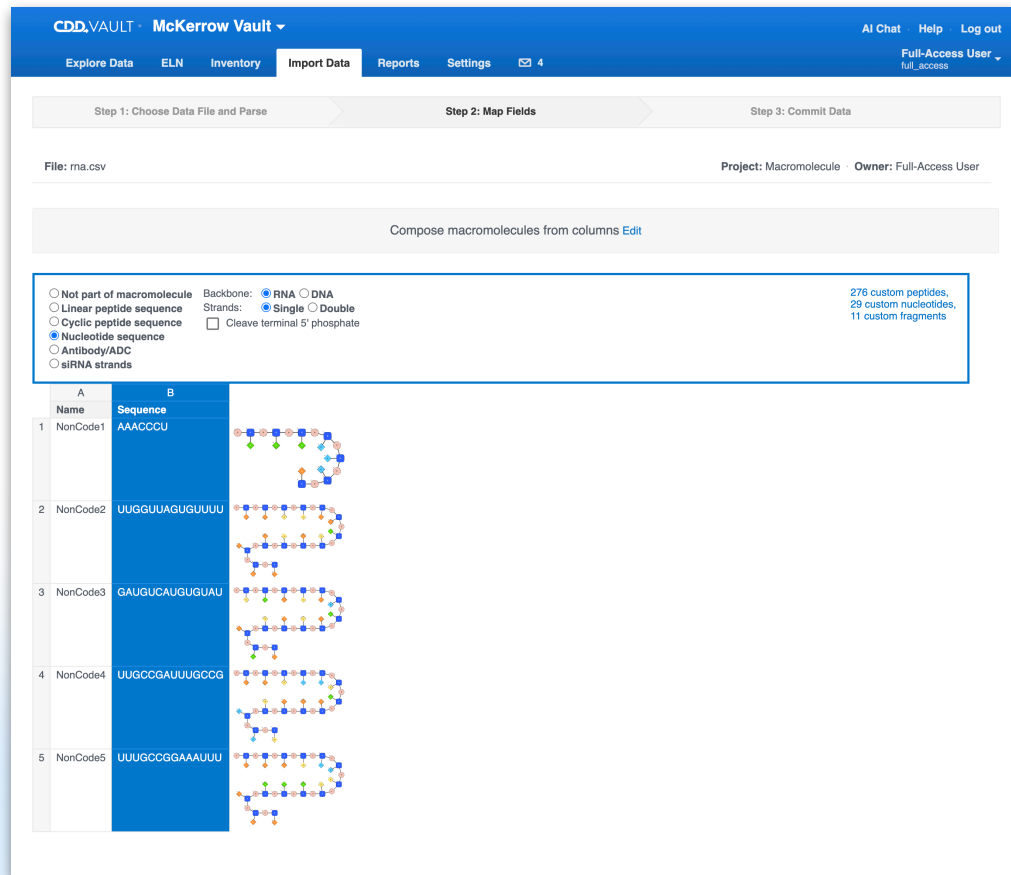
A	B
Name	Sequence
1 Synth1	Pro,Phe,Ile,Pro,Pro,g
2 Synth2	Pro,Ser,Leu,Tyr,Pro,1
3 Synth3	Pro,Phe4I,Gly,Pro,Th
4 Synth4	Pro,Phe,Gly,Pro,Gln,L
5 Synth5	Pro,Tyr,Asp,Pro,Leu,r

Download Copy

Close

Bulk import with Vault

- Incoming data:
 - sequences
 - metadata
- Markup to macromolecules
- Custom types
- Custom monomers



CDD.VAULT · McKerrow Vault

AI Chat Help Log out

Full-Access User
full_access

Explore Data ELN Inventory **Import Data** Reports Settings 4

Step 1: Choose Data File and Parse Step 2: Map Fields Step 3: Commit Data

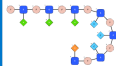
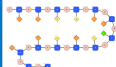
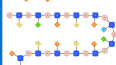
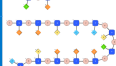
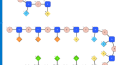
File: ma.csv Project: Macromolecule Owner: Full-Access User

Compose macromolecules from columns Edit

Not part of macromolecule
 Linear peptide sequence
 Cyclic peptide sequence
 Nucleotide sequence
 Antibody/ADC
 siRNA strands

Backbone: RNA DNA
Strands: Single Double
 Cleave terminal 5' phosphate

276 custom peptides,
29 custom nucleotides,
11 custom fragments

	A	B	
	Name	Sequence	
1	NonCode1	AAACCGU	
2	NonCode2	UUGGUUAGUUUUU	
3	NonCode3	GAUGUCAUGUGUAU	
4	NonCode4	UUGCCGAUUUGCCG	
5	NonCode5	UUUGCCGAAUUUU	

Bulk import with Vault

- Incoming data:
 - sequences
 - metadata
- Markup to macromolecules
- Custom types
- Custom monomers

CDD.VAULT · McKerrow Vault

AI Chat Help Log out

Full-Access User
full_access

Explore Data ELN Inventory **Import Data** Reports Settings 4

Step 1: Choose Data File and Parse Step 2: **Map Fields** Step 3: Commit Data

File: dna.csv Project: Macromolecule Owner: Full-Access User

Compose macromolecules from columns Edit

Not part of macromolecule Backbone: RNA DNA
 Linear peptide sequence Strands: Single Double
 Cyclic peptide sequence Cleave terminal 5' phosphate
 Nucleotide sequence
 Antibody/ADC
 siRNA strands

276 custom peptides,
29 custom nucleotides,
11 custom fragments

	A	B	
	Name	Sequence	
1	NonCode1	AAACCGT	
2	NonCode2	TTGGTTAGTGTTTT	
3	NonCode3	GATGTCATGTGAT	
4	NonCode4	TTGCCGATTGCCG	
5	NonCode5	TTGCCCGAAATT	

Bulk import with Vault

- Incoming data:
 - sequences
 - metadata
- Markup to macromolecules
- Custom types
- Custom monomers

