

Trilateral Project DR2
Biotechnology

**Trilateral Search Guidebook
in Biotechnology**

Ver.2

28 November, 2007

United States Patent and Trademark Office
European Patent Office
Japan Patent Office

Trilateral Search Guidebook in Biotechnology

I. Introduction

- What is “Trilateral Search Guidebook in Biotechnology”? -

The Trilateral Search Guidebook in Biotechnology is a non-binding reference manual comprised of two parts: the first part describing the databases/search tools most commonly used in Trilateral Office, and the second part providing some representative search strategies used by each office for common types of claim sets presented in biotechnology-related applications.

This Version 1 for publication of The Trilateral Search Guidebook in Biotechnology focuses on common databases and sequence search algorithms in its first part.

- Disclaimer -

A patent examiner is responsible for reviewing prior patent documents, both domestic and foreign, and other printed literature related to an application's subject matter during the examination process. This review, called the search, is performed by consulting the appropriate patent classification schemes, and classified and text searches of other patent document databases, and/or any other printed media (also known as "non-patent literature" or NPL), which might disclose the invention disclosed/claimed in a pending application for a patent. This search may include the use of various search tools or methodologies in the process. Once the search is performed, the examiner determines the patentability of a claimed invention in light of the prior art uncovered by this search.

When determining the appropriate field of search for an invention, the examiner must consider three sources of information: domestic patent documents, foreign patent documents, and NPL. Patent examiners are not required to search every listed resource in the examination of every application. However, an examiner should not eliminate any resource from consideration unless the examiner can justify to a reasonable certainty that no more pertinent references will be found in a further search. The examiner will determine the most appropriate resources to search for each application by relying upon their professional judgment and assessment of the disclosed and claimed subject matter in the application under consideration. What the search guidebook provides is a compilation of the resources found to most frequently discover the most relevant prior art disclosures for the particular subject matter. Searching is highly dependent upon the experience and job knowledge of the individual performing and/or reviewing the search. Prior art to be considered in a search can slightly vary depending on the national patent law. This aspect can also influence the sources used by examiners to perform their search.

II. Main Subject

A. Search Database / Search Tools

1. Patent literature, Bibliography & Reference databases and Search tools

[USPTO]

a. Patent literature and Patent related databases

Main USPTO Patent Data Resources	
Resource Name	Resource Description
U.S .Patents Full-Text (USPAT)	<ul style="list-style-type: none"> ○ Contains full text of most patents issued from 1971-present. ○ Updated weekly as patents are issued. ○ Also includes minimal information (patent number, publication date, and current classification) on all other US patents dating back to 1790.
U.S. Optical Character Recognition Full-Text (USOCR)	<ul style="list-style-type: none"> ○ Full text search of 2.5 million U.S. Patent Grants issued from 1920-1970. ○ Approximately 100,000 patents issued during 1971-1979 that were not electronically captured when the USPAT database was first created.
U.S. Pre-Grant Publication Full-Text (US-PGPUBS)	<ul style="list-style-type: none"> ○ Contains full text of published U.S. patent applications (2001 to the present). ○ Updated weekly as Pre-Grant Publications are published.
Patent Abstracts of Japan (PAJ)	<ul style="list-style-type: none"> ○ Contains English language abstracts of published unexamined Japanese patent applications, issued from 1976 to the present.
Derwent World Patents Index Abstracts (Derwent)	<ul style="list-style-type: none"> ○ Contains English language abstracts of published patent documents from more than 40 patent-issuing authorities, often with family information. ○ Years of coverage vary, normally 1970-present.
European Patent Office	<ul style="list-style-type: none"> ○ Contains published abstracts from the EPO, from Selected EPO member countries and from the World Intellectual Property Organization (WIPO). ○ Provides PDF images for the full text of patents from

Abstracts (EPO)	<p>a number of countries.</p> <ul style="list-style-type: none"> ○ Has text search capabilities in the title and abstract fields as well as retrieval by IPC and ECLA classification numbers.
<p>JPO Industrial Property Digital Library</p> <p>http://www.ipdl.inpit.go.jp/homepg_e.ipdl</p>	<ul style="list-style-type: none"> ○ Database contains Japanese patent information since 1976. ○ Has text search capabilities in the title and abstract fields as well as retrieval by FI and F terms.

b. Search tools for Patent –Related Databases

EAST (Examiner’s Automated Search Tool), a Windows TM-based client and WEST (Web Examiner’s Search Tool), a browser-based interface with a Web-like look and feel are search tools used by USPTO personnel to search for U.S. and Foreign Patents and Pre-Grant Publications. EAST/WEST may be used to perform general patent database searches or to perform searches bound to specific document sections.

PLUS (Patent Linguistics Utility System)

PLUS is a query by example search system for U.S. patents from 1971 to the present. PLUS produces a list of U.S. patents with frequency of word usage that most closely matches the application searched. PLUS searches U.S. Patents & U.S. PGPubs database. PLUS also provides a list of classes and subclasses where suggested patents are classified.

Benefits:

PLUS searches can be requested and received before the Examiner reviews the application.

Examiner can gain a better understanding of the application this way.

Saves Examiner time -- PLUS searches are conducted by library staff.

Image File Wrapper cases -- Receive results the same day.

Paper cases -- Receive the results within 24 hours.

Emerging Technologies -- PLUS is especially useful in areas of emerging technologies, areas without art-specific terminology, and in areas needing reclassification.

[JPO]

a. Patent literature and Patent related databases

Main JPO Patent Data Resources	
Resource Name	Resource Description
JP Patents	<ul style="list-style-type: none"> ○ Contains full text data of most patents issued from 1993- ○ Contains OCR full text data of most patents issued from 1986 to 1992. ○ Also includes minimal information (patent number, publication date, and current classification) on all other JP patents. ○ FI, F-term, keyword and full text (whole document) search is available
JPO Industrial Property Digital Library http://www.ipdl.inpit.go.jp/homepg_e.ipdl	<ul style="list-style-type: none"> ○ Database contains Japanese patent information since 1885. ○ Has text search capabilities in the title, claim and abstract fields as well as retrieval by FI and F terms.
Foreign Patent database (Internal databases)	<ul style="list-style-type: none"> ○ Contains CA, CH, DE, EP, FR, GB, US and WO documents ○ Contains gazette image/text (only English) data ○ ECLA, USC and full-text* search is available * full-text search (English): whole document of EP, US full-text search (Japanese): Japanese abstract of EP, US
WPI (World patent index) (Internal and commercial databases)	<ul style="list-style-type: none"> ○ Contains English language abstracts of published patent documents from more than 40 patent-issuing authorities, often with family information. ○ Years of coverage vary, normally 1970-present.
Patent Electronic Business Center U.S. Patent Full-Text and Full-Page Image Databases http://www.uspto.gov/patft/index.html	<ul style="list-style-type: none"> ○ Database contains U.S. patent information Issued Patents : full-text since 1976, full-page images since 1790 ○ Published Applications : published since 15 March 2001
European Patent Office Abstracts (EPO)	<ul style="list-style-type: none"> ○ Contains published abstracts from the EPO, from Selected EPO member countries and from the World Intellectual Property Organization (WIPO). ○ Provides PDF images for the full text of patents from a number of countries. ○ Has text search capabilities in the title and abstract fields as well as retrieval by IPC and ECLA classification numbers.

b. Search tools for Patent –Related Databases

Cluster search (domestic patent DB/ foreign patent DB/ Non-patent-literature DB)

Cluster search is a search tool used by JPO patent examiner to search for JP, Foreign Patents and Non-patent-literature. It is used to perform general patent database searches or to perform searches bound to specific document sections.

Cluster search is done by single query (FI, F-term, Computer Software-term, ECLA, USC, Keyword, Full-text) for three databases (domestic patent DB/ foreign patent DB/ Non-patent-literature DB). And importing the document list from WPI to Cluster search window, we can see the original documents

[EPO]

a. Patent literature and Patent related databases

Search collections at the EPO are in most cases accessible via the EPOQUE integrated search environment.

Main EPO Patent Data Resources	
Resource Name	Resource Description
EPODOC (EPO documentation)	<ul style="list-style-type: none"> ○ Contains references to patent documents which compose the systematically classified search documentation of the European Patent Office. ○ The documents consist of published applications, granted patents as well as classified and cited non-patent literature (XP documents). ○ The EPODOC database corresponds to the DocDB database which is the internal EPO master file used for the management of the search documentation. The bibliographic data (i.e. the publication, application and priority numbers and dates, the inventors and applicants data, the IPC classes, the EPO internal classification schemes ECLA and when available the US and JPO classification schemes, citation data, abstracts and title) are available for patent documents of most of the countries of the world. This database contains more than 55 million records that are usually linked to the image of the corresponding document. It covers more than 80 patent offices and organizations worldwide and is widely considered as the world patent master database. A simplified access to this data is available via esp@cenet for the external public.

