# Track Guidelines

## INTRODUCTION

'Machine reading of biomedical texts about Alzheimer's Disease' is a pilot task of the Question Answering for Machine Reading Evaluation (QA4MRE) at CLEF 2012. The task follows the same set up and principles as the QA4MRE, with the difference that it focuses on the biomedical domain. More information about the QA4MRE can be found in the QA4MRE2012 Track Guidelines to be downloaded from http://celct.fbk.eu/QA4MRE/.

This pilot task aims at exploring the ability of a machine reading system (Etzioni et al. 2006, Strassel et al. 2010) to answer questions about a scientific topic, namely the Alzheimer's Disease (AD). AD has been chosen as the focus of the task because there is a particular interest in more efficient processing of Alzheimer-related literature, as this condition constitutes a considerable health challenge for an aging population (Citron 2010). The increasing importance of AD is reflected in the recently approved US National Alzheimer's Project Act, which will result in considerable funding being made available for research on this disease and for financing better data infrastructure resources. The illness is being analyzed from various perspectives in a growing number of scientific studies (Al-Mubaid and Singh 2005, Li et al. 2009, Barbosa-Silva et al. 2011).

## TASK DESCRIPTION

As in the main task, participant systems are asked to read a document and identify the answers to a set of questions about information that is stated or implied in the text. Questions are in the form of multiple choice, each having five options, and only one correct answer. The detection of correct answers is specifically designed to require various kinds of inference and the consideration of previously acquired background knowledge from reference document collections provided by the organization. Although the additional knowledge obtained through the background collection may be used to assist with answering the questions, the principal answer is to be found among the facts contained in the test documents given.

Participants will be provided with a background collection and test documents about Alzheimer's Disease. To solve the task, participants can make use of existing resources, such as ontologies or databases, and tools, such as named entity taggers, event extractors, parsers, etc. In order to keep the task reasonably simple for systems, the task organizers will provide the texts of the background collection and the test documents processed at several levels of linguistic analysis (lemmas, part-of-speech, named entities, chunking, dependency parsing). Publicly available state of the art tools will be used for this purpose.

## BACKGROUND COLLECTION

The background collection is a collection of texts about Alzheimer's Disease called the Alzheimer's Disease Literature Corpus (ADLC corpus). Systems can use it to acquire reading capabilities and to obtain knowledge about Alzheimer's Disease that can help answering the questions about the test documents. It has been carefully selected to be as specific as possible for this topic and should constitute a comprehensive resource for this task in particular and for text mining efforts tailored to the Alzheimer's Diseases field in general.

The background collection will be released at the beginning of April, subject to signing a license agreement. Although the use of the background collection is recommended, it is not mandatory. It can be downloaded from the following link:
http://celct.fbk.eu/ResPubliQA/index.php?page=Pages/bg_collection_pilot.php

The collection contains the following sets of documents:

**(1) PubMed abstracts.**

66,222 abstracts obtained by performing the following search in PubMed:

((((("Alzheimer Disease"[Mesh] OR "Alzheimer's disease antigen"[Supplementary Concept] OR "APP protein, human"[Supplementary Concept] OR "PSEN2 protein, human"[Supplementary Concept] OR "PSEN1 protein, human"[Supplementary Concept]) OR "Amyloid beta-Peptides"[Mesh]) OR "donepezil"[Supplementary Concept]) OR ("gamma-secretase activating protein, human"[Supplementary Concept] OR "gamma-secretase activating protein, mouse"[Supplementary Concept])) OR "amyloid beta-protein (1-42)"[Supplementary Concept]) OR "Presenilins"[Mesh]) OR "Neurofibrillary Tangles"[Mesh] OR "Alzheimer's disease"[All Fields] OR "Alzheimer's Disease"[All Fields] OR "Alzheimer s disease"[All Fields] OR "Alzheimers disease"[All Fields] OR "Alzheimer's dementia"[All Fields] OR "Alzheimer dementia"[All Fields] OR "Alzheimer-type dementia"[All Fields] NOT "non-Alzheimer"[All Fields] NOT ("non-AD"[All Fields] AND "dementia"[All Fields]) AND (hasabstract[text] AND English[lang])

The abstracts are provided in XML format, and with the annotations described in Section Annotations.

