Visualization Tools for Biomedical Knowledge

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Outline

- Issues and Challenges
  - **SemNav** (UMLS Semantic Navigator)
    *Visualizing terminological knowledge*
  - **GenNav**
    *Visualizing gene annotations*
  - **RxNav**
    *Visualizing drug information*
Issues and Challenges
Issues

◆ Size
  ● Large number of concepts (>1 million)

◆ Complexity
  ● Polyhierarchical structures
  ● Multiple information sources
  ● Multiple properties

◆ Lack of formality
  ● Redundant relations
  ● Hierarchies vs. hierarchical relations
Challenges

◆ Restrict information space
  ● To selected information sources (SemNav)
  ● To selected organisms (GenNav)

◆ Reduce complexity (SemNav)
  ● Group concepts by semantic groups
  ● Transitive reduction on hierarchical relations
  ● Select co-occurring concepts

◆ Reduce the cognitive burden on the user
  ● Use graph-based rather than tree-based representations
UMLS Semantic Navigator

SemNav

http://umlsks.nlm.nih.gov*

► SN Resources ► Semantic Navigator

(* free UMLS registration required)
Unified Medical Language System®

◆ Developed at NLM since 1990
◆ 139 source vocabularies
  ● 17 languages
◆ Broad coverage of biomedicine
  ● 5.1M names
  ● 1.3M concepts
  ● 16M relations
◆ Integration
  ● Synonymous terms are clustered in a concept
  ● Hierarchies (trees) are combined in a graph structure
Terminology integration

Terms

- Duchenne muscular dystrophy
  - MeSH, SNOMED
  - CTV3, Jablonski, CRISP, DxPlain, MedDRA, LOINC
- Duchenne’s muscular dystrophy
  - COSTAR
- Duchenne de Boulogne muscular dystrophy
  - Jablonski
- Duchenne type progressive muscular dystrophy
  - SNOMED
- Pseudohypertrophic muscular dystrophy
  - MeSH, CTV3, SNOMED
- X-linked recessive muscular dystrophy
  - Jablonski
- Severe generalized familial muscular dystrophy
  - SNOMED
Terminology integration  Relationships

- Inter-concept relationships: hierarchies from the source vocabularies
- Redundancy: multiple paths
- One graph instead of multiple trees (multiple inheritance)
UMLS  A two-level structure

- Two-level structure
  - Semantic Network
    - 135 Semantic Types (STs)
    - 54 types of relationships among STs
  - Metathesaurus
    - >1M concepts
    - ~12 M inter-concept relationships
  - Link = categorization

Semantic Network

Semantic Type

categorization

Concept

Metathesaurus
Concept: Dystrophin

CUI: C0079259

Semantic Type: Amino Acid, Peptide, or Protein

Biologically Active Substance

Definition:
A muscle protein localized in surface membranes which is the product of the Duchenne/Becker muscular dystrophy gene. Individuals with Duchenne muscular dystrophy usually lack dystrophin completely while those with Becker muscular dystrophy have dystrophin of an altered size. It shares features with other cytoskeletal proteins such as SPECTRIN and alpha-actinin but the precise function of dystrophin is not clear. One possible role might be to preserve the integrity and alignment of the plasma membrane to the myofibrils during muscle contraction and relaxation. MW 400 kDa. (MeSH)

large, structural, spectrin-like protein expressed in skeletal muscle, genetic defect is linked to Duchenne and Becker muscular dystrophy. (CRISP Thesaurus)

Synonyms:
Dystrophin

Ancestors:

MeSH
MeSH Descriptors [1]
Index Medicus Descriptor [1]
Chemicals and Drugs (MeSH Category) [D]
Amino Acids, Peptides, and Proteins [I]

Done
### MeSH Descriptors

- **Index Medicus Descriptor**
- **Chemicals and Drugs (MeSH Category)**
- **Amino Acids, Peptides, and Proteins**
  - **Proteins**
    - **Contractile Proteins**
    - **Muscle Proteins**
    - **Dystrophin**

### CRISP Thesaurus

- **chemical**
  - **organic chemical**
  - **amide**
  - **peptide**
  - **protein**
    - **cytoskeletal protein**
      - **actin binding protein**
      - **dystrophin**

### Digital Anatomist

- **Anatomical entity**
  - **Physical anatomical entity**
  - **Material physical anatomical entity**
  - **Anatomical structure**
    - **Biological macromolecule**
      - **Protein**
        - **Actin-binding protein**
        - **Dystrophin**
Amino Acids, Peptides, and Proteins [D12]
Proteins [D12.776]
Contractile Proteins [D12.775.210]

- Actins [D12.776.210.500.100]

- Parvalbumin [D12.776.210.500.750]
- Profilin [D12.776.210.500.775]

Amino Acids, Peptides, and Proteins [D12]
Proteins [D12.776]
Membrane Proteins [D12.775.543]

- Ankyrin [D12.776.543.050]
- Arrestins [D12.776.543.050]
- Bacterial Outer Membrane Proteins [D12.776.543.100]
- Calnexin [D12.776.543.162]
- Connexins [D12.776.543.225]

