U.S. FOOD & DRUG ADMINISTRATION

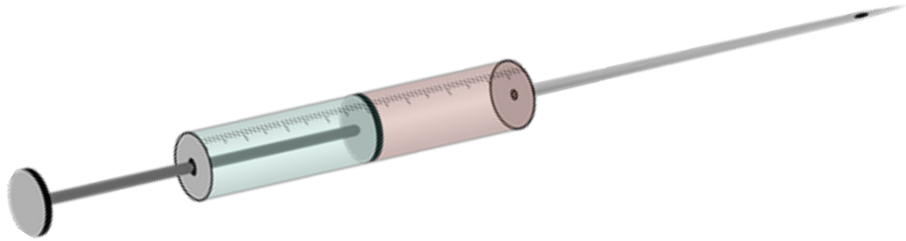
How InChI helps describe and identify biopharmaceuticals

Yulia Borodina

Office of Health Informatics

InChI symposium Fall 2019 San Diego, CA, August 23 - 24, 2019

Biopharmaceuticals



- Substances extracted from biological organism or synthesized on biological matrices
- Majority are proteins
- Often conjugated with small molecules, polymers or themselves in order to increase bioavailability, immunogenicity or to deliver an unspecific toxin
- Therapy of cancer, rheumatoid arthritis, migraine, asthma etc.
- Cost is very high

Biopharmaceuticals in clinical trials (Ex)

Clinical Trial

RINDOPEPIMUT

NCT Number:	NCT01480479	Recruitment:	Completed
Title:	Phase III Study of Rindopepimut/GM-CSF in Patients With Newly Diagnosed Glioblastoma	Conditions:	Glioblastoma|Small Cell Glioblastoma|Giant Cell Glioblastoma|Gliosarcoma|Glioblastoma With Oligodendroc Component
Sponsor:	Celldex Therapeutics Celldex Therapeutics|	Intervention:	Drug: Rindopepimut (CDX-110) with GM-CSF|Drug: Temozolomide|Drug: KLH
Funded Bys:	Industry	Phases:	Phase 3
Study Types:	Interventional	Age Groups:	18 Years and older (Adult, Older Adult)
Study Designs:	Allocation: Randomized|Intervention Model: Parallel Assignment|Masking: Triple (Participant, Care Provider, Investigator)|Primary Purpose: Treatment	Gender:	All
Study Results:	No Results Available	Enrollment:	745
Other Ids:	CDX110-04	Acronym:	ACT IV

RINDOPEPIMUT described by USAN

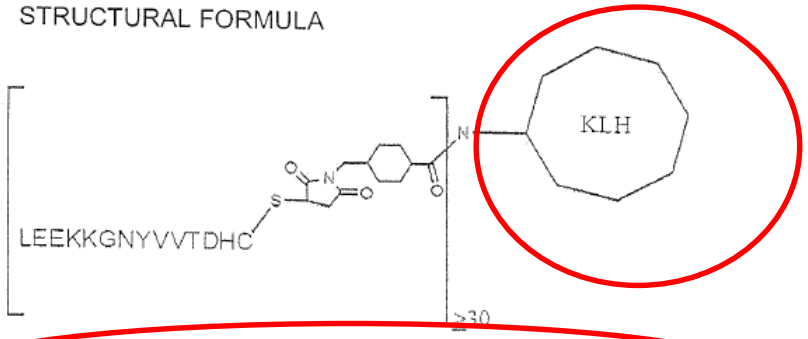
STATEMENT ON A NONPROPRIETARY NAME ADOPTED BY THE USAN COUNCIL

USAN (WW-34)	RINDOPEPIMUT
PRONUNCIATION	rin'' doe pep' i mut
THERAPEUTIC CLAIM	Immunotherapy for glioblastoma multiforme and potentially other tumor types

CHEMICAL NAME

L-leucyl-L-.alpha.-glutamyl-L-.alpha.-glutamyl-L-lysyl-L-lysylglycyl-L-asparaginyl-L-tyrosyl-L-valyl-L-valyl-L-threonyl-L-.alpha.-aspartyl-L-histidyl-L-cysteine thioether with *N*-[[4-[(3-mercapto-2,5-dioxo-1-pyrrolidinyl) methyl]cyclohexyl]carbonyl]-keyhole limpet hemocyanin (Megathura crenulata) (1:?)

STRUCTURAL FORMULA

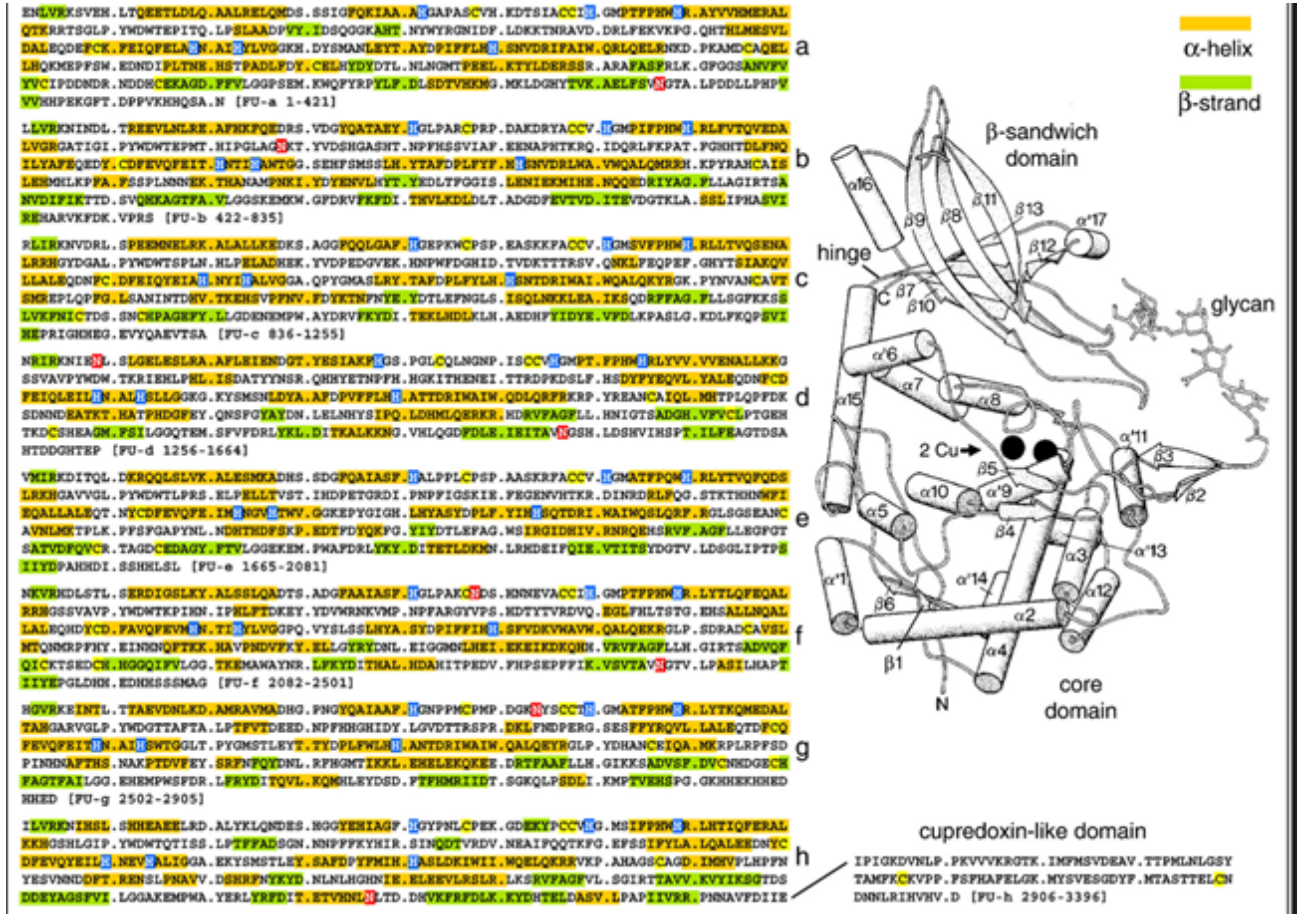


Hemocyanin

MOLECULAR WEIGHT	450-550 kDa
TRADEMARKS	Rintega™
MANUFACTURER	Celldex Therapeutics, Inc.
CODE DESIGNATION	CDX-110
CAS REGISTRY NUMBER	1108208-65-6

Hemocyanin

Keyhole limpet hemocyanin (KLH) is ***a very large, copper-containing protein*** molecule derived from the haemolymph of the inedible mollusc, Megathura crenulata. KLH is a highly immunogenic T-cell dependent antigen that is used increasingly in immunotoxicological studies, particularly in those involving animals.



