CLASSIFICATION, REPRESENTATION AND ANALYSIS OF CYCLIC PEPTIDES AND PEPTIDE-LIKE ANALOGS

ROGER SAYLE, DANIEL LOWE & NOEL O'BOYLE

NEXTMOVE SOFTWARE, CAMBRIDGE, UK

248th ACS National Meeting, San Francisco, CA. Sunday 10th August 2014

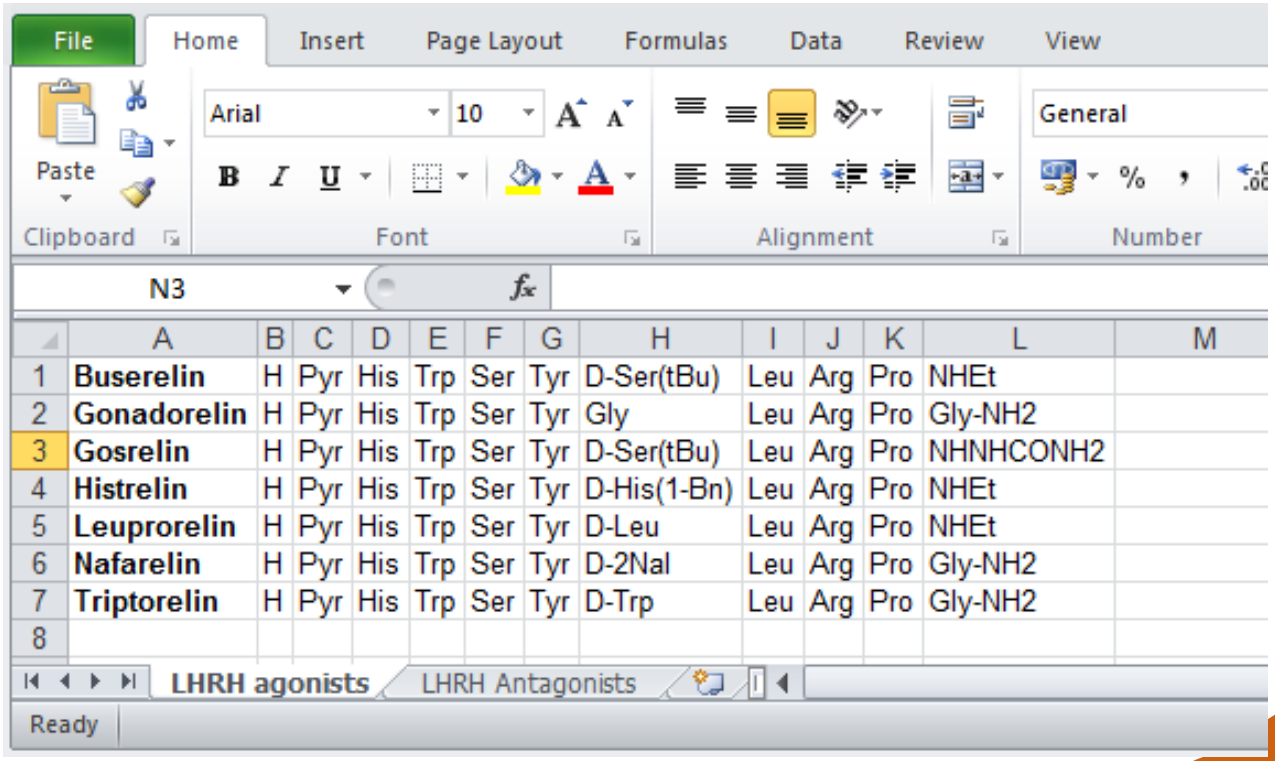
THE STORY SO FAR...

- Research & Development of biologics and biosimilars requires their representation/handling by compound registration systems.
- "Sugar & Splice" is NextMove Software's commercial toolkit for achieving this via perception and naming.
  - Perception: Recognition of biopolymers from "all atom" connection table representations (e.g. SMILES or V2000).
  - Naming: Assignment/generation of identifiers, such as line notations, IUPAC names, HELM, PLN, common names, etc.

LINE NOTATIONS FOR PEPTIDE QSAR

Comparing line notation representations simplifies the task of explaining variant/mutant biological activities/properties.

