

How InChI helps describe and identify biopharmaceuticals

Yulia Borodina

Office of Health Informatics

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





- 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			
NCT Number:	NCT01480479	Recruitment:	Completed
Title:	Phase III Study of Rindopepimut/GM-CSF in Patients With Newly Diagnosed Glioblastoma	Conditions:	Olioblastoma Small Cell Glioblastoma Giant Cell Glioblastoma Gliosarcoma Glioblastoma With Oligodendro 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

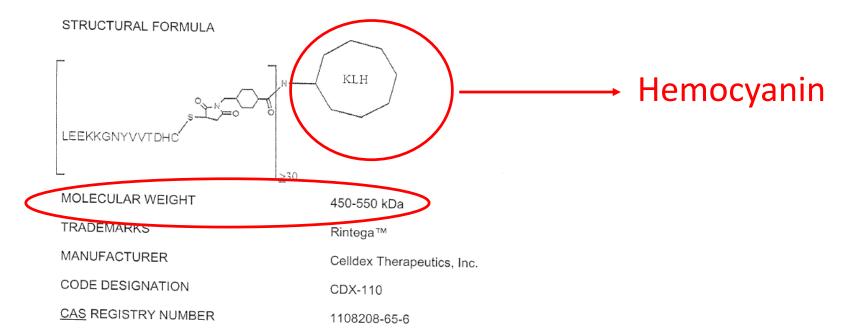
rin" doe pep' i mut

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:?)



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.

ENLVERSVEH.LTQEETLDLQ.AALRELQMDS.SSIGPQKIAA.MEGAPASCVH.KDTSIACCIE.GMPTPPHNER.AYVVEMERAL QTERRTSGLP.YMDNTEPITQ.LPSLAADPYY.IDSQOCKAHT.MYWYRGNIFF.LDKKTNRAVD.DRLFEKVKHG.GHTHMESVL DALEGDEPCK.FEIQFELA.N.ALWINGGKH.DYSMANLEYT.AYDPIPHEE.SNVDRIPAH.QRLGEHXMED.PKAMCAQEL LHCMKEPFEN ENND FLTHE.MSTPADLEDY.CELMYNDYTH.MINGTPELL.TYLDERSSR.ARAFASFRLK.GPGOSAWYFY YVCIPDDNDR.NDDECEMAGD.FFYLGGPSEM.KWQFYRPYLF.BLSDTVHKMG.MKLDGHYTVK.AELFSVGJTA.LPDDLLPHPY VVEHPEKGFT.DPPVKEHQSA.N [FU-a 1-421]

LLYRKNindl.Treevinler.aphicpodes.vdcyotary.gclparder.datocv.gcw.gcw.grupting.lliftoveda Lvyrgcatigi.pyndwtepht.hipglackt.yvdshgasht.nppessviaf.eenaphitko.idglyrpat.gcmhtdlyno Llyrkgcidy.cdpevorett.etteatgatg.cehysksik.ytrydlift.ycdlyrget.ks.vvgaloreet.rpyraicate Llernelkppa.gsslmnek.thatamenet.vdyinvleyt.ycdlypggis.lenteknine.nogedetyrg.flagirts Nvdifikto.svgekagta.vlggskench.gfds.vefdi.twvlkdldlt.adgdpevivd.itevdtkla.bselphasvi Kekarverdk.vvfsk.(vd.s.255)

RETENNORL, SPERNMER, K. ALALLKENKS, AGOPQOLGAP, MORPHOPSP, KASKYRACOV, MONSVPPHNIL RLITVOGENA LRENGYDGAL, PYNDWTSPLN, NLPELANGEN, VYDPEGGVEK, NNPMPDGHID, TVDKTTRKOV, ONKLPROPEP, GNYTŠIACOV LLALGODNEC, DERIGYETAL, NY BJALVGGA, OPYGNASLAV, TARDPLFYLM, "SNTDRINKI, NOALQKYRK, PYNNANCAPT C SMEEPLOPE, LEANINTONV, TKENSVPPNV, FDYKTNÝNÝE, YDTLEPNGLS, ISOLNNKLEM, IKSODEPFAG, FLLSGPKKSE LVEPRICTDS, SNCHPAGEFY, LLODENROPM, AYDNVEKTNÍ, TEKLNDIKLM, ARDNPHIDYE, VVDLKPASLG, KDLFKOPEVI MERTGNIEG, EVYGNEVICA, (740-08-06-1255)

