

CERN Workshop

on *Innovations in Scholarly Communication*

19-21 June 2013 University of Geneva, Switzerland

OAI8

CERN Workshop on Innovations
in Scholarly Communication (OAI8)
Geneva, Switzerland
June 20, 2013

Semantic indexing in PubMed



Olivier Bodenreider

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



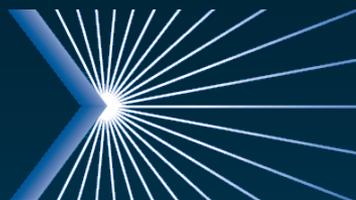
U.S. National Library of Medicine



Orientation



U.S. National Library of Medicine



PubMed.gov

US National Library of Medicine
National Institutes of Health

- ◆ NLM is the world's largest biomedical library
 - Located in Bethesda, Maryland, near Washington, DC
- ◆ PubMed provides access to MEDLINE, NLM's bibliographic database of over 20M citations
 - MEDLINE covers 5600 journals and adds almost 1M new citations each year
 - PubMed is part of the Entrez system of the National Center for Biotechnology Information (NCBI)



Outline

- ◆ Anatomy of a MEDLINE citation
- ◆ Types of PubMed searches
 - Simple text search
 - Search based on MeSH indexing
- ◆ Automatic indexing
- ◆ Beyond topics



Anatomy of a MEDLINE citation



US National Library of Medicine
National Institutes of Health

J Am Podiatr Med Assoc. 2003 Jul-Aug;93(4):333-5.

Fluoroquinolone therapy and Achilles tendon rupture.

Title

Vanek D, Saxena A, Boqqs JM.

California School of Podiatric Medicine, San Francisco, CA, USA.

Abstract

Fluoroquinolones have been associated with tendinopathies. The authors present three cases of Achilles tendinopathy in which the patients' symptoms were preceded by treatment for unrelated bacterial infections with ciprofloxacin. Although the exact mechanism of the relationship is not understood, those who engage in sports or exercise should be advised of the risk of quinolone-induced tendinopathy.

Abstract

PMID: 12869605 [PubMed - indexed for MEDLINE]

Publication Types, MeSH Terms, Substances

Indexing

LinkOut - more resources



Publication Types, MeSH Terms, Substances

Publication Types

Case Reports

Review

MeSH Terms

Achilles Tendon/drug effects*

Adult

Aged

Anti-Infective Agents/adverse effects*

Anti-Infective Agents/therapeutic use

Ciprofloxacin/adverse effects*

Ciprofloxacin/therapeutic use

Female

Humans

Male

Middle Aged

Muscular Diseases/chemically induced

Risk Factors

Rupture

Tendon Injuries/chemically induced*

MeSH main heading[/subheading(s)]
[+ * for major topic]

Substances

Anti-Infective Agents

Ciprofloxacin



Types of PubMed searches

<http://www.ncbi.nlm.nih.gov/pubmed>



Non-semantic search

- ◆ PubMed does not require the use of MeSH for querying
 - Supports “Google-like” text searches
 - “no librarian required”
 - But can identify MeSH terms even if they are not labeled as such



Non-semantic search Example

- ◆ Find articles about the cheese Gruyère
 - Gruyère

[Show additional filters](#)

[Display Settings:](#) Summary, 20 per page, Sorted by Recently Added

[Send to:](#)

[Filters:](#) [Manage Filters](#)

Article types

Review

More ...

★ Did you mean: [gruytere](#) (1 items)

Results: 1 to 20 of 27

<< First < Prev Page 1 of 2 Next > Last >>

Titles with your search terms

[Symptoms and respiratory function in a group of Gruyère c [Rev Mal Respir. 1990]

[J Dairy Sci.](#) 2008 Jan;91(1):29-38.

Occurrence of the angiotensin-converting enzyme inhibiting tripeptides Val-Pro-Pro and Ile-Pro-Pro in different cheese varieties of Swiss origin.

[Bütikofer U.](#), [Meyer J.](#), [Sieber R.](#), [Walther B.](#), [Wechsler D.](#)

Agroscope Liebefeld-Posieux Research Station ALP, Schwarzenburgstrasse 161, 3003 Bern, Switzerland.