CDD.VAULT · McKerrow Vault

Explore Data ELN Inventory **Import Data** Reports Settings 4

Step 1: Choose Data File and Parse Step 2: Map Fields Step 3: Commit Data

File: dna.csv

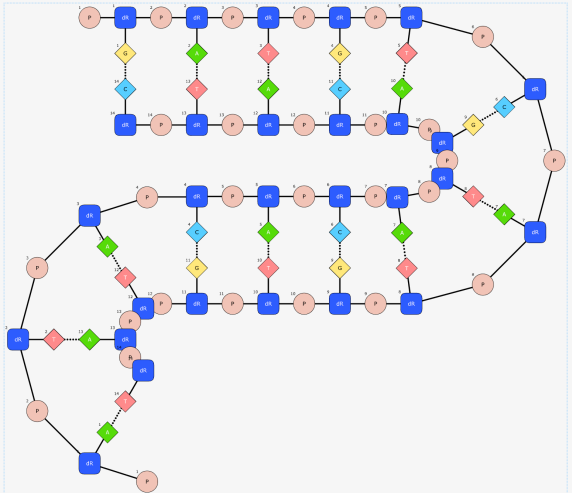
Owner: Full-Access User

276 custom peptides, 29 custom nucleotides, 11 custom fragments

Preview Macromolecule

SVG PNG Molfile Original Identifiers

Size 940 x 820 Full atom layout



○ Not part of macromolecule
○ Linear peptide sequence
○ Cyclic peptide sequence
● Nucleotide sequence
○ Antibody/ADC
○ siRNA strands

A	B
Name	Sequence
1 NonCode1	AAACCGT
2 NonCode2	TTGGTTAGTGT
3 NonCode3	GATGTCATGT
4 NonCode4	TTGCCGATTG
5 NonCode5	TTGCCGAAAT

Download Copy

scale 81%

Close

Custom monomers

Insert Structure: Molecule name, common name, or identifier

Molecules 100%

Create Monomer

Type: Nucleotide (preset)

Attributes

Select all atoms that form the sugar. Mark as sugar

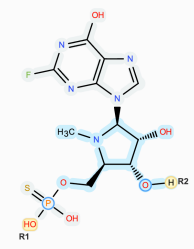
Code: H4S

Name: e.g. Diethylene Glycol

Attachment points

Aliases

Highlight



Discard Submit

Predictions

H	MW	385.309
	LogD	-3.6
C	LogP	-1.9
N	LogS	-1.1
O	HBD	5
	HBA	11
S	RB	4
	TPSA	157.2
P	BBB	1.7
F	CNS-MPD	3.7
	pKa	5
Cl	pKaAcidic	5
Br	pKaBasic	7.9
I	Exact MW	385.046469
	Fsp3	0.55
PT	HA	25
	Ru5	1

Launch CDD AI Tools

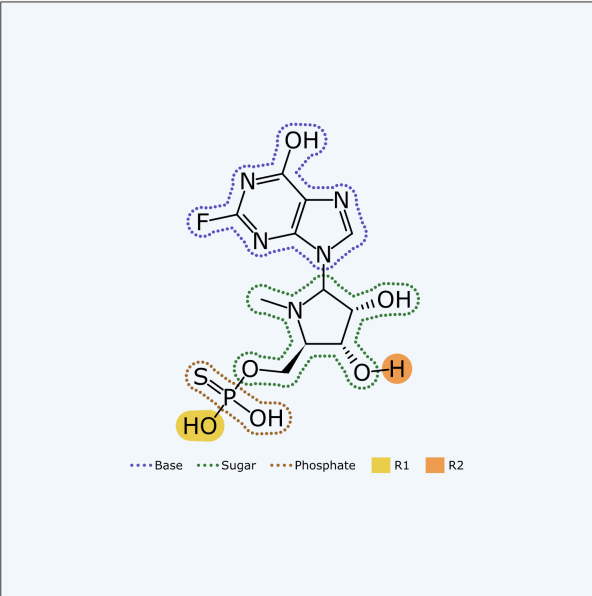
Use this structure or cancel

Custom monomers

Edit Monomer

Type **Peptide** **Nucleotide** Fragment Natural Analog

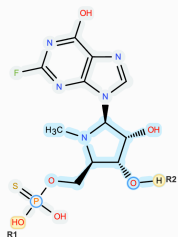
Synonyms Other codes



•••Base •••Sugar •••Phosphate ■ R1 ■ R2

[Edit Orient](#) [Deprotect](#)

[Remove monomer information](#) [Cancel](#) [Apply](#)



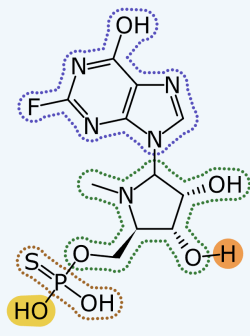
Custom monomers

Edit Monomer

Type: Peptide **Nucleotide** Fragment

Code: Natural Analog:

Synonyms: Other codes:

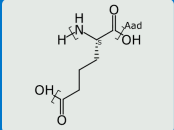
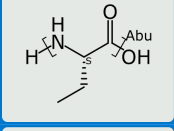
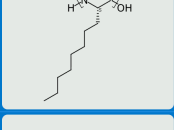
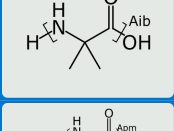
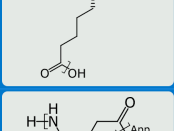
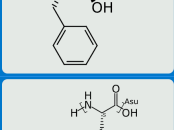
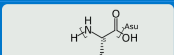


..... Base
 Sugar
 Phosphate
 R1
 R2

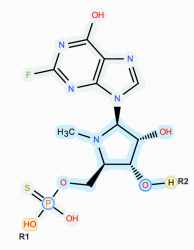
[Edit Orient](#) [Deprotect](#)

[Remove monomer information](#)
 [Cancel](#)
 [Apply](#)

Not part of monomer
 Default Type: Any Peptide Nucleotide Fragment
 Structure
 Deprotection: Amines Acids
 Monomer Type
 Orient monomer coordinates
 Natural Analog

	A ⇒ Structure	B	C ⇒ Natural
	SDF Molfile	Name	Natural
1		Aad	K
2		Abu	A
3		Aca	X
4		Aib	X
5		Apr	X
6		App	X
7		Asu	X