**(2) Open Access full articles PMC.**

8,249 Open Access full articles from PubMed Central in PDF format. These articles have been selected by performing the search indicated above and selecting the full articles that belong to the PubMed Central Open Access subset and that were available on 1.03.2012. 7,512 of these articles are provided in text format, which has been obtained by converting the PDF files into text by using the tool LA-PDFText (http://code.google.com/p/lapdftext/)(Ramakrishnan et al. 2012). 7,447 of these articles are also provided with annotations.

**(3) Open Access full articles PMC, smaller collection.**

1,041 full text articles from PubMed Central in HTML and text format. The articles are also provided with annotations. For this articles the text version has been converted from the PubMed HTML version. To select these documents a search was performed on PubMed using Alzheimer's Disease related keywords and restricting the search to the last three years. The search was performed on 3.02.2012. Only a subset of the articles obtained by the search has been included in the collection.

**(4) Elsevier full articles.**

379 full text articles from Elsevier and 103 abstracts. The documents are provided in XML and text format. They are also provided with annotations. The text files have been obtained by converting the XML files into text. The articles in this subset have been selected from a list of articles provided by Professor Tim Clark from the Massachusetts Alzheimer's Disease Research

Center, USA. The list contains bibliographic records representing 45 core hypotheses in Alzheimer's Disease. Elsevier kindly provided the articles from this list that were Elsevier publications.

The XML-based text content of articles is represented as a block of two or three lines in the plain text files (.txt), where the first line always contains the actual content, while the second and third line contain useful keys and qualifiers to identify the corresponding source in the original file. Each block is separated from the next by an empty newline. A block (commonly, a paragraph, or a caption, etc.) consists of one to three lines:

<SECTION_TYPE> [text content]
KEY [optional keywords]
QUALIFIER [optional qualifiers]

Where <SECTION_TYPE> can be one of the following values:

ABSTRACT, APPENDIX, ARTICLE-CLASS, BODY, DEFINITION, FIGURE, KEYWORD, TABLE, TEXTBOX, TITLE, GLOSSARY
The main article content is found in the TITLE, ABSTRACT, and BODY sections, plus some FIGURE captions. Other sections are optional and not every article will contain them. 103 articles do not have a BODY section. These articles do have unstructured raw text content. See "Elsevier_raw_text_files.tar.gz" in the download site for the list of XML files that contain a raw-text element.
All text content appears in exactly the same order as it was found in the XML file. The qualifiers contain the section and subsection headings, while KEYs are only added for DEFINITION, KEYWORD, and GLOSSARY sections. By default, only the TITLE section does not have any qualifier at all. Author names and bibliographic references are not extracted from the XML into the raw text files.

## TEST DATA

The test set will be composed of 4 reading tests. Each reading test will consist of one document, with 10 questions and a set of five choices per question. So, there will be in total 40 questions and 200 choices/options. Participating systems will be required to answer these 40 questions by choosing in each case one answer from the five alternatives. There will always be one and only one correct option. Systems should answer all questions.

The test documents are selected from a list of bibliographic records provided by professor Tim Clark from the Massachusetts Alzheimer's Disease Research Center, USA. The records represent 45 core hypotheses in Alzheimer's disease. The records were compiled in 2011.

The test documents are provided in text format. We perform first an automatic conversion from pdf and then we correct conversion errors manually, paying attention to symbols that express relevant information about Alzheimer's Disease. The captions of figures and tables are also included, but the figures and tables not. Participants are not expected to process the contents of tables and figures. A sample of a test document with questions can be downloaded from the QA4MRE website. The sample test is provided in .txt and .pdf format, and preprocessed.

An example test can be downloaded from the main task website, as well as the preprocessed version.

