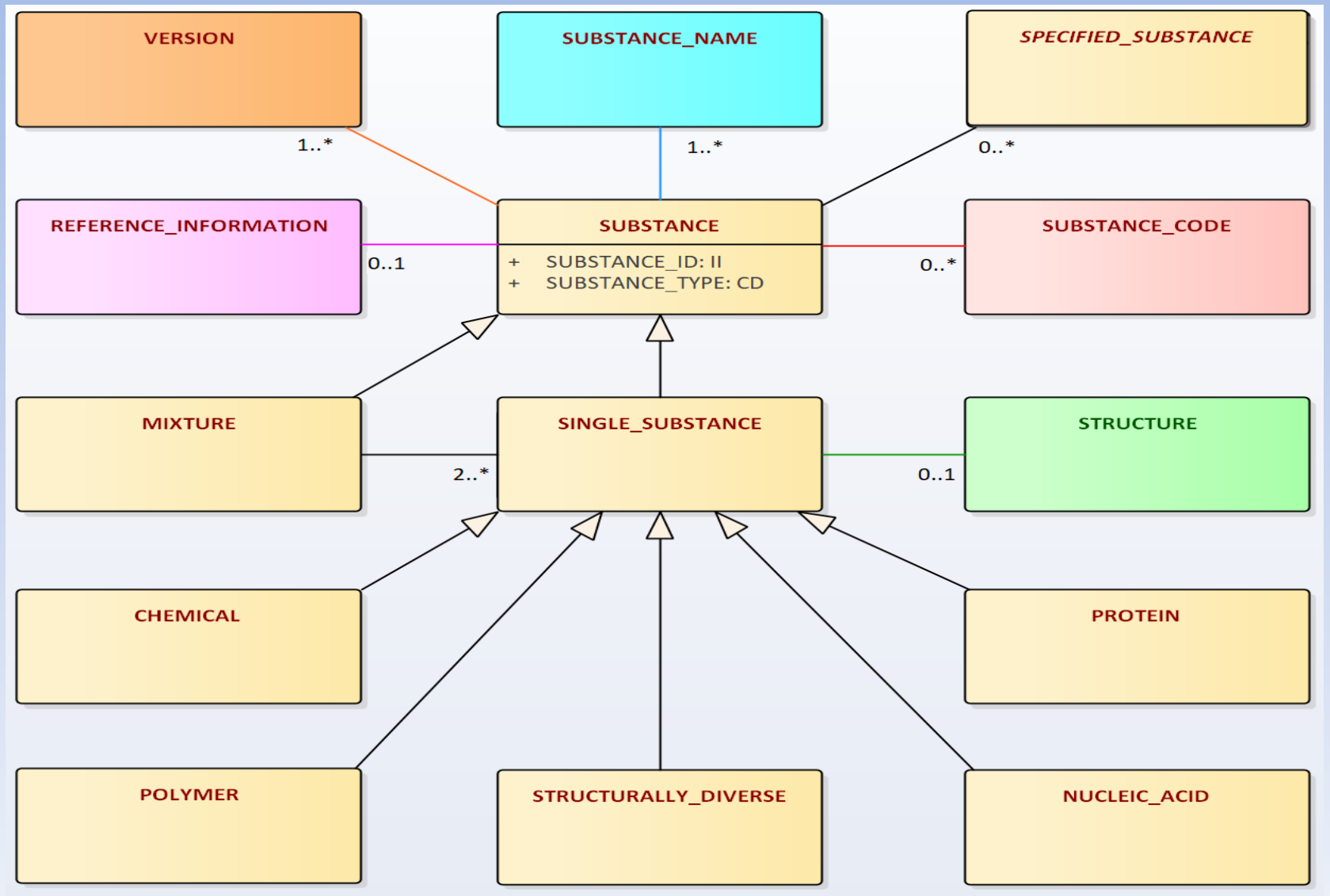
Signature fields (Minimum requirements) for a Protein Substance Registration