← Create Monomer

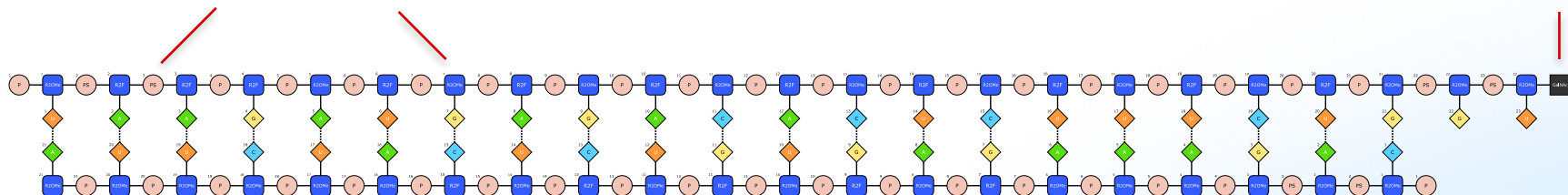


[Launch CDD AI Tools](#)

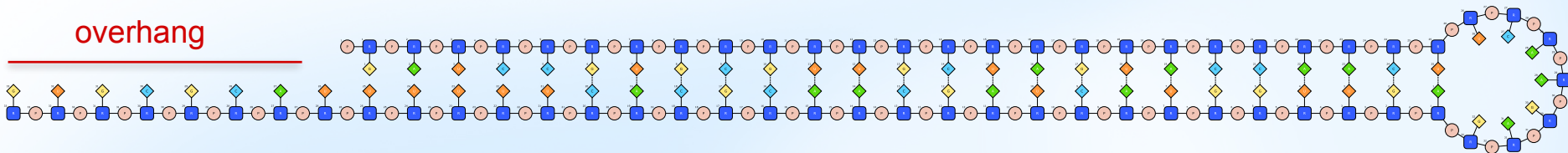
- Two-stranded RNA/DNA: more control

custom phosphates & sugars

end-cap



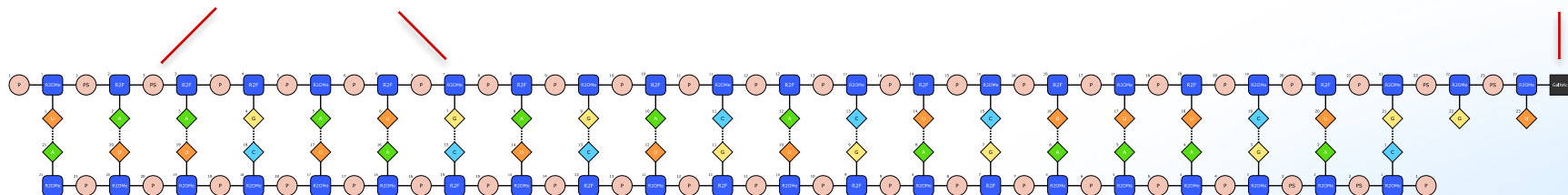
overhang



- Two-stranded RNA/DNA: more control

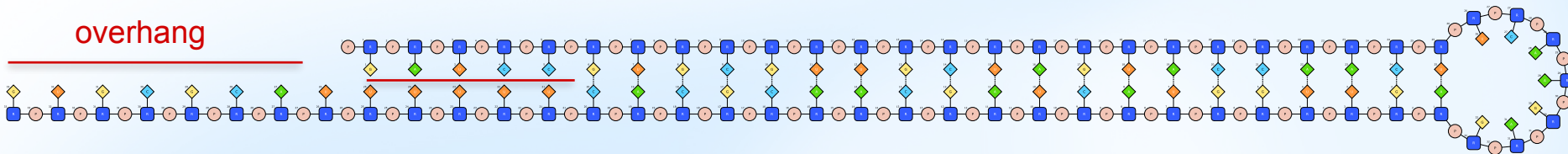
custom phosphates & sugars

end-cap



mismatch

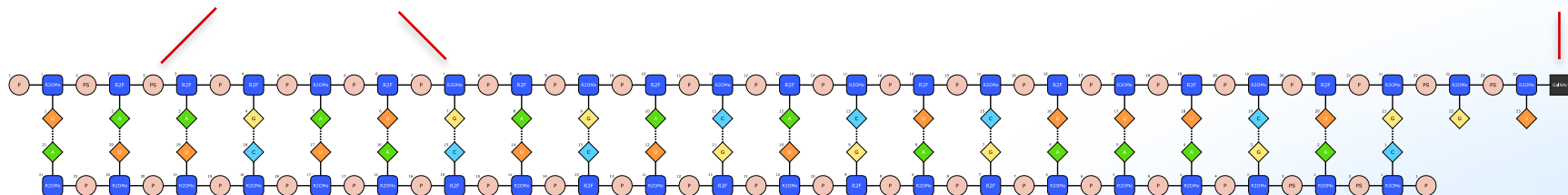
overhang



- Two-stranded RNA/DNA: more control

custom phosphates & sugars

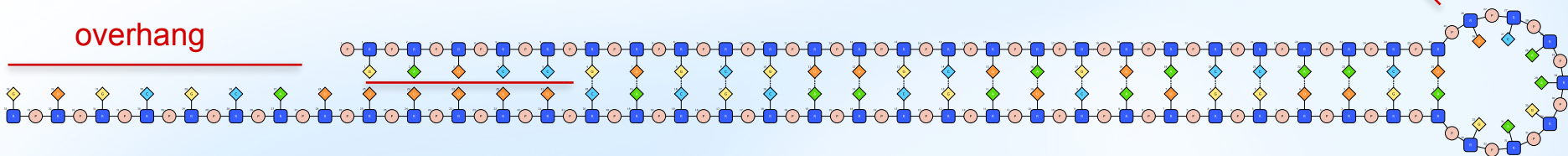
end-cap



hairpin loop

mismatch

overhang



Light

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

Heavy

EVQLVESGGGLVQPGGSLRLSCAASGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVK
GRFTISADTSKNTAYLQMNSLRAEDTAVYYCSRWGGDGFYAMDYWGQGLTVTVSSASTKGPSVF
PLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVTVTPSSS
LGTQTYICNVNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPE
VTCVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKC
KVSNAKALPAPIEKTISKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEWESNGQP
ENNYKTTTPVLDSDGSFFLYSKLTVDKSRWQQGNV FSCSVMHEALHNHYTQKSLSLSPGK
{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