Abstract

The contents of the 2 antihypertensive peptides Val-Pro-Pro (VPP) and Ile-Pro-Pro (IPP) were determined in 101 samples from 10 different Swiss cheese varieties using HPLC with subsequent triple mass spectrometry. In the category of extra hard and hard cheeses, the Protected Denomination of Origin cheeses Berner Alpkäse and Berner Hobelkäse, L'Etivaz à rebibes, Le [Gruyère](#), Sbrinz, Emmentaler (organic and conventional) and in the category of semihard cheeses, the varieties Tilsiter, Appenzeller 1/4 fat and full fat, Tête de Moine, and Vacherin fribourgeois were screened in the study. The average concentration of the sum of VPP and IPP in the screened cheese varieties varied to a large extent, and substantial variations

MeSH (semantic) search

◆ Medical Subject Headings (MeSH)

- Controlled vocabulary developed at NLM for indexing and retrieval of MEDLINE citations
- ~26,000 descriptors (main headings)
- <100 qualifiers (subheadings)
- 214,000 supplementary concept records

◆ Hierarchical structure (“tree numbers”)

- Supports query expansion (“explosion”)
 - Search for a descriptor or any of its descendants

http://www.nlm.nih.gov/mesh/2013/mesh_browser/MBrowser.html



Simple MeSH search Example

- ◆ Find articles about drug-induced psychoses
 - "Psychoses, Substance-Induced"[Mesh]

PubMed

[RSS](#) [Save search](#) [Advanced](#)

[Display Settings:](#) Summary, 20 per page, Sorted by Recently Added [Send to:](#)

Results: 1 to 20 of 6563 << First < Prev Page of 329 Next > Last >>

- [Substance-induced psychoses converting into schizophrenia: a register study of 18,478 Finnish inpatient cases.](#)
1. Niemi-Pynttari JA, Sund R, Putkonen H, Vormaa H, Wahlbeck K, Pirkola J Clin Psychiatry. 2013 Jan;74(1):e94-9. doi: 10.4088/JCP.12m07822.
PMID: 23419236 [PubMed - indexed for MEDLINE]
[Related citations](#)
- [Monitoring the prevalence of methamphetamine-related presentations in psychiatric hospitals in Cape Town, South Africa.](#)
2. Plüddemann A, Dada S, Parry CD, Kader R, Parker JS, Temmingh H, van S, de Clercq C, Lewis I.
Afr J Psychiatry (Johannesbg). 2013 Jan;16(1):45-9. doi: http://dx.doi.org/10.4314/PMID: 23417636 [PubMed - indexed for MEDLINE] **Free Article**
[Related citations](#)

[Methamphetamine/adverse effects*](#)
[Middle Aged](#)
[Mood Disorders/chemically induced](#)
[Mood Disorders/epidemiology](#)
[Patient Admission/statistics & numerical data*](#)
[Patient Readmission](#)
[Psychoses, Substance-Induced diagnosis](#)
[Psychoses, Substance-Induced epidemiology](#)
[Schizophrenia/chemically induced](#)

Search with “Explosion”

- ◆ By default, PubMed retrieves articles indexed with a descriptor or any of its descendants
- ◆ Use **mesh:noexp** to prevent “explosion” from happening

[Quinolones \[D03.438.810.835\]](#)

[4-Quinolones \[D03.438.810.835.055\]](#) +

[Carteolol \[D03.438.810.835.188\]](#)

▶ [Fluoroquinolones \[D03.438.810.835.322\]](#)

[Ciprofloxacin \[D03.438.810.835.322.186\]](#) +

[Enoxacin \[D03.438.810.835.322.280\]](#)

[Norfloxacin \[D03.438.810.835.322.374\]](#)

[Ofloxacin \[D03.438.810.835.322.500\]](#)

[Pefloxacin \[D03.438.810.835.322.750\]](#)

[PQQ Cofactor \[D03.438.810.835.661\]](#)

“Explosion” Example

- ◆ Find articles about fluoroquinolones (or desc.)
 - "fluoroquinolones"[Mesh]

[Can J Neurol Sci](#). 2013 Jan;40(1):127-8.