GInAS Meeting, 30 October 2019

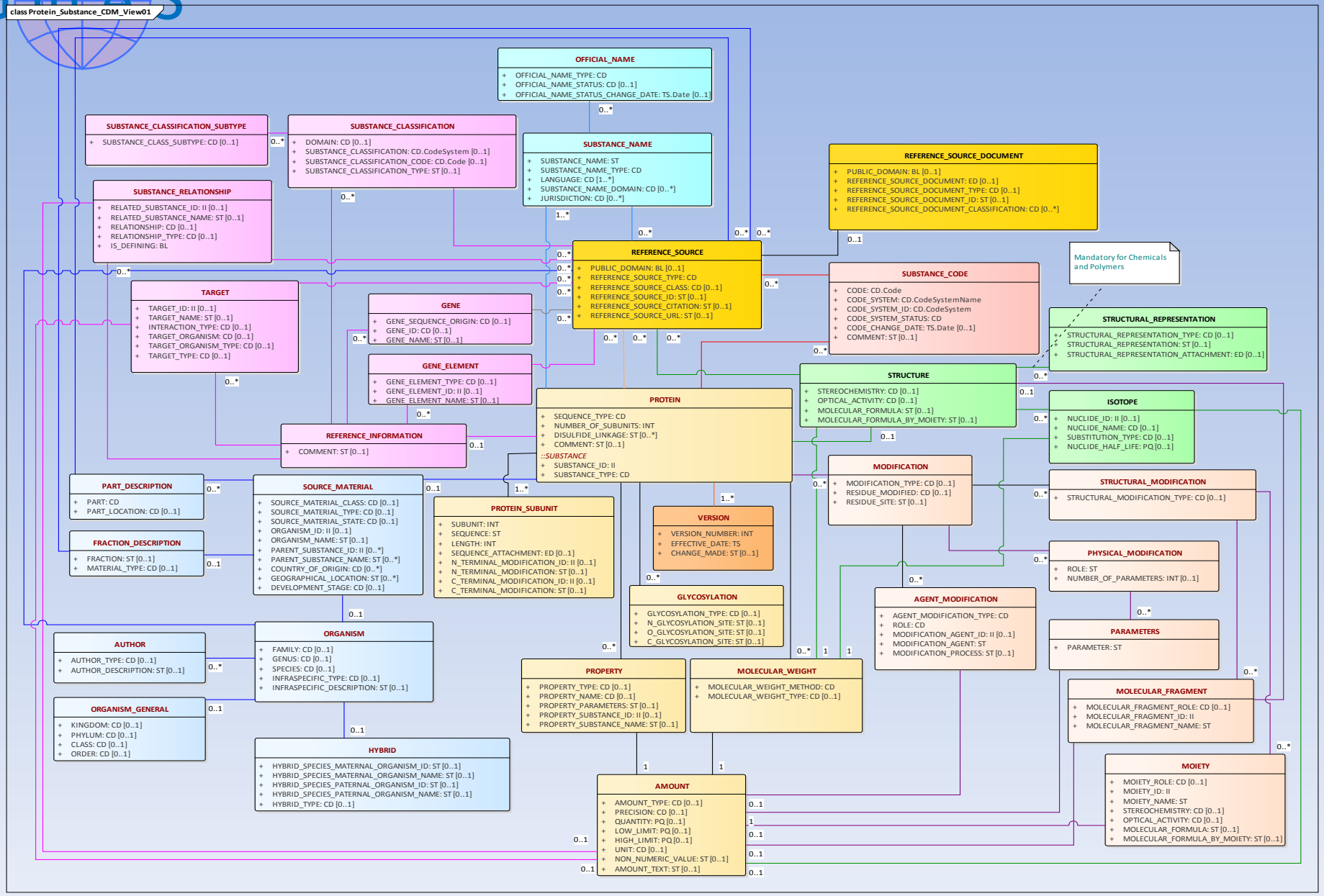
Herman Diederik

(h.diederik@cbg-meb.nl; hdiederik@planet.nl)

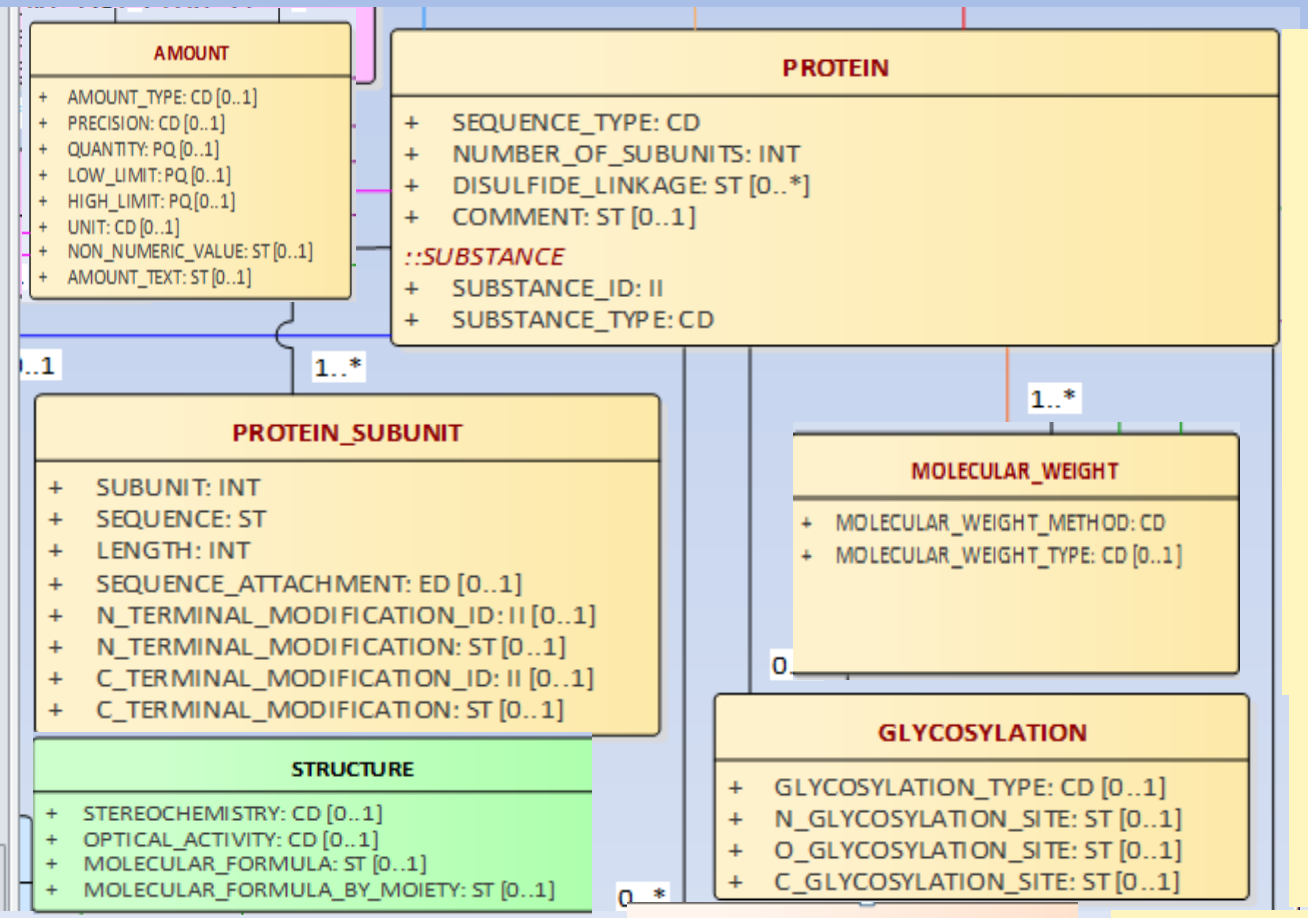
ISO 11238/19844 Substance Classification Overview (high-level)