We do not use InChI to define primary protein structure

Code	Amino acid	Code	Amino acid
A	alanine	I	isoleucine
R	arginine	L	leucine
N	asparagine	K	Lysine
D	aspartic acid	M	methionine
B	asparagine or aspartic acid	F	phenylalanine
C	cysteine	P	proline
E	glutamic acid	S	serine
Q	glutamine	T	threonine
Z	glutamine or glutamic acid	W	tryptophan
G	glycine	Y	tyrosine
H	histidine	V	Valine
X	(wildcard / placeholder)		

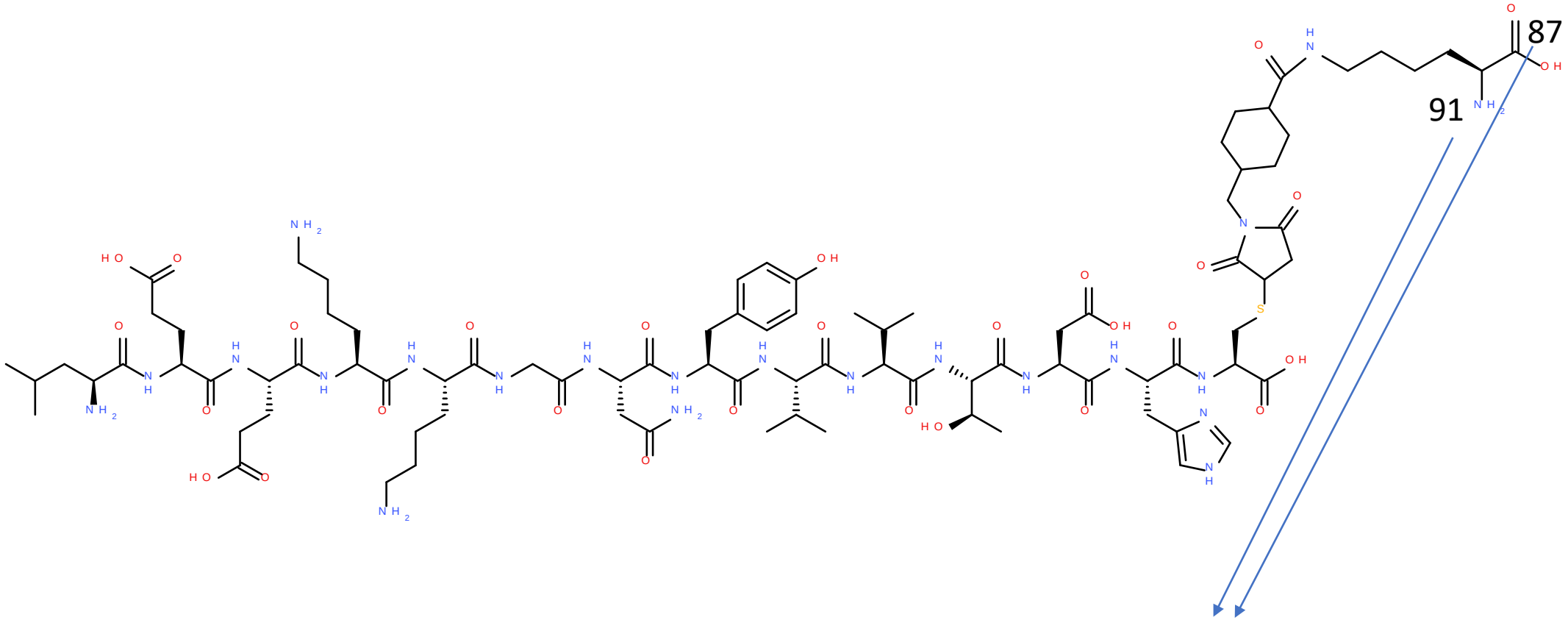
Primary protein structure is defined by amino acid single letter code