Ac-D-2Nal-D-Phe(4-Cl)-D-3Pal-Ser-Phe(4-S-dihydroorotamido)-D-Phe(4-ureido)-Leu-Lys(iPr)-Pro-D-Ala-NH2 **Degarelix**
Ac-D-2Nal-D-Phe(4-Cl)-D-3Pal-Ser-N(Me)Tyr-D-Asn-Leu-Lys(iPr)-Pro-D-Ala-NH2 **Abarelix**
Ac-D-2Nal-D-Phe(4-Cl)-D-3Pal-Ser-Tyr-D-Cit-Leu-Arg-Pro-D-Ala-NH2 **Cetrorelix**

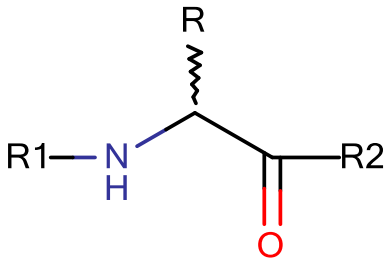


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Buserelin**	H	Pyr	His	Trp	Ser	Tyr	D-Ser(tBu)	Leu	Arg	Pro	NHEt	
2	**Gonadorelin**	H	Pyr	His	Trp	Ser	Tyr	Gly	Leu	Arg	Pro	Gly-NH2	
3	**Gosrelin**	H	Pyr	His	Trp	Ser	Tyr	D-Ser(tBu)	Leu	Arg	Pro	NHNHCONH2	
4	**Histrelin**	H	Pyr	His	Trp	Ser	Tyr	D-His(1-Bn)	Leu	Arg	Pro	NHEt	
5	**Leuprorelin**	H	Pyr	His	Trp	Ser	Tyr	D-Leu	Leu	Arg	Pro	NHEt	
6	**Nafarelin**	H	Pyr	His	Trp	Ser	Tyr	D-2Nal	Leu	Arg	Pro	Gly-NH2	
7	**Triptorelin**	H	Pyr	His	Trp	Ser	Tyr	D-Trp	Leu	Arg	Pro	Gly-NH2	
8													

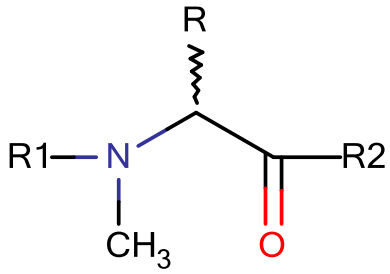
THE STORY TODAY

- At the Dallas ACS in March, I described systematic monomer naming for use in line notations.
- Today, I'll cover recent advances:
  - Unusual peptide backbones: non-alpha amino acids.
  - Simple cyclization: Disulfide bridges and homodetic cycles.
  - Heterodetic cyclization: isopeptide (sidechain and backbone), ester crosslinking and thioether bridges.
  - Derivatives of named (cyclic) peptides.
  - Antibody perception: Isotype and CDR loop identification.