<p style="text-align: center;">TXT (Full-text databases)</p>	<ul style="list-style-type: none"> ○ Represents the collection of the patent full-text databases available at the EPO. This collection contains more than 17 million records covering following countries and organizations: AU, AT, BE, CH, DE, EP,FR, GB, US and WO. An similar collection exists for major NPL collections.
<p style="text-align: center;">WPI (World patent index)</p>	<ul style="list-style-type: none"> ○ Includes the bibliographic data and abstracts of patents from more than 40 countries as well as European and PCT patents.
<p>Patent Abstracts of Japan (PAJ)</p>	<ul style="list-style-type: none"> ○ Contains English language abstracts of published unexamined Japanese patent applications, issued from 1976 to the present.

2. Non Patent literature, Bibliography and Reference databases

[USPTO , JPO, EPO]

The following resources are primarily used as bibliography and reference databases by USPTO, JPO and EPO biotech examiners. This list is non-exhaustive and does not include all other NPL sources used in other fields. Some Trilateral Offices may have additional sources at their disposal.

<p style="text-align: center;">BIOSIS</p>	<ul style="list-style-type: none"> ○ Biological and biomedical Abstracts ○ Includes citation references to research articles and other source publications that include primary journals, books, reference texts, monographs, papers and abstracts from meetings of interest to the biological community and research reports. ○ The subjects covered by the BIOSIS database concerns the following technical fields: <ul style="list-style-type: none"> . Botany . Zoology . Microbiology . Agriculture . Pharmacology . Biochemistry . Biophysics . Ecology . Bioengineering . Experimental and clinical medicine
--	---

MEDLINE	<ul style="list-style-type: none"> ○ Abstracts of the US National Library of Medicine ○ The NLM's premier bibliographic database covering the fields of medicine, nursing, dentistry, veterinary medicine, the health care system, and the preclinical sciences. ○ Contains bibliographic citations and author abstracts from more than 4,000 biomedical journals published in the United States and 70 other countries. ○ Contains over 12 million citations dating back to the 1950's . Coverage is worldwide, but most records are from English language sources or have English abstracts.
Science Direct Collection (Elsevier)	<ul style="list-style-type: none"> ○ Contains over 25% of the world's science, technology and medicine full text and bibliographic information. ○ Apart from online reference works, handbooks and book series Science Direct offers a rich journal collection of over 2,000 titles. ○ In addition, the Back files program offers the ability to search a historical archive of over 6.75 million articles directly from your desktop, back to Volume 1, Issue 1. The collections contain 4 million articles prior to 1995, and 2.75 million articles from after 1994.
CA, CAPLUS (Chemical Abstracts)	<ul style="list-style-type: none"> ○ Abstracts of the chemical literature back to 1907. ○ File CAplus covers the same material as File CA but contains early access to bibliographic information and patent family information for the documents in the process of being indexed. ○ Useful in text, inventor and registry number searching and a broad range of areas including chemical engineering, electrochemistry, biochemistry, and applied, macromolecular, organic, physical, inorganic, environmental and analytical chemistry.
MRCK (Merck Index Online)	<ul style="list-style-type: none"> ○ 14th edition of The Merck Index, with descriptions of important chemicals, drugs, and biological, agricultural and natural products. ○ Coverage Late 19th century – present.
DIALOG	<ul style="list-style-type: none"> ○ Dialog offers access to more than 15 Terabytes (Tb) of searchable content, including articles and reports from thousands of real-time news feeds, newspapers, broadcast transcripts and trade publications; In-depth repositories of scientific and technical data, patents, trademarks and other intellectual property data. ○ Additional content areas include government regulations, social sciences, food and agriculture, reference, energy and environment, chemicals, pharmaceuticals and medicine.

<p style="text-align: center;">Registry (CAS Registry)</p>	<ul style="list-style-type: none"> ○ Chemical structure and dictionary database that contains unique substance records. ○ Protein and nucleic acid sequences are retrieved using codes for the amino acids and nucleotides. ○ Useful for searching sequences, registry numbers and chemical compounds/structures. ○ Coverage 1957 to date.
<p style="text-align: center;">STN International</p>	<ul style="list-style-type: none"> ○ The online scientific and technical information service providing access to current and archival information from over 200 scientific, technical, business, and patent databases with links to the literature, patents, and chemical catalogs. ○ STN is operated by Chemical Abstracts Service (CAS), a division of the American Chemical Society, in North America; FIZ Karlsruhe (FIZ-K), in Europe; and Japan Science and Technology Agency, Information Center for Science and Technology (JST), in Japan. STN databases cover a wide range of scientific and technical topics. ○ STN allows access to the CAS Registry® database of chemistry drawings. ○ STN on the Web offers web-browser access to the complete content and functionality of STN for information professionals and advanced end-users. ○ STN on the Web combines the STN command functionality with web-browser technology. ○ STN express allows database searching via command-line interface.

3. Nucleic acid and Protein sequence repositories

The USPTO sequence repositories reside on an internal system called the Automated Biotechnology Sequence Search System (ABSS System).

The JPO sequence repositories reside on an internal system called Genetic Information Analysis Services (GIAS).

The EPO sequence search is performed in the EPO-EBI Environment.

(1) Type of databases and their content

a. EMBL / GenBank / DDBJ

The EMBL, GenBank or DDBJ Nucleotide Sequence Databases constitute USA, Europe's or Japanese primary nucleotide sequence resources. Main sources for DNA and RNA sequences are direct submissions from individual researchers, genome sequencing projects and patent applications.

The database is produced in an international collaboration with EMBL, GenBank and DDBJ. Each of the three groups collects a portion of the total sequence data reported worldwide, and all new and updated database entries are exchanged between the groups on a daily basis.

b. Uniprot

UniProt (Universal Protein Resource) is the world's most comprehensive catalogue of information on proteins. It is a central repository of protein sequence and function created by joining the information contained in Swiss-Prot, TrEMBL, and PIR. UniProt is comprised of three components, each optimized for different uses. The UniProt Knowledgebase (UniProt) is the central access point for extensive protein information, including function, classification, and cross-reference. The UniProt Non-redundant Reference (UniRef) databases combine closely related sequences into a single record to speed searches.

c. Geneseq

GENESEQ is the most comprehensive source of information on nucleic and amino acid sequences from Derwent WPI World patent index. It contains sequence information taken from patent applications and granted basic patents (all nucleic acids which are 10 or more bases in length, amino acids which are four or more residues in length plus all PCR primers and probes, of any length are included in the database). Sequences from patents from 1981 onwards are included in GENESEQ

d. PDB

The Protein Data Bank (PDB) is an archive of experimentally determined three-dimensional structures of biological macromolecules, serving a global community of researchers, educators, and students.

The archives contain atomic coordinates, bibliographic citations, primary and secondary structure information, as well as crystallographic structure factors and NMR experimental data. The PDB Newsletter and CD ROM are published quarterly.

(2) The search algorithms

a. The Smith-Waterman algorithm

The Smith and Waterman algorithm is a biological sequence comparison algorithm. It runs a search on cluster, using single and parallelized versions of the software. It allows an rigorous search in a reasonable computational time. It is an exhaustive algorithm, which is recognized as the most sensitive sequence comparison method available in Blast, Fasta and it. As a consequence, it is capable of identifying hits in cases where Blast and Fasta fail and also reports fewer false-positive hits.

b. FASTA

FASTA stands for FAST-All, reflecting the fact that it can be used for a fast protein comparison or a fast nucleotide comparison. This program achieves a high level of sensitivity for similarity searching at high speed. This is achieved by performing optimized searches for local alignments using a substitution matrix, in this case a DNA identity matrix.

The high speed of this program is achieved by using the observed pattern of word hits to identify potential matches before attempting the more time consuming optimized search. The trade-off between speed and sensitivity is controlled by the ktup parameter, which specifies the size of the word. Increasing the ktup decreases the number of background hits. Not every word hit is investigated but instead it initially looks for segment's containing several nearby hits. This program is much more sensitive than blast programs, which is reflected by the length of time required to produce results. Fasta produces optimal local alignment scores for the comparison of the query sequence to every sequence in the database. The majority of these scores involves unrelated sequences, and therefore can be used to estimate lambda and K values. These are statistical parameters estimated from the distribution of unrelated sequence similarity scores. This approach avoids the artificiality of a random sequence model by employing real sequences, with their natural correlations.

c. BLAST

BLAST is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. It emphasizes speed over sensitivity. This emphasis on speed is vital to making the algorithm practical on the huge genome databases currently available.

d. GenePast / Fragment Search

GenePast does not produce local alignments. Instead it reports a best-fit alignment where the largest possible number of residues of the smallest sequence is aligned to the largest sequence with the minimal number of errors. GenePAST algorithm uses an approximate string-matching (or purely mathematical) approach. This algorithm try to find an alignment between 2 sequences with edit cost at most k , i.e. can we transform part of text string to pattern string using at most k additions, deletions, and substitutions.

