

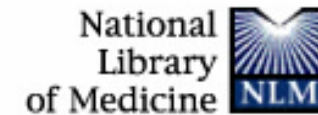
I due centri maggiori, EBI e NCBI hanno sviluppato sistemi dedicati di **RETRIEVAL** allo scopo di ottenere il massimo delle informazioni con il minimo sforzo da parte dell'utente

⇒ **SRS** (Sequence Retrieval System) della EBI che mette a disposizione anche dello spazio sul server per memorizzare le ricerche.

⇒ **Entrez** della NCBI, molto più eterogenea ma più semplice da utilizzare, è il vero riferimento per le ricerche bibliografiche perchè perfettamente integrata con MedLine, database di letteratura bio-medica, tanto da far nascere

Entrez + MedLine => **PubMed**

**PubMed è la principale fonte di informazione delle banche dati, tutti i dati nascono da lavori pubblicati e presenti in PubMed.**



PubMed è la principale risorsa della rete per quanto riguarda le pubblicazioni scientifiche di natura biomedica. Presenta una QUERY FORM in cui immettere le parole da cercare

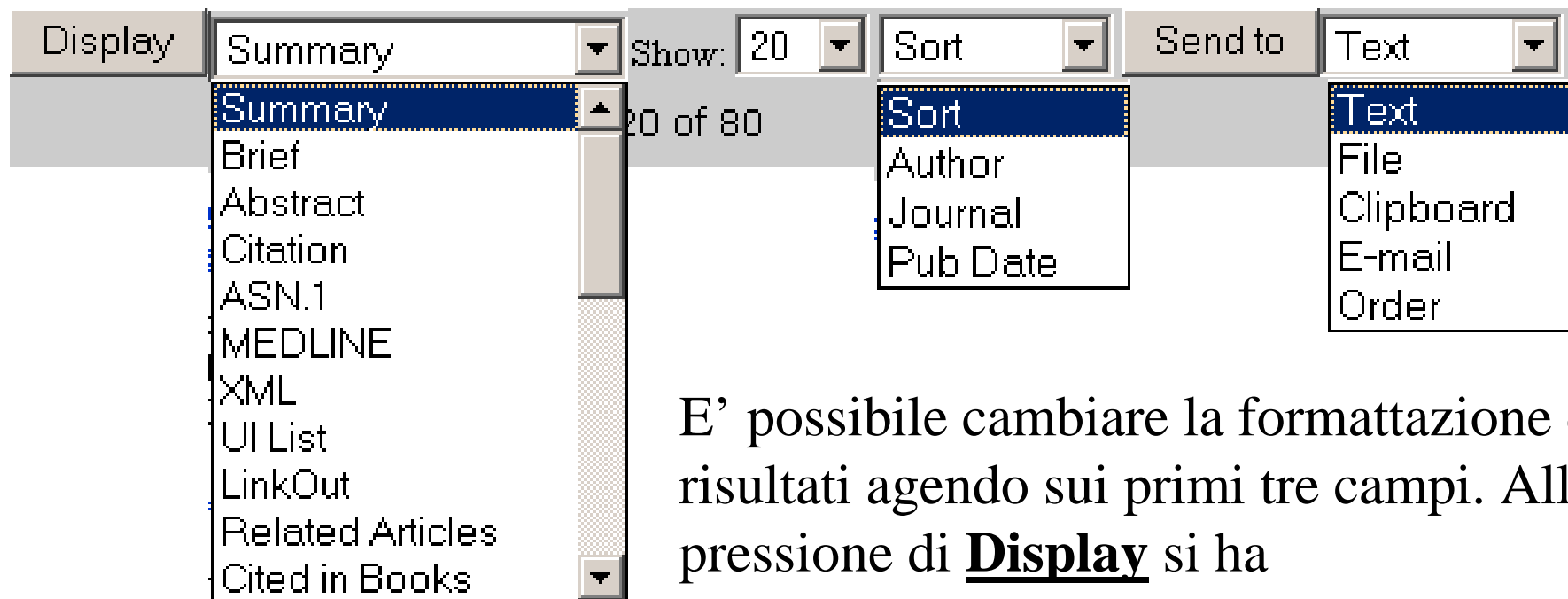
Search PubMed  for

La ricerca per parole chiave comporta la possibilità di combinare varie queries utilizzando gli operatori booleani. Con **Go** si procede con la ricerca. E' possibile fornire alla query un tag di campo (es. query [AU] cerca tra gli autori...) per limitare la ricerca.