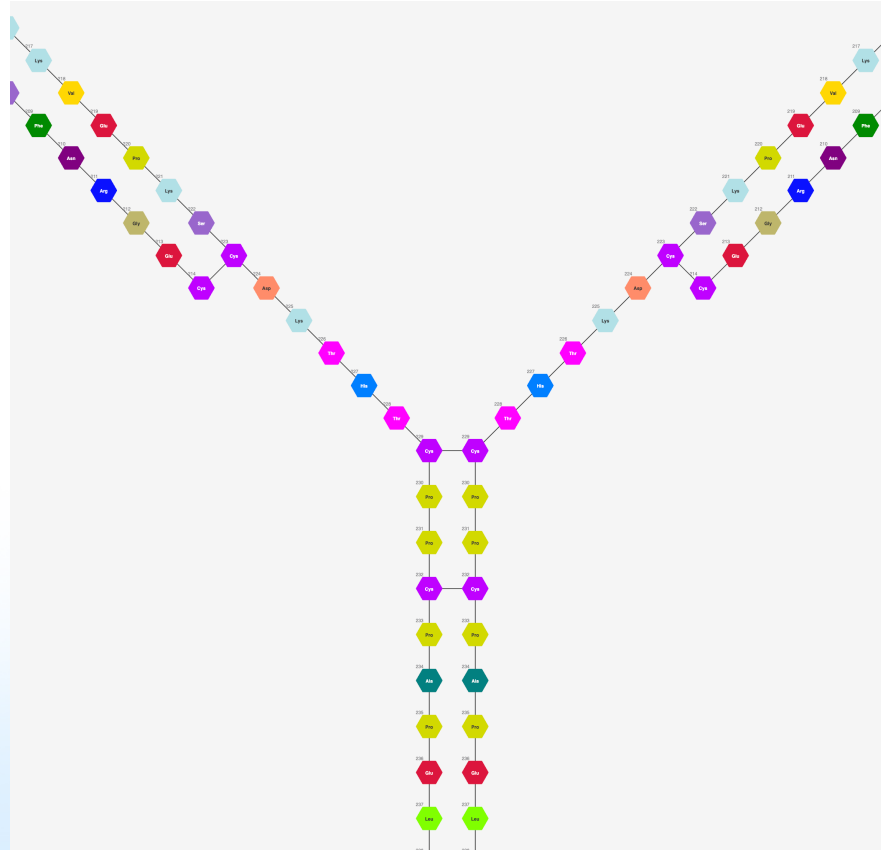


Hinge Region

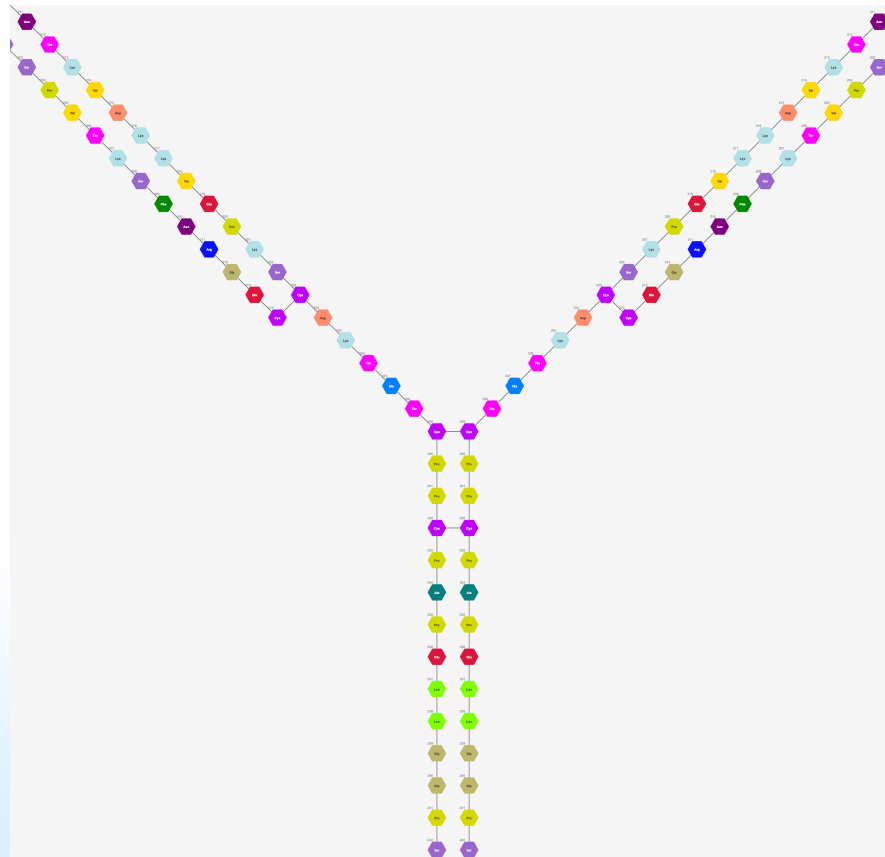


Drug Conjugate Site

Antibodies

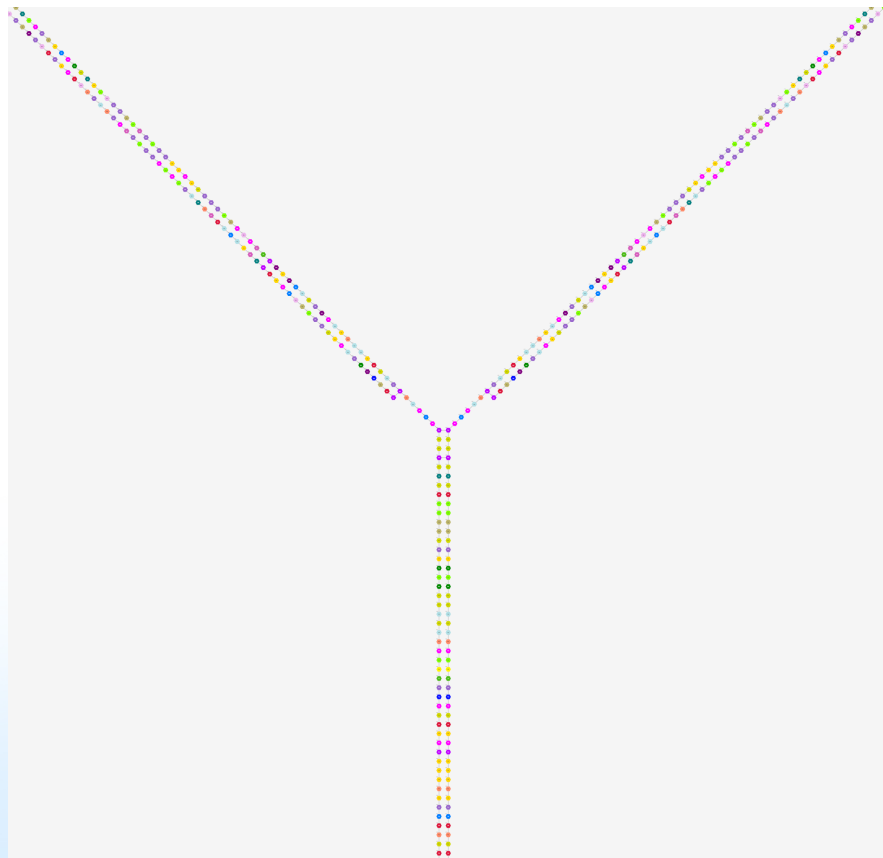


Antibodies