## ANNOTATIONS

The documents in the background collection and the test documents are provided with annotations in a column format as shown in Figure 1.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 22319430 | 1 | 1 | HD | HD | B-NP | NN | B-protein | 4 | NMOD | B-XHD_amino_acid_duplex | B-PROTEIN |
| 22319430 | 1 | 2 | amino | amino | I-NP | JJ | I-protein | 4 | NMOD | I-XHD_amino_acid_duplex | I-PROTEIN |
| 22319430 | 1 | 3 | acid | acid | I-NP | NN | I-protein | 4 | NMOD | I-XHD_amino_acid_duplex | I-PROTEIN |
| 22319430 | 1 | 4 | duplex | duplex | I-NP | NN | I-protein | 5 | SUB | I-XHD_amino_acid_duplex | I-PROTEIN |
| 22319430 | 1 | 5 | has | have | B-VP | VBZ | O | 0 | ROOT | O | O |
| 22319430 | 1 | 6 | been | be | I-VP | VBN | O | 5 | VC | O | O |
| 22319430 | 1 | 7 | found | find | I-VP | VBN | O | 6 | VC | O | O |
| 22319430 | 1 | 8 | in | in | B-PP | IN | O | 7 | VMOD | O | O |
| 22319430 | 1 | 9 | the | the | B-NP | DT | O | 11 | NMOD | O | O |
| 22319430 | 1 | 10 | active | active | I-NP | JJ | O | 11 | NMOD | O | O |
| 22319430 | 1 | 11 | center | center | I-NP | NN | O | 8 | PMOD | O | O |
| 22319430 | 1 | 12 | of | of | B-PP | IN | O | 11 | NMOD | O | O |
| 22319430 | 1 | 13 | many | many | B-NP | JJ | O | 12 | PMOD | B-Xmany_different_enzyme | O |
| 22319430 | 1 | 14 | different | different | I-NP | JJ | O | 15 | NMOD | I-Xmany_different_enzyme | O |
| 22319430 | 1 | 15 | enzymes | enzyme | I-NP | NNS | B-protein | 13 | NMOD | I-Xmany_different_enzyme | B-PROTEIN |
| 22319430 | 1 | 16 | . | . | O | . | O | 5 | P | O | O |

Figure 1. Example sentence of an annotated abstract.

The annotations have been obtained with the dependency parser GDep (Sagae and Tsujii 2007), a UMLS-based NE tagger developed at CLiPS, and the ABNER NE tagger (Settles 2005). The content of the columns is the following:

- Column 1: Document identifier.
- Column 2: Sentence number in the document.
- Column 3: Token number in the sentence.
- Column 4: Word (GDep parser).
- Column 5: Lemma (GDep parser).
- Column 6: Chunk tag (GDep parser).
- Column 7: Part-of-speech tag (GDep parser).
- Column 8: Named entity (GDep parser).
- Column 9: Parent node in the dependency syntaxt tree (GDep parser).
- Column 10: Dependency syntax label (GDep parser).
- Column 11: UMLS named entity (CLiPS NE Tagger).
- Column 12: Named entity (ABNER tagger).

## LANGUAGE

All the materials will be provided in English.

## QUESTIONS

As in the QA4MRE, questions are in multiple choice format and focus on testing the comprehension of one single document. Example questions can be found in Appendix I.

The questions posed for this task should address aspects that are of biomedical relevance and that have been proven to be of importance in the context of previous efforts such as BioCreative [http://www.biocreative.org], Genomics TREC track [http://ir.ohsu.edu/genomics] or the BioNLP shared tasks [http://sites.google.com/site/bionlpst]. This will enable participants to make use of resources developed for these competitions and will establish a link between this pilot task and previous efforts. Additionally, since machine reading of biomedical texts is a new task, it is better to restrict the types of questions somehow. Therefore a restricted set of named entity types associated to the questions has been defined as well as a list of question types. The expected answer types will depend on allowed entity types.

## Named Entities

The categories of named entities considered for this task are

| | |
|---|---|
| GENE_PROT | Genes and gene products (proteins, mRNA). |
| CHEM_DRUG | Chemicals/drugs/pharmacological agents. |
| DIS_SYMPT | Disease/symptoms |
| EXP_METHOD | Experimental method/qualifier |
| SPEC_ORG | Species/organism |
| PATH_PROC | Pathway/Biological process |
| ANAT_CELL | Anatomical/cellular/subcellular structures |
| MUT_PTM | Mutations/genetic variations/posttranslational modifications |
| ADV_TOXIC | Adverse effect/toxic endpoints |
| DOSE | Dose of a given treatment |
| TIMING | Schedule of treatments (timing) |
| PAT_CHAR | Patient characteristics: age, gender, sex, race, population, animal strain |
| MOL_MARKER | Molecular marker |

In order to identify the named entities above, the following lexico-semantic resources and tools can/will be used (among others): ABNER, BANNER, Genia Tagger, BioThesaurus, BioLexicon,UMLS, LINNAEUS tagger, OrganismTagger, MeSH, Gene Ontology (and other ontologies from OBO), etc... .