Il numero di entries trovate viene indicato in una results table con varie opzioni a disposizione

Display Summary  Show: 20  Sort  Send to Text

Items 1-20 of 80 Page 1 of 4 [Next](#)



E' possibile cambiare la formattazione dei risultati agendo sui primi tre campi. Alla pressione di **Display** si ha l'aggiornamento

E' inoltre possibile spedire i risultati non al browser (Text) ma ad una diversa destinazione, tra cui il file con la funzione **Send to**.

Per scorrere tra i risultati si utilizza la sezione **Page**, che si trova in alto e in basso a destra della schermata



Sotto la query form è situata una Feature Table, molto utile per le ricerche perché fornisce strumenti di controllo sulle interrogazioni

**Limited to:**

All Fields  only items with abstracts

Publication Types Languages Subsets

Ages Human or Animal Gender

Entrez Date

Publication Date From    To

Use the format YYYY/MM/DD; month and day are optional.

**Limits** permette di restringere la query cercando solo in certi tipi di pubblicazioni, solo in certe zone delle varie queries (sfruttando l'univocità dei tags con cui viene compilata ogni entry), entro dei limiti di tempo e secondo molti altri criteri.

Limits

Preview/Index

History

Clipboard

Details

Search	Most Recent Queries	Time	Result
#5 Search	ramazzotti m [au]	12:55:26	<a href="#">3</a>
#4 Search	manao g [au]	12:55:04	<a href="#">80</a>
#3 Search	acylphosphatase	12:54:08	<a href="#">192</a>

**Preview/Index** permette di visualizzare solo il numero di risultati che una query totalizza, così da non attendere il parsing delle varie entries. E' un modo per snellire la ricerca e trovare le keywords più adatte.

**History** permette di monitorare tutte le queries fatte in precedenza, entro le ultime otto ore. Osservando la History si possono combinare varie queries mediante gli indici segnati con il '#'





Nella **Clipboard** è possibile inserire (grazie al Send to visto prima) gli articoli trovati più interessanti, per riguardarli in seguito (entro otto ore).

Con details è possibile vedere come il sistema di retrieval ha interpretato e scomposto la nostra query.

Le regole di interpretazione **del sistema PubMed** sono:

- Per ogni parola viene consultato l'indice MeSH (un vocabolario creato e controllato da MedLine con tutti i termini medici e i loro sinonimi, per non perdere informazioni).
- Se non si trova nulla, viene consultato l'indice con tutti i giornali e le riviste, coi i codici ISSN e le altre informazioni.
- Se non si trova nulla, si cerca nel campo autori di tutte le entries possibili.

# Un esempio di results page di PubMed

- 1: [Jacob J, Krantz E, Dothager RS, Thiyagarajan P, Sosnick TR.](#) [Related Articles, Links](#)
-  Early Collapse is not an Obligate Step in Protein Folding.  
J Mol Biol. 2004 Apr 23;338(2):369-82.  
PMID: 15066438 [PubMed - in process]
- 2: [Degl'Innocenti D, Marzocchini R, Malentacchi F, Ramazzotti M, Raugei G, Ramponi G.](#) [Related Articles, Links](#)
-  ACYP1 gene possesses two alternative splicing forms that induce apoptosis.  
IUBMB Life. 2004 Jan;56(1):29-33.  
PMID: 14992377 [PubMed - in process]
- 3: [Degl'Innocenti D, Taddei N, Ramazzotti M, Stefani M, Chiti F, Ramponi G.](#) [Related Articles, Links](#)
-  Selection of antibody fragments specific for an alpha-helix region of acylphosphatase.  
J Mol Recognit. 2004 Jan-Feb;17(1):62-6.  
PMID: 14872538 [PubMed - in process]
- 4: [Monti M, Garolla di Bard BL, Calloni G, Chiti F, Amoresano A, Ramponi G, Pucci P.](#) [Related Articles, Links](#)
-  The regions of the sequence most exposed to the solvent within the amyloidogenic state of a protein initiate the aggregation process.  
J Mol Biol. 2004 Feb 6;336(1):253-62.  
PMID: 14741220 [PubMed - indexed for MEDLINE]

Entrez PubMed

Overview

Help | FAQ

Tutorial

New/Noteworthy

E-Utilities

Nella parte sinistra della home page di PubMed, si trovano i link a dei tutorial e a delle FAQ per imparare ad usare al meglio PubMed. **Saper cercare in PubMed è importantissimo perché tutte le ricerche partono sempre da lì!**

# SRS

## (Sequence Retrieval System)

<http://srs.ebi.ac.uk>



Library Page

Query Form

Tools

Results

Projects

Views

Databanks

SRS permette la ricerca contemporanea in molti databases ad essa collegati, permettendo l'analisi di tutte le entries di ogni database in modo appropriato integrando strumenti bioinformatici di vario tipo

# La ricerca in SRS è un sistema complesso ma facilitato dall'approccio multi-step proposto

- 1) Scelta dei databases da utilizzare per la ricerca
- 2) Immissione di una o più query concatenabili
- 3) Visualizzazione dei risultati in modo personalizzabile
- 4) Applicazione di programmi di analisi ai risultati ottenuti
- 5) Possibilità di salvare nel server EBI i risultati di una ricerca e di richiamarli successivamente.