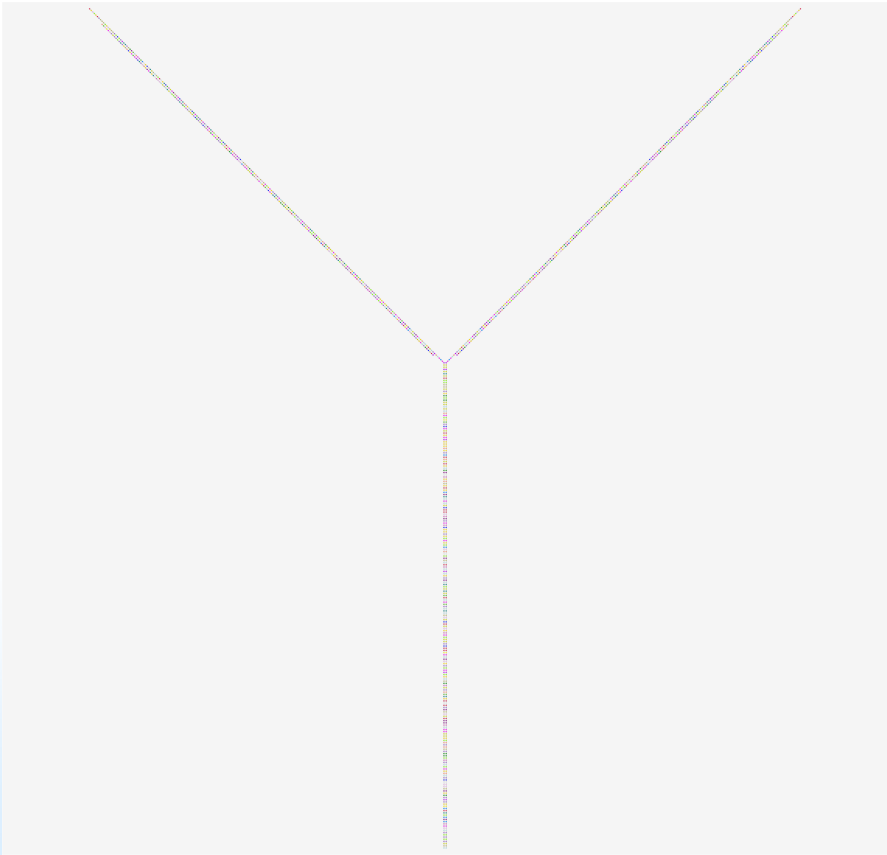


Antibodies

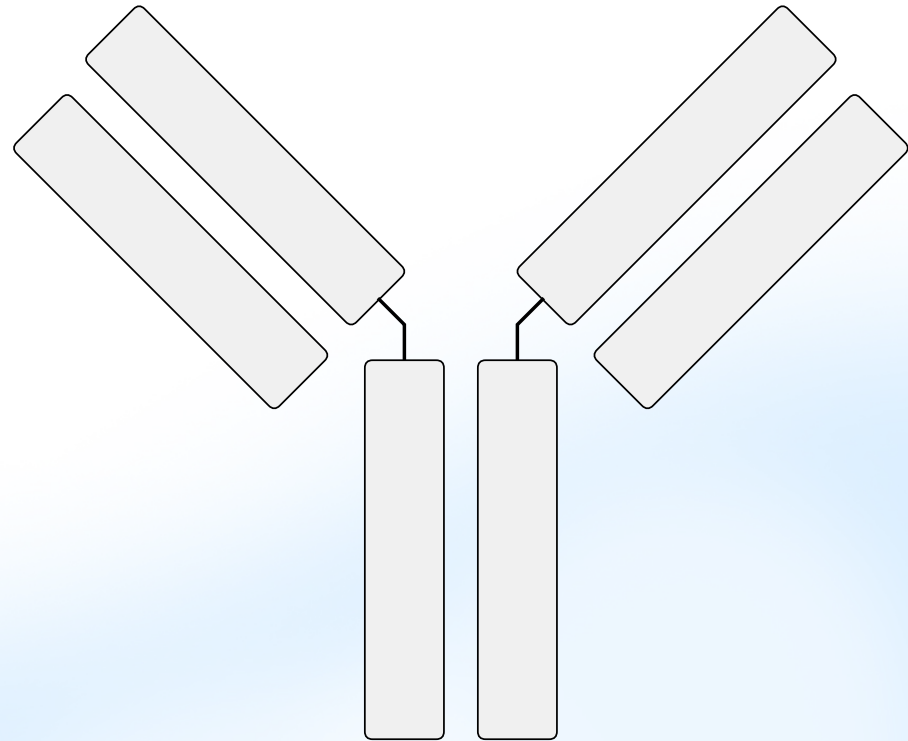




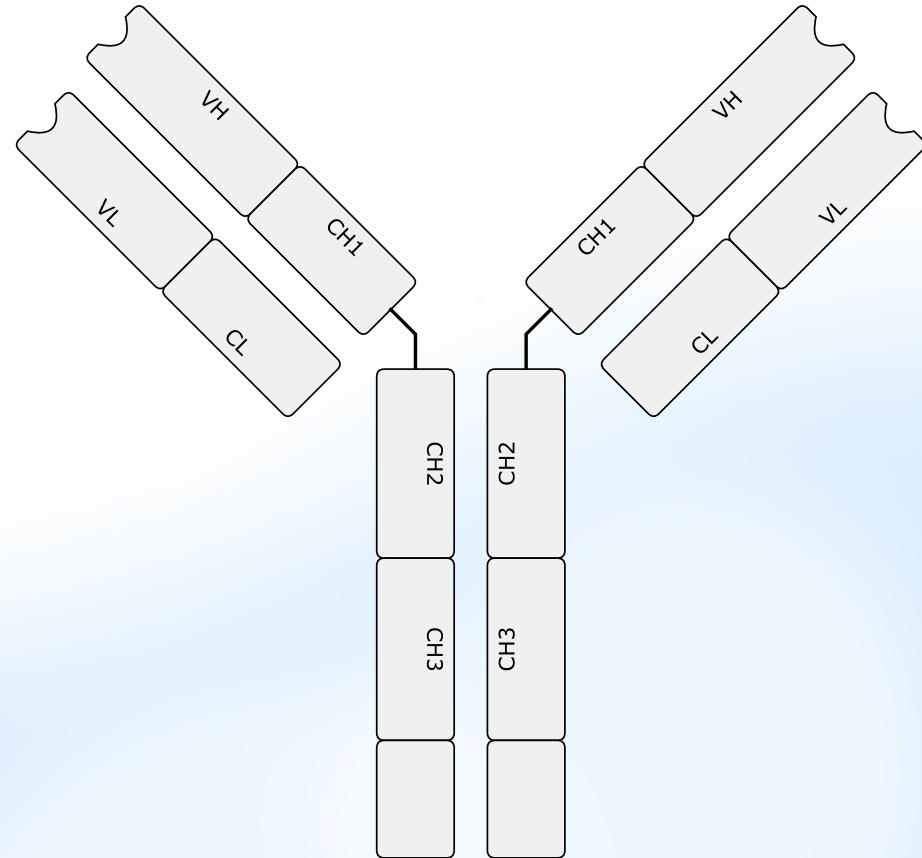
Antibodies



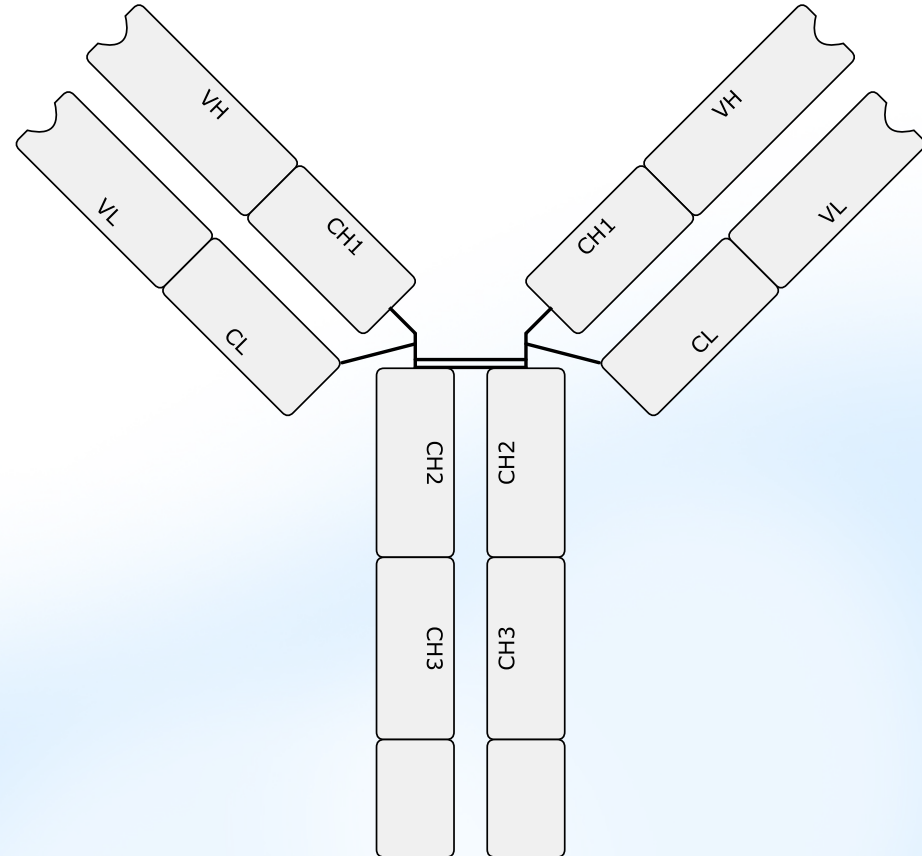