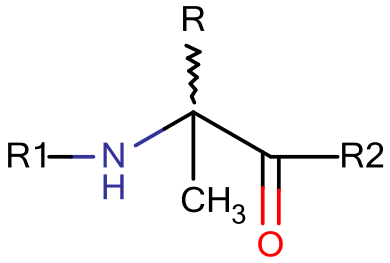
ACYCLIC PEPTIDE BACKBONES



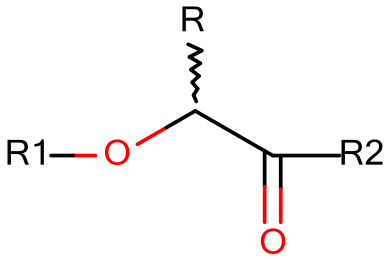
Aaa



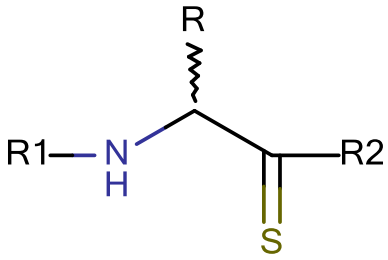
N(Me)Aaa



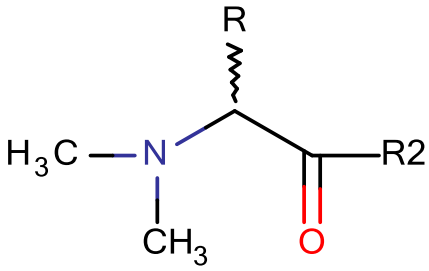
aMeAaa



OAaa

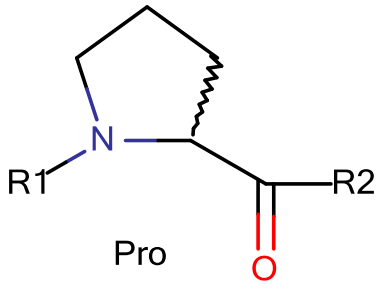


thionoAaa

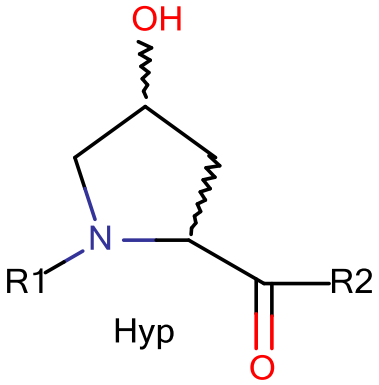


N(Me2)Aaa

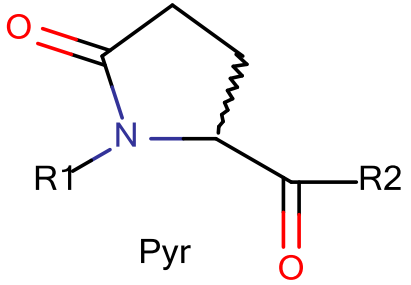
CYCLIC PEPTIDE BACKBONES #1