Ciprofloxacin induced acute small fibre neuropathy. Case report.

[Jumma OK](#), [Dick J](#), [Marshall A](#), [Mellor K](#).

PMID: 23427360 [PubMed - indexed for MEDLINE]

MeSH Terms

[Anti-Infective Agents/adverse effects*](#)

[Ciprofloxacin/adverse effects*](#)

[Humans](#)

[Leukocyte Count](#)

[Male](#)

[Middle Aged](#)

[Nerve Fibers/pathology](#)

[Peripheral Nervous System Diseases/chemically induced*](#)

[Peripheral Nervous System Diseases/physiopathology](#)

[Sensory Thresholds/physiology](#)



Search leveraging synonymy in MeSH

- ◆ MeSH descriptors include related concepts (Entry terms)
 - Synonyms
 - Closely related (and clustered or indexing and retrieval purposes)
- ◆ All terms from a descriptor and its entry terms are used for retrieval in PubMed



PubMed

primary adrenocortical insufficiency

Search

RSS Save search Advanced

Help

Show additional filters

Display Settings: Summary, 20 per page, Sorted by Recently Added

Send to:

Filters: Manage Filters

Article types

Clinical Trial

Review

More ...

Text availability

Abstract available

Free full text available

Full text available

Publication dates

5 years

10 years

Custom range...

Species

Humans

Other Animals

Clear all

Results: 21 to 40 of 4560

<< First < Prev Page 2 of 228 Next > Last >>

[Autoimmune Addison's disease.](#)

21. Napier C, Pearce SH. Presse Med. 2012 Dec;41(12 P 2):e626-35. doi: 10.1016/j.lpm.2012.09.010. Epub 2012 Nov 21. Review. PMID: 23177474 [PubMed - indexed for MEDLINE] [Related citations](#)

22. [Mitotane therapy in adrenocortical cancer induces CYP3A4 and inhibits 5 \$\alpha\$ -reductase, explaining the need for personalized glucocorticoid and androgen replacement.](#)

Chortis V, Taylor AE, Schneider P, Tomlinson JW, Hughes BA, O'Neil DM, Libé R, Allolio B, Bertagna X, Bertherat J, Beuschlein F, Fassnacht M, Karavitaki N, Mannelli M, Mantero F, Opocher G, Porfiri E, Quinkler M, Sherlock M, Terzolo M, Nightingale P, Shackleton CH, Stewart PM, Hahner S, Arit W. J Clin Endocrinol Metab. 2013 Jan;98(1):161-71. doi: 10.1210/jc.2012-2851. Epub 2012 Nov 16. PMID: 23162091 [PubMed - indexed for MEDLINE] [Related citations](#)

23. [Idiopathic intracranial hypertension: a unifying neuroendocrine hypothesis through the adrenal-brain axis](#)

Results by year



Download CSV

Find related data

Database: Select

Find items

Search details

"addison disease"[MeSH Terms] OR ("addison"[All Fields] AND "disease"[All Fields]) OR "addison disease"[All Fields] OR ("primary"[All Fields]



Entry terms for “Addison Disease”

MeSH Heading	Addison Disease
Tree Number	C19.053.500.263
Tree Number	C20.111.163
Scope Note	An adrenal disease characterized by the progressive destruction of the ADRENAL CORTEX , resulting in insufficient production of ALDOSTERONE and HYDROCORTISONE . Clinical symptoms include ANOREXIA ; NAUSEA ; WEIGHT LOSS ; MUSCLE WEAKNESS ; and HYPERPIGMENTATION of the SKIN due to increase in circulating levels of ACTH precursor hormone which stimulates MELANOCYTES .
Entry Term	Addison's Disease
Entry Term	Primary Adrenal Insufficiency
Entry Term	Primary Adrenocortical Insufficiency
Entry Term	Primary Hypoadrenalism

Search leveraging UMLS Synonymy

- ◆ Unified Medical Language System (UMLS)
 - Terminology integration system
 - ~130 biomedical terminologies
 - Synonymous terms clustered into concepts
- ◆ UMLS synonymy used in PubMed
 - Query translation happens “behind the scenes”
 - E.g., search on “primary adrenocortical insufficiency”
 - Retrieves articles about “Addison’s disease”



[Show additional filters](#)**Display Settings:** Summary, 20 per page, Sorted by Recently Added**Send to:** **Article types**

Clinical Trial

Review

More ...