- Draw basic Y-shape outline



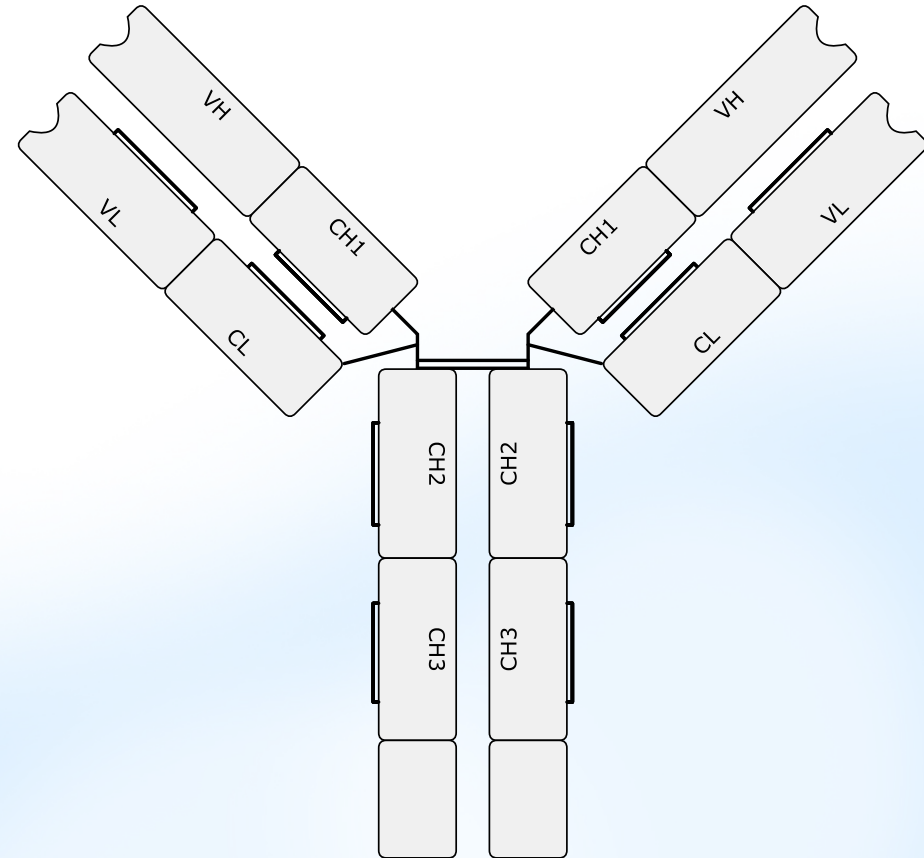
- Draw basic Y-shape outline
- Segment domain regions



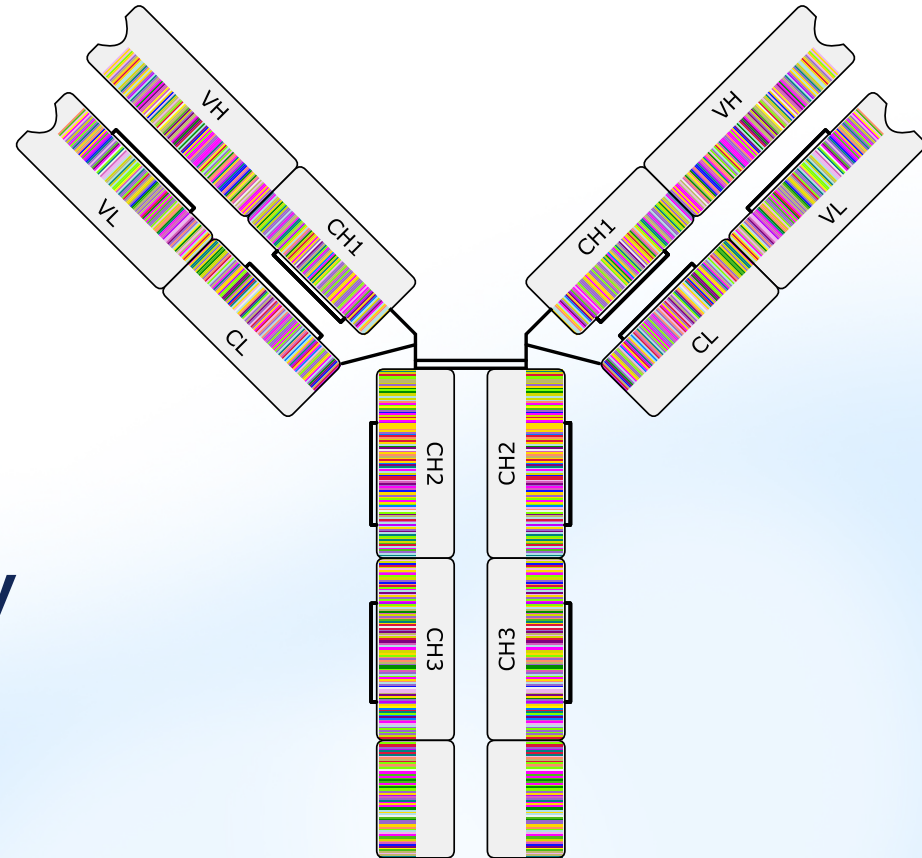
- Draw basic Y-shape outline
- Segment domain regions