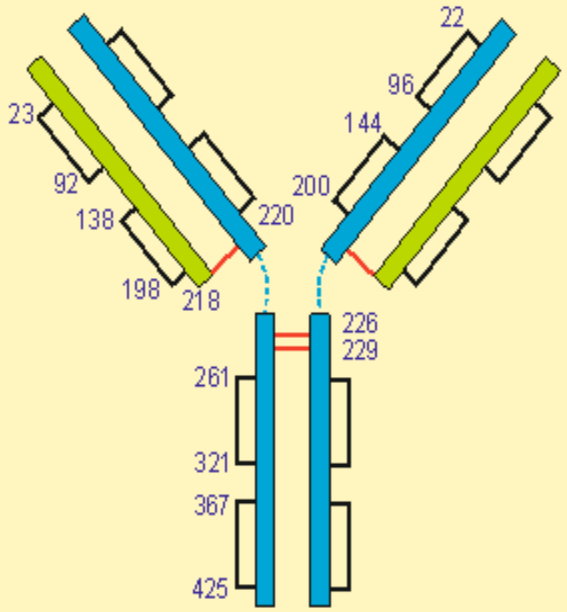


Protein Information model



Characterising elements at the Substance info level for an IGG1

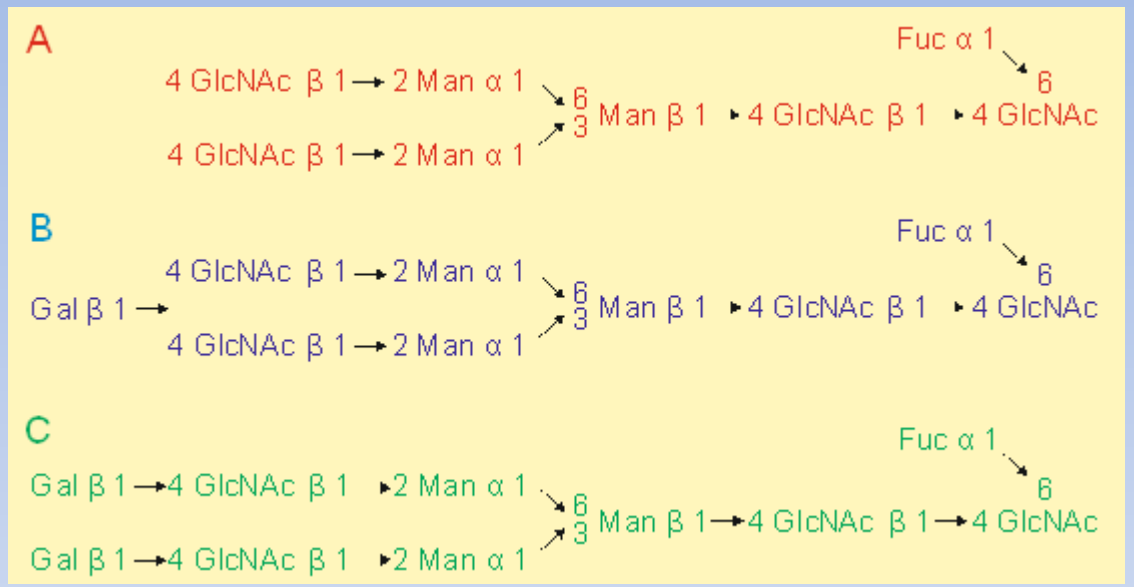


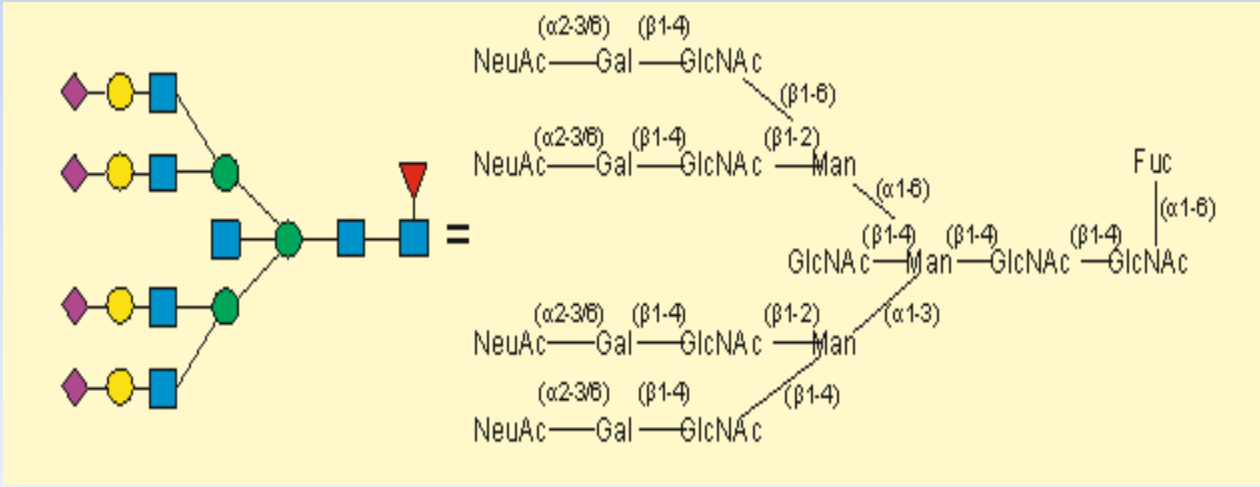


1. Green bars denote light chain
2. Blue bars represent heavy chain
3. Dashed blue line represents hinge region
4. Black lines represent intra-chain disulfide bonds
5. Red lines represent inter-chain disulfibe bonds
6. Numbers indicate cysteine associated with disulfide bonds

Predominant Glycan Information at the Substance level (A)

Characterising elements at the Specified Substance Group 1 info level: Glycan Information: Structural representation of common Glycans