- Dystrophin [D12.776.543.250]
- Dystrophin-Associated Proteins [D12.776.543.268]
- Ektins [D12.776.543.287]
Dystrophin

140-kDa dystrophin
Dp260 protein, rat
apo-dystrophin 1

Muscle Proteins
Actin-Binding Protein

Contractile Proteins

Protein, Organized by Location

Filament Proteins

Binding Protein

Proteins

Disorders
Muscular Dystrophies
Muscular Dystrophy
Duchene

Other Related Concepts

Living Beings
Mice, Inbred

Siblings
Chemicals & Drugs

Co-occurring Concepts

Anatomy
Astrocytes
Brain
Cell
Cytoskeleton
Discharge (Anatomy)
Heart

Highlight vocabulary
JMLS data:
UMLS 2006AA


**Siblings**

**Chemicals & Drugs**
- 120-kDa hemocyte-specific membrane protein, flesh fly
- 15a protein, Aedes aegypti
- 22.6-kDa antigen, Schistosoma japonicum
- 22kDa polypeptide, Nicotiana tabacum
- 3D3-lyric protein, human
- 3D3-lyric protein, mouse
- 4.1B protein, mouse
- 5-lipoxygenase-act protein
- 7 kDa protein, carlavus
- A14.5L protein,

---

---
SemNav Visualization options

- **Restrict to vocabulary:** Show all
- **Highlight vocabulary:** Nothing
- **UMLS data:** UMLS_2002
- **Type of hierarchical rel.:**
  - All
  - Parent/Child only
  - Broader/Narrower only
- **Transitive reduction:**
  - Yes
  - No
**Siblings**

**Chemicals & Drugs**
- Actinin
- Actins
- Actomyosin
- Adenomatous Polyposis Coli Protein
- Ankyrins
- Arrestins
- Bacterial Outer Membrane Proteins
- Calnexin
- Calnequesin
- CapZ Actin Capping Protein
- Catenins
- Caveolin 3
- Cofilin 2
- Connexins
- Dystrophin-Associated Proteins
- Ephrins
- Heterotrimeric GTP-Binding Proteins

**Other Related Concepts**

**Disorders**
- Muscular Dystrophy
- Duchenne

**Living Beings**
- Mice, Inbred mdx

(2 other related concepts)

**Proteins**

- Contractile Proteins
- Cytoskeletal Proteins
- Membrane Proteins
- Muscle Proteins
- Dystrophin

**Similar Concepts**

- (none)

**Closest MeSH Terms**

- (none)

**Co-occurring Concepts**

- (not displayed: Restrict to vocabulary: MeSH, Highlight vocabulary: Nothing, UMLS data: UMLS 2006AA)
SemNav Relationships

Metathesaurus Relationships

C1 otherwise related to C2
- not defined
- MeSH

C1 co-occurs with C2
- Frequency = 190
- MEDLINE

Semantic Network Relationships

<table>
<thead>
<tr>
<th>Amino Acid, Peptide, or Protein</th>
<th>affects</th>
<th>causes</th>
<th>Disease or Syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biologically Active Substance</td>
<td>affects</td>
<td>causes</td>
<td>complicates</td>
</tr>
</tbody>
</table>

Dystrophin

Amino Acid, Peptide or Protein

Biologically Active Substance

Muscular Dystrophy, Duchenne

Disease or Syndrome

Concepts

Semantic Types

Amino Acid, Peptide or Protein

Muscular Dystrophy, Duchenne

Biologically Active Substance

Disease or Syndrome
Technical details

- Simple web/cgi technology (apache, Perl)
- `dot` (GraphViz)
  - PNG file (-Tpng)
  - Client-side map (-Tcmap)
- Precompute the transitive closure on hierarchical relations to perform the transitive closure fast
- Remove cycles (UMLS)
Gene Ontology browser

Gene Ontology™

- Developed by the GO Consortium
- Several components (GO database)
  - Ontology (~17,000 concepts)
    - Molecular functions
    - Cellular components
    - Biological processes
  - Gene products (~1.6M)
  - Associations between Gene products and GO concepts (~6.8M)
Search GO

- GO:0008150: biological_process (127417)
- GO:0005575: cellular_component (116421)
  - GO:0005623: cell (85785)
    - GO:0005622: intracellular (69449)
      - GO:0005856: cytoskeleton (2376)
    - GO:0043229: intracellular organelle (62778)
      - GO:0043232: intracellular non-membrane-bound organelle (7453)
        - GO:0005856: cytoskeleton (2376)
    - GO:0043226: organelle (62799)
      - GO:0043229: intracellular organelle (62778)
        - GO:0043232: intracellular non-membrane-bound organelle (7453)
          - GO:0005856: cytoskeleton (2376)
    - GO:0043228: non-membrane-bound organelle (7453)
      - GO:0043232: intracellular non-membrane-bound organelle (7453)
        - GO:0005856: cytoskeleton (2376)
  - GO:0003674: molecular_function (121803)
    - GO:obsolete Biological Process: obsolete biological process (0)
    - GO:obsolete Cellular Component: obsolete cellular component (0)
    - GO:obsolete Molecular Function: obsolete molecular function (0)
| Molecular functions          | • actin binding [TAS]  
|                            | • structural constituent of cytoskeleton [TAS]  
|                            | • protein binding [IPI]  
|                            | • structural constituent of muscle [IDA, TAS]  

| Biological processes        | • muscle contraction [NR]  
|                            | • muscle development [NAS]  
|                            | • peptide biosynthesis [IDA]  

| Cellular components         | • cytoskeleton [TAS]  
|                            | • dystrophin-associated glycoprotein complex [TAS]  
|                            | • costamere [IDA]  