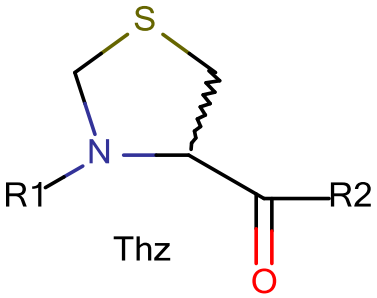
Pro



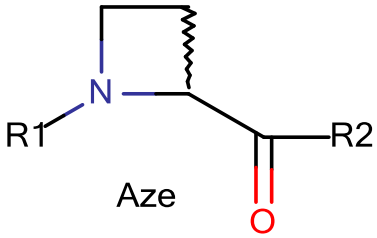
Hyp



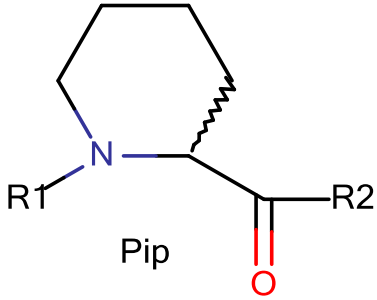
Pyr



Thz

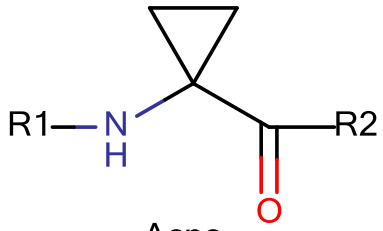


Aze

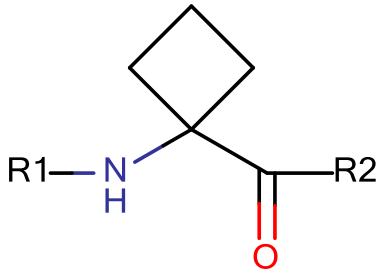


Pip

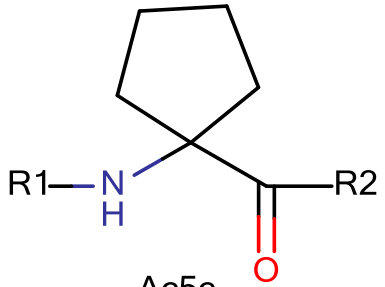
CYCLIC PEPTIDE BACKBONES #2



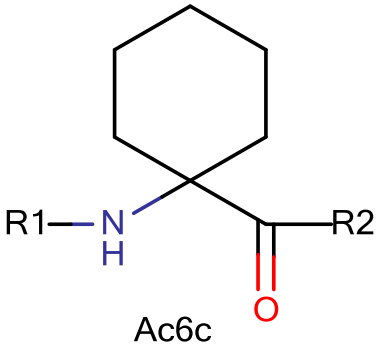
Acpc



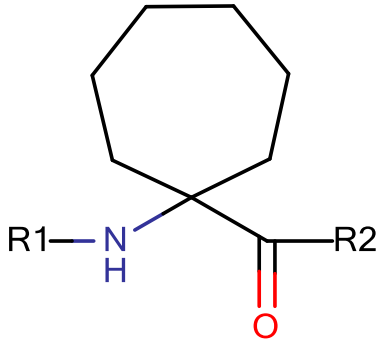