The test documents will be processed with these tools before making the questions, so that questions refer only to entities that can be automatically identified with existing resources. The background collection will also be processed with these tools.

## Question Types

Based on examination of the relationships between the various entity types we compiled a collection of biomedically relevant question types.

- Experimental evidence/qualifier. This question type refers to experimental techniques, methods or models used to generate or validate a given discovery. Examples include animal models used for a given in vivo study, interaction detection methods used to detect protein interactions, imaging techniques for visualization or localization of a particular protein.

- Protein-protein interaction. This questiontype refers to the detection of an interaction partner of a given protein. Examples include physical binding of two proteins in a protein-protein complex or more transient interaction in phosphorylation of one protein by another.

- Gene synonymy relation. This question type tries to establish relations between two entity mentions of genes or proteins that refer actually to the same biological entity. For instance this relation exists between 'APP' and 'amyloid beta (A4) precursor protein'. Here alternative aliases of a gene name or symbol are included, as well as typographical variants and acronym and their corresponding expanded forms.

- Organism source relation. This question type refers to the actual organism source for a given protein or gene. Example includes genes encoded in the human genome or expressed in humans.

- Regulatory relation: This question type refers to gene regulatory relationships between two bio-entities (protein and gene), i.e. whether one bio-entity affects the gene expression of another entity (e.g. transcription factor target gene relation).

- Increase (improvement, higher expression). This is a more specific question type of the regulatory relation. It refers to cases where one bio-entity causes the upregulation (increased expression) of another bio-entity.

- Decrease (depletion, reduction). This is a more specific question type of the regulatory relation. It refers to cases where one bio-entity causes the downregulation (decreased expression) of another bio-entity.

- Inhibition/disruption/impaired. This question type refers to cases were one bio-entity blocks or inhibits another bio-entity. Examples include drugs blocking a given protein or enzyme, or proteins that inhibit a particular biological process or pathway.

## Degrees of difficulty

Questions can be assigned a degree of difficulty: simple, medium and complex. A collection of criteria for question difficulty classification has been followed.

- *Simple*: factual questions that can be answered using information from the target document and whose textual evidence is contained multiple times in the paper, e.g. several text snippets are supporting the correct answer. The answer is found almost verbatim in the paper.

- *Medium*: The correct answer is phrased in a way that requires the use of lexico-semantic dictionaries and name alias recognition capabilities to be able to handle lexico-semantic alienations of keywords and entities.
- *Complex*: Reasoning must be applied to answer this question. Choosing the correct answer requires combining pieces of evidence. Such questions might need ad hoc axiomatic knowledge and abductive processes.

Other aspects that influence question difficulty include:

— Are the ontological relations encoded in the question? If they are encoded the question should be easier.
— If keyword-based indexing and conceptual indexing are required the question is less easy.
— Script like questions such as 'how is an anatomical structure assembled?' should be more difficult since answering them requires combining several units of information.
— Template questions about successive temporal events (biological processes, disease stages) should be more difficult since it also requires several units of information.
— Is it necessary to process morphological alternations such as 'phosphorylate' lexicalized as the nominalization 'phosphorylation'? In this case the degree of difficulty should be simple/medium, depending on other characteristics of the question.
— Is it necessary to process lexical alternations? The usage of synonyms or semantically related terms derived from ontologies is necessary to increase the recall.
— Is it necessary to process semantic alternations and paraphrases? This includes relations between multi-term paraphrases and single terms, textual patterns, complex examination between words building terms within the ontology.
— Is it necessary to process terminological variants and high level indexes comprising terms and their variants for retrieval? A variant recognition module is required as well as weighting of matching between questions and documents.
— Paragraph window size of the evidence text and whether it is a continuous span of text. The bigger the window size, the more difficult is the question. Non continuous is more difficult than continuous.