- Inter-chain disulfide links



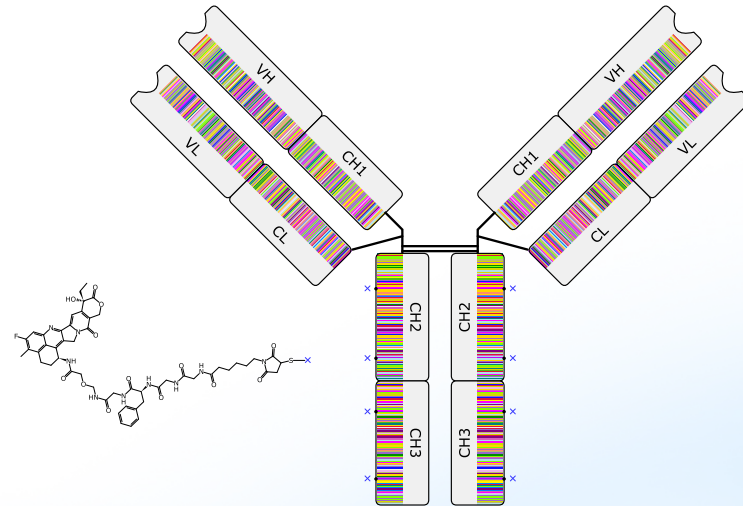
- Draw basic Y-shape outline
- Segment domain regions
- Inter-chain disulfide links
- Intra-chain disulfide links