Acbc



Ac5c

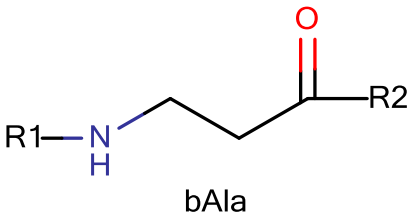


Ac6c

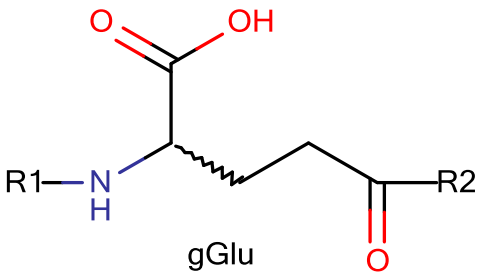


Ac7c

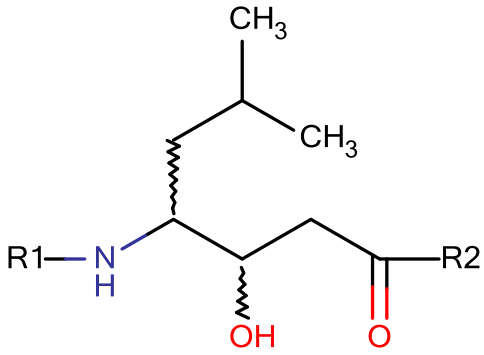
NON-ALPHA AMINO BACKBONES



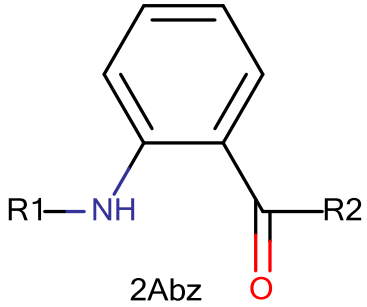
bAla



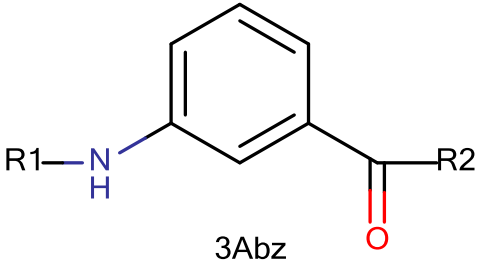
gGlu



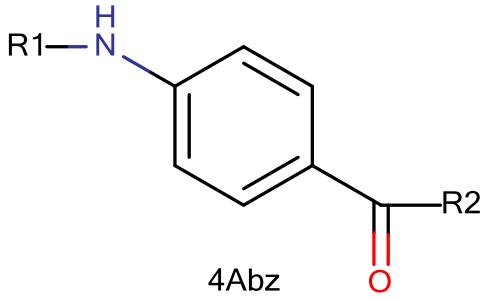
Sta



2Abz

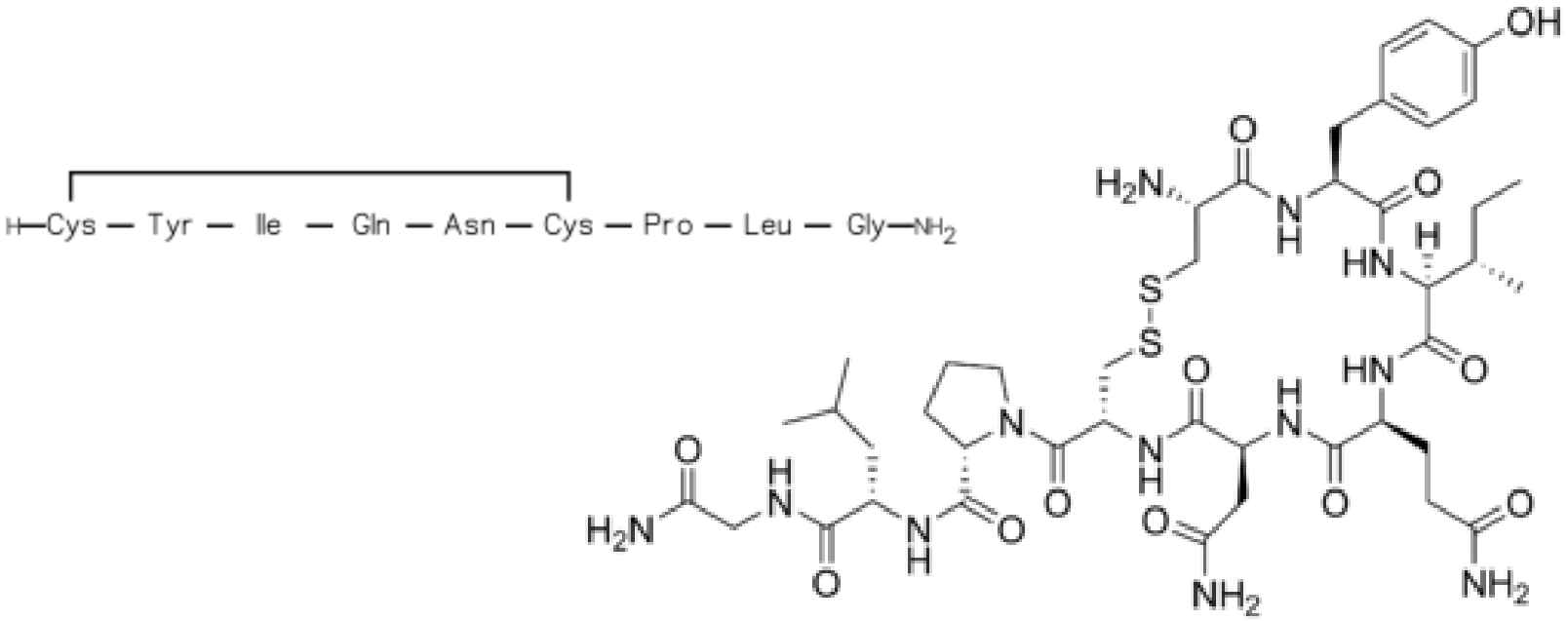


3Abz



4Abz