Text**availability**

Abstract available

Free full text available

Full text available

Publication dates

5 years

10 years

Custom range...

Species

Humans

Other Animals

Results: 21 to 40 of 192998

<< First

< Prev

Page 2

of 9650

Next >

Last >>

- [Prophylactic amiodarone and lidocaine improve survival in an ovine model of large size myocardial infarction.](#)

21. Li T, Wei X, Watkins AC, Sanchez PG, Wu ZJ, Griffith BP.

J Surg Res. 2013 Jun 7. doi:pii: S0022-4804(13)00498-8. 10.1016/j.jss.2013.05.050. [Epub ahead of print]

PMID: 23773712 [PubMed - as supplied by publisher]

[Related citations](#)

- [Drug-Eluting Stents Versus Bare-Metal Stents in Patients With Decreased GFR: A Meta-analysis.](#)

22. Wang ZJ, Harjai KJ, Shenoy C, Gao F, Shi DM, Liu YY, Zhao YX, Zhou YJ.

Am J Kidney Dis. 2013 Jun 14. doi:pii: S0272-6386(13)00811-1. 10.1053/j.ajkd.2013.04.014. [Epub ahead of print]

PMID: 23773369 [PubMed - as supplied by publisher]

[Related citations](#)

- [Targeting inflammatory pathways in myocardial infarction.](#)

23. Christia P, Frangogiannis NG.

No entry term for “Heart attack”

MeSH Heading	Myocardial Infarction
Tree Number	C14.280.647.500
Tree Number	C14.907.585.500
Annotation	do not coordinate with ACUTE DISEASE for "acute infarct"
Scope Note	NECROSIS of the MYOCARDIUM caused by an obstruction of the
Entry Term	Myocardial Infarct
See Also	Heart Rupture, Post-Infarction

Search details

```
"myocardial  
infarction"[MeSH Terms]  
OR ("myocardial"[All  
Fields] AND  
"infarction"[All Fields])  
OR "myocardial
```

Search

See more...



Query translation

Search Details

Query Translation:

```
"myocardial infarction"[MeSH Terms] OR ("myocardial"[All Fields] AND "infarction"[All Fields]) OR "myocardial infarction"[All Fields] OR ("heart"[All Fields] AND "attack"[All Fields]) OR "heart attack"[All Fields]
```

Search

URL

Subheading restrictions

- ◆ Subheadings represent the context of use of a particular descriptor
 - Ciprofloxacin/Adverse effects
 - Mood Disorders/Chemically induced
- ◆ Assigned during indexing
- ◆ Can be queried in PubMed

Subheading restrictions Example

- ◆ Find articles about drugs involved in adverse events
 - "Chemicals and Drugs Category"/adverse effects[MeSH]

[Can J Neurol Sci](#). 2013 Jan;40(1):127-8.

Ciprofloxacin induced acute small fibre neuropathy. Case report.

[Jumma OK](#), [Dick J](#), [Marshall A](#), [Mellor K](#).

PMID: 23427360 [PubMed - indexed for MEDLINE]

MeSH Terms

[Anti-Infective Agents/adverse effects*](#)

[Ciprofloxacin/adverse effects*](#)

[Humans](#)

[Leukocyte Count](#)

[Male](#)

[Middle Aged](#)

[Nerve Fibers/pathology](#)

[Peripheral Nervous System Diseases/chemically induced*](#)

[Peripheral Nervous System Diseases/physiopathology](#)

[Sensory Thresholds/physiology](#)



Recapitulative example

- ◆ Find articles about drugs involved in adverse events and drug-induced manifestations
 - ("Chemicals and Drugs Category"[Mesh]) AND (adverse effects[sh] OR contraindications[sh] OR mortality[sh])) AND (chemically induced[sh] OR ("Drug-Induced Liver Injury"[Mesh:noexp]) OR ("Drug Eruptions"[Mesh:noexp]) OR ("Epidermal Necrolysis, Toxic"[Mesh]) OR ("Drug-Induced Liver Injury, Chronic"[Mesh]) OR ("Erythema Nodosum"[Mesh]) OR ("Serotonin Syndrome"[Mesh]) OR ("Hand-Foot Syndrome"[Mesh]) OR ("Neuroleptic Malignant Syndrome"[Mesh]) OR ("MPTP Poisoning"[Mesh]) OR ("Dyskinesia, Drug-Induced"[Mesh]) OR ("Neurotoxicity Syndromes"[Mesh:noexp]) OR ("Psychoses, Substance-Induced"[Mesh]) OR ("Akathisia, Drug-Induced"[Mesh]))) AND (medline[sb])

Automatic indexing



Lister Hill National Center for Biomedical Communications

Automatic indexing Motivation

- ◆ Indexing by humans is costly and has limited reproducibility
- ◆ Natural language processing can effectively support named entity recognition
- ◆ Automatic indexing can produce
 - Suggestions for human indexers
 - Final indexing for some journals



Automatic indexing Principles

◆ Hybrid approach

- Concepts extracted from title and abstract
 - Mapped from UMLS to MeSH
- MeSH descriptors extracted from related citations

◆ Post-processing

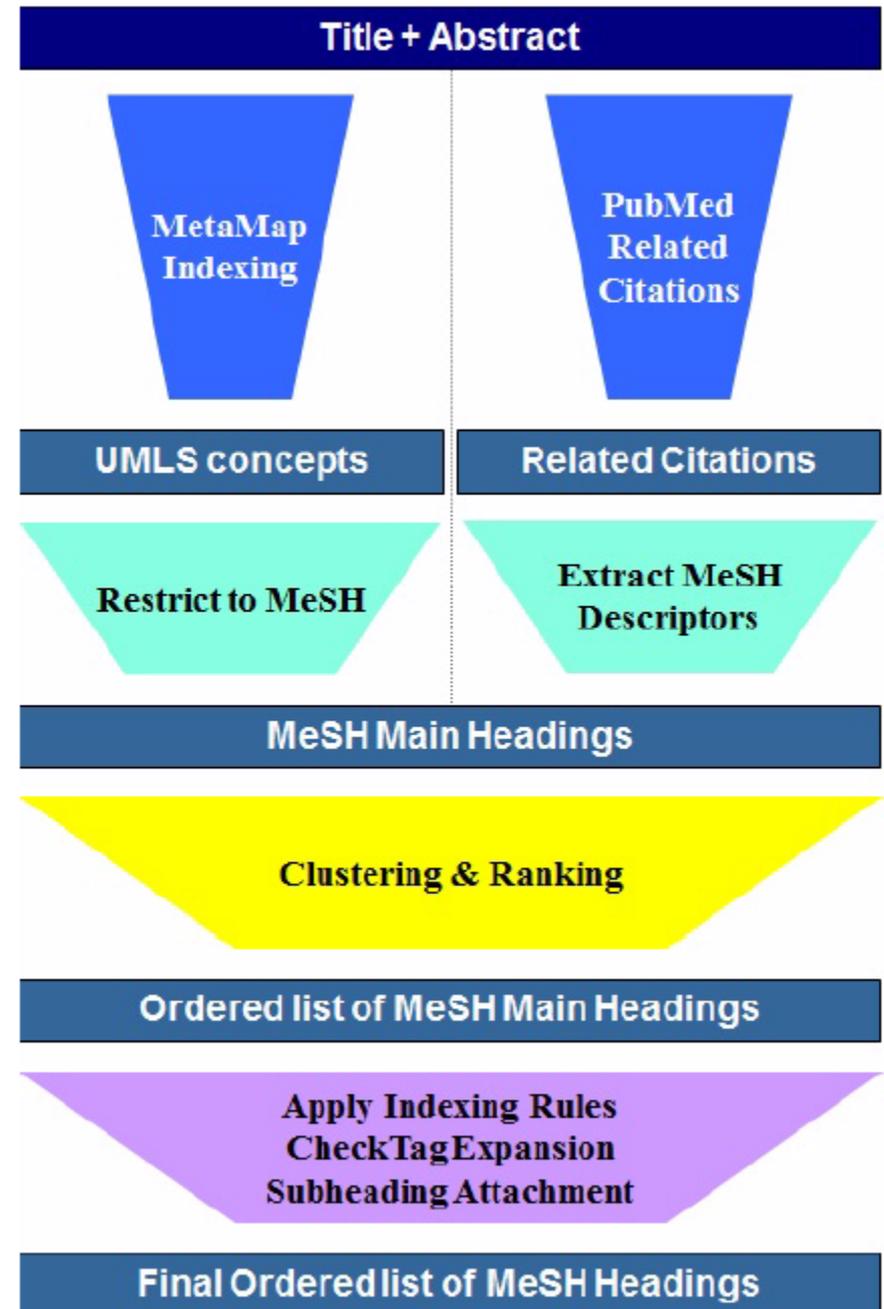
- Clustering and ranking
- Integrate indexing rules
 - E.g., “rule of 3”
 - Index with a higher-level descriptor rather than with 3 or more lower-level descriptors



Automatic in

◆ Medical Text Indexer

<http://ii.nlm.nih.gov/mti.shtml>



Automatic indexing Applications

◆ MEDLINE indexing

- Support MEDLINE indexing at NLM
 - 3600 new citations processed every weeknight
 - Suggestions displayed in the indexing environment
- “First-line” indexing
 - For 75 journals
 - MTI recommendations are used as an indexer
 - Simply reviewed by a senior indexer

◆ Cataloging and History of Medicine

- Assisted indexing



Beyond topics



Lister Hill National Center for Biomedical Communications

Beyond concepts... relations

◆ Also known as

- Facts
- Predications
- Nano-publications
- ...

◆ Relation extraction

- Usually based on natural language processing (NLP)
 - E.g., SemRep
- Relations stored in (*subject, predicate, object*) form
 - With provenance information



Experimental application Semantic MEDLINE

- ◆ Multi-document summarization
- ◆ Based on a database of 60M predications extracted from MEDLINE
- ◆ Entities normalized to the UMLS Metathesaurus
- ◆ Relations aligned with the UMLS Semantic Network
- ◆ Interfaced with PubMed (for retrieving PMIDs) on a given topic
 - Forms the basis for summarization

<http://skr3.nlm.nih.gov/SemMedDemo/>



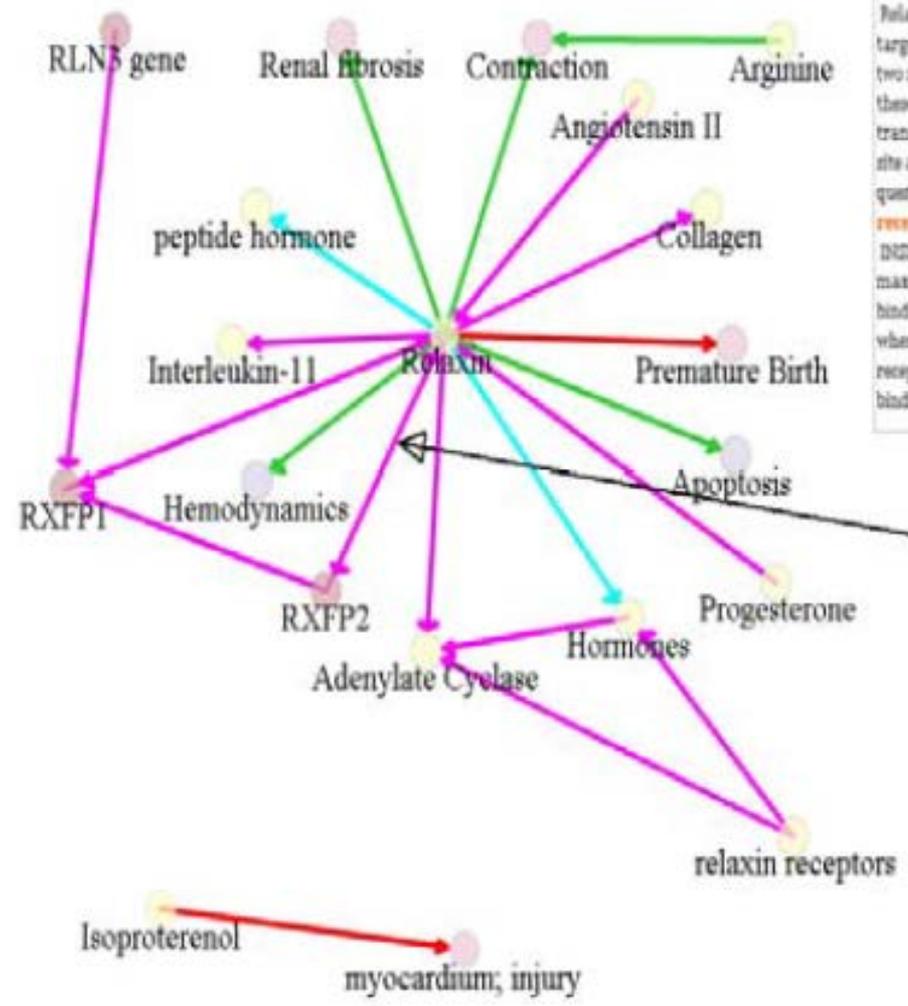
ID:15649866

Title: Multiple binding sites revealed by interaction of relaxin family peptides with native and chimeric relaxin family peptide receptors 1 and 2 (LGR7 and LGR8).

Abstract:

Relaxin family peptide 1 (RXFP1) receptor (LGR7) and RXFP2 receptor (LGR8) were recently identified as the receptor targets for H₂ relaxin and insulin-like peptide 3 (INL3), respectively. In this study, we define the pharmacology of these two receptors by using a number of receptor chimeras and relaxin family peptides. We have identified two binding sites on these receptors: one primary, high-affinity site within the ectodomain