## ANSWERS

As in the main task, systems are not required to answer every question, since the c@1 measure (Peñas and Rodrigo 2011) will be used. This measure encourages systems to reduce the number of incorrect answers while maintaining the number of correct ones by leaving some questions unanswered. More information about the nswers can be found in the Track Guidelines of the main task.

## RUNS

As in the main task, participants are allowed to submit a maximum of 10 runs. Each run must be categorized as one of the following types, depending on the resources that have been used to assist in asnwering the questions:

- No external resource is used (only test document);
- Only the test document and the associated background collection is used;
- The test document and other resources are used, but not the background collection;

- The test document together with the background collection and other resources are used.

The type used can be indicated in the submission file (see Output Format in the Track Guidelines of the main task).

## FORMATS

The DTD for the input and output format is the same as for the main taks and it can be downloaded from the QA4MRE website [http://celct.fbk.eu/QA4MRE/]. More information about the format can be found in the Track Guidelines of the main task.

## EVALUATION

Evaluation will follow the same automatic procedure as in the main task. More information can be found in the Track Guidelines. The c@1 (Peñas and Rodrigo 2011) measure will be used.

## IMPORTANT DATES

| | |
|---|---|
| Release of backgound collections | April 19 |
| Test set release | June 5 |
| Run submissions | June 15 |
| Individual results to participants | June 25 |
| Submissions of Working Notes Papers | August 17 |
| CLEF Workshop | September 17-20, Rome, Italy |

As in the main task, test datasets will be available on the QA4MRE website [http://celct.fbk.eu/QA4MRE/] on June 5, and submissions will be due within 5 days from the first test set download and not later than June 15 by 11:59 p.m. (CEST). Late submissions will not be considered.

Participant runs will be submitted following the same procedure as for the main task (see Track Guidelines from the main task).

## ORGANISERS

- Roser Morante, Walter Daelemans - CLiPS, University of Antwerp, Belgium
- Martin Krallinger and Alfonso Valencia - CNIO, Madrid, Spain

## TECHNICAL SUPPORT

- Vincent Van Asch - CLiPS, University of Antwerp, Belgium
- Florian Leitner - CNIO, Madrid, Spain
- Cartic Ramakrishnan - Information Sciences Institute of the University of Southern California, USA
- Gully A.P.C. Burns - Information Sciences Institute of the University of Southern California, USA

## DOMAIN ADVISOR

- Tim Clark, Massachusetts Alzheimer's Disease Research Center, USA

## DATA PROVIDERS

- Elsevier, PubMed, PubMed Central, Medline

## TECHNICAL MANAGEMENT AND DATA COLLECTION INFRASTRUCTURE

- Anselmo Peñas - IR&NLP Group, UNED, Madrid, Spain
- Eduard Hovy - Information Sciences Institute of the University of Southern California, USA

## TECHNICAL MANAGEMENT AND DATA COLLECTION INFRASTRUCTURE

- Pamela Forner - Giovanni Moretti, CELCT , Italy

## References

Al-Mubaid, H. and R. K. Singh (2009) A New Text Mining Approach for Finding Protein-to-Disease Associations. *American Journal of Biochemistry and Biotechnology* 1(3), pages 145-152.

Barbosa-Silva, A., J.-F. Fontaine, E. R. Donnard, F. Stussi, J. M. Ortega and M. A. Andrade-Navarro (2011) PESCADOR, a web-based tool to assist text-mining of biointeractions extracted from PubMed queries. *BMC Bioinformatics* 2011, 12:435

Chambliss, M.L. and Conley, J. and others (1996) Answering clinical questions. *The Journal of Family Practice* 43(2).

Citron, M. (2010) Alzheimer's disease: strategies for disease modification. *Nature Reviews Drug Discovery* 9, 387-398.

Etzioni, O., M. Banko, and M. J. Cafarella (2006) Machine reading. *Proceedings of the 21st National Conference on Artificial intelligence (AAAI 2006)*.