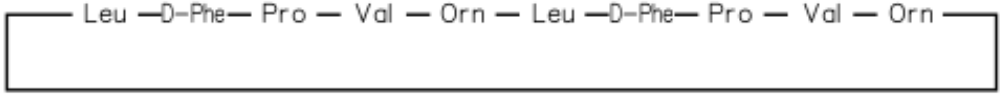
DISULFIDE BRIDGES

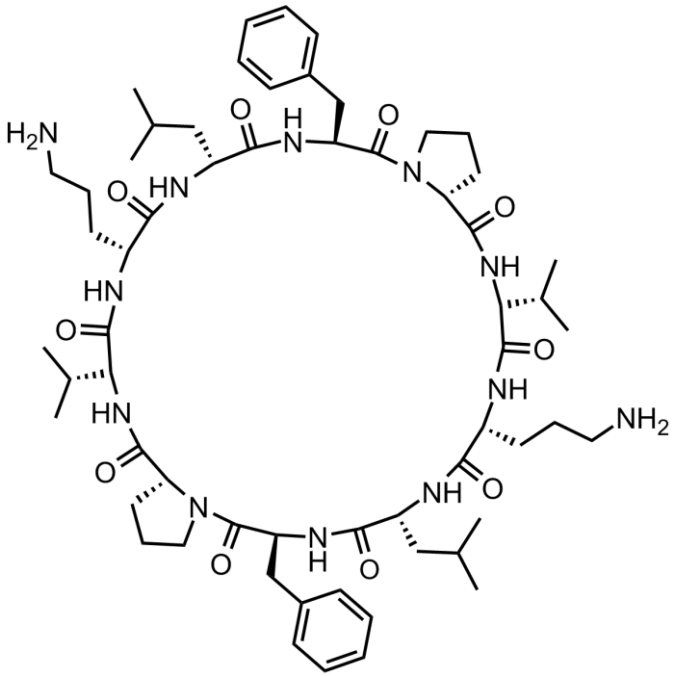


H-Cys(1)-Tyr-Ile-Gln-Asp-Cys(1)-Pro-Leu-Gly-NH2

PEPTIDE1{C.Y.I.Q.D.C.P.L.G.[am]}$PEPTIDE1,PEPTIDE1,1:R3-6:R3$$$
H-C(1)YIQDC(1)PLG-[NH2]
[5-L-aspartic acid]oxytocin

HOMODETIC CYCLES #1

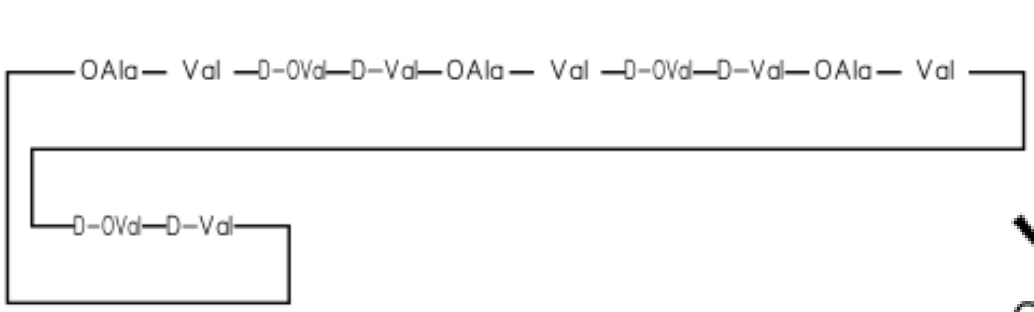


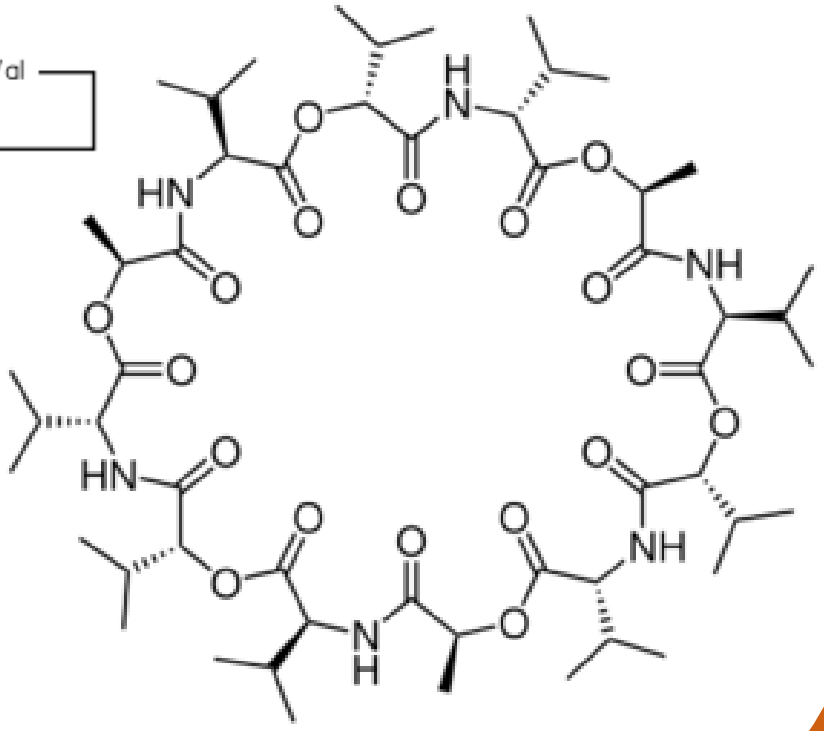


cyclo[Leu-D-Phe-Pro-Val-Orn-Leu-D-Phe-Pro-Val-Orn]