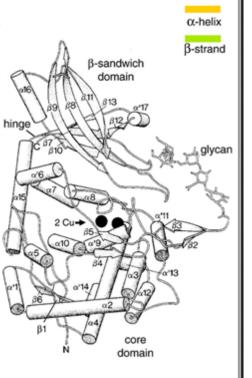
NRIEWNIED. SLOELESLEA.AFLEIENDOT.YHSIAKKOS.POLOOLNONP.ISCYGOMPT.FPHHIRLYYY.YYENALLKYG SUVAFYNDW.TKRIENDFEL.IGATYTNSK.QHHYETNFFH.HOKITHENEI.TTEDHKOLG.HSDYYEQVL.YALBODNCO FEIGHEIEN ALBELLGAUGK.KYSMSHIDYA.REPYFFHH.AKTITHENEI.TTEDHKOLG.UGAFEKEP.YESANCAIGL.MHFDQPFDK SUNNEATXT.HATPHDOFEY.QHSPOXATON.LELNHYSIPO.LOHMLGEEKE.HDEVFAGELL.HNIGTSÄDGH.VEYCEPTGEN TKDCSHEAM.F9LOQTEM.SFYFDRJKL.DITKALKON.YKLQDPDLE.IEITAVNOSH.LDSHVIHSPT.IEFEAGTDSA HTDDOHTEP [FU-d 1256-1664]

VMERKDITCL.DROQUSEVX.ALESKRADKS.SDOPONIASP.EALPPLOPSP.AASKFACCV.EGMATEPONE.RLTTOGPODS LREGGAVVGL.PVMWTLPS.RLPELTVST.HDPETGRD.PNPFIGSHE.PEGENVHKR.DINRDELGS.STKINNFI GALLALGOT.NYCDEFVPT.F.HENGUENV.GKEPYGIGH.LHYASYDELF.YHEGOTDFI.HAINGLQF.RGLGGENAC AVMLEKTPLK.PFSFGAPYNL.NDHTENFSKP.EDEFDYGGPG.YIYOTLEFAG.MSTRGIDHIV.ENNOBLGF.AGHLEGOT SATVDFOVCR.TAGDCEDAGY.FFVLGGEEM.PMAPDRLKY.DITETLDXDN.LHHDEIFGIE.VTITEYDGTV.LDGGLIPFPS (ITTOPAHDI.SSHNISL [FU-e 1655-2081]

NEVERDLSTL.SERDIGSLEY.ALSSLOADTS.ADGFAATASF. GLPAKGEDS.HONEVACTE.GMPTFPHNER.LYTLOFEDAL RENGSSVAVP.YHDWTKFINH.IPHLFTOKET.TUVNRNKVPP.NFPARGYVPS.HDTTTVRUOV.BGEFHLTSTG.EHSALDADL LALGONGOFN.FXVDEVNEN.TIRKLVGOP.YYSLSLOXENKA.STDEIFFINE.SEVDKVMAVN.QALGUEKGUP.SGAALCANSI HTQNSRPFHY.EINGOPTKK.HAVPNDVFXY.ELLOXEYDEL.EIGGOGINEI.EKEIKDKOMN.YRVFAGPLLH.GIRTSADNOF GICKTSEDCH.HOGGIFFLOG.TKEMANYNR.LFKKDITHLA.HDAHITPEDV.FHDSEPFFIK.VEVTAVQDTV.LPASILHAPT IYTEGLINK.EHNENSSKOM (FU-1:022-2031)