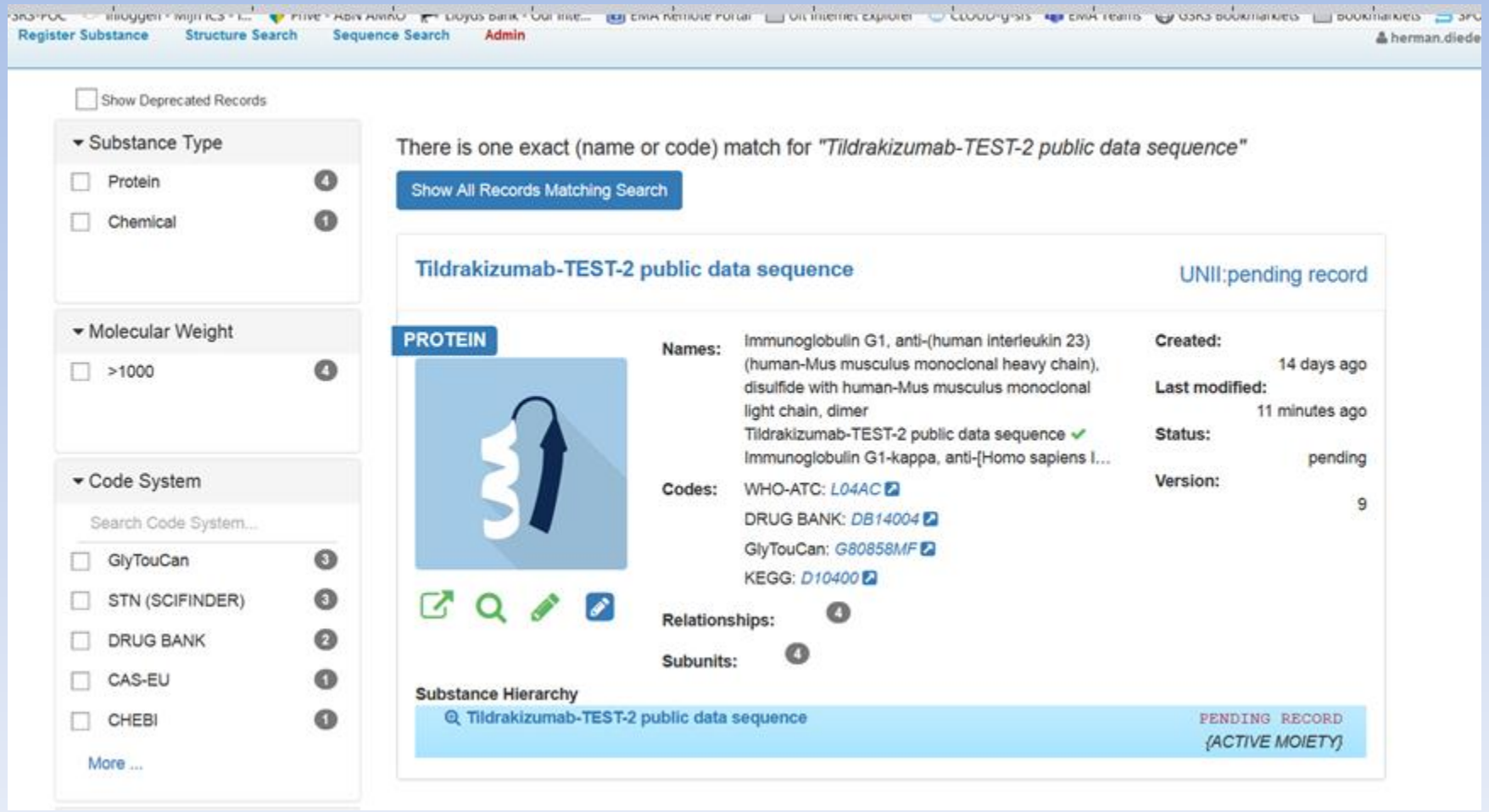




- N-Acetylneuraminic acid (NANA)
- Galactose (Gall)
- N-Acetylglucosamine (GlcNAc)
- Mannose (Man)
- Fucose (Fuc)

Minimum fields based on Public data Example Tildrakizumab

- Record Overview prepared in NCATS 2.3.3. version Tildrakizumab-TEST-2-public data sequence



Overview and Names/ Codes

- Definition information:
  Public data Sources: INN List 108 **WHO Drug Information, Vol. 26 , No. 4, 2012 Proposed INN: List 108**
  Immunoglobulin G1-kappa, anti-[Homo sapiens IL23A (interleukin 23 alpha subunit (p19), IL-23A)], humanized monoclonal antibody; gamma1 heavy chain (1-446) [humanized VH (Homo sapiensIGHV1-18*01 (81.60%) -(IGHD)-IGHJ4*01)) [8.8.9] (1-116) -Homo sapiens IGHG1*01 (CH1 (117-214, hinge (215-229), CH2 (230-339), CH3 (340-444), CHS (445-446)) (117-446)], (219-214')-disulfide with kappa light chain (1'-214') [humanized V-KAPPA (Homo sapiensIGKV1-39*01 (85.30%) -IGKJ1*01) [6.3.9] (1'-107') -Homo sapiens IGKC*01 (108'-214')]; dimer (225-225":228-228")-bisdisulfide immunomodulator
- Other Sources: WHO-ATC: *L04AC* ; **DRUG BANK: *DB14004*** ; **GlyTouCan: *G80858MF*** ; **KEGG: *D10400***

  Web: https://newdrugapprovals.org/2018/06/01/tildrakizumab-asmn/

Detailed Overview

Tildrakizumab-TEST-2 public data sequence

UNII:pending record

PROTEIN

Names: Immunoglobulin G1, anti-(human interleukin 23) (... Tildrakizumab-TEST-2 public data sequence ✔ Immunoglobulin G1-kappa, anti-[Homo sapiens IL23A (interleukin 23 alpha subunit (p19), IL-23A)], humanized monoclonal antibody; gamma1 heavy chain (1-446) [humanized VH (Homo sapiensIGHV1-18*01 (81.60%) -(IGHD)-IGHJ4*01)) [8.8.9] (1-116) -Homo sapiens IGHG1*01 (CH1 (117-214, hinge (215-229), CH2 (230-339), CH3 (340-444), CHS (445-446)) (117-446)], (219-214')-disulfide with kappa light chain (1'-214') [humanized V-KAPPA (Homo sapiensIGKV1-39*01 (85.30%) -IGKJ1*01) [6.3.9] (1'-107') -Homo sapiensIGKC*01 (108'-214')]; dimer (225-225":228-228")-bisdisulfide immunomodulator