B. Search Strategies

1. Views on Basic Search (General)

[JPO]

In the examination of the invention relating to biotechnology, the features of search, major search tools and know-how are different in accordance with the technical contents of the invention to be the object of the examination. Accordingly, the basic search method (general) which can be considered to be the most efficient in the invention relating to biotechnology in general is shown by the flow chart below. The detailed explanation on each technical content is described in 5. "Views on search strategy (Detailed discussion).

1. Search of inventors by commercial DB or in-house DB

In JDreamII , BIOSIS (DIALOG) or the in-house DB, a non-patent document corresponding to this invention is searched using the names of inventors and the key words that express the invention best.

<Explanation> The technical background, technical trends, common sense and the like to the invention are described in the introduction of related document, and effective information to help technical understanding of the invention can be obtained. JDreamII in particular is extremely high in its efficiency as the names of inventors can be searched in Japanese and also summaries of presented lectures at scientific societies can be searched.

2. Amino acid sequence or nucleic acid sequence search

When it is characterized by amino acid sequence or nucleic acid sequence

- When homology search is required (many in comparatively long sequence) :
In DNA search system in Agency, amino acid sequence or nucleic acid sequence are searched, and documents describing the sequence information concerned or documents describing the sequence concerned and the sequence with high homology and the identical function are discovered.
- When homology search is not required (many in comparatively short sequence such as probe or epitope):
It is effective to perform a partial sequence search in REGISTRY (STN).
 - ❖ The partial sequence search is the same as the terminology used in REGISTRY (STN) and means the identical search of characteristic sequence.
 - ❖ <USPTO> The USPTO uses an in-house system whether or not a homology search is required. The USPTO also does a REGISTRY search when a search of modified residues is required, or if there is a limited and defined substitutions of a sequence.

3. key word search by commercial DB or in-house DB

In WPI (STN, DIALOG), BIOSIS (STN, DIALOG), MEDLINE (STN), JDreamII or the in-house DB, related technical documents are searched using key words that characterize the invention.

<points to be noticed>

- The cross search to link WPI (STN, DIALOG) and BIOSIS (STN, DIALOG) is the most effective.
- Care must be exercised on the inflection of keywords, synonyms and the like.
In particular, pay special attention to protein as it may have multiple names in spite of the identical material.

4. Cluster search, full text search by electronic journal (Elsevier, bio-related web site)

- When the related documents are neither properly focused nor found by the keyword search by commercial DB, as there are no proper keywords or proper keywords are seldom described in abstracts:
The full text search of patent documents and non-patent documents is performed using cluster search, electronic journal (Elsevier, bio-related web site) and the like.
- When well-known art is searched : Cluster search (/FW) is effective.
- When unpublished patent specifications are searched: Cluster search (all document designation, no theme designation, /TX) is effective.

[USPTO]

Each claim set under examination in each application before the USPTO is searched on a case-by-case basis, dependent on the claim content. The specification is read and claims under consideration are analyzed and interpreted to determine what is being claimed, giving each term its broadest reasonable interpretation consistent with the disclosure. Each claim is then searched in both in-house and external databases, including sequence searches, text and keyword searches, inventorship searches, and classification searches as prescribed by the interpretation of the claim.

From MPEP 904.02:

Planning a thorough search of the prior art requires three distinct steps by the examiner:

(A) Identifying the field of search

Three reference sources must be considered -domestic patents (including patent application publications), foreign patent documents, and non-patent literature (NPL).

The search should cover the claimed subject matter and should also cover the disclosed features which might reasonably be expected to be claimed.

The field of search should be prioritized, starting with the area(s) where the invention would most likely be found in the prior art.

(B) Selecting the proper tool(s) to perform the search

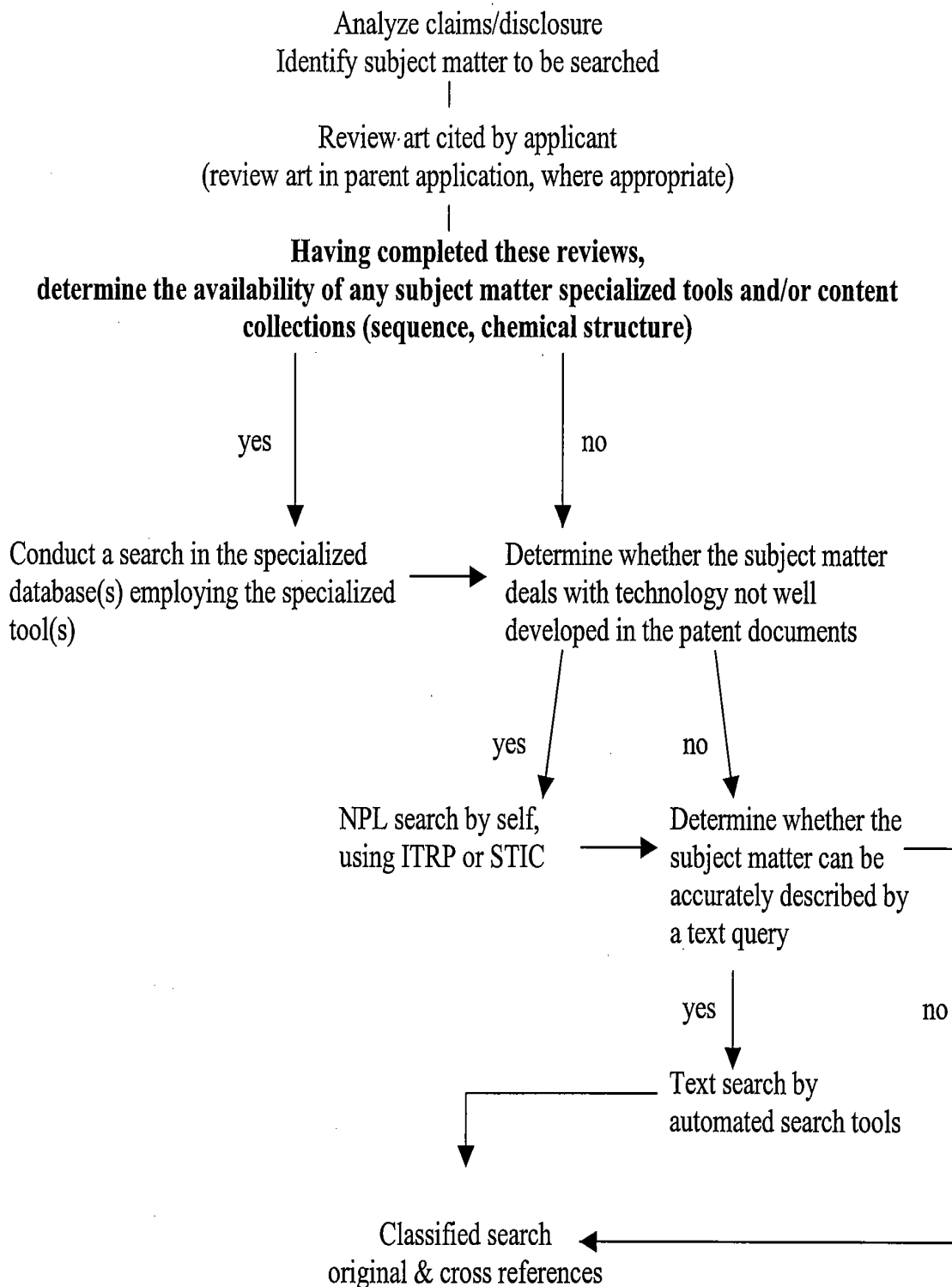
Choice of search tools is a key factor in ensuring that the most relevant prior art is found during the search.

The choice of search tools to be used is based on the examiner's knowledge of the coverage, strengths and weaknesses of the available search tools that are appropriate for use in an examiner's assigned art.

Templates describing search resources for the classified areas of science and technology found in the USPTO **Manual of Classification** are available on the USPTO website,

<http://www.uspto.gov/web/patents/searchtemplates/searchtemplates.htm>

Detailed guidance on the choice and use of specific search tools can be established only within the context of the special requirements of each Technology Center (TC). However, a general methodology following a "decision tree" process, set forth below, for making broad decisions in search tool selection is suggested.



(C) Determining the appropriate search strategy for each search tool selected

Having determined what search tool(s) should be used to conduct the search, the examiner should then determine the appropriate search strategy for each search tool selected.

The appropriate search strategy should be determined by the examiner on a case-by-case basis.

[EPO]

1. Views on Basic Search

Applications in biotechnology relate to many different subjects e.g. genetically modified organisms, gene sequences and their applications, devices, antibodies, screening methods involving molecules, target proteins, enzymes or receptors and the like; a common basic search method does not exist.