We use InChI to define modifications to the primary protein structure

- Modified amino acids
- Links
- Prosthetic groups

Example: definition of a modified amino acid

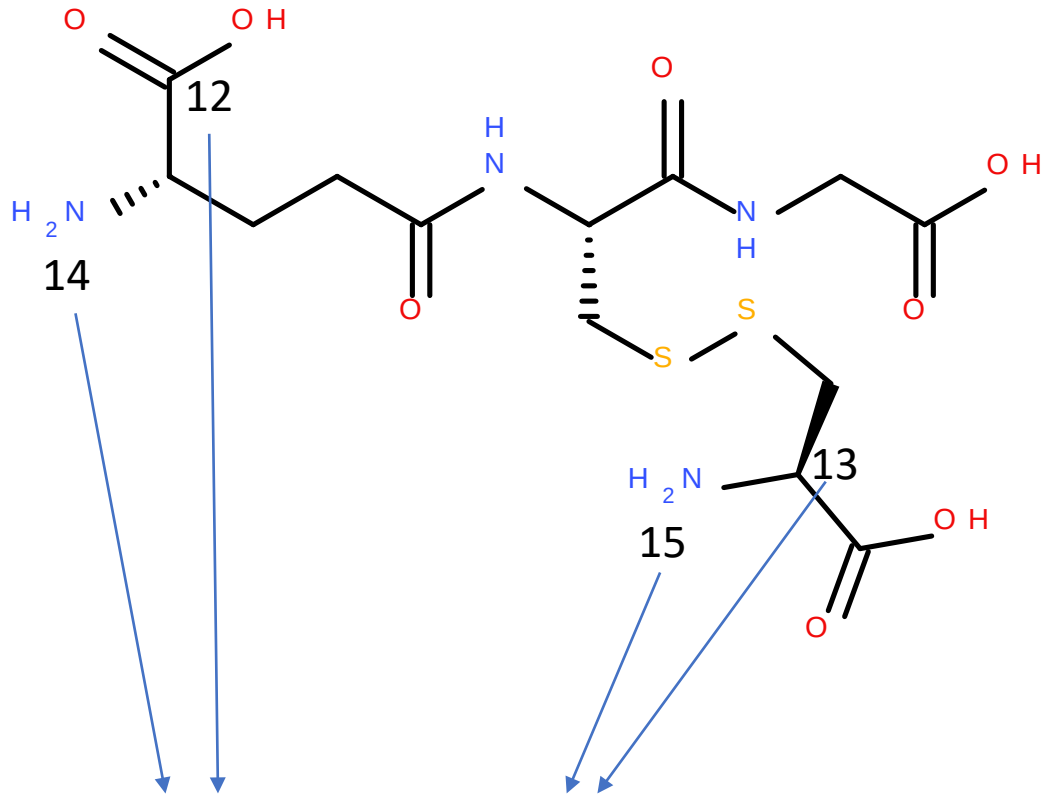
Lysine linked to Epidermal Growth Factor Receptor variant III



InChI canonical atom numbers for atoms that ***substitute*** N of amino group and C of carboxyl group of a natural amino acid