Codes: WHO-ATC: *L04AC*

DRUG BANK: *DB14004*

GlyTouCan: *G80858MF*

KEGG: *D10400*

Relationships: 4

Created: 14 days ago

Last modified: 15 minutes ago

Status: pending

Version: 9

Walk through of the Addition Public Information, besides the INN Definitional information (1)

HOME » ANTIBODIES » TILDRAKIZUMAB-ASMN

Tildrakizumab-asmn

Heavy chain:

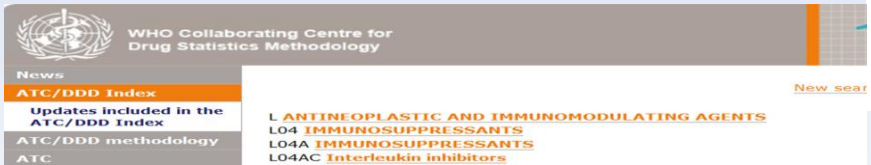
QVQLVQSGAEVKKPGASVKVSCKASGYIFITYWMTWVRQAP

EWMGQIFPASGSADYNEKFEGRVTMTTDTSTSTAYMELRSL

TAVYYCARGGGGFAYWGQGTLVTVSSASTKGPSVFPLAPSS

GGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS

WHO-ATC Code: https://www.whocc.no/atc_ddd_index/?code=L04AC



Immunoglobulin G1, anti-(human interleukin 23) (human-Mus musculus monoclonal heavy chain), disulfide with human-Mus musculus monoclonal light chain, dimer

CAS 1326244-10-3, **BLA 761067**

Tildrakizumab (SCH 900222/MK-3222)

ILUMYA; MK-3222; SCH-900222; SUNPG 1622; SUNPG 1622 I; SUNPG 1623 I; SUNPG 1623 II; SUNPG 1623 III; SUNPG 1623 IV; SUNPG1623; Tildrakizumab-asmn

DRUG BANK https://www.drugbank.ca/drugs/DB14004

Company Sun Pharmaceuticals

Approval Status **FDA** Approved March 2018 FOR Psoriasis, plaque

Treatments plaque psoriasis

Protein chemical formula $C_{6426}H_{9918}N_{1698}O_{2000}S_{46}$

Protein average weight144400.0 DaSequences

Walk through of the Addition Public Information, besides the INN Definitional information (2)

https://www.kegg.jp/dbget-bin/www_bget?dr:D10400

KEGG

DRUG: Tildrakizumab

Help

Entry	D10400 Drug
Name	Tildrakizumab (USAN); Tildrakizumab (genetical recombination) (JAN); Tildrakizumab-asmn; Ilumya (TN)
Product	ILUMYA (Sun Pharmaceutical Industries)
Formula	C6426H9918N1698O2000S46
Exact mass	144345.3728
Mol weight	144434.6737
Sequence Type	Peptide
Class	Other DG02019 Interleukin inhibitor
Remark	ATC code: L04AC17 Product: D10400<US>
Efficacy Disease	Antipsoriatic, Immunosuppressant, Anti-IL-23 antibody Plaque psoriasis [DS:H01656]
Comment	Monoclonal antibody Treatment of immunologically mediated inflammatory disorders
Target Pathway	IL23A [HSA:51561] [KO:K05426] hsa04060 Cytokine-cytokine receptor interaction
Brite	Anatomical Therapeutic Chemical (ATC) classification [BR:br08303] L ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS L04 IMMUNOSUPPRESSANTS L04A IMMUNOSUPPRESSANTS L04AC Interleukin inhibitors L04AC17 Tildrakizumab D10400 Tildrakizumab (USAN) <US> USP drug classification [BR:br08302] Immunological Agents Immunomodulators Tildrakizumab-asmn D10400 Tildrakizumab (USAN)

Walk through of the Addition Public Information, besides the INN Definitional information (3)

https://www.genome.jp/dbget-bin/www_bget?hsa:51561+D06556+D09214+D09588+D1

Search for details !!

Kegg Homo sapiens (human): 51561

interleukin 23 subunit alpha **Information Interleukin inhibitor**

https://www.genome.jp/dbget-bin/www_bget?hsa:51561+D06556+D09214+D09588+D1

Sequences = Target

```
>Tildrakizumab Sequence
MLGSRAVMLLLLLPWTAQGRAVPGGSSPAWTQCQQLSQKLCTLAWSAHPLVGHMDLREEG
DEETTNDVPHIQCGDGCDPQGLRDNSQFCLQRIHQGLIFYEKLLGSDIFTGEPSLLPDSP
VGQLHASLLGLSQLLQPEGHHWETQQIPSLSPSQPWQRLLLRFKILRSLQAFVAVAARVF
AHGAATLSP
```

AA seq 189 aa [AA seq] [DB search]