For any technical field, however, the general principles for search at the EPO apply, i.e. a search is carried out on the basis of the subject-matter claimed with due regard to the technical information provided by the description and the drawings. Before starting to search for the most pertinent prior art documents relevant for deciding on the patentability of the invention claimed, the examiner has to analyse/understand the technical content of the application. In this regard the critical analysis of the background art cited in the application is an essential early step in every search.

A further step to be taken early in every search is the search for further relevant documents published by the inventors or applicants of the patent application, which has to be performed in all available databases of interest.

2. Search with preparations

EPOQUE provides predefined search-routines/tools, so called "preparations". These preparations automatically collect information linked to the application number, the inventors' names and the applicant. They also retrieve possibly relevant documents in the most pertinent EC and ICO or IPC classes in the internally available databases DOSYS, EPODOC, WPI and ECLA.

3. Search of inventor(s) / applicant in combination with the most relevant keywords

In general the examiner will, after running the relevant preparations, conduct a search using the names of the inventors and/or applicant of the application. This information can be combined with the keywords and their synonyms which best cover the subject matter of the application. Alternatively the examiner may start with a search as described in point 4, below. The search of inventors / applicants in combination with the most relevant keywords, conducted in selected internally available databases, either as single databases or clusters, forms the start of the search. Index searches on STN in the allbib-index and in exceptional cases the internet may be performed subsequently.

4. Search of classes

As a result of the extensive classification activities a search of ECLA and ICO will often generate relevant hits. The ECLA and ICO systems at the EPO have been developed by experienced examiners over many years and are a very useful search tool. Starting in EPODOC the results are transferred to e.g. WPI where a further limitation with proper keywords is done.

5. Search with keywords / sequences / chemical structures

Keywords are searched in all relevant in-house databases as well as in the STN index (allbib). Special attention has to be paid to the generation of the list of synonyms. Synonyms can most often be found in the internet, but also in the internal databases. If chemical names are involved, the Registry file and the like on STN may be consulted. Sequences and chemical structures are searched in the relevant external databases or the internet (for published sequences and structures only). Sequences e.g. dependent on homologies, modification or length, chemical structures, with or without a Markush formula, etc. will be discussed in more detail later.

Whatever strategy is followed, the claimed subject matter in its broadest reasonable interpretation consistent with the disclosure has to be determined and the strategy of the search has to be adapted accordingly.

2. Views on Search Know-How (Particulars)

(1) Protein [including Peptide]

[JPO]

a. Understanding of technique, features of search

- ① Even if protein relating to this invention is specified by amino acid sequence, and the amino acid sequence is new, it is judged that it has no novelty when **the protein is isolated and purified prior to this application.** Accordingly, a search to determine if **the protein is isolated and purified already by the keyword search of the protein is required.** At this time, the similarity in physical and chemical properties such as molecular weight of protein and appropriate pH become an important base for estimating identity of the protein.
- ② The peptide as a partial fragment of protein is frequently used as an antigen relating to the original protein, and the effect is likely to be predicted in many cases if the original protein is known publicly. Accordingly, **in the examination of antigen peptide used as the antigen relating to original protein, it is important to survey not only the publicity of the antigen peptide itself but also the publicity of protein including the antigen peptide.**
- ③ On **peptide to be obtained changing a part of the amino acid sequence** of publicly known protein or peptide, and peptide composed of specified epitope, **cross search of the partial sequence search paying attention to the identity of characteristic amino acid sequence and the keyword search paying attention to the function (function or antigen property of protein) is effective.**
- ④ The peptide has various functions (e.g. a sweetener, a nerve poison, hormone, and the like) different from protein that is a high molecular compound in many cases. Accordingly, **it is necessary to consider examination technique for low molecular compounds.** (The detailed search technique is described in the search strategy file of Peptide (tentative title) published in FY 2001 in Japanese. Refer to the file.)

b. Major search tools, search formula, keywords, and know-how to be used.

(In case of necessity), search of journals corresponding to this invention announced by the inventor after the application:

To be searched using the keywords expressing the names of inventors and the invention most suitable by JDreamII or BIOSIS (STN, DIALOG).

(i) When protein (including peptide) is specified by amino acid sequence:

① The homology search of amino acid sequence using in-Agency DNA search system is performed.

② When the specified (short) amino acid sequence is not specified in the invention of protein fragment, and the optionally obtained partial fragment is specified in all amino acid sequence (specified in this application) of the protein:

The homology search by in-Agency DNA search system is carried out using the total length sequence as a search formula to examine if there is any part (fragment) with the same amino acid sequence as the partial fragment in the long amino acid sequence.

③ When a part of publicly known amino acid sequence of protein is varied in the invention: The partial sequence search in REGISTRY (STN) for amino acid sequence including the varied part is effective. In case of necessity, additional keyword search by keywords expressing the function of the protein is carried out using CA (STN). The cross search of both results is also effective.

④ When amino acid sequence of comparatively short peptide, and amino acid sequence including characteristic amino acid sequence are searched: The partial sequence search of the amino acid sequence directly by REGISTRY (STN) is effective. In this case, even when the amino acid is a special and non-defined one, the partial sequence search by REGISTRY (STN) is possible. In case of necessity, additional keyword search by the keyword expressing the function of the nucleotide is carried out using CA (STN). The cross search of both results is also effective.

(ii) Keyword search of protein (including peptide):

① The cross search of WPI (STN), BIOSIS (STN) and MEDLINE (STN), JDreamII, cluster search in theme code 4B024 and 4B050 and 4H045

in case of necessity, electronic journal (Elsevier, bio-related web site) is used.

* The search by IPC is usually not carried out .

② Keyword search of peptide:

In the search of peptide, it is rare that the peptide is searched only by keyword search. In many cases, when the amino acid sequence of peptide is searched by partial sequence search in REGISTRY (STN), many amino acid sequences with the partial sequence can be hit. Accordingly, it is effective to carry out keyword search by CA (STN) and the like using keywords expressing the features of peptide best for the purpose of focusing the results of partial sequence, and to carry out the cross search of both results. The cluster search in theme code 4H045 and the search by IPC may be effective depending on the contents of invention (Aspartame and the like).

- ③ As keywords, words expressing the name, activity and function of protein, and the names of hosts expressing the protein are effective. At that time, care must be exercised on inflection of each keyword (singular, plural / noun form / past form / progressive form and the like), synonyms and thesaurus.

[Typical words]

It is possible to focus the required technical documents by taking logical products of these words and keywords characteristic of this invention.

Isolation, purification: S isolat? or purif?

[USPTO]

The USPTO now offers Search Templates at <http://www.uspto.gov/web/patents/searchtemplates/>.

The search templates define a general field of search, search tools, and search methodologies that should be considered when examining patent applications in each area of technology. The search templates are organized by the U.S. class and subclass in which the claimed subject matter is classified and/or searched. The search tools for U.S. and foreign patents and non-patent literature, as well as general Internet search tools used by patent examiners are listed within each template. For a protein, peptide example, see <http://www.uspto.gov/web/patents/searchtemplates/class930-060.htm>.

Views on Search (Particulars) -(1) Protein [including peptide]-

Planning a thorough search of the prior art requires three distinct steps by the examiner:

a. Identifying the field of search

- Domestic patents: issued and published applications (PGPubs) should be searched.
- Foreign patent documents should be searched including databases such as Derwent World Patents Index, EPO Abstracts, FPAS3, JPO Abstracts, JPO Industrial Property Digital Library.
- Commercially available sequence containing databases such as:
 - A-Geneseq – contains sequences out of patents indexed in Derwent World Patents Index
 - PIR - Protein Information Resource; produced by the National Biomedical Research Foundation at Georgetown University in collaboration with Martinsried Institute for Protein Sequences, Max Planck Institute for Biochemistry, Martinsried, Germany and the Japan International Protein Information Database, Tokyo University of Science. This database merged with SwissProt to form UniProt. No new sequences have been added to PIR since that time. However, we have found that there are sequences in PIR that are not in UniProt, so we continue to search this database.
 - UniProt - Universal Protein Resource. It is a central repository of protein sequence and function created by merging the information contained in Swiss-Prot, TrEMBL, and PIR. The UniProt Consortium, which began in 2002, is comprised of the European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, UK; the Swiss Institute of Bioinformatics/ExPASy, Geneva, Switzerland; and the Protein Information Resource.

