

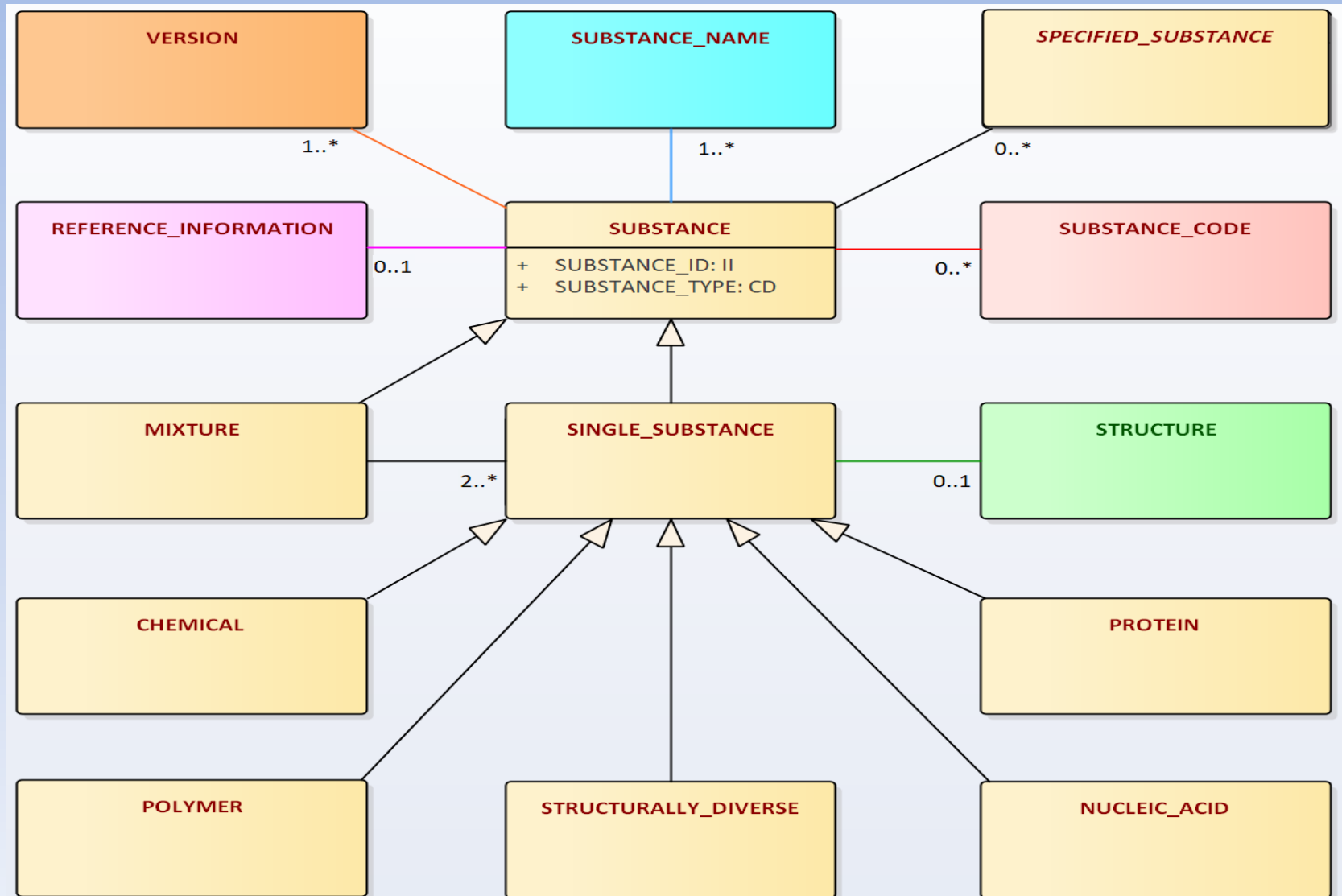


Signature fields (Minimum requirements) for a Protein Substance Registration

GInAS Meeting, 30 October 2019
Herman Diederik

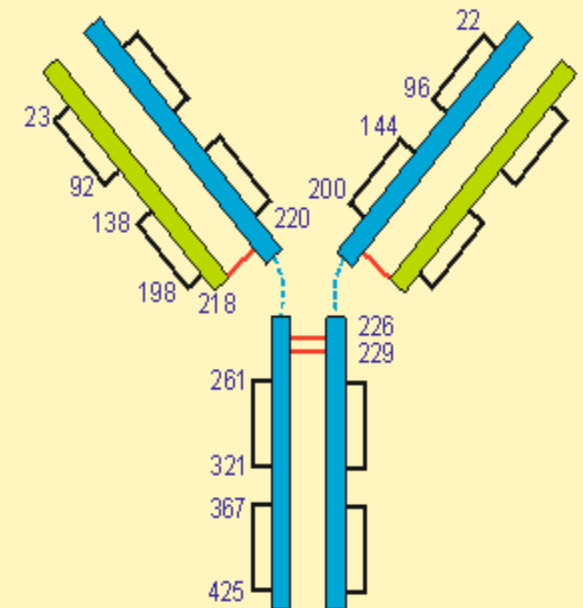
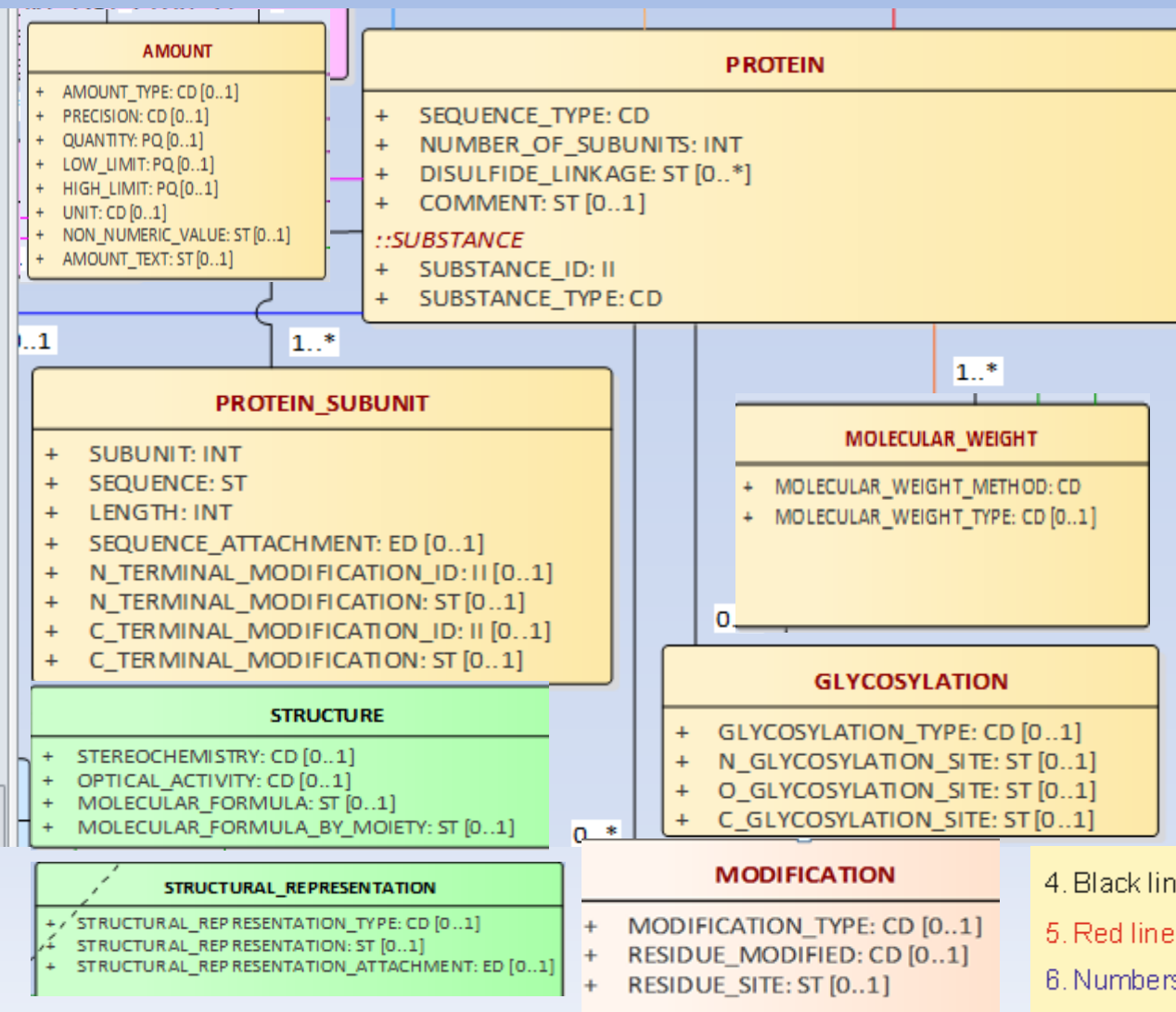
(h.diederik@cbg-meb.nl; hdiederik@planet.nl)

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





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 disulfide bonds
6. Numbers indicate cysteine associated with disulfide bonds

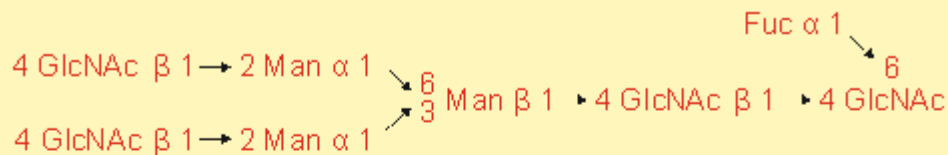


Predominant Glycan Information at the Substance level (A)

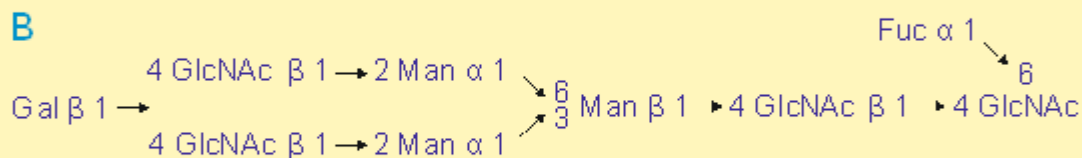
Characterising elements at the Specified Substance Group 1 info level:

Glycan Information: Structural representation of common Glycans

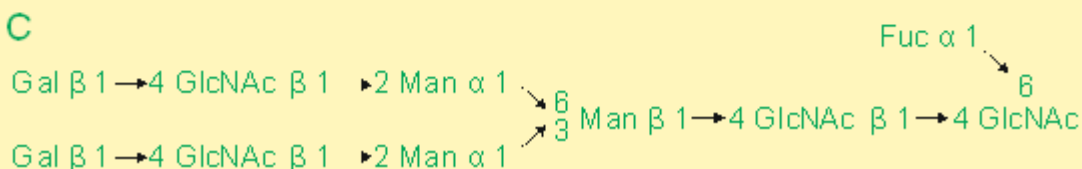
A








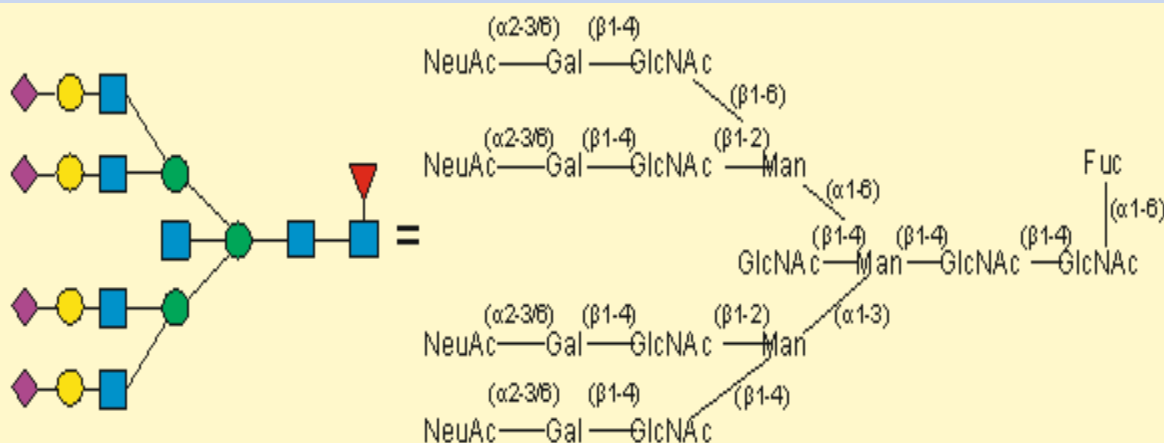
B



C



-  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

Register Substance Structure Search Sequence Search Admin

☐ Show Depreciated Records

▼ Substance Type

☐ Protein 4

☐ Chemical 1

▼ Molecular Weight

☐ >1000 4

▼ Code System

Search Code System...

☐ GlyYouCan 3

☐ STN (SCIFINDER) 3

☐ DRUG BANK 2

☐ CAS-EU 1

☐ CHEBI 1


More ...

There is one exact (name or code) match for "Tildrakizumab-TEST-2 public data sequence"

Show All Records Matching Search

Tildrakizumab-TEST-2 public data sequence UNII:pending record

PROTEIN



Names: Immunoglobulin G1, anti-(human Interleukin 23)
(human-Mus musculus monoclonal heavy chain),
disulfide with human-Mus musculus monoclonal
light chain, dimer
Tildrakizumab-TEST-2 public data sequence ✓
Immunoglobulin G1-kappa, anti-(Homo sapiens I...

Codes: WHO-ATC: L04AC
DRUG BANK: DB14004
GlyYouCan: G80858MF
KEGG: D10400

Relationships: 4

Subunits: 4

Created: 14 days ago

Last modified: 11 minutes ago

Status: pending

Version: 9

Substance Hierarchy

Tildrakizumab-TEST-2 public data sequence PENDING RECORD (ACTIVE MOIETY)

- **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/>

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

G In A S Walk through of the Addition Public Information, besides the INN Definitional information (1)



HOME » ANTIBODIES » TILDRAKIZUMAB-ASMN

Tildrakizumab-asmn

Heavy
chain:

```

QVQLVQSGAEVKKPGASVKVSCKASGYIFITYWMTWVRQAE
EWMGQIFPASGSADYNEKFEGRVTTTDTSTSTAYMELRSI
TAVYYCARGGGGFAYWGQGLVTVSSASTKGPSVFPLAPSS
GGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSS
    
```

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 weight 144400.0 Da Sequences

WHO-ATC Code:

https://www.whocc.no/atc_ddd_in dex/?code=L04AC

WHO Collaborating Centre for Drug Statistics Methodology

News

ATC/DDD Index

Updates included in the ATC/DDD Index

ATC/DDD methodology

ATC

L ANTINEOPLASTIC AND IMMUNOMODULATING AGENTS


L04 IMMUNOSUPPRESSANTS

L04A IMMUNOSUPPRESSANTS

L04AC Interleukin inhibitors

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

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

| | |
|--|--|
|  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

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

```
MLGSRVAMLLLLLPWTAQGRAVPGGSSPAWTQCQQLSQKLCTLAWSAHPLVGHMDLREEG
DEETTNDVPHIQCGDGDQPQLRDN SQFCLQRIHQGLIFYEKL LGSDIFTGEP SLLPDSP
VGQLHASLLGLSQLLQPEGHHWETQQIPSLSPSQPWQRLLLRFKILRSLQAFVAVAAARVF
AHGAATLSP
```

| | | | |
|--------|---|--------|-----------|
| AA seq | 189 aa | AA seq | DB search |
| | MLGSRVAMLLLLLPWTAQGRAVPGGSSPAWTQCQQLSQKLCTLAWSAHPLVGHMDLREEG DEETTNDVPHIQCGDGDQPQLRDN SQFCLQRIHQGLIFYEKL LGSDIFTGEP SLLPDSP VGQLHASLLGLSQLLQPEGHHWETQQIPSLSPSQPWQRLLLRFKILRSLQAFVAVAAARVF AHGAATLSP | | |
| NT seq | 570 nt | NT seq | 0 |
| | atgctggggagcagagctgtaatgctgctgttgctgctgccctggacagctcagggcaga gctgtgcctgggggcagcagccctgcctggactcagtgccagcagctttcacagaagctc tgcacactggcctggagtgcacatccactagtgggacacatggatctaagagaagagga gatgaagagactacaaatgatgtttcccatatccagtgtggagatggctgtgaccccaa ggactcagggacaacagtcagttctgcttgcaaaggatccaccagggctcgatttttat gagaagctgctaggatcggatattttcacaggggagccttctctgctccctgatagccct gtgggccagcttcattgcctccctactgggcctcagccaactcctgcagcctgagggcac cactgggagactcagcagattccaagcctcagtcgccagccagccatggcagcgtctcctt ctccgcttcaaaatccttcgcagcctccaggcctttgtggctgtagccgccgggtcttt gcccatggagcagcaaccctgagtccttaa | | |



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

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

UniProtKB

Advanced

Search

BLAST Align Retrieve/ID mapping Peptide search Help Contact

UniProtKB - Q9NPF7 (IL23A_HUMAN)

Display

Entry

Publications

Feature viewer

Feature table

Protein

Gene

Organism

Status

Interleukin-23 subunit alpha

IL23A

Homo sapiens (Human)

Reviewed - Annotation score: ●●●●● - Experimental evidence at protein levelⁱ

None

☒ Function

☒ Names & Taxonomy

☒ Subcellular location

☒ Pathology & Biotech

Functionⁱ

Associates with IL12B to form the IL-23 interleukin, a heterodimeric cytokine which functions in innate and adaptive immunity. IL-23 may constitute with IL-17 an acute response to infection in peripheral tissues. IL-23 binds to a heterodimeric receptor complex composed of IL12RB1 and IL23R, activates the Jak-Stat signaling cascade, stimulates memory rather than naive T-cells and promotes production of proinflammatory cytokines. IL-23 induces autoimmune inflammation and thus may be responsible for autoimmune inflammatory diseases and may be important for tumorigenesis.