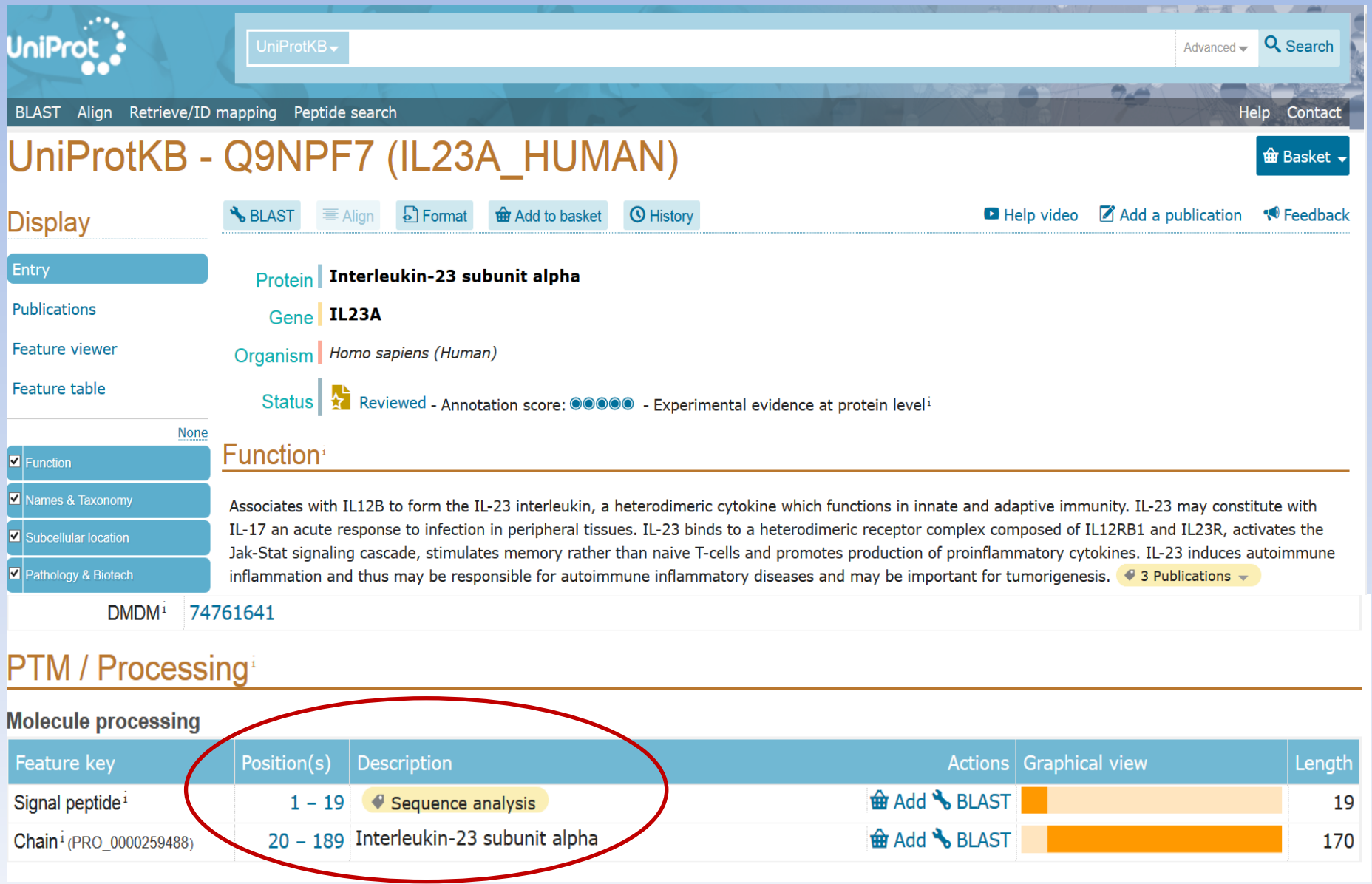
```
MLGSRAVMLLLLLPWTAQGRAVPGGSSPAWTQCQQLSQKLCTLAWSAHPLVGHMDLREEG
DEETTNDVPHIQCGDGCDPQGLRDNSQFCLQRIHQGLIFYEKLLGSDIFTGEPSLLPDSP
VGQLHASLLGLSQLLQPEGHHWETQQIPSLSPSQPWQRLLLRFKILRSLQAFVAVAARVF
AHGAATLSP
```

NT seq 570 nt [NT seq] 0

```
atgctggggagcagagctgtaatgctgctgttgctgctgccctggacagctcagggcaga
gctgtgcctgggggcagcagccctgcctggactcagtgccagcagctttcacagaagctc
tgcacactggcctggagtgcacatccactagtgggacacatggatctaagagaagaggga
gatgaagagactacaaatgatgttccccatatccagtgtggagatggctgtgaccccaa
ggactcagggacaacagtcagttctgcttgcaaaggatccaccagggtctgatttttat
gagaagctgctaggatcggatattttcacaggggagccttctctgctccctgatagccct
gtgggccagcttcatgcctccctactgggcctcagccaactcctgcagcctgagggtcac
cactgggagactcagcagattccaagcctcagtcccagccagccatggcagcgtctcctt
ctccgcttcaaaatccttcgcagcctccaggcctttgtggctgtagccgcccgggtcttt
gcccatggagcagcaaccctgagtccctaa
```

Walk through of the Addition Public Information, besides the INN Definitional information (4)

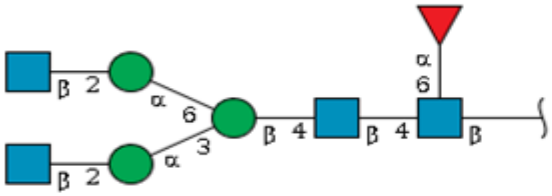
https://www.uniprot.org/uniprot/Q9NPF7#ptm_processing

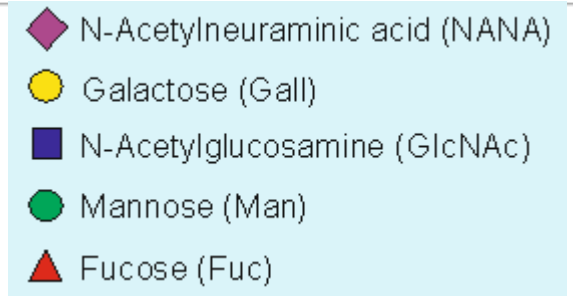


MAB, Predominant Glycans

- **Characterising elements at the Specified Substance Group 1 info level:**
  1) Structural representation of Glycans in the GlyTouCan data base Accession Number
  2) Display of GlyTouCan Accession Number G80858MF Fragment, IUPAC Extended Name and Calculated Monoisotopic Mass (1462.54)
  https://glytoucan.org/Structures/Glycans/G80858MF