gramicidin S

HOMODETIC CYCLES #2





cyclo[OAla-Val-D-OVal-D-Val-OAla-Val-D-OVal-D-Val-OAla-Val-D-OVal-D-Val]

valinomycin

EXTENDING DISULFIDE BRIDGING

- The postfix parenthesized digit(s) notation can be extended beyond cysteine to allow bridges between other amino acids containing terminal sulfanes.
- Cysteine-like residues include Cys, Pen, Hcy and Sec.

• H-Ala-Pen(1)-Ala-Hcy(1)-Ala-OH	Non-cystine disulfide bridge
• H-Ala-Sec(1)-Ala-Sec(2)-Ala-OH	Diselenide bridge

NOTATION FOR ISOPEPTIDE BONDING

- The syntax of disulfide bridges (and N-modifications) and be used to represent/encode isopeptide bonds.
- Acid residues include Asp, Glu, Aad, Asu, Gly(CO2H).
- Amine residues include Lys, Orn, Dab, Dap, hLys, Agl.

• H-Ala-Lys(1)-Ala-Asp(1)-Ala-OH	Sidechain acid-sidechain amine
• H-Ala-Lys(1)-Ala-Glu(1)-Ala-OH	Sidechain acid-sidechain amine
• H-Ala-N(1)Ala-Ala-Asp(1)-Ala-OH	Sidechain acid-backbone amine
• H-N(1)Ala-Ala-Glu(1)-Ala-OH	Sidechain acid-backbone amine
• Me-N(1)Gly-Ala-Ala-Asp(1)-OH	Sidechain acid-backbone amine
• H-Ala-Lys(1)-Ala-Ala-(1)	Backbone acid-sidechain amine

REAL ISOPEPTIDE BONDING EXAMPLES

Anantin

H-N(1)Gly-Phe-Ile-Gly-Trp-Gly-Asn-Asp(1)-Ile-Phe-Gly-His-Tyr-Ser-Gly-Asp-Phe-OH

Sungsanpin

H-N(1)Gly-Phe-Gly-Ser-Lys-Pro-Ile-Asp(1)-Ser-Phe-Gly-Leu-Ser-Trp-Leu-OH

ADDITIONAL CROSS LINK NOTATIONS

- Thioether linkages, between a cysteine and an N-terminal (chloro)acetyl group, are increasingly popular, perhaps thanks to PeptiDream technology.
- Ac(1)-Ala-Cys(1)-Ala-OH

- Ester linkages are between acids and hydroxyl sidechains, including Ser, Tyr and Thr.
- H-Ala-Asp(1)-Ala-Ser(1)-Ala-OH

PEPTIDE NAMES IMPLY ARCHITECTURE

- Named peptides imply not only sequence but also N-terminal acetylation, C-terminal amidation and disulfide bridge topology.
- Example named derivatives:
  - gastrin (14-17)
  - motilin amide
  - oxytocin free-acid
  - acetyl-oxytocin
  - deacetyl-abarelix
  - oxytocin reduced
  - endothelin-1 (1→3),(11 → 15)-bis(disulfide)

NAMED CYCLIC PEPTIDE DERIVATIVES

- Mutants of named cyclic peptides can be identified by comparing against all “rotational” permutations.

Example line notation query (CHEMBL478596)

cyclo[Ala-Gly-Thr-Phe-Val-Tyr]

Reference database line notations:

cyclo[Gly-Thr-Phe-Leu-Tyr-Thr] dichotomin B

cyclo[Ala-Gly-Thr-Phe-Leu-Tyr] dichotomin C

Resulting Sugar & Splice peptide name:

[5-L-valine]dichotomin C

ISOTYPING ANTIBODY CHAINS

- Antibody chains may be classified into isotypes using global (NWS) sequence alignment [*n.b.* not blastp].

	UNIPROT	**ChEMBL19 count**
• **Light chains**		
– Ig kappa [human]	IGKC_HUMAN	131
– Ig lambda [human]	LAC2_HUMAN	12
– Ig kappa [mouse]	IGKC_MOUSE	7
• **Heavy chains**		
– Ig gamma-1 [human]	IGHG1_HUMAN	105
– Ig gamma-2 [human]	IGHG2_HUMAN	13
– Ig gamma-4 [human]	IGHG4_HUMAN	21
– Ig gamma-1 [mouse]	IGHG1_MOUSE	2
– Ig gamma-2A [mouse]	GCAA_MOUSE	2
– Ig gamma-2B {secreted} [mouse]	IGG2B_MOUSE	1
– Ig mu {secreted} [human]	IGHM_HUMAN	1 (Panobacumab)

ANTIBODY CDR LOOP PERCEPTION

```
LC light chain: 215 AA
Isotype: Ig kappa [human]
Aligning against Adalimumab_L (214 AA)
Identity:    81.78% (175/214)
Similarity:  91.59% (196/214)

Ref DIQMTQSPSSLSASVGDRVTITCRASQGIRN-YLAWYQQKPGKAPKLLIYAASTLQSGVP 59
    :| :||||::|| | |:| |::|||||| : : |||||||||:||:||:: :|:  :|:|
Qry EIVLTQSPATLSLSPGERATLSCRASQIVSSAYLAWYQQKPGQAPRLLMFGSSSRATGIP 60

Ref SRFSGSGSGTDFTLTISSLQPEDVATYYCQRYNRAPYTFGQGTKVEIKRTVAAPSVFIFP 119
     |||||||||||||||| |:||| | ||||:|  :  ||| ||||:|||||||||||||
Qry DRFSGSGSGTDFTLTISRLEPEDFAVYYCQQYGSSQGTFGPGTKVDIKRTVAAPSVFIFP 120

Ref PSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTL 179
    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qry PSDEQLKSGTASVVCLLNNFYPREAKVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTL 180

Ref TLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC 214
    |||||||||||||||||||||||||||||||||
Qry TLSKADYEKHKVYACEVTHQGLSSPVTKSFNRGEC 215

Anchor positions: 25, 33, 49, 53, 88, 98
IMGT CDR lengths: [7.3.9]
..SCRAS[QIVSSAY]LAWYQ..RLLMF[GSS]SRATG..AVYYC[QQYGSSQGT]FGPGT..
```

ANTIBODY PERCEPTION EXAMPLES

DavesADC.mol
SciTegic10231309212D
Courtesy of Keith Taylor, Ladera
0 0 0 0 0 0 999 V3000
M V30 BEGIN CTAB
M V30 COUNTS 10256 10536 0 0 1
M V30 BEGIN ATOM

SNSIsotyper can be used to check that an antibody has been correctly registered [bridges, glycans, etc.].

DavesADC_H: 450 AA Isotype: Ig gamma-1 [human] IMGT CDR lengths: [8.8.13]

..SCAAS[GFNIKDTY]IHWVR..EWVAR[IYPTNGYT]RYADS..AVYYC[SRWGGDGFYAMDY]WGQGT..

DavesADC_L: 213 AA Isotype: Ig kappa [human] IMGT CDR lengths: [6.3.9]

..TCRAS[QDVNTA]VAWYQ..KLLIY[SAS]FLYSG..ATYYC[QQHYTTPPT]FGQGT..

HC heavy chain: 446 AA Isotype: Ig gamma-1 [human] IMGT CDR lengths: [8.7.10]

..TCTVS[GGSISGYY]WSWIR..EWIGR[IYTSGST]NYNPS..AVYYC[ARGRFTYFDY]WGQGT..

LC light chain: 215 AA Isotype: Ig kappa [human] IMGT CDR lengths: [7.3.9]

..SCRAS[QIVSSAY]LAWYQ..RLLMF[GSS]SRATG..AVYYC[QQYGSSQGT]FGPGT..

ACKNOWLEDGEMENTS

- Lisa Sach-Peltason, Hoffmann-La Roche, Basel.
- Joann Prescott-Roy, Novartis, Boston, MA.
- Greg Landrum, Novatis, Basel, Switzerland.
- Evan Bolton, NCBI PubChem project, Bethesda, MD.