HOVEKEINTL. TAREVONLED. AMERAVARDHG. PHOYOAIARF. COMPFMCPHP. DOR YSCCT. GRATPPINER. LYTKOMEDAL TARGARVGLP. YNDGTTATA. LPFFYDEED. NPFHHHHIDY. LGYDTTRSFR. DKLINDPERG. SESPFYRQUL. KALEGYDFCG FROGRIFN. AXISTYGGUT. PYCMSTLEY. TYDPLEMIAN. ANDTDRIKAK. GALGERHGLP. YNHANCEIGA. MKRPLRFFSD PINHNAPTHS. NAK PTDVFEY. SEPRFOYDNL. RFHONTIKKL. EHELEKOKEE. DRTFAAFLLH. GIKKSADVSF. DVCNHDGECH FAGTATIGG. EHEMPMSFDR. LFRYDITGVL. KOMLEYDSD. PTFHMILIET. SGRQLPEDEI. KMPTVEHSPG. GKHEKKHED BIED (FVG 2502-2905)

ILVERGINSE. SHHEAEELRD. ALYKLONDES. BOGYENIAGF. MOYPNLOPEK. GDEKYPCCYMG. MSIFFEN R. LHTIQFERAL KERGSHLGIP. YNDWTOTISS. LEYFFADSON. NNFFYKYHIR. SINGDYVRDV. NRAIFQCTKFG. EFSSFYLA. LQALEMINY DFEVQYELL. NEVSALIQA. EKYSMSTLEY. SAFDYTYMIN MABLOKIWII. MQELQKEVYP. ANAGCAGD. IMMYPLHFPN DFEVQYELL. NEVSALIQA. EKYSMSTLEY. SAFDYTYMIN MABLOKIWII. MQELQKEVYP. ANAGCAGD. IMMYPLHFPN PISYNNODFT. EKNSLFNNY. DESERNYKYD. NLAHCHNIE. KLEVLESLE, LKSVFACHVL. SJIRTTAVV. KVYIKEGTOS DDEYMGEVYI. LGGAREMMA. YERLEMEDT. ETYBNEN TO. DMYKFRFDLK. KYMFTELDASY. LPAPITYER. PNNAVPDIIE



cupredoxin-like domain

IPIGKDVNLP.PKVVVKRGTK.IMFMSVDEAV.TTPMLNLGSY TAMFKCKVPP.PSPHAFELGK.MYSVESGDYF.MTASTTELCN DNNLRIHVHV.D [FU-h 2906-3396]

We do not use InChI to define primary protein structure

Code	Amino acid	Code	Amino acid
А	alanine	I	isoleucine
R	arginine	L	leucine
Ν	asparagine	K	Lysine
D	aspartic acid	М	methionine
В	asparagine or aspartic acid	F	phenylalanine
С	cysteine	Р	proline
Е	glutamicacid	S	serine
Q	glutamine	Т	threonine
Z	glutamine or glutamic acid	W	tryptophan
G	glycine	Y	tyrosine
Н	histidine	V	Valine

Primary protein structure is defined by amino acid single letter code

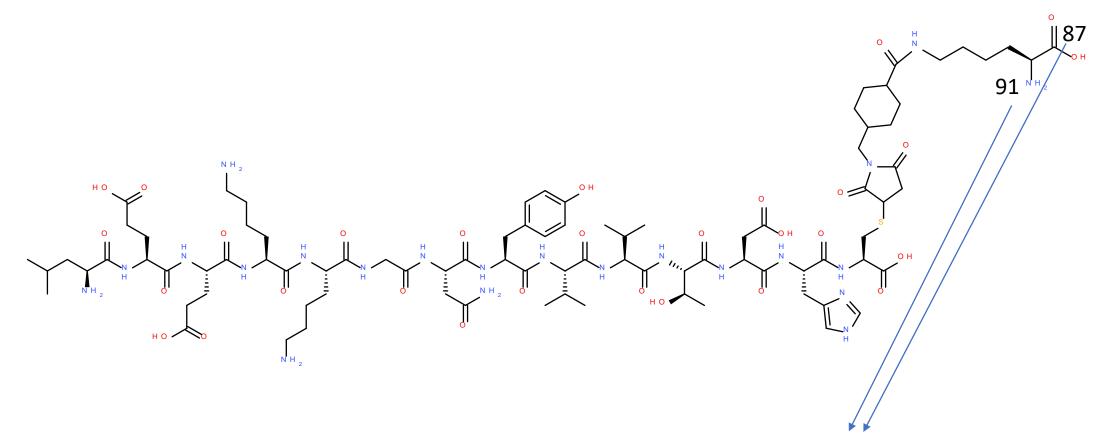
X (wildcard/placeholder)