Etzioni, O., A. Fader, J. Christensen, and S. Soderland (2011) Mausam: Open information extraction: The second generation. In *Proceedings IJCAI 2011* (pp. 3-10).

Lee, M., J. Cimino, H.R. Zhu, C. Sable, V. Shanker, J. Ely, and H. Yu (2006) Beyond information retrieval—medical question answering. *AMIA Annual Symposium Proceedings*. American Medical Informatics Association

Peñas, A. and A. Rodrigo (2011) A simple measure to assess the non-response. In *Proceedings of 49th Annual Meeting of the Association for Computational Linguistics - Human Language Technologies (ACL-HLT 2011)*, Portland, Oregon, USA, June 19-24, 2011.

Ramakrishnan, C., A. Patnia, E. Hovy and G. APC Burns (2012) Layout-Aware Text Extraction from Full-text PDF of Scientific Articles. Accepted for publication in the *Open Research Computation* Journal.

Sagae, K., Tsujii, J. (2007) Dependency parsing and domain adaptation with LR models and parser ensembles. *Proceedings of the CoNLL 2007 Shared Task. Joint Conferences on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL'07)*. Prague, Czech Republic.

Settles, B. (2005) ABNER: An open source tool for automatically tagging genes, proteins and other entity names in texts. *Bioinformatics* 21(14):3191-2.

Strassel, S., D. Adams, H. Goldberg, J. Herr, R. Keesing, D. Oblinger, H. Simpson, R. Schrag, and J. Wright (2010) The DARPA Machine Reading Program - Encouraging Linguistic and Reasoning Research with a Series of Reading Tasks. *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*.

## APPENDIX I: EXAMPLE QUESTIONS

A collection of questions with multiple choice answers is provided below. The questions refer to the following example document:

- C. McAllister et al. (2010) Genetic Targeting Aromatase in Male Amyloid Precursor Protein Transgenic Mice Down-regulates ß-Secretase (BACE1) and Prevents Alzheimer-Like Pathology and Cognitive Impairment. *The Journal of Neuroscience*, 30(21):7326 –7334. PMID 20505099.

The correct answer is the first one. Apart from the question, text snippets are provided that contain evidence for the answer.

**Q1 Which enzyme is responsible for the transformation of testosterone into estrogen?**
A1 aromatase
A2 AD
A3 androgen
A4 BACE1
A5 actinomycin D

*As brain testosterone plays both androgenic and estrogenic actions due to its conversion into estrogen via aromatase naturally, it is unclear that the age-related reduction of testosterone increased risk of Alzheimer's disease (AD) in men is mediated through androgen alone or both androgen and estrogen mechanisms.*

*Our previous studies using a gene-based approach in mouse model to block the conversion of testosterone into estrogen (aromatase gene knock-out, ArKO), found a depletion of estrogen and increase in testosterone endogenously in males.*

*Because testosterone can be converted into estrogen by aromatase in vivo*

*aromatase gene knock-out prevented the conversion of endogenous testosterone into estrogen*

**Q2 What experimental approach is useful to create an in vivo system where conversion of testosterone into estrogen is blocked?**
A1 knock-out of the aromatase gene
A2 immunohistochemical staining
A3 ELISA analysis
A3 Western blot
A5 Testosterone radioimmunoassay

*Our previous studies using a gene-based approach in mouse model to block the conversion of testosterone into estrogen (aromatase gene knock-out, ArKO), found a depletion of estrogen and increase in testosterone endogenously in males.*

*aromatase gene knock-out prevented the conversion of endogenous testosterone into estrogen*

**Q3 What experimental approach was successful to inhibit in vivo the production of amyloid β?**
A1 knock-out of BACE1 gene
A2 amyloid plaque

A3 immunohistochemical analysis
A4 amyloid deposition
A5 immunohistochemistry

*BACE1 gene knock-out has been shown to prevent Aβ synthesis in mice*

**Q4 Which assay was used to determine cognitive performance?**
A1 hole-board memory test
A2 spatial memory
A3 double-transgenic mice
A4 APP23 mice
A5 genetic knock-out