DMDMⁱ 74761641

PTM / Processingⁱ

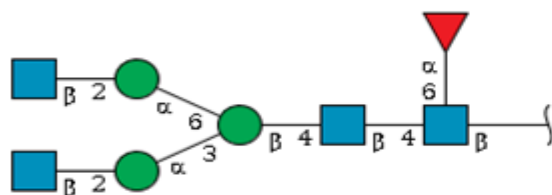
Molecule processing

| Feature key | Position(s) | Description | Actions | Graphical view | Length |
|-------------------------------------|-------------|------------------------------|------------|----------------|--------|
| Signal peptide ⁱ | 1 – 19 | Sequence analysis | Add BLAST | | 19 |
| Chain ⁱ (PRO_0000259488) | 20 – 189 | Interleukin-23 subunit alpha | Add BLAST | | 170 |

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



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

| | |
|------------------|------------|
| Accession number | G80858MF |
| Created Date | 2014-07-27 |

Accession number: G80858MF

Calculated Monoisotopic Mass : 1462.5444

IUPAC Extended:

beta-D-GlcNAc-(1->2)-alpha-D-Manp-(1->3)[beta-D-GlcNAc-(1->2)-alpha-D-Manp-(1->6)]-beta-D-Manp-(1->4)-beta-D-GlcNAc-(1->4)[alpha-L-Fucp-(1->6)]-beta-D-GlcNAc(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

Tildrakizumab-TES...

pending record

Overview

Names 3

Identifiers 5

Subunits 4

Disulfide Links 16

Glycosylation 2

Relationships 2

Active Moiety 2

Modifications 3

Characteristic Attributes 1

Overview

Substance Class

Protein

Protein Type

MONOCLONAL ANTIBODY

Protein Sub Type

IGG1

Sequence Origin

HUMANIZED MOUSE

Sequence Type

COMPLETE

Record UNII

pending record

Record Status

pending

Record Version

9

Show Definitional References ▼

Names

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 entries

| 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 entries

| 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 | |  | |
| Source: | | | | | |
| 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)

▼ Subunits

Subunit 1

| | | | | |
|------------|------------|------------|------------|------------|
| 10 | 20 | 30 | 40 | 50 |
| QVQLVQSGAE | VKKPGASVKV | SCKASGYIFI | TYWMTWVRQA | PGQGLEWMGQ |
| 60 | 70 | 80 | 90 | 100 |
| IFPASGSADY | NEKFEGRVTM | TTDTSTSTAY | MELRSLRSD | TAVYYCARGG |
| 110 | 120 | 130 | 140 | 150 |
| GGFAYWGQGT | LVTVSSASTK | GPSVFPLAPS | SKSTSGGTAA | LGCLVKDYFP |
| 160 | 170 | 180 | 190 | 200 |
| EPVTVSWNSG | ALTSGVHTFP | AVLQSSGLYS | LSSVVTVPSS | SLGTQTYICN |

Light Chain (Partial representation)

Subunit 3

| | | | | |
|------------|-------------|------------|------------|------------|
| 10 | 20 | 30 | 40 | 50 |
| DIQMTQSPSS | LSASVGDRVT | ITCRTSENIY | SYLAWYQQKP | GKAPKLLIYN |
| 60 | 70 | 80 | 90 | 100 |
| AKTLAEGVPS | RFSGSGSGTD | FTLTISSLQP | EDFATYYCQH | HYGIPFTFGQ |
| 110 | 120 | 130 | 140 | 150 |
| GTKVEIKRTV | AAPSVFI FPP | SDEQLKSGTA | SVVCLLNIFY | PREAKVQWKV |
| 160 | 170 | 180 | 190 | 200 |

▼ 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  <i>Degree of N-Glycan Occupation</i> | 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)

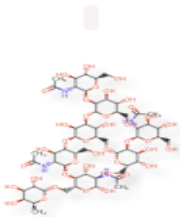

| ▼ Characteristic Attributes | | | | | | |
|----------------------------------|---------------|------------------------------------|----------------------|----------|------------|------------|
| ▼ Properties | | | | | | |
| Name | Property Type | Amount | Referenced Substance | Defining | Parameters | 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 | | |
|  <p>Glycan-G80858MF-A2G0F</p> | N-GLYCAN->PARENT | | show | view 1 reference(s) | | |
|  <p>INTERLEUKIN-2 RECEPTOR SUBUNIT ALPHA</p> | 08TF01H1UT 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 Sequence information Interleukine Inhibitor | KEGG | PUBLIC_DOMAIN_RELEASE | | | |