We use InChI to define modifications to the primary protein structure

- Modified amino acids
- Links
- Prosthetic groups

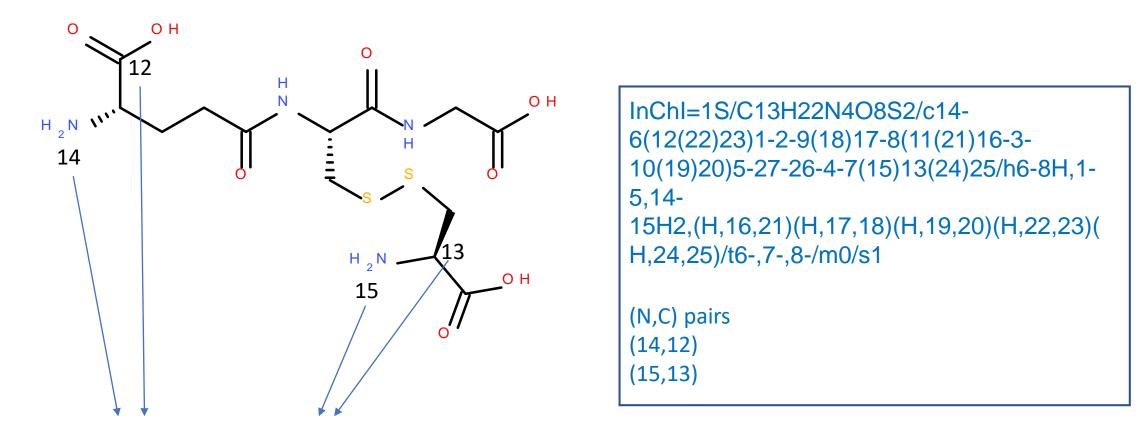
Example: definition of a modified amino acid

Lysine linked to Epidermal Growth Factor Receptor variant III



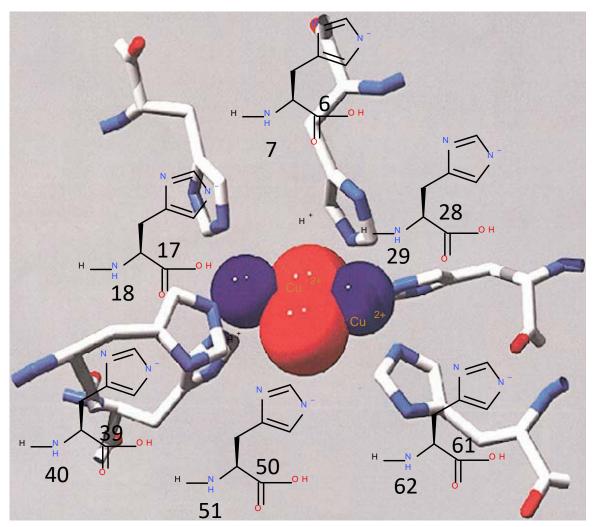
InChI canonical atom numbers for atoms that *substitute* N of amino group and C of carboxyl group of a natural amino acid

Example: definition of a link CYSTEINE-GLUTATHIONE DISULFIDE



InChI 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



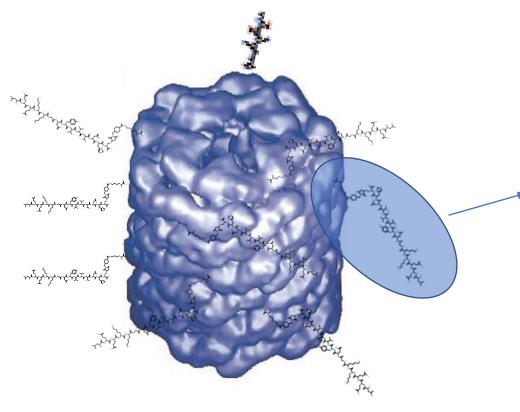
InChI=1S/6C6H9N3O2.2Cu/c6*7-5(6(10)11)1-4-2-8-3-9-4;;/h6*2-3,5H,1,7H2,(H2,8,9,10,11);;/q;;;;;2*+2/p-4/t6*5-;;/m000000../s1