<table>
<thead>
<tr>
<th>Organisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>[restrict] Arabidopsis thaliana (thale cress)</td>
</tr>
<tr>
<td>[restrict] Bos taurus (cow)</td>
</tr>
<tr>
<td>[restrict] Caenorhabditis elegans</td>
</tr>
<tr>
<td>[restrict] Danio rerio (zebrafish)</td>
</tr>
<tr>
<td>[restrict] Dictyostelium discoideum</td>
</tr>
<tr>
<td>[restrict] Drosophila melanogaster (fruit fly)</td>
</tr>
<tr>
<td>[restrict] Gallus gallus (chicken)</td>
</tr>
<tr>
<td>[restrict] Homo sapiens (human)</td>
</tr>
<tr>
<td>[restrict] Human immunodeficiency virus 1</td>
</tr>
<tr>
<td>[restrict] Human immunodeficiency virus type 1 (BRU ISOLATE)</td>
</tr>
<tr>
<td>[restrict] Human immunodeficiency virus type 1 (CLONE 12)</td>
</tr>
<tr>
<td>[restrict] Human immunodeficiency virus type 1 (PV22 ISOLATE)</td>
</tr>
<tr>
<td>[restrict] Human immunodeficiency virus type 1 lw12.3 isolate</td>
</tr>
<tr>
<td>[restrict] Macaca fascicularis (crab-eating macaque)</td>
</tr>
<tr>
<td>[restrict] Mus musculus (house mouse)</td>
</tr>
<tr>
<td>[restrict] Plasmodium falciparum (malaria parasite P. falciparum)</td>
</tr>
<tr>
<td>[restrict] Rattus norvegicus (Norway rat)</td>
</tr>
<tr>
<td>[restrict] Saccharomyces cerevisiae (baker’s yeast)</td>
</tr>
<tr>
<td>[restrict] Schizosaccharomyces pombe (fission yeast)</td>
</tr>
<tr>
<td>[restrict] Trypanosoma brucei</td>
</tr>
<tr>
<td>[restrict] Trypanosoma brucei TREU927</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Full name</th>
<th>Symbol</th>
<th>Organism</th>
<th>DB</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>nucleoside diphosphate kinase, NDP kinase</td>
<td>ndkC</td>
<td>Dictyostelium discoideum</td>
<td>DDB</td>
<td>IDA</td>
</tr>
</tbody>
</table>
RxNorm browser

<table>
<thead>
<tr>
<th>Strength</th>
<th>Ingredient</th>
<th>Dose form</th>
</tr>
</thead>
<tbody>
<tr>
<td>4mg/ml</td>
<td>Fluoxetine</td>
<td>Oral Solution</td>
</tr>
</tbody>
</table>

Semantic clinical drug component

Semantic clinical drug form

Semantic clinical drug
Generic vs. Brand

- **Generic**
  - Ingredient (IN)
  - Clinical drug form (SCDF)
  - Clinical drug component (SCDC)
  - Clinical drug (SCD)

- **Brand**
  - Brand name (BN)
  - Branded drug form (SBDF)
  - Branded drug component (SBDC)
  - Branded drug (SBD)

\[ \text{tradename}_o \]
Relations among drug entities
RxNorm database

- **Data sources**
  - Master Drug Data Base
  - Multum MediSource Lex.
  - Micromedex DRUGDEX
  - FDA National Drug Code Directory
  - National Drug Data File
  - Plus Source Vocabulary
  - VA National Drug File
  - SNOMED Clinical Terms

- **Content**
  - 5,570 ingredients
  - 10,788 brand names
  - 22,724 clinical drug comp.
  - 29,734 clinical drugs
  - 17,149 branded drugs
  - 16,447 branded drug comp.
  - 13,516 clinical drug forms
  - 13,035 branded drug forms
  - 140 dose forms

*(as of February 28, 2006)*
Relations among drug entities
Retrieved "Sertraline 20 MG/ML Oral Solution".
208993|C0710662|prednisolone 0.005 MG/ML / Sulfacetamide Sodium 0.1 MG/ML Ophthalmic Ointment [Metimyd]
<table>
<thead>
<tr>
<th>Branded Drug</th>
<th>Dose Form</th>
<th>Clinical Drug Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>[List of Branded Drugs]</td>
<td>[List of Dose Forms]</td>
<td>[List of Clinical Drug Forms]</td>
</tr>
</tbody>
</table>
Retrieval Status or Detailed View of an RxNorm Entry (RXCUI | UMLSCUI | STR)

Please check spelling!
Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA