PRECIS

An automated pipeline for producing concise reports about proteins

Phillip Lord

p.lord@russet.org.uk

Department of Computing Science, University of Manchester

RoadMap

- What is annotation?
- Where does annotation come from?
- What does PRECIS do?
- Results of PRECIS
- Conclusions.

Biology builds on sequence data

Most biological data is built on top of sequence data.

- Sequence data is what we have most of!
- Its the simplest data type. Its easy to model, as a string.
- Sequence is fairly incontrovertible.

Sequence data is opaque

Therefore it is common to attach large amounts of data to the sequence which helps with its interpretation.

- Data about the experimental conditions.
- Data interpreted from the sequence.
- Data about other related proteins.

This data is usually described as “annotation”.

A SWISS-PROT entry

```
ID   PRIO_HUMAN     STANDARD;      PRT;   253 AA.
AC   P04156;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   Major prion protein precursor (PrP) (PrP27-30) (PrP33-35C) (ASCR).
GN   PRNP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86300093; PubMed=3755672;
RA   Kretzschmar H.A., Stowring L.E., Westaway D., Stubblebine W.H.,
RA   Prusiner S.B., Dearmond S.J.;
RT   "Molecular cloning of a human prion protein cDNA.";
RL   DNA 5:315-324(1986).
RN   [2]
RP   SEQUENCE OF 8-253 FROM N.A.
RX   MEDLINE=86261778; PubMed=3014653;
RA   Liao Y.-C.J., Lebo R.V., Clawson G.A., Smuckler E.A.;
RT   "Human prion protein cDNA: molecular cloning, chromosomal mapping,
RT   and biological implications.";
RL   Science 233:364-367(1986).
RN   [3]
RP   SEQUENCE OF 58-85 AND 111-150 (VARIANT AMYLOID GSS).
RX   MEDLINE=91160504; PubMed=1672107;
RA   Tagliavini F., Prelli F., Ghiso J., Bugiani O., Serban D.,
RA   Prusiner S.B., Farlow M.R., Ghetti B., Frangione B.;
RT   "Amyloid protein of Gerstmann-Straussler-Scheinker disease (Indiana
RT   kindred) is an 11 kd fragment of prion protein with an N-terminal
RT   glycine at codon 58.";
RL   EMBO J. 10:513-519(1991).
RN   [4]
RP   STRUCTURE BY NMR OF 118-221.
RX   MEDLINE=20359708; PubMed=10900000;
RA   Calzolai L., Lysek D.A., Guntert P., von Schroetter C., Riek R.,
RA   Zahn R., Wuethrich K.;
RT   "NMR structures of three single-residue variants of the human prion
RT   protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:8340-8345(2000).
CC   -!- FUNCTION: THE FUNCTION OF PRP IS NOT KNOWN. PRP IS ENCODED IN THE
CC       HOST GENOME AND IS EXPRESSED BOTH IN NORMAL AND INFECTED CELLS.
CC   -!- SUBUNIT: PRP HAS A TENDENCY TO AGGREGATE YIELDING POLYMERS CALLED
CC       "RODS".
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- POLYMORPHISM: THE FIVE TANDEM OCTAPEPTIDE REPEATS REGION IS HIGHLY
CC       UNSTABLE. INSERTIONS OR DELETIONS OF OCTAPEPTIDE REPEAT UNITS ARE
CC       ASSOCIATED TO PRION DISEASE.
CC   -!- DISEASE: PRP IS FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND
CC       ANIMALS INFECTED WITH NEURODEGENERATIVE DISEASES KNOWN AS
CC       TRANSMISSIBLE SPONGIFORM ENCEPHALOPATHIES OR PRION DISEASES, LIKE:
CC       CREUTZFELDT-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME
CC       (GSS), FATAL FAMILIAL INSOMNIA (FFI) AND KURU IN HUMANS; SCRAPIE
CC       IN SHEEP AND GOAT; BOVINE SPONGIFORM ENCEPHALOPATHY (BSE) IN
CC       CATTLE; TRANSMISSIBLE MINK ENCEPHALOPATHY (TME); CHRONIC WASTING
CC       DISEASE (CWD) OF MULE DEER AND ELK; FELINE SPONGIFORM
CC       ENCEPHALOPATHY (FSE) IN CATS AND EXOTIC UNGULATE ENCEPHALOPATHY
CC       (EUE) IN NYALA AND GREATER KUDU. THE PRION DISEASES ILLUSTRATE
CC       THREE MANIFESTATIONS OF CNS DEGENERATION: (1) INFECTIOUS (2)
CC       SPORADIC AND (3) DOMINANTLY INHERITED FORMS. TME, CWD, BSE, FSE,
CC       EUE ARE ALL THOUGHT TO OCCUR AFTER CONSUMPTION OF PRION-INFECTED
CC       FOODSTUFFS.
DR   EMBL; M13667; AAA19664.1; -.
DR   EMBL; M13899; AAA60182.1; -.
DR   EMBL; D00015; BAA00011.1; -.
DR   PIR; A05017; A05017.
DR   PIR; A24173; A24173.
DR   PIR; S14078; S14078.
DR   PDB; 1E1G; 20-JUL-00.
DR   PDB; 1E1J; 20-JUL-00.
DR   PDB; 1E1P; 20-JUL-00.
DR   PDB; 1E1S; 21-JUL-00.
DR   PDB; 1E1U; 20-JUL-00.
DR   PDB; 1E1W; 20-JUL-00.
DR   MIM; 176640; -.
DR   MIM; 123400; -.
DR   MIM; 137440; -.
DR   MIM; 245300; -.
DR   MIM; 600072; -.
DR   MIM; 604920; -.
DR   InterPro; IPR000817; Prion.
DR   Pfam; PF00377; prion; 1.
DR   PRINTS; PR00341; PRION.
DR   SMART; SM00157; PRP; 1.
DR   PROSITE; PS00291; PRION_1; 1.
DR   PROSITE; PS00706; PRION_2; 1.
KW   Prion; Brain; Glycoprotein; GPI-anchor; Repeat; Signal;
KW   3D-structure; Polymorphism; Disease mutation.
FT   SIGNAL        1     22
FT   CHAIN        23    230       MAJOR PRION PROTEIN.
FT   PROPEP      231    253       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   LIPID       230    230       GPI-ANCHOR (BY SIMILARITY).
FT   CARBOHYD    181    181       N-LINKED (GLCNAC...) (PROBABLE).
FT   DISULFID    179    214       BY SIMILARITY.
FT   DOMAIN       51     91       5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-
FT                                Q.
FT   REPEAT       51     59       1.
FT   REPEAT       60     67       2.
FT   REPEAT       68     75       3.
FT   REPEAT       76     83       4.
FT   REPEAT       84     91       5.
FT   VARIANT     102    102       P -> L (IN GSS AND EOAD).
FT                                /FTId=VAR_006464.
FT   VARIANT     105    105       P -> L (IN GSS).
FT                                /FTId=VAR_006465.
FT   VARIANT     117    117       A -> V (LINKED TO DEVELOPMENT OF
FT                                DEMENTING GSS).
FT                                /FTId=VAR_006466.
FT   VARIANT     129    129       M -> V (DETERMINES THE DISEASE PHENOTYPE
FT                                IN PATIENTS WHO HAVE A PRP MUTATION AT
FT                                CODON 178: PATIENTS WITH MET DEVELOP FFI,
FT                                THOSE WITH VAL DEVELOP CJD).
FT                                /FTId=VAR_006467.
FT   VARIANT     171    171       N -> S (IN SCHIZOAFFECTIVE DISORDER).
FT                                /FTId=VAR_006468.
FT   VARIANT     178    178       D -> N (IN FFI AND CJD).
FT                                /FTId=VAR_006469.
FT   VARIANT     180    180       V -> I (IN CJD).
FT                                /FTId=VAR_006470.
FT   VARIANT     183    183       T -> A (IN FAMILIAL SPONGIFORM
FT                                ENCEPHALOPATHY).
FT                                /FTId=VAR_006471.
FT   VARIANT     187    187       H -> R (IN GSS).
FT                                /FTId=VAR_008746.
FT   VARIANT     188    188       T -> K (IN EOAD; DEMENTIA ASSOCIATED TO
FT                                PRION DISEASES).
FT                                /FTId=VAR_008748.
FT   VARIANT     188    188       T -> R.
FT                                /FTId=VAR_008747.
FT   VARIANT     196    196       E -> K (IN CJD).
FT                                /FTId=VAR_008749.
FT                                /FTId=VAR_006472.
SQ   SEQUENCE   253 AA;  27661 MW;  43DB596BAAA66484 CRC64;

     MANLGCWMLV LFVATWSDLG LCKKRPKPGG WNTGGSRYPG QGSPGGNRYP PQGGGGWGQP
     HGGGWGQPHG GGWGQPHGGG WGQPHGGGWG QGGGTHSQWN KPSKPKTNMK HMAGAAAAGA
     VVGGLGGYML GSAMSRPIIH FGSDYEDRYY RENMHRYPNQ VYYRPMDEYS NQNNFVHDCV
     NITIKQHTVT TTTKGENFTE TDVKMMERVV EQMCITQYER ESQAYYQRGS SMVLFSSPPV
     ILLISFLIFL IVG

//
```

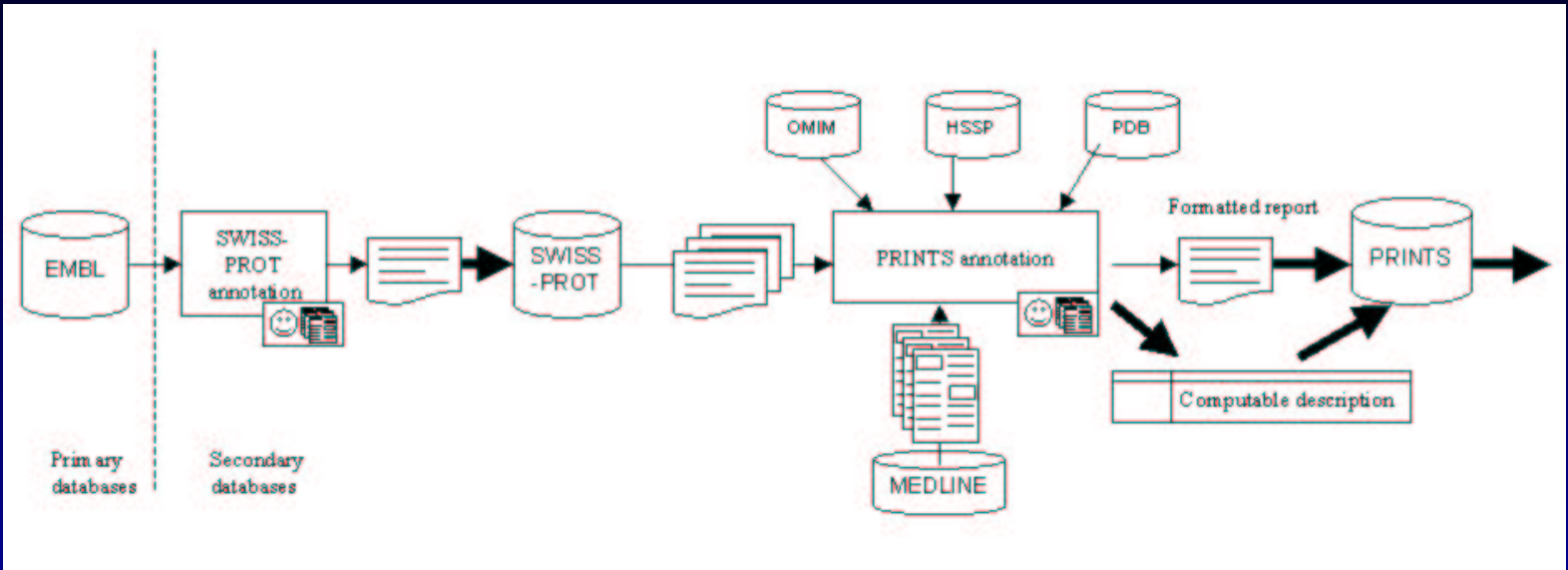
A SWISS-PROT entry

```
ID   PRIO_HUMAN     STANDARD;      PRT;   253 AA.
AC   P04156;
DT   01-NOV-1986 (Rel. 03, Created)
DT   01-NOV-1986 (Rel. 03, Last sequence update)
DT   20-AUG-2001 (Rel. 40, Last annotation update)
DE   Major prion protein precursor (PrP) (PrP27-30) (PrP33-35C) (ASCR).
GN   PRNP.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=86300093; PubMed=3755672;
RA   Kretzschmar H.A., Stowring L.E., Westaway D., Stubblebine W.H.,
RA   Prusiner S.B., Dearmond S.J.;
RT   "Molecular cloning of a human prion protein cDNA.";
RL   DNA 5:315-324(1986).
RN   [2]
RP   SEQUENCE OF 8-253 FROM N.A.
RX   MEDLINE=86261778; PubMed=3014653;
RA   Liao Y.-C.J., Lebo R.V., Clawson G.A., Smuckler E.A.;
RT   "Human prion protein cDNA: molecular cloning, chromosomal mapping,
RT   and biological implications.";
RL   Science 233:364-367(1986).
RN   [3]
RP   SEQUENCE OF 58-85 AND 111-150 (VARIANT AMYLOID GSS).
RX   MEDLINE=91160504; PubMed=1672107;
RA   Tagliavini F., Prelli F., Ghiso J., Bugiani O., Serban D.,
RA   Prusiner S.B., Farlow M.R., Ghetti B., Frangione B.;
RT   "Amyloid protein of Gerstmann-Straussler-Scheinker disease (Indiana
RT   kindred) is an 11 kd fragment of prion protein with an N-terminal
RT   glycine at codon 58.";
RL   EMBO J. 10:513-519(1991).
RN   [4]
RP   STRUCTURE BY NMR OF 118-221.
RX   MEDLINE=20359708; PubMed=10900000;
RA   Calzolai L., Lysek D.A., Guntert P., von Schroetter C., Riek R.,
RA   Zahn R., Wuethrich K.;
RT   "NMR structures of three single-residue variants of the human prion
RT   protein.";
RL   Proc. Natl. Acad. Sci. U.S.A. 97:8340-8345(2000).
CC   -!- FUNCTION: THE FUNCTION OF PRP IS NOT KNOWN. PRP IS ENCODED IN THE
CC       HOST GENOME AND IS EXPRESSED BOTH IN NORMAL AND INFECTED CELLS.
CC   -!- SUBUNIT: PRP HAS A TENDENCY TO AGGREGATE YIELDING POLYMERS CALLED
CC       "RODS".
CC   -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC   -!- POLYMORPHISM: THE FIVE TANDEM OCTAPEPTIDE REPEATS REGION IS HIGHLY
CC       UNSTABLE. INSERTIONS OR DELETIONS OF OCTAPEPTIDE REPEAT UNITS ARE
CC       ASSOCIATED TO PRION DISEASE.
CC   -!- DISEASE: PRP IS FOUND IN HIGH QUANTITY IN THE BRAIN OF HUMANS AND
CC       ANIMALS INFECTED WITH NEURODEGENERATIVE DISEASES KNOWN AS
CC       TRANSMISSIBLE SPONGIFORM ENCEPHALOPATHIES OR PRION DISEASES, LIKE:
CC       CREUTZFELDT-JAKOB DISEASE (CJD), GERSTMANN-STRAUSSLER SYNDROME
CC       (GSS), FATAL FAMILIAL INSOMNIA (FFI) AND KURU IN HUMANS; SCRAPIE
CC       IN SHEEP AND GOAT; BOVINE SPONGIFORM ENCEPHALOPATHY (BSE) IN
CC       CATTLE; TRANSMISSIBLE MINK ENCEPHALOPATHY (TME); CHRONIC WASTING
CC       DISEASE (CWD) OF MULE DEER AND ELK; FELINE SPONGIFORM
CC       ENCEPHALOPATHY (FSE) IN CATS AND EXOTIC UNGULATE ENCEPHALOPATHY
CC       (EUE) IN NYALA AND GREATER KUDU. THE PRION DISEASES ILLUSTRATE
CC       THREE MANIFESTATIONS OF CNS DEGENERATION: (1) INFECTIOUS (2)
CC       SPORADIC AND (3) DOMINANTLY INHERITED FORMS. TME, CWD, BSE, FSE,
CC       EUE ARE ALL THOUGHT TO OCCUR AFTER CONSUMPTION OF PRION-INFECTED
CC       FOODSTUFFS.
DR   EMBL; M13667; AAA19664.1; -.
DR   EMBL; M13899; AAA60182.1; -.
DR   EMBL; D00015; BAA00011.1; -.
DR   PIR; A05017; A05017.
DR   PIR; A24173; A24173.
DR   PIR; S14078; S14078.
DR   PDB; 1E1G; 20-JUL-00.
DR   PDB; 1E1J; 20-JUL-00.
DR   PDB; 1E1P; 20-JUL-00.
DR   PDB; 1E1S; 21-JUL-00.
DR   PDB; 1E1U; 20-JUL-00.
DR   PDB; 1E1W; 20-JUL-00.
DR   MIM; 176640; -.
DR   MIM; 123400; -.
DR   MIM; 137440; -.
DR   MIM; 245300; -.
DR   MIM; 600072; -.
DR   MIM; 604920; -.
DR   InterPro; IPR000817; Prion.
DR   Pfam; PF00377; prion; 1.
DR   PRINTS; PR00341; PRION.
DR   SMART; SM00157; PRP; 1.
DR   PROSITE; PS00291; PRION_1; 1.
DR   PROSITE; PS00706; PRION_2; 1.
KW   Prion; Brain; Glycoprotein; GPI-anchor; Repeat; Signal;
KW   3D-structure; Polymorphism; Disease mutation.
FT   SIGNAL       1    22
FT   CHAIN       23   230       MAJOR PRION PROTEIN.
FT   PROPEP     231   253       REMOVED IN MATURE FORM (BY SIMILARITY).
FT   LIPID       230   230       GPI-ANCHOR (BY SIMILARITY).
FT   CARBOHYD    181   181       N-LINKED (GLCNAC...) (PROBABLE).
FT   DISULFID    179   214       BY SIMILARITY.
FT   DOMAIN       51    91       5 X 8 AA TANDEM REPEATS OF P-H-G-G-G-W-G-
FT                                Q.
FT   REPEAT       51    59       1.
FT   REPEAT       60    67       2.
FT   REPEAT       68    75       3.
FT   REPEAT       76    83       4.
FT   REPEAT       84    91       5.
FT   VARIANT     102   102       P -> L (IN GSS AND EOAD).
FT                                /FTId=VAR_006464.
FT   VARIANT     105   105       P -> L (IN GSS).
FT                                /FTId=VAR_006465.
FT   VARIANT     117   117       A -> V (LINKED TO DEVELOPMENT OF
FT                                DEMENTING GSS).
FT                                /FTId=VAR_006466.
FT   VARIANT     129   129       M -> V (DETERMINES THE DISEASE PHENOTYPE
FT                                IN PATIENTS WHO HAVE A PRP MUTATION AT
FT                                CODON 178: PATIENTS WITH MET DEVELOP FFI,
FT                                THOSE WITH VAL DEVELOP CJD).
FT                                /FTId=VAR_006467.
FT   VARIANT     171   171       N -> S (IN SCHIZOAFFECTIVE DISORDER).
FT                                /FTId=VAR_006468.
FT   VARIANT     178   178       D -> N (IN FFI AND CJD).
FT                                /FTId=VAR_006469.
FT   VARIANT     180   180       V -> I (IN CJD).
FT                                /FTId=VAR_006470.
FT   VARIANT     183   183       T -> A (IN FAMILIAL SPONGIFORM
FT                                ENCEPHALOPATHY).
FT                                /FTId=VAR_006471.
FT   VARIANT     187   187       H -> R (IN GSS).
FT                                /FTId=VAR_008746.
FT   VARIANT     188   188       T -> K (IN EOAD; DEMENTIA ASSOCIATED TO
FT                                PRION DISEASES).
FT                                /FTId=VAR_008748.
FT   VARIANT     188   188       T -> R.
FT                                /FTId=VAR_008747.
FT   VARIANT     196   196       E -> K (IN CJD).
FT                                /FTId=VAR_008749.
FT                                /FTId=VAR_006472.
SQ   SEQUENCE   253 AA;  27661 MW;  43DB596BAAA66484 CRC64;

     MANLGCWMLV LFVATWSDLG LCKKRPKPGG WNTGGSRYPG QGSPGGNRYP PQGGGGWGQP
     HGGGWGQPHG GGWGQPHGGG WGQPHGGGWG QGGGTHSQWN KPSKPKTNMK HMAGAAAAGA
     VVGGLGGYML GSAMSRPIIH FGSDYEDRYV RENMHRYPNQ VYYRPMDEYS NQNNFVHDCV
     NITIKQHTVT TTTKGENFTE TDVKMMERVV EQMCITQYER ESQAYYQRGS SMVLFSSPPV
     ILLISFLIFL IVG

//
```

The annotation pipeline



A PRINTS entry therefore represents a collation of multiple related SWISS-PROT entries (and multiple other data sources)

THE UNIVERSITY of MANCHESTER

Problem

- Annotation generation is labour intensive
- It’s often hard to trace lines of evidence.

Therefore, whilst recognising that human annotation is of higher “quality”, automated annotation systems are still required.

PRECIS Objectives

- Analyse how PRINTS is currently formed
- Generate software tools to mimic this.

Therefore we are trying to generate an annotation transformation tool, examining and extracting commonality between multiple SWISS-PROT entries. We want to use simple techniques, and see how far we can get with these.

PRINTS:- Highs and Lows

Low Level Annotation

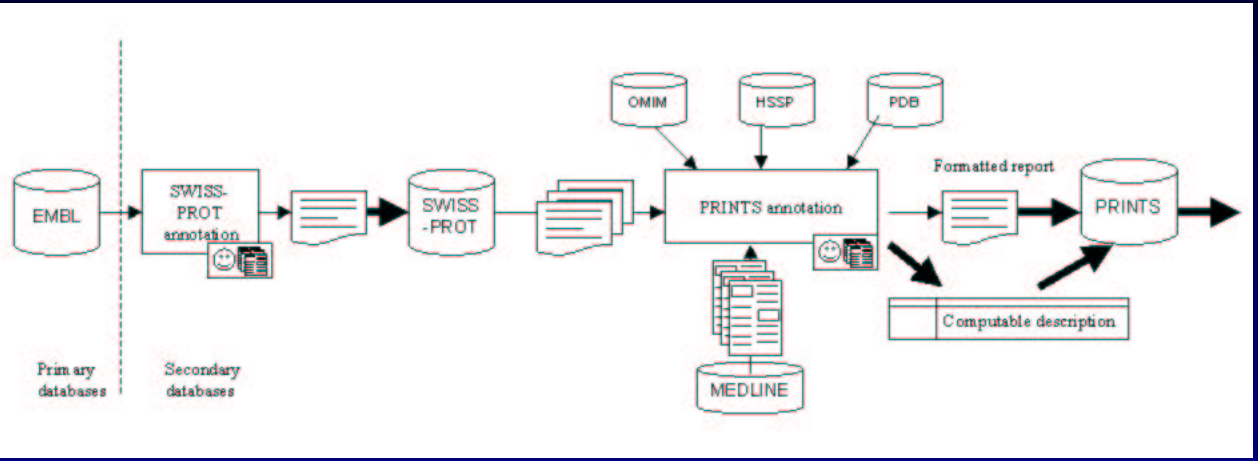
Prion protein signature
PROSITE; PS00291 PRION_1; PS00706 PRION_2
BLOCKS; BL00291
PFAM; PF00377 prion
INTERPRO; IPR000817
1. STAHL, N. AND PRUSINER, S.B.
Prions and prion proteins.
FASEB J. 5 2799-2807 (1991).

High Level Annotation

Prion protein (PrP) is a small glycoprotein found in high quantity in the brain of animals infected with certain degenerative neurological diseases, such as sheep scrapie and bovine spongiform encephalopathy (BSE), and the human dementias Creutzfeldt-Jacob disease (CJD) and Gerstmann-Straussler syndrome (GSS).

THE UNIVERSITY of MANCHESTER

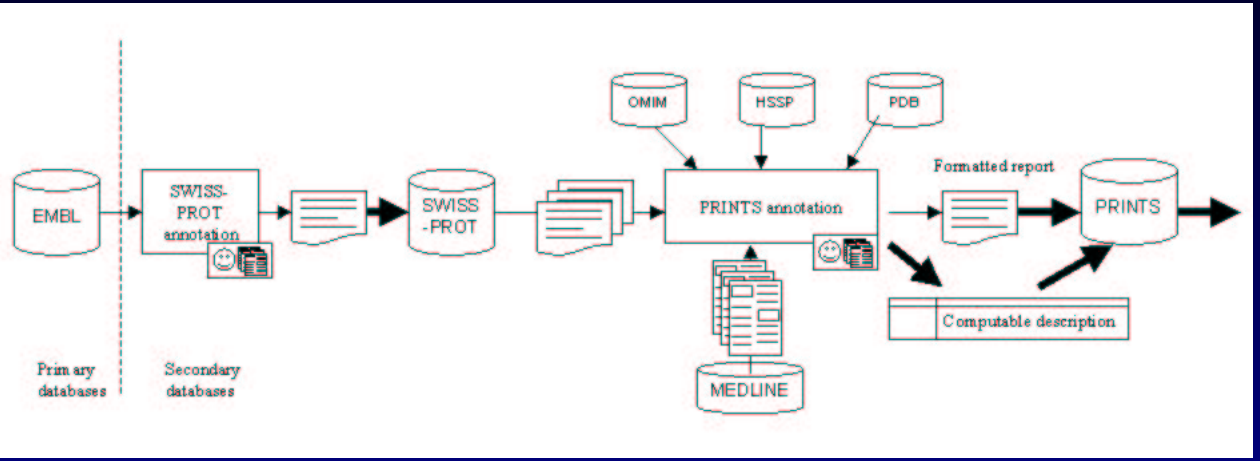
Other systems



- GeneQuiz (Hoersch et al. 2000)
- MAGPIE (Gaasterland et al. 1996)
- PEDANT (Frishman et al. 2001)
- EditToTrembl (Moller et al. 1999)

THE UNIVERSITY of MANCHESTER

Other systems



- GeneQuiz (Hoersch et al. 2000)
- MAGPIE (Gaasterland et al. 1996)
- PEDANT (Frishman et al. 2001)
- EditToTrembl (Moller et al. 1999)

These systems are mostly concerned with the first part of this pipeline.

THE UNIVERSITY of MANCHESTER

Knowledge

Where can we get knowledge from?

- Knowledge from Database structure.
- Knowledge from Words.
- Knowledge from Domain knowledge.

AC P04156;
DR InterPro; IPR000817; Prion.
CC -!- DISEASE: PRP IS FOUND IN HIGH QUANTITY

Knowledge

Where can we get knowledge from?

- Knowledge from Database structure.
- Knowledge from Words.
- Knowledge from Domain knowledge.

Easy (Selley et al. 2001)
Protein Annotators Assistant (Wise et al. 2001)
AbXtract (Andrade and Valencia, 1998)

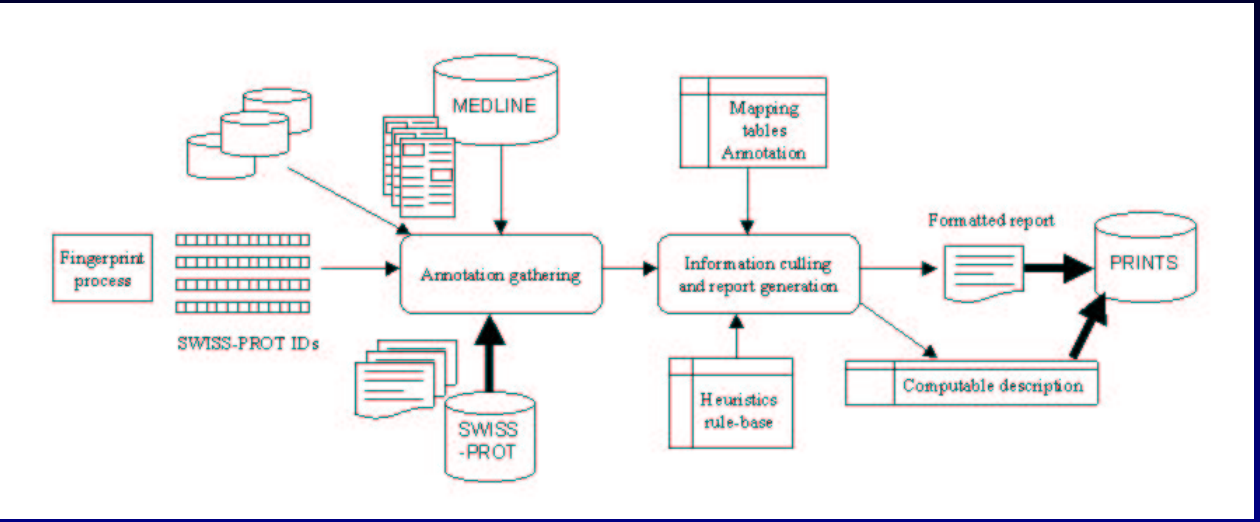
Knowledge

Where can we get knowledge from?

- Knowledge from Database structure.
- Knowledge from Words.
- Knowledge from Domain knowledge.

Implicit within the application
Explicit in an ontology
Combined automated and human annotation

Precis Phases



- Fingerprint formation, SWISS-PROT ID gathering.
- Annotation Gathering
- Information Harvesting
- Report Generation

THE UNIVERSITY of MANCHESTER

Harvesting and Generation

- Ranking
- Redundancy Checks
- Heuristics

Weighting:- majority or golden voting

References weighted by keywords.
Some databases preferred over others.
Newer links are preferred over older.

Harvesting and Generation

- Ranking
- Redundancy Checks
- Heuristics

OPSD_SHEEP	DR PRINTS; PR00237; GPCRRHODOPSN.
OPSD_HUMAN	DR PRINTS; PR00237; GPCRRHODOPSN.
OPSD_MOUSE	DR PRINTS; PR00237; GPCRRHODOPSN.

Harvesting and Generation

- Ranking
- Redundancy Checks
- Heuristics

OPSD_SHEEP	VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT MEDIATE VISION
OPSD_HUMAN	VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT MEDIATE VISION
OPSD_MOUSE	VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT MEDIATE VISION

Harvesting and Generation

- Ranking
- Redundancy Checks
- Heuristics

ACM1_HUMAN	Primary transducing effect is *pi turnover.*
ACM4_HUMAN	Primary transducing effect is *inhibition of adenylate cyclase.*
ACM2_HUMAN	Primary transducing effect is *adenylate cyclase inhibition.*

Harvesting and Generation

- Ranking
- Redundancy Checks
- Heuristics

SWISS-PROT identifiers

The first half of the identifier is the same in similar proteins. The second half records the species.

PRECIS Output:- Databases

Majority Voting

Major prion protein precursor (PRP)

PRINTS; PR00341 PRION
PROSITE; PS00291 PRION_1; PS00706 PRION_2
PFAM; PF00377 prion
INTERPRO; IPR000817
PDB; 1B10; 1AG2

PRECIS Output:- References

Date Ranking, Heuristics

1. CERVENAKOVA, L., [...]
Infectious amyloid precursor gene sequences in pri-
mates used for experimental transmission of
human spongiform encephalopathy.
PROC.NATL.ACAD.SCI.USA 91 12159-12162 (1994).

2. LOWENSTEIN, D.H., [...]
Three hamster species with different scrapie incuba-
tion times and neuropathological features
encode distinct prion proteins.
MOL.CELL.BIOL. 10 1153-1163 (1990).

3. KALUZ, S., [...]
Sequencing analysis of prion genes from red deer and camel.
GENE 199 283-286 (1997).

THE UNIVERSITY of MANCHESTER

PRECIS Output:- Description

Majority Voting

The function of prp is not known. Prp is encoded in the host genome and is expressed both in normal and infected cells.

Attached to the membrane by a gpi-anchor.

PRECIS Output:- Disease

Heuristic, Golden Voting

```
(PRIO_HUMAN; P04156): Prp is found in high quan-
tity in the brain of humans and animals infected
with neurodegenerative diseases known as transmissible spongi-
form encephalopathies or prion
diseases [...]

(PRIO_HUMAN; P04156): Kuru is transmitted during ritualistic can-
nibalism, among natives of the new
guinea highlands. [...]

(PRIO_SHEEP; P23907): Polymorphism at position 171 may be re-
lated to the alleles of scrapie [...]
```

PRECIS Output:- The rest

Majority Voting

The structure has been determined, e.g. "NMR characterization of the full-length recombinant murine prion protein, mPrP(23-231)" [5].

Belongs to the prion family.

Keywords: GPI-anchor; Repeat; Signal; Prion; Brain; Glycoprotein; Polymorphism; Disease mutation; 3D-structure.

PRECIS strengths

- Clear English-like reports
- Retains context information
- Provides some provenance information.
- Results are update-able easily.

PRECIS weaknesses

- Inherits SWISS-PROT errors.
- Only uses SWISS-PROT as a data source
- Many problems with free text. Particularly redundancy decisions.

Future directions

- Improve reference selection. Perhaps automatic term recognition and clustering.
- Improve implementation. Make more “pluggable”.
- Structured metadata layer within PRECIS output.

Acknowledgements

- Jacqueline Reich
- Alex Mitchell
- Robert Stevens
- Terri Attwood
- Carole Goble