*We evaluated the effect of aging and testosterone levels on the cognitive function of male APP23/Ar$^{+/-}$, APP23, and WT mice using a hole-board memory task as described in previous publications (Dodart et al., 2002; He et al., 2007). Cognitive performance was evaluated by ability of the mice to quickly and successfully to recognize the target hole.*

**Q4 What is the product of the transformation of testosterone carried out by aromatase?**
A1 estrogen
A2 testosterone
A3 aromatization
A4 BACE1
A5 APP

*Because testosterone can be converted into estrogen by aromatase in vivo*

**Q5 Where is BACE1 particularly enriched in the case of AD?**
A1 neurons around Aβ plaques
A2 mice
A3 fluoride membranes
A4 neocortex
A5 hand

*BACE1 activity levels are significantly increased in sporadic AD cases (Fukumoto et al., 2002; Holsinger et al., 2002; Yang et al., 2003; Li et al., 2004), particularly in neurons around Aβ plaques*

**Q6 Which protein is known to remove Aβ from the brain?**
A1 IDE
A2 APP
A3 BACE1
A4 brain
A5 testosterone

*IDE degrades the intracellular domain of APP and is involved in clearing Aβ from the brain*

**Q7 Which protein is thought to be the main one responsible for degrading Aβ?**
A1 NEP
A2 testosterone

A3 BACE1
A4 aromatase
A5 APP

*NEP, whose mRNA and protein levels are significantly decreased in AD brains compared with age-matched normal brains (Yasojima et al., 2001a,b), may be the principal Aβ-degrading enzyme*

**Q8 Which molecular marker is useful to detect neuronal loss in the hippocampus?**
A1 NeuN
A2 Aβ
A3 APP23
A4 CA3 region
A5 AICD

*Eight to 10 sections (~120 μm apart) were immunostained for Aβ (rabbit anti-Aβ-peptide; 1:250; Zymed), neuronal marker NeuN (mouse anti-NeuN; 1:500; Millipore Bioscience Research Reagents), or APP intracellular c-terminal fragment AICD (rabbit anti-APP-CT20; 1:1000; Calbiochem).*

*Immunostaining of NeuN (1:1000) in the brain from an 18-month-old APP23 male mouse showed extensive neuronal loss in the CA3 region of the hippocampus (indicated in box) compared with the hippocampi from WT and APP23/Ar$^{+/-}$ male mice at the same age.*

**Q9 Which hormone is able to inhibit the transcription of BACE1?**
A1 testosterone
A2 IDE
A3 NEP
A4 pB1P-A
A5 APP

*Testosterone downregulates BACE1 both in vivo and in vitro.*

*testosterone-induced reduction of BACE1 protein and activity is mediated through transcriptional mechanisms.*

**Q10 At which age was the expression of BACE1 mRNA significantly lower?**
A1 eighteen month
A2 six month
A3 old mice
A4 age
A5 quickly

*To investigate whether reduced BACE1 protein levels in APP23/Ar$^{+/-}$ mice are partially caused by changes in BACE1 mRNA abundance, we performed RT-PCR and found that BACE1 mRNA levels were significantly decreased in 18-month-old APP23/Ar$^{+/-}$ mice compared with age-matched,*

**Q11 How does NEP promote the decrease of amyloid plaques in the brain?**
A1 Through clearance of Aβ
A2 Through enhancement of Aβ expression
A3 Through increasing expression of IDE
A4 Through standard scanning densitometry

A5 Through development of development of AD-like neuropathology

*reduction of amyloid plaque formation in APP23/Ar$^{+/-}$ male mice is partially mediated by the enhancement of Aβ clearance by NEP and IDE*

**Q12 What substance described in the papers could be used potentially used to treat elderly male patients suffering from AD?**
A1 testostrone
A2 actinomycin D
A3 BACE1
A4 293 cells
A5 estrogen

*Recent studies have suggested that testosterone reduction might be a risk factor for AD in aged males (Hogervorst et al., 2001; Moffat et al., 2004; Rosario et al., 2004; Rosario and Pike, 2008). Consistent with this hypothesis, research data has shown that testosterone treatment can reduce AD neuropathology both in clinical observation (Moffat et al., 2006) and in cell culture research*