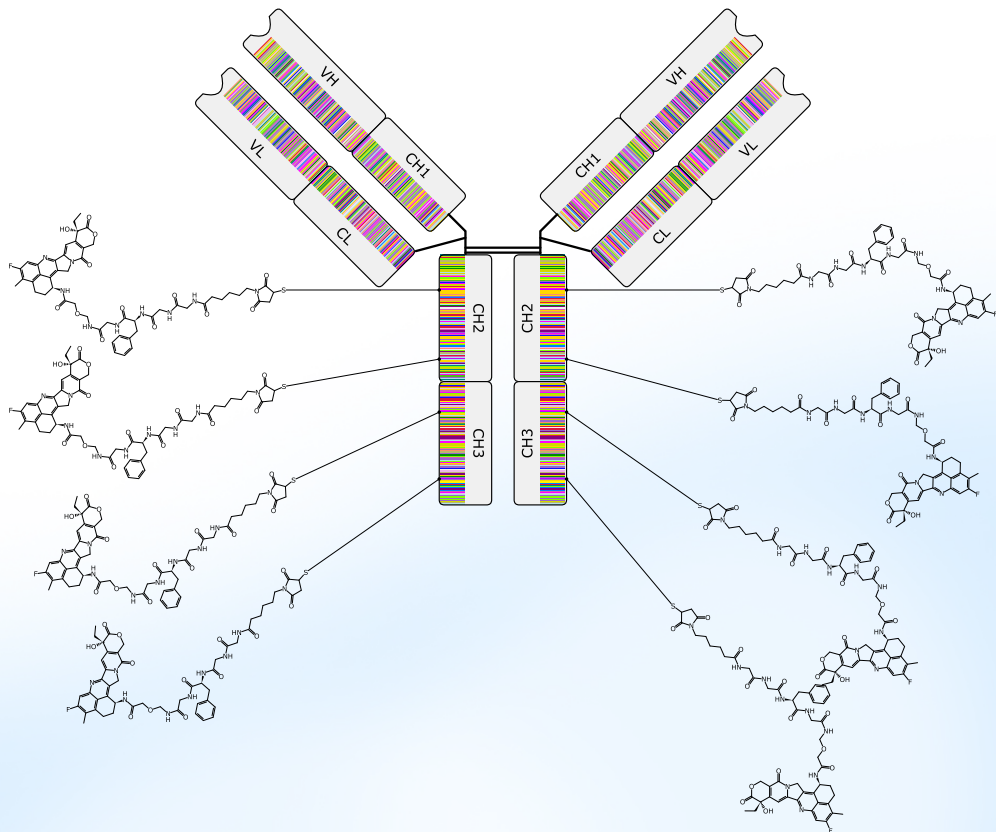
- Draw basic Y-shape outline
- Segment domain regions
- Inter-chain disulfide links
- Intra-chain disulfide links
- Sequence barcoding
- ... **chemistry, biochemistry**
and **biology**



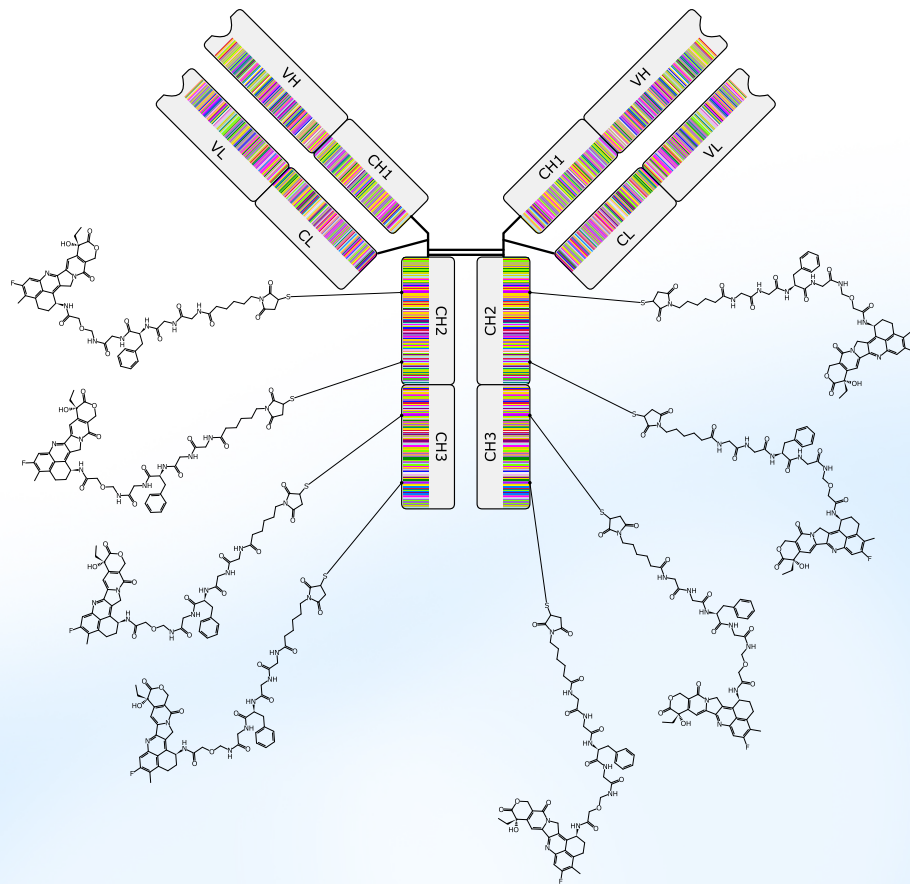
- Anchor point is indicated



- Anchor point is indicated
- Tethered initial placements



- Anchor point is indicated
- Tethered initial placements
- Rigid body position optimization
- Underlying chemistry is fully preserved



- Adopted an industry standard protocol, funded Ketcher editor, enhanced RDKit
- Creation, viewing, registration & searching in Vault
- Management of custom monomers
- Automated antibody markup: *in progress*
- Sequence searching and analytics: *in progress*
- Macromolecule property calculation: *in progress*
- All with open standards & technology that's under our control