- NPL literature: Commercially available databases such as Biosis, CAOLD, Caplus, Dissertation Abstracts Online, Dissertations and Theses - Full Text, EMBASE, MEDLINE, REGISTRY, SciSearch, CANCERLIT, CONFSCI, Derwent Biotechnology Resource, Inside Conferences, should be searched.

b. Selecting the proper tool(s) to perform the search

- EAST/WEST and PLUS should be used for searching domestic patents both issued and pre-grant published.
- ABSS (Automated Biotechnology Sequence Search System) should be used to search protein or peptide sequence-containing databases (listed above) and also patents and published applications (PGPubs). Parameters used are:
 - **Search algorithm** - Smith-Watermann
 - **Matrix** - Blosum62
 - **Gap penalty** - (aka Gap Open) 10.0 - penalty for inserting a gap in either the query seq or database seq in order to improve the alignment
 - **Gap size penalty** - (aka Gap Extend) 0.5 - penalty for how large (how many residues) the gap is, once it has been inserted
- At least one of STN International, DIALOG, LexisNexis®, Questel*Orbit should be used to search NPL literature and commercial databases.

c. Determining the appropriate search strategy for each search tool selected

Search strategy for a protein/peptide claim should be performed in the field of search established and should include:

A sequence search by sequence identifier.

A text search using the protein/peptide name and synonyms, protein function, source of isolation.

An inventor name search should be performed.

[EPO]

a. General aspects

The search of proteins depends entirely on the information provided in respect thereof. Usually in applications, proteins are characterised by their primary structure, and are hence primarily searched using the provided amino acid sequences. If such information is not available, or as a supplement to a sequence search, protein searches can also be based on keywords and/or classes.

b. Sequence search

For searches relating to amino acid sequences, the most appropriate sequence identity search tools are selected, roughly as follows:

- from 9 amino acid residues upwards, proteins are searched with the FASTA algorithm at EBI. The European Patent Office has a long standing collaboration with the European Bioinformatics Institute (EBI) through which the exchange of biological sequence patent data is carried out. The EBI provides search services to EPO examiners in order for them to compare patent claim sequences against the public nucleotide and protein sequence databases, amongst others. Shorter proteins are searched differently because of sensitivity issues of the FASTA algorithm and limited database coverage for shorter lengths.
- peptides of up to 4 amino acid residues are searched by chemical structure search with STN-Express, and optionally also Beilstein.
- peptides of 4 - 8 amino acid residues and more are searched in peptide sequence databases, most commonly the BIOSEQ file of Registry and the Gene-IT fragment algorithm via the custom EPO-EBI interface.
- the MPsrch tool, also available via the EPO EBI interface, due to its underlying Smith-Watermann protocol, is particularly suited for work requiring high detection sensitivity, such as short sequences and sequences with relatively low homology.
- variants of peptides, proteins and/or fragments thereof with a partly defined sequence and defined variation or mutation possibilities can be effectively searched in Registry. Variants with only conservative substitutions are conveniently searched using the family function in Registry.
- for queries retrieving large numbers of highly homologous, but late published sequences in EBI-FASTA, it can be useful to use the GenePast engine from Gene-IT, also available via the EPO EBI interface, as this groups identical sequences from the same patent family in a single hit, thereby allowing access also to somewhat less homologous, but possibly timely published sequences.

- both EBI and GenomeQuest offer fragment search capabilities; both also allow specification of length interval for the retrieved sequences. If a minimum percentage of identity is critical, the GenomeQuest Fragment Search engine is the tool of choice. EBI has the advantage that sub-ranges of the sequence can be specified as query.

- if so desired, BLAST-based sequence search tools are also available, via EBI, NCBI via the internet (for published sequences only), or STN. The latter allows for combination of the sequence search results with a regular STN search. BLAST searches are not considered suitable for very short sequences.

- the interproscan tool, finally, allows for the analysis of polypeptides, characterised by signature sequences or consensus domains.

Default FASTA protein searches cover UniProt, Geneseq, PDB, and the European, Japanese and US published patent applications. For claims directed to proteins encoded by defined nucleic acid sequences, various translations can be searched.

c. Keyword search

In addition to a sequence search, a keyword-based search is often used to complement the results. In cases where no sequence information is provided, keywords (and classes, see below) are the only search means. A keyword search can be performed independently, or in combination with any sequence search results already obtained.

The EPO-EBI search tool allows for the extraction of accession numbers, patent application numbers and MEDLINE accession numbers from the sequence databases. This allows for the use of the sequence search results, optionally after a selection, in further keyword searches in literature abstracts, patent applications (both abstracted and in full-text) and the headers of the retrieved entries in the sequence databases.

When performed as a search, separate from the sequence search, if any, an appropriate selection can be made from the available abstract and full-text patent and literature databases. Due consideration is given to the use of synonyms, alternative spellings, registry numbers, and various indexing codes. The use of specific fields, such as restriction of a search to the chemical name field, can be particularly useful when searching for e.g. particular manners in which a protein is modified, such as covalent modifications of a particular type, the fact that it comprises non-natural amino acids, etc.

d. Searches by classes

As a further alternative or supplementary search tool for proteins, the ECLA classes and ICO codes can be used. The ECLA classes allow the structural homology information from a sequence search to be combined with the information provided by the classes, which is predominantly based on function and/or source of the sequence. ICO codes are particularly useful for finding additional aspects to a protein, such as it being a chimeric protein, or having unusual residues or modifications.

(2) Polynucleotide [including Polynucleotide partial fragment (ex. probe, primer, anti-sense)]

[JPO]

a. Understanding of technique, features of search

- ① Polynucleotide is often shown by a base sequence. When the homology of the base sequence between two different Polynucleotides is high, there is the technical common sense that the probability in similarity of protein function to be coded by both is high. In addition, Polynucleotide has a property to hybridize other Polynucleotide similar in base sequence. Therefore, it is possible to obtain Polynucleotide that codes protein similar to and with the similar function in base sequence from a Polynucleotide utilizing aforesaid property. Accordingly, it is important in search to discover Polynucleotide with high base sequence in homology with Polynucleotide in this invention.
- ② Even when Polynucleotide relating to this invention is specified by base sequence and the base sequence is new, there may be a case that the gene (no description of amino acid or base sequence) of protein to be coded by the Polynucleotide and DNA fragment or chromosome fragment including the gene is obtained prior to the application. Accordingly, it is important to perform not only sequence search but also keyword search combining the keyword expressing the name, activity and function of protein to be coded by the Polynucleotide and the keyword expressing gene, DNA fragment and chromosome.
- ③ If protein to be coded by Polynucleotide in this invention is isolated and purified prior to the application, it is easy for a person skilled in the art to obtain Polynucleotide to code protein, applying well-known art, analyzing the amino acid sequence of a part (end part) of protein, preparing a probe based on the sequence and using the probe. Accordingly, it is also important to perform the keyword search to discover if protein to be coded by Polynucleotide in this invention has been already isolated and purified.
- ④ The Polynucleotide partial fragment to code a part of a useful gene is useful for cloning the gene in the same class as a primer or probe and the one to decrease the expression of the gene by hybridizing is useful as an anti-sense medicine. Therefore, the function as a primer, probe, anti-sense and the like is not described in the document for prior art but the Polynucleotide including the identical base sequence is described, the Polynucleotide partial fragment has high possibility to be used as a material possessing the above described functions. So, the search noticing the homology of the base sequence is important.

b. Major search tools, search formulas, key words and know-how to be used.

(In case of necessity), search of journals corresponding to this invention announced by the inventor after the application:

To be searched using the keywords expressing the names of inventors and the invention most suitable by JDreamII or BIOSIS (STN, DIALOG).

< Polynucleotide >

(i) When Polynucleotide is specified by base sequence:

① The homology search of base sequence using DNA search system is performed.

② When a part of publicly known base sequence of Polynucleotide is varied:

The partial sequence search in REGISTRY (STN) for base sequence including the varied part is effective. In case of necessity, additional keyword search by keywords expressing the function of the nucleotide is carried out using CA (STN). The cross search of both results is also effective.

(ii) Key word search of Polynucleotide

① The cross search of WPI (STN), BIOSIS (STN) and MEDLINE (STN), JDreamII, cluster search in theme code 4B024 and 4B050 and 4H045 in case of necessity, electronic journal (Elsevier, bio-related web site) is used.

* The search by IPC is usually not carried out.

② As keywords, the names of protein coded by Polynucleotide, words expressing the activity and function of protein, and the names of host expressing the protein are effective. At that time, care must be exercised on inflection of each keyword (singular, plural / noun form / past form / progressive form and the like), synonyms and thesaurus.

[Typical words] It is possible to focus the required technical documents by taking logical products of these words and keywords characteristic of this invention.

Gene : S DNA or cDNA or RNA or mDNA or nucleotid? or clon? or gene or genom? or sequenc?

Fusion DNA : S fusion? or fused? or chimer?

Plasmid: S plasmid? or vector? or vesicle?

Isolation, purification: S isolat? or purif?

< Polynucleotide partial fragment >

(i) When the base sequence of Polynucleotide partial fragment is searched:

① When comparatively short base sequence (e.g. shorter base sequence than 50 bases) and the sequence including characteristic base sequence, are searched :

The partial sequence search of the Polynucleotide partial fragment directly in REGISTRY (STN), without homology search, is effective. As a result of partial sequence search, if many sequences are hit, additional keyword search is carried out by the keywords expressing the function of the Polynucleotide in CA (STN) and also the cross search of both results is carried out to focus the search results.

② When the specified short sequence as a fragment of the Polynucleotide partial fragment is not obviously specified, and the partial fragment to be obtained optionally in whole nucleotide sequence (specified in this specification) of the fragment is searched:

The homology search of the base sequence using in-Agency DNA search system is carried out using the total length sequence as search formula. As the result, examination is made to check if there is the sequence that has the identical amino acid sequence (as a standard, more than 12 - 15 base sequences in minimum) with the fragment in the long base sequence.

(ii) **Keyword search of Polynucleotide partial fragment:**

In the search of Polynucleotide partial fragment, it is rare to carry out only keyword search. Generally, when the partial sequence search of the base sequence of Polynucleotide partial fragment is carried out in REGISTRY (STN), the base acid sequence with the partial sequence may be hit in many cases. Accordingly, the keyword search is carried out in CA (STN) and the like using the keywords that express the features of Polynucleotide partial fragment best in order to focus the partial sequence result, and also the cross search of both results is carried out.

[Typical words]	Probe: S prob? or hybridiz?
	Primer: S primer?
	Detection: S detect? or assay? or test? or ident?

[USPTO]

The USPTO now offers Search Templates at
<http://www.uspto.gov/web/patents/searchtemplates/>.

The search templates define a general field of search, search tools, and search methodologies that should be considered when examining patent applications in each area of technology. The search templates are organized by the U.S. class and subclass in which the claimed subject matter is classified and/or searched. The search tools for U.S. and foreign patents and non-patent literature, as well as general Internet search tools used by patent examiners are listed within each template.

A representative search template for polynucleotide may be found at:
http://www.uspto.gov/web/patents/searchtemplates/class536-023_1.htm

[EPO]

a. General aspects

As with proteins, the search of nucleic acids depends on the information provided in respect thereof, although it is far less common to find nucleic acids being described without a primary structure having been determined (and provided) than it is for proteins. Normally, therefore, a search for a nucleic acid will comprise a sequence search. If the sequence is not available, or as a supplement to a sequence search, nucleic acid searches can also be based on keywords and/or classes.

b. Sequence search

For the selection of the most useful nucleic acid homology search tool, the following criteria serve as an initial guide:

- nucleic acids defined by their primary sequence are routinely searched with the FASTA protocol at EBI. BLAST is also available at EBI, via the internet at NCBI (for published sequences only), and via STN, but use thereof is determined on a case-by-case basis on practical grounds, and is not routine.
- nucleic acids defined by the protein sequence they encode can be searched using various reverse translation tools available under the EBI interface, allowing a protein sequence to be screened against the nucleic acid databases.
- the EBI-FASTA interface allows for the definition of intervals within the query sequence to use for the search, as well as an interval defining the length of the retrieved sequences, thereby allowing global fragment searches as well as searches for short sequences derived from a predefined area of the query sequence (primers, anti-sense). FASTA is however not very reliable for retrieving sequences shorter than about 20 nucleotides.
- the Fragment Search algorithm of Gene-IT, and to a lesser extent also the GenePast program, can produce reliable results with shorter sequences than FASTA, in practice essentially only limited by the database coverage for such short sequences.
- due to its more extensive database coverage, especially for short sequences, and the reduced reliability of the other protocols for shorter sequences, a supplementary search in Registry for sequences up to about 50 nucleotides is strongly recommended.
- variants of nucleic acid sequences or fragments thereof with a partly defined sequence and defined variation or mutation possibilities can be effectively searched in Registry.

- the same consideration as outlined above for the polypeptide sequence search also apply at the nucleic acid level, in that for queries which generate large numbers of highly homologous late-published sequences or where a minimum percentage of identity is critical, the Gene-IT tools are to be preferred.

- no MPsrch equivalent is currently used for nucleic acid searches, but Smith-Watermann-based search tools are available, although only used for specialist work.

Default FASTA nucleic acid searches cover EMBL and GeneSeq. If so desired, various sub-databases of EMBL can be expressly in- or excluded.

c. Keyword search

In general, the same considerations apply as for keyword-based searches for proteins, as detailed above in paragraph 3. In addition, the specific role or function of a nucleic acid can be filtered through the use of keywords, such as anti-sense, ribozyme, probe, primer, promoter, iRNA, etc.

d. Searches by classes

As for proteins, the classes allow structural information obtained in a sequence search to be combined with the functional information, provided by the ECLA classes. Not only coding nucleic acids, but also associated polynucleotides, such as promoters, ribozymes, iRNA's, anti-sense molecules, etc. are, to the extent that they are specific for a particular sequence, grouped with protein they encode or with which they are affiliated. ICO codes are used for classifying additional aspects to a nucleic acid, such as its function or use, or additional structural aspects (PNA's, non-standard nucleotides, etc.)

(3) Antibody

[JPO]

a. Understanding of technique, features of search

- ① The basic structure of antibody originated from the identical species is the same, and the specificity and affinity to antigen to react are determined by variable region (super variable region in particular) constituting the antibody. Accordingly, the invention relating to the antibody is usually specified by the reactivity of the antibody and is searched noticing the reactivity.
- ② It is a basic work to notice the reactivity and to carry out the keyword search of the keyword expressing the features of the antigen and antibody (including polyclonal antibody and monoclonal antibody).
- ③ Generally the production method by polyclonal antibody or monoclonal antibody is a well-known art and the search only to antigen is often carried out (as the work to obtain the antibody applying the manufacturing method of the antibody, this is a well-known art to the antigen, can be judged to devise easily). When only the antigen is searched, refer to (1) Protein related invention or (5) Search technique of peptide (C07K) for details.
- ④ As epitope may be identical or similar in (micro)organisms of closest kinds, wider search is required so as to include antigen originated from the closest kinds of (micro)organisms when the keyword search to the antigen with the epitope is carried out.
- ⑤ The antibody specified by the antibody produced by specified hybridoma to which a deposit number is given or by amino acid sequence of the antibody (in variable region) is thought to have a high possibility of possessing an inventive step but there is a case that has no inventive step. There is a case that the antibody possessing similar combining ability may exist or the antibody possessing the similar combining ability can be obtained easily (by well-known art) using the identical antigen. Accordingly, it is important to search keeping this point in mind.
- ⑥ On specified operation on antibody related invention, refer to “Examination memorandum of monoclonal antibody related invention (960801 version, in Japanese)” and “Present problems in examination of monoclonal antibody and suggested solutions (as of March 19, 1998, in Japanese) (Reference).

b. Major search tools, search formulas, keywords and know-how to be used.

(In case of necessity), search of journals corresponding to this invention announced by the inventor after the application:

To be searched using the keywords expressing the names of inventors and the invention most suitable by JDreamII or BIOSIS (STN, DIALOG).

(i) When the antigen and antibody are specified by keywords:

① The cross search of WPI (STN), BIOSIS (STN) and MEDLINE (STN), JDreamII, cluster search in theme code 4B024 and 4B050 and 4H045 in case of necessity, electronic journal (Elsevier, bio-related web site) is used.

*The search by IPC (C12P 21/08, C07K 16/00 and the like) is also effective

② When the antigen with the identical or similar epitope is searched by keywords, the wider search is required so as to include the antigen originated from the closest kinds of (micro)organisms with high probability possessing the identical or similar epitope.

③ As keywords, the names of materials to react with antibody, words expressing the activity and function of the material are effective. At that time, care must be exercised on inflection of each keyword (singular, plural / noun form / past form / progressive form and the like), synonyms and thesaurus.

[Typical words] Antibody to react to material A in DIALOG, and the search formula to be used when monoclonal antibody is searched, are described.

Antibody : S (antibod? or antiser?) (9n) Name of material A

Monoclonal antibody :

S (((mono-clonal or monoclonal) (5n) antibod?) or mab) (9n)
Name of material A

(ii) When those of which antigen and epitope or variable region of the antibody is specified by amino acid sequence (or base sequence) are searched.

(Basically, the same as the method in (1) Protein related invention (including peptide)).

① When the whole antigen protein is specified by amino acid sequence (or base sequence):

The homology search of amino acid sequence using in-Agency DNA search system is carried out. As the part to be epitope in the antigen protein is multiple parts of protein, the homology search using in-Agency DNA search system is carried out using the total length sequence as a search formula. As the result, examination is made to check if there is a

part that has the identical amino acid sequence with the whole antigen protein in the amino acid sequence that is hit.

② When the antigen protein varies a part of amino acid sequence of well-known protein in the invention:

The partial sequence search is carried out in REGISTRY (STN) for amino acid sequence including the varied part. In case of necessity, additional keyword search is carried out by the keywords expressing the function of the antigen protein in CA (STN). The cross search of both results is also effective.

③ When the variable region of epitope or antibody is specified by amino acid sequence: The partial sequence search is carried out directly in REGISTRY (STN) for amino acid sequence including epitope or variable region of the antibody. In case of necessity, additional keyword search is carried out by the keywords expressing the function of the antigen in CA (STN) and also the cross search of both results is carried out.

[USPTO]

The USPTO now offers Search Templates at
<http://www.uspto.gov/web/patents/searchtemplates/>.

The search templates define a general field of search, search tools, and search methodologies that should be considered when examining patent applications in each area of technology. The search templates are organized by the U.S. class and subclass in which the claimed subject matter is classified and/or searched. The search tools for U.S. and foreign patents and non-patent literature, as well as general Internet search tools used by patent examiners are listed within each template.

A representative search template for antibodies may be found at:
<http://www.uspto.gov/web/patents/searchtemplates/class530-386.htm>

[EPO]

a. General aspects

Antibodies may be defined by structural features, by their specificity or by their process of production.

b. Sequence search

If the antibody is defined by structural features like the sequence of its domains, the sequence of the antigen or the sequence of a small epitope, said features are searched in sequence databases (EBI, STN Registry, DGENE, PCTGEN). The choice is made according to the length of the sequences. If sequences have to be searched on STN which are longer than the command line, STN on the web is used. There sequences up to a length of 2000 peptides/nucleotides can be searched.

c. Keyword search

In addition to the sequence search, a keyword search for the specificity of the antibody is carried out. The keyword search has to be performed in all available, relevant abstract databases (WPI, EPODOC, MEDLINE, BIOSIS, EMBASE) as well as other sources of information (technical handbooks, catalogues, etc.). The internet site of the applicant is also screened for indications of early disclosure.

Of special importance for the present technical area is the search in full text databases using various combinations/strings of keywords considered to define the technical teaching the examiner intends to find.

However, the generally appreciated fact that claims directed to methods are often difficult to cover by a combination/string of keywords has to be taken into consideration. A keyword search in many cases either cannot cover the technical field represented by the method due to the existence of a variety of synonymous expressions for said keywords taken from the claims, or may on the other hand produce a lot of background noise due to the "non-specificity" of the keywords, i.e. the method features, as the contribution to the art often lies in the *combination/sequence* of features.

If the antibody is defined by its specificity or by the process of production, the classification based keyword search is the only tool available to carry out a complete search.

d. Searches by Classes

d-1. ECLA classification

As a consequence of said drawbacks in keyword searching, the EPO has established a function- and specificity-oriented classification system covered by the ECLA codes C07K16/00 - 46D.

C07K16/00 is subdivided in three major groupings which represent the main fields of search.

- (1) C07K16/00A - 10N1, relating to the origin of the antibody
- (2) C07K16/12 - 44, relating to the antigen for which the antibody is specific
- (3) C07K16/46, relating to hybrid immunoglobulins

An internal keyword database, in which most of the common cell surface receptors and the possible synonyms are listed, allows to determine in which class the receptor is classified and to get the most common synonyms.

d-2. ICO Classes

In addition to the main ECLA classes, so-called indexing codes have been established for additional relevant and sufficiently disclosed aspects of the application as found in the examples.

The ECLA classes together with the indexing codes allow a sophisticated keyword search in abstract and full text databases. In combination with a sequence search this results in a complete search.

(4) Expression system

[JPO]

a. Understanding of technique, features of search

- ① The invention relating to the expression system includes expression adjustment region (promoter and the like) of the gene or the invention of other functional gene (signal sequence and the like), the invention of vector and the invention relating to specific host for improvement of expression efficiency.
- ② Generally, the functional constitution is the major constitution of the invention, which makes the sequence search, and the like, difficult, and the keyword search is obliged in many cases. However, the functional constitution makes the proper focusing by the keyword search difficult in many cases. Accordingly, it is important to search for the proper keywords after understanding the features of the art well.
- ③ In order to search for the proper keywords after understanding the invention exactly or to understand the technical background and technical trends, the search of corresponding theses by the names of inventors by JDreamII, BIOSIS (STN, DIALOG), or MEDLINE (STN), is useful. In addition, in order to understand the technical features of this invention, it is effective to read relating documents written in Japanese carefully. Therefore, the search of patent specification in Japanese by cluster search and of non-patent document in Japanese by JDreamII is effective. In addition, it is also useful to read related parts in the reference books on expression system and to utilize the reference document introduced here.
- ④ After the proper keywords based on the technical features in this invention are found, patent and non-patent documents other than Japanese are searched using BIOSIS (STN, DIALOG), WPI (STN, DIALOG), MEDLINE (STN), electronic journal (Elsevire, bio-related web site) by the proper keywords.
- ⑤ When the base sequence is specified in the expression adjustment region (promoters and the like) of the gene or the invention of other functional gene (signal sequence and the like), the same examination technique as in (2) Polynucleotide partial fragment related invention is taken.
- ⑥ When there is the feature in host to be used in expression, refer to “Fermentation, microorganism search strategy file (tentative name)” completed in FY 2001 in Japanese.

b. Major search tools, search formulas, keywords and know-how to be used.

(In case of necessity), search of journals corresponding to this invention announced by the inventor after the application:

To be searched using the keywords expressing the names of inventors and the invention most suitable by JDreamII or BIOSIS (STN, DIALOG).

(i) Keyword search in expression system.

The search of patent specification in Japanese by cluster search and non-patent document in Japanese by JDreamII is performed for the purpose of reading Japanese documents relating to this invention carefully in order to understand the technical features in this invention.

- ① After the proper keywords based on the technical features in this invention are found, patent and non-patent documents other than Japanese are searched using BIOSIS (STN, DIALOG), WPI (STN, DIALOG), MEDLINE (STN), electronic journal (Elsevier, bio-related web site) by the proper keywords.
- ② When proper keywords based upon the technical features to focus the document exactly are found, it is useful to search using “concrete words” used in actual working examples of this specification (BIOSIS (STN, DIALOG), WPI (STN, DIALOG), MEDLINE (STN), electronic journal (Elsevier, bio-related web site)).
- ③ (The search by ECLA classification (C12N15/63-90) in which rather detailed classification on the expression system is developed, may be also effective in some cases.
- ④ The whole text file search in cluster search and Elsevier is useful for searching the names of specific constitution (specific promoter, specific expression vector, specific host and the like). Also, it is effective for search of well-known art and prior art

(ii) When the base sequence is specified in the expression regulatory region (promoters and the like) of the gene or the invention of other functional gene (signal sequence and the like).

When the base sequence is short, the sequence search by REGISTRY (STN) and the keyword search (functional words) is carried out.

When the base sequence is comparatively long, the homology search of the base sequence using the DNA search system, and the keyword search (functional words) is carried out.

[Typical words]

Promoter: S Promoter? or gdn or control? or leader? or transcript?

Linker: S linker? or spacer?

Enhancer: S enhancer?

Signal sequence: S signal?

Anti-sense: S antisense?

Suppressor: S suppressor?

Terminator: S terminator?

[USPTO]

The USPTO now offers Search Templates at <http://www.uspto.gov/web/patents/searchtemplates/>.

The search templates define a general field of search, search tools, and search methodologies that should be considered when examining patent applications in each area of technology. The search templates are organized by the U.S. class and subclass in which the claimed subject matter is classified and/or searched. The search tools for U.S. and foreign patents and non-patent literature, as well as general Internet search tools used by patent examiners are listed within each template.

A representative search template for an expression system may be found at: http://www.uspto.gov/web/patents/searchtemplates/class435-069_1.htm

[EPO]

a. General aspects

Applications relating to expression systems may cover structural and functional features like expression vectors and their regulatory elements (e.g. promoters, enhancers, etc.), as well as methods for the expression of genes in recombinant host cells. Hence, the search may have to be focussed on structural and/or functional aspects.

b. Keyword search

A keyword search has to be performed in all available, relevant abstract databases (WPI, EPODOC, MEDLINE, BIOSIS, EMBASE). Of special importance for the present technical area is the search in full text databases using various combinations/strings of keywords considered to define the technical teaching the examiner intends to find.

However, in the case of a method for the expression of a gene in a recombinant host cell, which may be characterized by a combination or interaction of diverse structural elements of an expression system the generally appreciated fact that claims directed to methods are often difficult to cover by a combination/string of keywords has to be taken into consideration. A keyword search in many cases either cannot cover the technical field represented by the method due to the existence of a variety of synonymous expressions for said keywords taken from the claims, or may on the other hand produce a lot of background noise due to the "non-specificity" of the keywords, i.e. the method features, as the contribution to the art often lies in the *combination/sequence* of features.

c. Searches by Classes

c-1. ECLA Classification

As a consequence of these drawbacks in keyword searching, the EPO has established a function- and application-oriented classification system for expression systems covered by the ECLA codes C12N15/63-869.

This part of the classification system is organized as follows:

- according to general aspects of the expression of genes in recombinant cells and the preparation of vectors per se, covered by the classes C12N15/63-69
- according to the recombinant host cell used, prokaryotic (i.e. bacterial, C12N15/70-78) and eukaryotic (i.e. fungal, plant, and animal, C12N15/79-869) expression systems can be distinguished

c-2. ICO Classes

In addition to classes from the main ECLA scheme, so-called Indexing codes (ICO) have been established. Indexing codes are used to specify/classify relevant/essential features in combination with main ECLA classes (e.g. special characteristics of the vectors). For searching relevant state of the art, these indexing codes can either be used alone or in combination with other ICO and/or the main ECLA codes.

d. Sequence search

In the case of a new expression vector or novel elements thereof (e.g. promoters), a search strategy based on structural features may be chosen (EBI, STN REGISTRY, DGENE, PCTGEN).

The search strategy, i.e. the parameters to chose for the best sequence search, needs to be adapted to the sequence to be searched. Of importance with respect to sequence searches are also "post-processing" tools which allow the further analysis of sequence search results in abstract and full text databases.

(5) Gene amplification method, detection method

[JPO]

a. Understanding of techniques, features of search

- ① The development and improvement of the art in the gene amplification and detection methods have spread widely in scope and details. Accordingly, it is necessary to understand the basic matters and the flow of art in development and improvement of the gene amplification and detection method correctly and to grasp the position of this invention in the prior art.
- ② Therefore, the understanding of “ECLA classification (C12Q1/68-70) developing the detailed classification along the technical development of the gene amplification and detection methods is quite effective. The understanding by a collection of standard art “Amplification and detection of nucleic acid” (JPO home page, in Japanese) describing the art for gene amplification and detection methods in Japanese along with the ECLA classification is useful. Also, the understanding by Basic Book is helpful.
- ③ For understanding of this invention, only reading claims is not enough to understand the invention. There scarcely are appropriate drawings showing the intention of the invention described in the claim. Accordingly, it is essential to schematize (to prepare caricatures) the flow of art contained in the claim to understand this invention.
- ④ When the constitution of the prior art and this invention is compared for examination after the search, (as the technical development and improvement have been made widely and in detail), there are many things at which a person skilled in the art could have easily concluded in the constitution of the invention. However, even a slight difference in constitution may cause a remarkable effect that a person skilled in the art cannot predict by the difference many times in this field. Accordingly, it is important to search noticing not only the difference in the constitution but also the effect to be caused by adopting this constitution.

b. Major search tools, search formulas, keywords and know-how to be used.

(In case of necessity), search of journals corresponding to this invention announced by the inventor after the application:

To be searched using the keywords expressing the names of inventors and the invention most suitable by JDreamII or BIOSIS (STN, DIALOG).

Effective search for the gene amplification method and detection method:

- ① The “search by ECLA classification (C12Q1/68-70) developing the detailed classification along the technical development of the gene amplification and detection methods is the most effective.
- ② The utilization of reference books and a collection of standard art “Amplification and detection of nucleic acid” (JPO home page, in Japanese) owned by Life Engineering shown below. The re-quotation from them is also useful.
- ③ On a keyword search, the same words (PCR, gene amplification and the like) are used in many cases as keywords, so it is quite difficult to focus the art by the keywords. Even if a full text search is carried out, it is difficult to focus on the art as much noise can be caused. However, when keyword search is planned, effective database to be used includes BIOSIS (STN, DIALOG), WPI (STN, DIALOG), MEDLINE (STN), JDreamII, cluster search, electronic journal (Elsevier, bio-web site), and the like. The cluster search and electronic journal (Elsevier, bio-web site) are effective mainly for the search of well-known art and prior basic art.

[USPTO]

The USPTO now offers Search Templates at <http://www.uspto.gov/web/patents/searchtemplates/>.

The search templates define a general field of search, search tools, and search methodologies that should be considered when examining patent applications in each area of technology. The search templates are organized by the U.S. class and subclass in which the claimed subject matter is classified and/or searched. The search tools for U.S. and foreign patents and non-patent literature, as well as general Internet search tools used by patent examiners are listed within each template.

A representative search template for an amplification method may be found at: http://www.uspto.gov/web/patents/searchtemplates/class435-091_1.htm

A representative search template for a DNA detection method may be found at: <http://www.uspto.gov/web/patents/searchtemplates/class435-003.htm>

[EPO]

a. General aspects

Applications of this technical field are concerned with the provision of methods for nucleic acid analysis in general (e.g. hybridization methods, amplification methods, sequencing of nucleic acids etc.), and/or the provision/use of specific nucleic acids as probes/analytes.

The search for methods for nucleic acid analysis is mainly based on a "keyword search" and the use of ECLA and ICO codes. The generally appreciated fact that claims directed to methods are often difficult to cover by a combination/string of keywords has to be taken into consideration. For searching specific nucleic acid probes and/or their respective use, the search strategy must also encompass a sequence search.

b. Keyword search

A keyword search has to be performed in all available, relevant abstract databases (WPI, EPODOC, MEDLINE, BIOSIS, EMBASE). Of special importance for the present technical area is the search in full text databases using various combinations/strings of keywords considered to define the technical teaching the examiner intends to find.

A keyword search in many cases either cannot cover the technical field represented by the method due to the existence of a variety of synonymous expressions for said keywords taken from the claims, or may on the other hand produce a lot of background noise due to the "non-specificity" of the keywords, i.e. the method features, as the contribution to the art often lies in the *combination/sequence* of features.

c. Searches by Classes

c-1. ECLA Classification

As a consequence of said drawbacks in keyword searching, the EPO has established a function- and application-oriented classification system covered by the ECLA codes C12Q1/68-70.

C12Q1/68 is subdivided in two major groupings which represent the main fields of search in this area: Method groups (C12Q1/68-68E, C12Q1/68P) and Analyte/Probe groups (C12Q1/68M+).

The Method groups cover the areas in the field of nucleic acid analysis (hybridization reactions, amplification reactions, e.g. for the analysis of polymorphisms, sequencing of nucleic acids, detection involving reporter genes).

The Analyte groups are used for the retrieval of patent documentation relating to specific nucleic acid sequences for use as probes. Said groups may be used for search together with the tools for sequence searching (EBI, STN Registry).

C12Q1/70 groups are directed to methods for nucleic acid analysis specifically directed to viruses or bacteriophages.

c-2. ICO Classes

In addition to classes from the main ECLA scheme, so-called Indexing codes (ICO) have been established. Indexing codes are used to specify/classify relevant/essential features in combination with main ECLA classes. For searching relevant state of the art, these indexing codes can either be used alone or in combination with other ICO and/or the main ECLA codes.

d. Sequence search

A sequence search (EBI, STN REGISTRY, DGENE, PCTGEN) becomes necessary when the claimed invention relates to sequences. The search strategy, i.e. the parameters to choose for the best sequence search, needs to be adapted to the sequence to be searched. Of importance with respect of sequence searches are also "post-processing" tools which allow the further analysis of sequence search results in abstract and full text databases.

(6) SNPs / Haplotypes

[USPTO]

The USPTO now offers Search Templates at <http://www.uspto.gov/web/patents/searchtemplates/>.

The search templates define a general field of search, search tools, and search methodologies that should be considered when examining patent applications in each area of technology. The search templates are organized by the U.S. class and subclass in which the claimed subject matter is classified and/or searched. The search tools for U.S. and foreign patents and non-patent literature, as well as general Internet search tools used by patent examiners are listed within each template.

A representative search template for SNPs/Haplotypes may be found at: http://www.uspto.gov/web/patents/searchtemplates/class536-023_52.htm

(7) 3D Structures/Crystals

[USPTO]

The USPTO now offers Search Templates at <http://www.uspto.gov/web/patents/searchtemplates/>.

The search templates define a general field of search, search tools, and search methodologies that should be considered when examining patent applications in each area of technology. The search templates are organized by the U.S. class and subclass in which the claimed subject matter is classified and/or searched. The search tools for U.S. and foreign patents and non-patent literature, as well as general Internet search tools used by patent examiners are listed within each template.

A representative search template for polypeptides with or without 3D structures may be found at: <http://www.uspto.gov/web/patents/searchtemplates/class530-350.htm>

(8) Treatment methods (ex. gene therapy and anti-sense therapy)

[USPTO]

The USPTO now offers Search Templates at
<http://www.uspto.gov/web/patents/searchtemplates/>.

The search templates define a general field of search, search tools, and search methodologies that should be considered when examining patent applications in each area of technology. The search templates are organized by the U.S. class and subclass in which the claimed subject matter is classified and/or searched. The search tools for U.S. and foreign patents and non-patent literature, as well as general Internet search tools used by patent examiners are listed within each template.

A representative search template for gene therapy and anti-sense therapy may be found at:

<http://www.uspto.gov/web/patents/searchtemplates/class514-044.htm>

(9) Others