G80858MF





Accession number	G80858MF
Created Date	2014-07-27

Accession number: G80858MF Calculated Monoisotopic Mass : 1462.5444

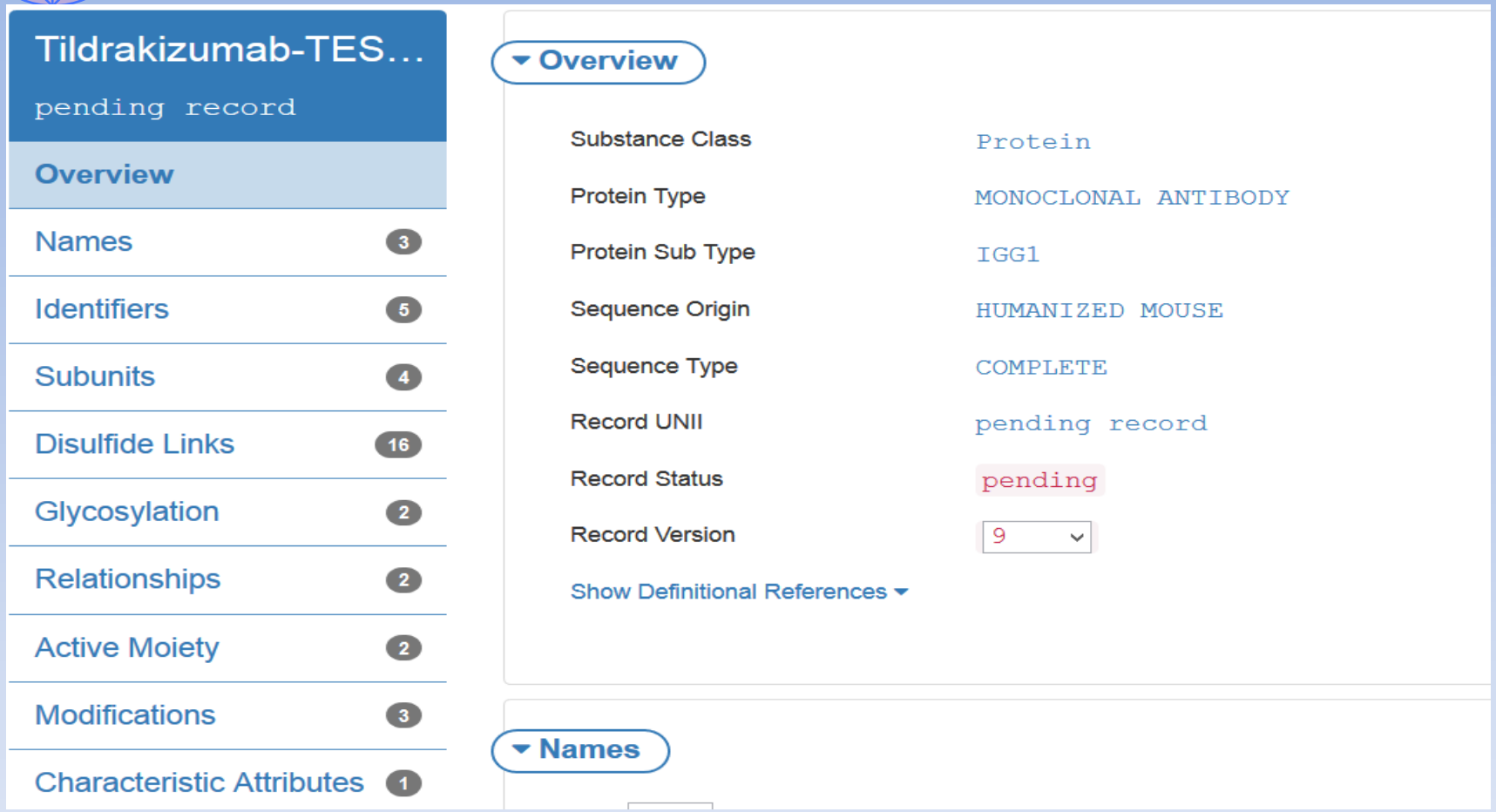
IUPAC Extended:
beta-D-GlcpNAc-(1->2)-alpha-D-Manp-(1->3)[beta-D-GlcpNAc-(1->2)-alpha-D-Manp-(1->6)]-beta-D-Manp-(1->4)-beta-D-GlcpNAc-(1->4)[alpha-L-Fucp-(1->6)]-beta-D-GlcpNAc(1->

Signature Fields

- What is to be considered as Minimal Definitional Information (=Signature fields) for a protein enclosed in the complete Record, which is a subset (view) of the complete record.

- This information could be used to be the basic information on which a Global ID is created

Definitional Facets selected on the Left side



Overall Info: Subst. Class/ Protein Type/ Protein Subtype/ Sequence Origin/ Sequence Type

Facet information: Names, ID/ Codes, Subunits, Disulfide Links, Glycosylation, Modifications, Minimum relationship as Active Moiety/ Basis of Strength/ Reference Strength and Some properties like Molecular Weight (Molecular Formula)

Name & Identifiers (Codes)

Names

Show 5 entries

Search ...

Name	Type	Language	References
Immunoglobulin G1, anti-(human interleukin 23) (human-Mus musculus monoclonal heavy chain), disulfide with human-Mus musculus monoclonal light chain, dimer ✔	Preferred Name	English	view
Tildrakizumab-TEST-2 public data sequence	Common Name	English	view
Immunoglobulin G1-kappa, anti-[Homo sapiens IL23A (interleukin 23 alpha subunit (p19), IL-23A)], humanized monoclonal antibody; gamma1 heavy chain (1-446) [humanized VH (Homo sapiensIGHV1-18*01 (81.60%) -(IGHD)-IGHJ4*01)) [8.8.9] (1-116) -Homo sapiens IG ✔	Preferred Name	English	view

Identifiers

Show 5 entries

Search ...

Code System	Code	Type	Description	References
STN (SCIFINDER)	1326244-10-3	PRIMARY		view
GlyTouCan	G80858MF	PRIMARY		hide

Source:

Index	Source Text / Citation	Source Type	Tags	Document	Date Accessed
4	Glycon G0F	GlyTouCan			

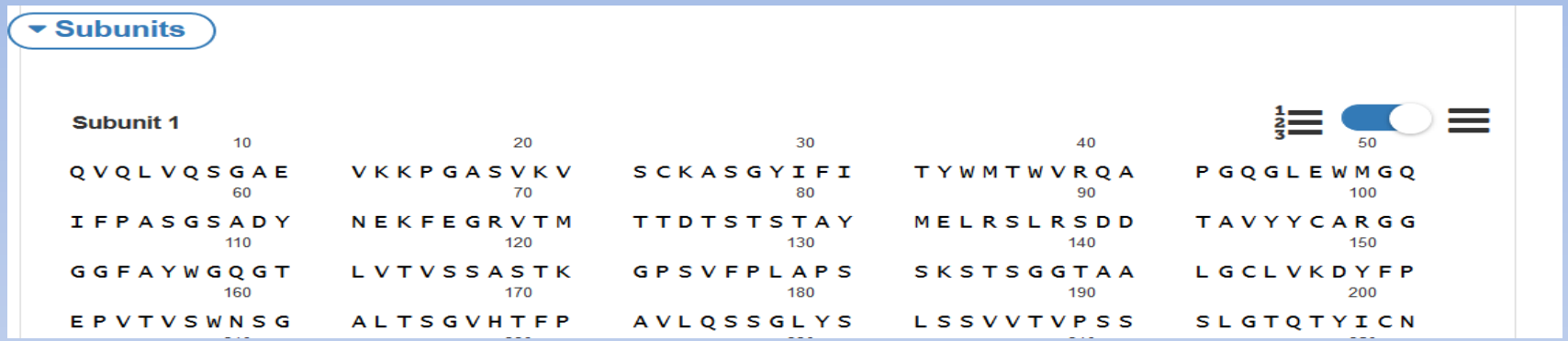
Code System	Code	Type	Description	References
DRUG BANK	DB14004	PRIMARY		hide

Source:

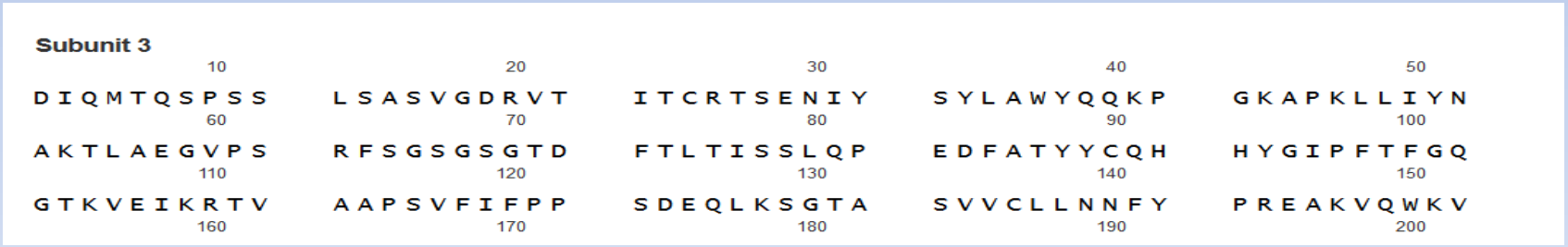
Index	Source Text / Citation	Source Type	Tags	Document	Date Accessed

Subunit Sequence and disulfide bonds

Heavy Chain (Partial representation)



Light Chain (Partial representation)



Disulfide Links

From	To
1_22	1_96
1_143	1_199
1_260	1_320
1_366	1_424
1_219	3_214

Partial representation of the Disulfide bonds

Glycosylation & Modification

Glycosylation Type and Side

Glycosylation

Glycosylation Type	MAMMALIAN

Glycosylation Link Type	Site
N	1_296
N	2_296

Modification Type and Side

Modifications

Structural Modifications

Modification Type	Location Site	Location Type	Residue Modified	Extent	Modification Name	Modification ID
AMINO_ACID_REMOVAL	[1_446] [2_446]	C-TERMINUS			LYSINE	K3Z4F929H6
N-GLYCAN FORMATION		RESIDUE_SPECIFIC	asparagine	**Amount:** DEGREE OF SUBSTITUTION [redacted] *Degree of N-Glycan Occupation*	Glycan-G80858MF-A2G0F	3a617623
AMINO_ACID_SUBSTITUTION	[1_1] [2_1]	N-TERMINUS			PIDOLIC ACID	SZB83O1W42

Degree of substitution is not Public data

Properties & Relationships

Characteristic Attribute (Property like Mol. Weight)

Characteristic Attributes

Properties

Name	Property Type	Amount	Referenced Substance	Defining	Parameters	References
MOL_WEIGHT: WEIGHT AVERAGE	CHEMICAL	EXACT 144465.04 Mol/g (average)				0

Relationships (Active Moiety/ Basis or strength/ Reference strength

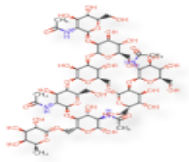
Active Moiety

Related Record	Type	Details	References
 Tildrakizumab-TEST-2 public data sequence	ACTIVE MOIETY REFERENCE STRENGTH	show	none
 Tildrakizumab-TEST-2 public data sequence	ACTIVE MOIETY	show	none

Other Relationships

Glycan -> Parent; Target -> Inhibitor and References

Relationships

Related Record	Type	Details	References
 Glycan-G80858MF-A2G0F	N-GLYCAN->PARENT	show	view 1 reference(s)
08TF01H1UT  INTERLEUKIN-2 RECEPTOR SUBUNIT ALPHA	TARGET -> INHIBITOR	show	hide 3 reference(s)

Index	Source Text / Citation	Source Type	Tags	Document	Date Accessed
2	Q9NPF7 (IL23A_HUMAN) Interleukin-23 subunit alpha	UNIPROT			
7	Kegg Tildrakizumab Substance	KEGG	PUBLIC_DOMAIN_RELEASE		
10	Target Reference Seqience information Interleukine Inhibitor	KEGG	PUBLIC_DOMAIN_RELEASE		