and a secondary, lower affinity site within the transmembrane region. The primary site was found to dictate receptor binding characteristics, although the lower affinity site also exerts some influence and modulates ligand affinity for the primary site in a manner dependent upon the peptide in question. **Not all relaxin peptides were able to bind to the RXFP2 receptor, indicating that the relaxin-RXFP2 receptor interaction is species-specific.**

INL3 was found to exhibit characteristics of a partial agonist at the RXFP2 and chimeric RXFP1/2 receptors, with low maximal cAMP response but high potency in coupling to this pathway. cAMP accumulation studies also revealed that the binding sites couple to cAMP signaling pathways with differing efficiency: the high-affinity site signals with high efficiency, whereas the lower affinity site signals with little to no efficiency. Comparisons between RXFP1, RXFP2, the chimeric receptors, and the truncated receptors revealed that the interaction between receptor sites is critical for optimal ligand binding and signal transduction and that the ectodomain is essential for signaling. Evidence obtained in this study supports



Relaxin INTERACTS_WITH RXFP2

Subject:	Relaxin
Relation:	INTERACTS_WITH
Object:	RXFP2
No. Predictions:	2
No. Citations:	2

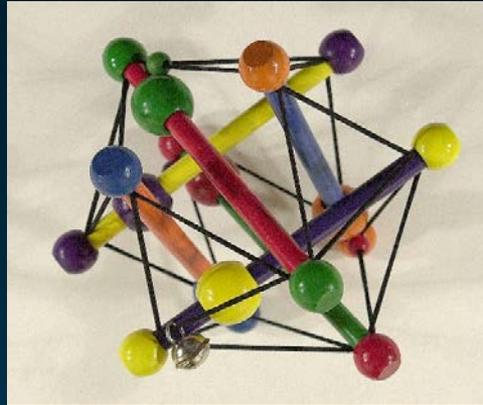
Citations

Filters Search

Relation extraction Applications

- ◆ Enhanced information retrieval
 - Indexing on relations in addition to concepts or association main heading/subheading
- ◆ Multi-document summarization
 - Extract and visualize the facts extracted from 250 recent abstracts on the treatment of Parkinson's disease
- ◆ Question answering
 - Clinical and biological questions
- ◆ Knowledge discovery
 - Connect facts from heterogeneous resources





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