Example: definition of a link

CYSTEINE-GLUTATHIONE DISULFIDE

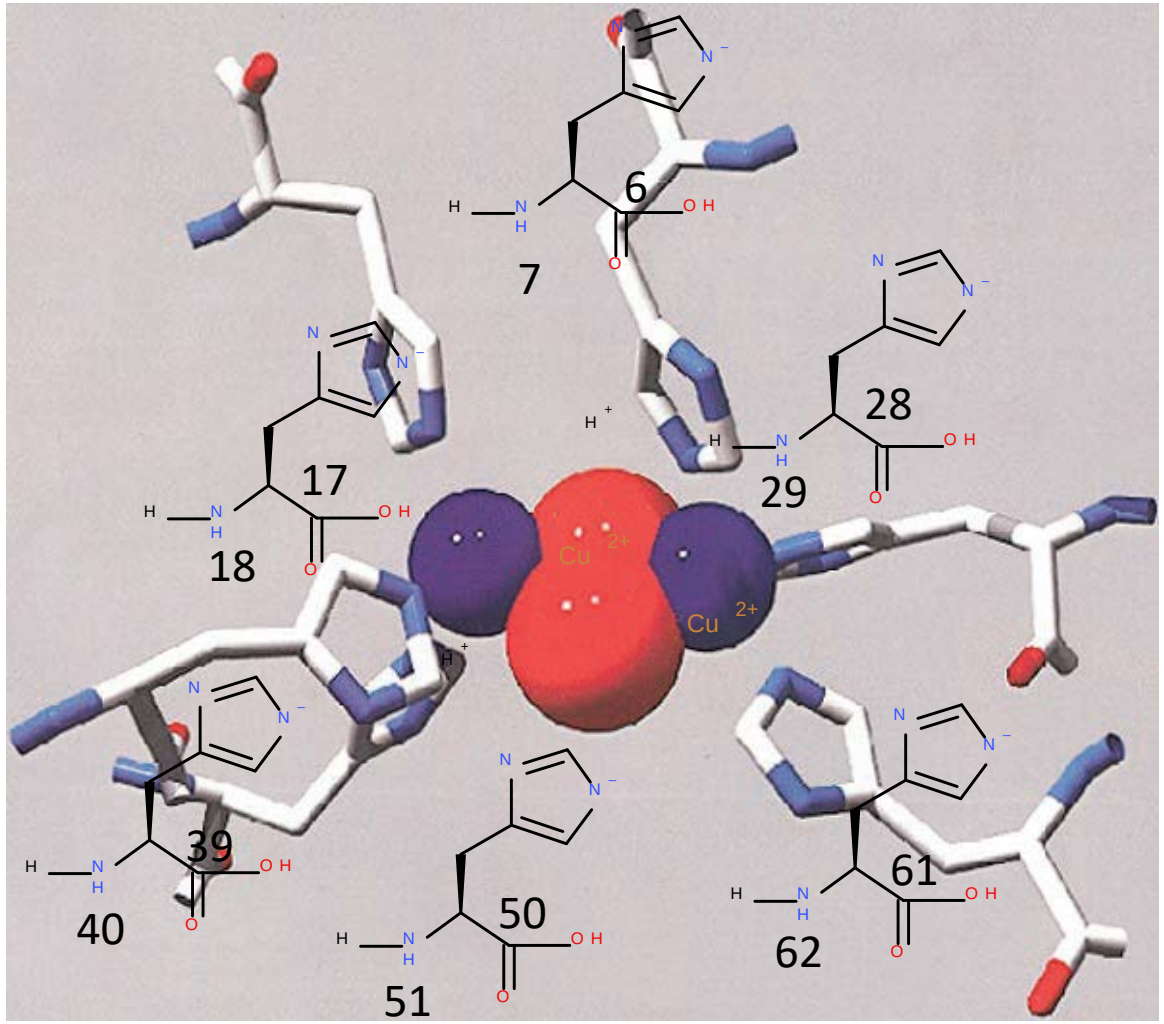


InChI=1S/C13H22N4O8S2/c14-6(12(22)23)1-2-9(18)17-8(11(21)16-3-10(19)20)5-27-26-4-7(15)13(24)25/h6-8H,1-5,14-15H2,(H,16,21)(H,17,18)(H,19,20)(H,22,23)(H,24,25)/t6-,7-,8-/m0/s1

(N,C) pairs
(14,12)
(15,13)

InChI canonical atom numbers for atoms that ***substitute*** N of amino group and C of carboxyl group of a natural amino acid

Example: definition of type 3 copper center



InChI=1S/6C6H9N3O2.2Cu/c6*7-5(6(10)11)1-4-2-8-3-9-4;;/h6*2-3,5H,1,7H2,(H2,8,9,10,11);;/q;;;;;;2*+2/p-4/t6*5-;;/m000000../s1

(N,C) pairs
(7,6)
(18,17)
(29,28)
(40,39)
(51,50)
(62,61)

Identification

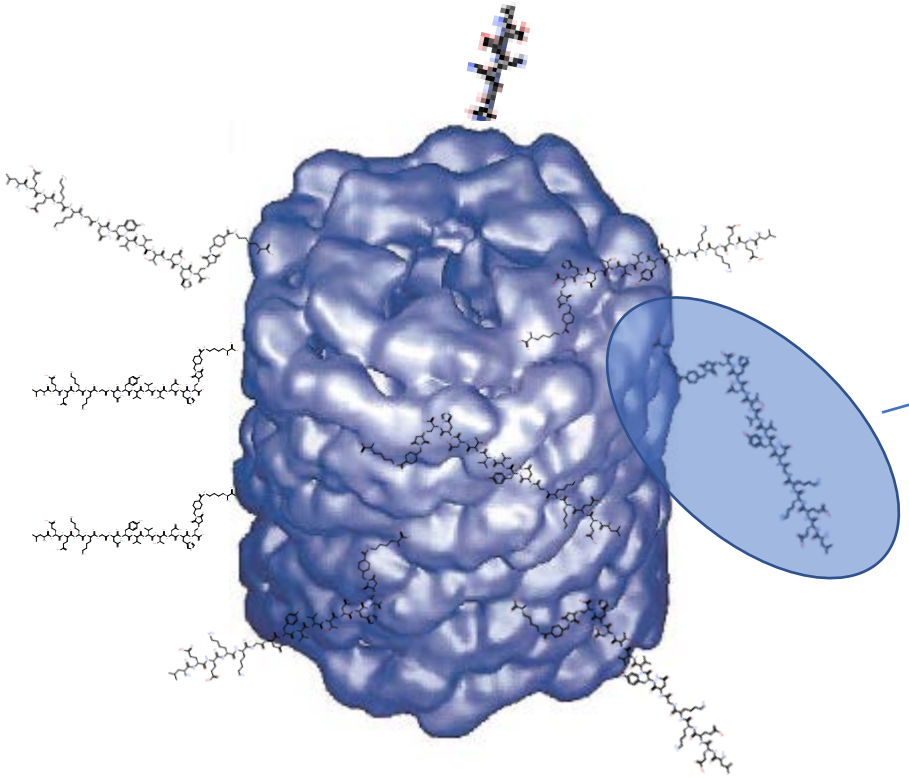
- InChI and ordered list of canonical (N,C) pairs provide that every modification is uniquely indentified

- Location of a modification in the protein is specified by positions of substituted amino acids

- For example, Type 3 copper center position in hemocyanin is (2940, 2959, 2968, 3069, 3073, 3100)

- Complete 2D chemical structure of the modified protein can be reconstructed if needed

Uncertain modifications are annotated by empirical probabilities



```
<quantity >
  <numerator xsi:type="URG_PQ" value="30" unit="mol" >
   <denominator value="170" unit="mol" />
</quantity>
```

Data on (some) proteins with UNIIs are published

- Substance structures in XML format

https://dailymed.nlm.nih.gov/dailymed/spl-resources-all-indexing-files.cfm

- Substance names in text format

https://fdasis.nlm.nih.gov/srs/

- Other product-related information in XML format

https://dailymed.nlm.nih.gov/dailymed/

- SPL Substance Implementation Guide is #14 in

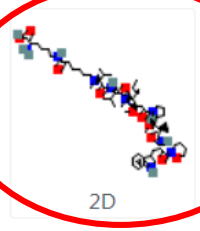
https://www.fda.gov/media/84201/download

Protein descriptions provided by PubChem are misleading

PubChem About Blog Submit Contact

COMPOUND SUMMARY

Aprutumab ixadotin

PubChem CID:	71471223
Structure:	 2D Find Similar Structures
Molecular Formula:	$C_{57}H_{95}N_9O_{11}$
Chemical Names:	UNII-OP0KZ535MV OP0KZ535MV Ixadotin Aprutumab ixadotin DDD2AB5TWK

This is not Aprutumab ixadotin

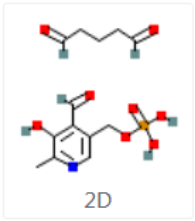
Protein descriptions provided by PubChem are misleading

PubChem About Blog Submit Contact

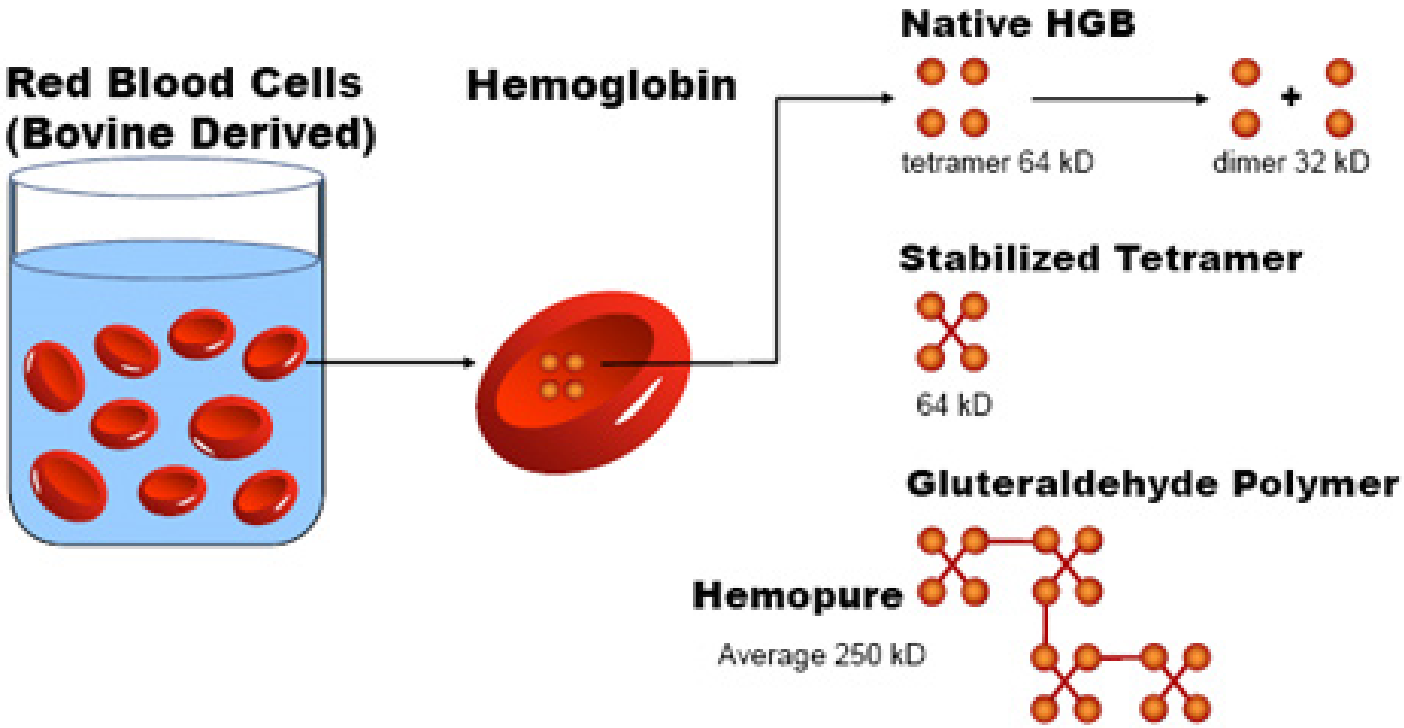
COMPOUND SUMMARY

Hemoglobin glutamer-256 (Human)

kDa

PubChem CID:	72941846
Structure:	 2D Find Similar Structures
Molecular Formula:	$C_{13}H_{18}NO_8P$
Chemical Names:	Polyheme Hemoglobin glutamer-256 UNII-353O60Z7Q1 Hemoglobin glutamer-256 (human) 353O60Z7Q1

Hemoglobin Glutamer



[2] Clinical Investigator's Brochure, Version: IBHEM-003.20 Sept05