> (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-indexingfiles.cfm
- Substance names in text format https://fdasis.nlm.nih.gov/srs/
- Other product-related information in XML format <u>https://dailymed.nlm.nih.gov/dailymed/</u>
- SPL Substance Implementation Guide is #14 in https://www.fda.gov/media/84201/download

Protein descriptions provided by PubChem are misleading

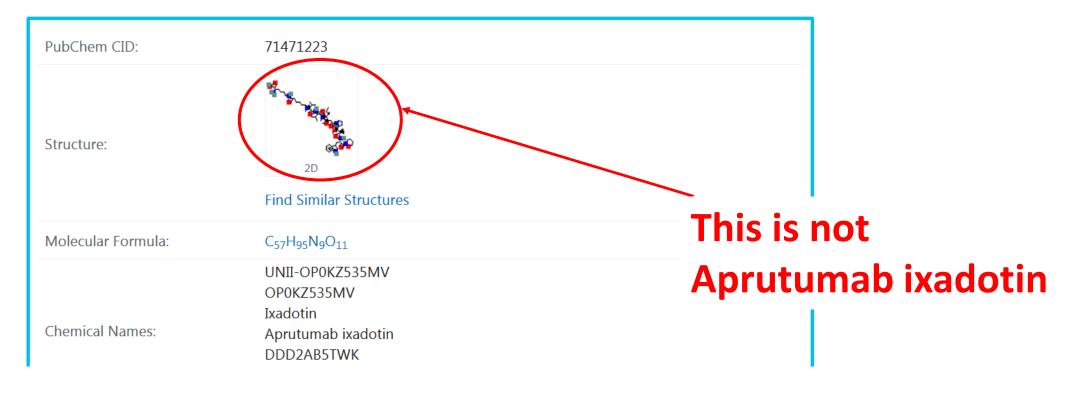


M About

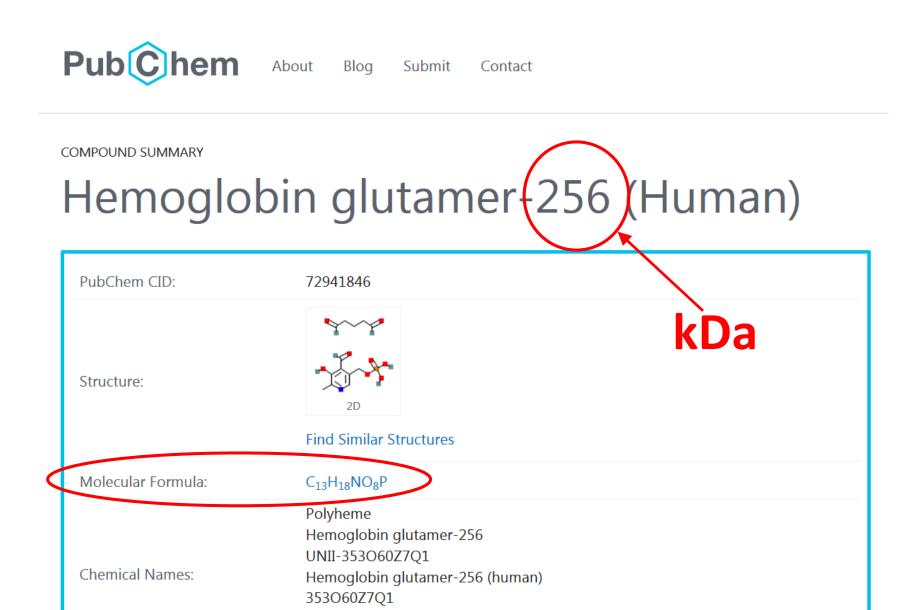
Blog Submit Contact

COMPOUND SUMMARY

Aprutumab ixadotin



Protein descriptions provided by PubChem are misleading



Hemoglobin Glutamer