## Fase 1: Scelta dei databases per la ricerca

### Available Databanks





[Expand all](#)     [Collapse all](#)

- Literature, Bibliography and Reference Databases
- Nucleotide sequence databases
- UniProt Universal Protein Resource
- Other protein sequence databases
- Deprecated Protein Databases
- Nucleotide related databases
- Protein function databases
- Protein structure databases
- Enzymes, reactions and metabolic pathway databases
- Mutation and SNP databases
- Gene ontology resources
- Biological Resources Catalogues (CABRI)
- Mapping databases
- Other databases
- User owned databases
- Application result databases
- EMBOSS result databases

## Fase 2: Immissione di una o più query concatenabili

Questa è la standard query page

Fields you can search	Your search terms
In a single field, you can separate multiple values by &,  , !	
	

 AllText	
 AllText	
 AllText	
 AllText	

- AllText
- ID
- Division
- Accession Number
- SeqVersion
- Molecule
- Description
- Keywords
- Organism Name
- Organelle
- Entry Creation Date

Ma è possibile tramite

**Standard Query Form**

**Extended Query Form**

impostare una ricerca mediante una extended query form, molto più articolata

## Fase 3: Visualizzazione dei risultati in modo personalizzabile

**Result Display Options**

View results using:  
EMBLSeqSimpleView ▾

or

Create a view

---

Show  ▾  
results per page

E' possibile utilizzare sia metodi di visualizzazione preimpostati, sia crearne di nuovi, facendo risaltare solo alcuni aspetti della ricerca. Quando si è impostato tutto si parte con la



### Create a view

Select the fields you want displayed in your view and choose the format

Choose 1 or more fields:

ID	<input type="checkbox"/>
Division	<input type="checkbox"/>
Accession Number	<input type="checkbox"/>
SeqVersion	<input type="checkbox"/>
Molecule	<input type="checkbox"/>
Description	<input type="checkbox"/>
Keywords	<input type="checkbox"/>

Display As:  Table  List

Sequence Format:  ▾

## Come si presentano (di default) i risultati

Query "[embl-AllText:acylphosphatase\*]" found 245 entries

next

<u>EMBL</u>	<u>Accession</u> (Links to SVA)	<u>Description</u>	<u>SeqLength</u>
<input type="checkbox"/> <a href="#">EMBL:AE017132</a>	<a href="#">AE017132</a> ; <a href="#">AE017042</a> ;	Yersinia pestis biovar Mediaevails str. 91001 section 6 of 16 of the complete genome.	290155
<input type="checkbox"/> <a href="#">EMBL:AE017204</a>	<a href="#">AE017204</a> ; <a href="#">AE017198</a> ;	Lactobacillus johnsonii NCC 533, section 5 of 7 of the complete genome.	303289
<input type="checkbox"/> <a href="#">EMBL:AE017299</a>	<a href="#">AE017299</a> ; <a href="#">AE016823</a> ;	Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130, chromosome I, section 13 of 14 of the complete sequence.	300363
<input type="checkbox"/> <a href="#">EMBL:AE017301</a>	<a href="#">AE017301</a> ; <a href="#">AE017221</a> ;	Thermus thermophilus HB27 section 1 of 7 of the complete genome.	280558
<input type="checkbox"/> <a href="#">EMBL:AK004176</a>	<a href="#">AK004176</a> ;	Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:1110039O14 product:ACYLPHOSPHATASE, ORGAN-COMMON TYPE ISOZYME (EC 3.6.1.7) (ACYLPHOSPHATE PHOSPHOHYDROLASE) (ACYLPHOSPHATASE, ERYTHROCYTE/TESTIS ISOZYME) homolog [Bos taurus], full insert sequence.	640

1. Nome del database
2. Accession number
3. Descrizione
4. Feature principale

Cliccando sui vari AN si va direttamente alla pagina EMBL con tutte le features. Si possono selezionare più entries per successive analisi

## Fase 4: Applicazione di programmi di analisi ai risultati ottenuti

Se si sono selezionate alcune sequenze si possono far partire moltissimi programmi che le analizzano. Sono programmi residenti sul server EBI e che producono risposte o visualizzabili in internet oppure scaricabili sul proprio computer appositamente formattate

**Apply Options to:**

selected results only

unselected results only

**Result Options**

Launch analysis tool:

BlastN **Launch**

**Tools**

Link to related information: **Link**

Save results: **Save**

FastX

FastY

NClustalW

Restrictionmap

BiosedN

Btwisted

Cai

Chaos

Chips

CompseqN

**Result Options**

Launch analysis tool:

BlastN **Launch**

Show tools relevant to these results: **Tools**

Link to related information: **Link**

Save results: **Save**

## Fase 5: Salvataggio di progetti e riapertura

### Contents of your temporary project

Queries		Views
Name	Query Expression	Name
Q1	[libs={uniprot uniparc uniref100 uniref90 uniref50 swissprot sptrembl}-AllText:acylphosphatase*]	No views

La pagina Project mostra tutto quello che è stato fatto in quella sessione, le queries, i risultati ottenuti, le visualizzazioni tipiche o personalizzate ecc.

**SRS Project Options**

Save to desktop:

---

























Open from desktop:

Con il campo SRS Project Options si può salvare il progetto su disco, per riaprirlo in un secondo momento.

# Entrez

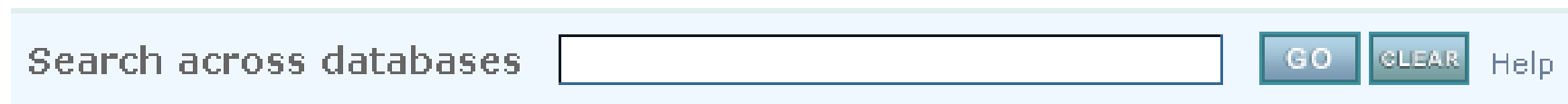
<http://www.ncbi.nlm.nih.gov/Entrez/>

## Welcome to the new Entrez cross-database search page

 <b>PubMed:</b> biomedical literature citations and abstracts		 <b>Books:</b> online books	
 <b>PubMed Central:</b> free, full text journal articles		 <b>OMIM:</b> Online Mendelian Inheritance in Man	
		 <b>Site Search:</b> NCBI web and FTP sites	
 <b>Nucleotide:</b> sequence database (GenBank)		 <b>UniGene:</b> gene-oriented clusters of transcript sequences	
 <b>Protein:</b> sequence database		 <b>CDD:</b> conserved protein domain database	
 <b>Genome:</b> whole genome sequences		 <b>3D Domains:</b> domains from Entrez Structure	
 <b>Structure:</b> three-dimensional macromolecular structures		 <b>UniSTS:</b> markers and mapping data	
 <b>Taxonomy:</b> organisms in GenBank		 <b>PopSet:</b> population study data sets	
 <b>SNP:</b> single nucleotide polymorphism		 <b>GEO:</b> expression and molecular abundance profiles	
 <b>Gene:</b> gene-centered information		 <b>GEO DataSets:</b> experimental sets of GEO data	
 <b>HomoloGene:</b> Eukaryotic homology groups			
 <b>Journals:</b> detailed information <i>about</i> the journals indexed in PubMed and other Entrez databases		 <b>MeSH:</b> detailed information about NLM's controlled vocabulary	

Entrez è un sistema di retrieval relativamente semplice ma estremamente potente perché poggiato su un consolidato motore di ricerca, identico nella forma a quello visto in precedenza per PubMed

Come in SRS è possibile scegliere il database su cui operare le ricerche, ma si lavora o su un database per volta o su tutti insieme.



Search across databases    [Help](#)

Inserendo una query direttamente nella home page è possibile effettuare uno screening dei risultati in tutti i databases disponibili, con una elegante grafica per la risposta.











Accanto ad ogni database viene infatti indicato il numero di entries ad esso associate per quella query.

192 entries in PubMed

330 entries in Nucleotide

244 entries in Protein

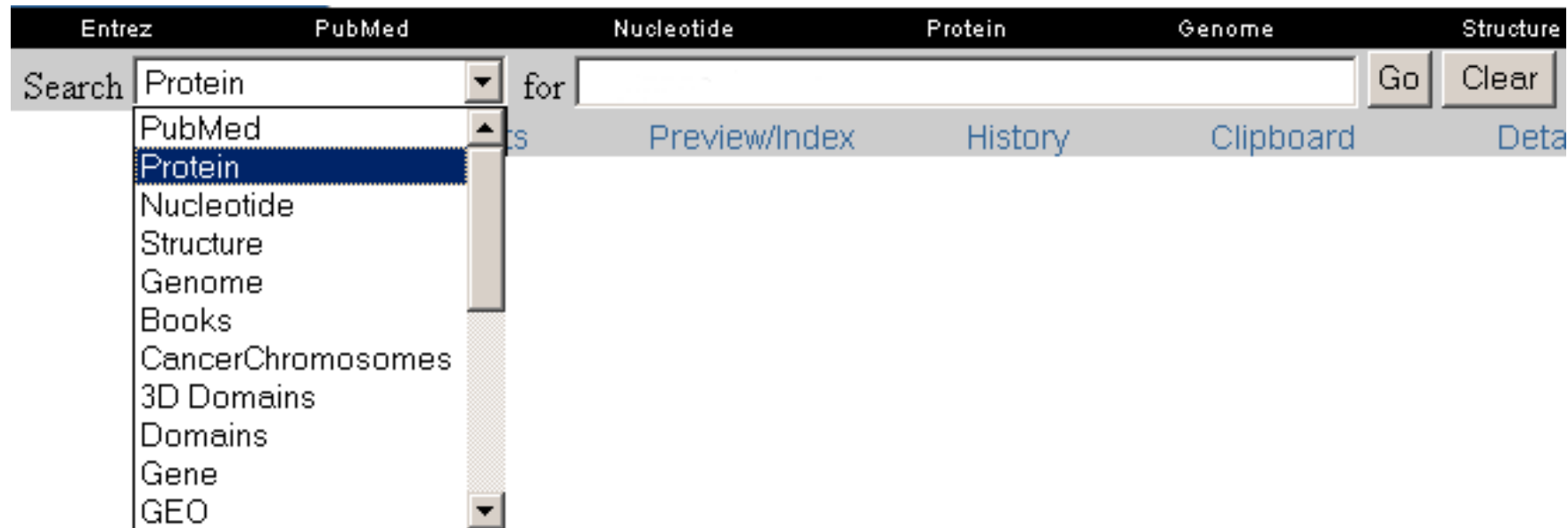
Cliccando sul database  
si accede così alle varie  
entries

192		<b>PubMed:</b> biomedical literature citations and abstracts		none		<b>Books:</b> online books	
9		<b>PubMed Central:</b> free, full text journal articles		2		<b>OMIM:</b> Online Mendelian Inheritance in Man	
none				none		<b>Site Search:</b> NCBI web and FTP sites	
330		<b>Nucleotide:</b> sequence database (GenBank)		16		<b>UniGene:</b> gene-oriented clusters of transcript sequences	
244		<b>Protein:</b> sequence database		2		<b>CDD:</b> conserved protein domain database	
71		<b>Genome:</b> whole genome sequences		4		<b>3D Domains:</b> domains from Entrez Structure	
4		<b>Structure:</b> three-dimensional macromolecular structures		9		<b>UniSTS:</b> markers and mapping data	
none		<b>Taxonomy:</b> organisms in GenBank		none		<b>PopSet:</b> population study data sets	
582		<b>SNP:</b> single nucleotide polymorphism		427		<b>GEO:</b> expression and molecular abundance profiles	
84		<b>Gene:</b> gene-centered information		none		<b>GEO DataSets:</b> experimental sets of GEO data	
99		<b>HomoloGene:</b> Eukaryotic homology groups					
none		<b>Journals:</b> detailed information <i>about</i> the journals indexed in PubMed and other Entrez databases		none		<b>MeSH:</b> detailed information about NLM's controlled vocabulary	



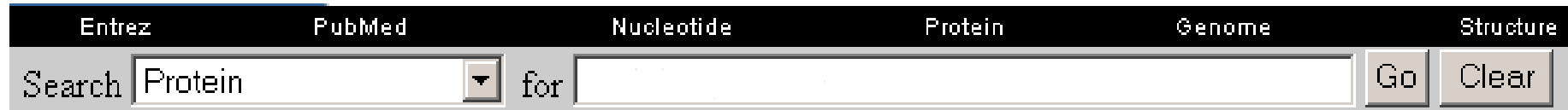
Il database protein di Entrez è una collezione di molti databases, sono più o meno gli stessi di SRS, ma non si possono scegliere.

Per navigare tra i databases non serve ritornare ogni volta alla home page: la lista completa è disponibile direttamente nelle query form di ogni database.



Il sistema di ricerca è identico a quello descritto per PubMed, tranne che per la modalità di ricerca, visto che quella valeva per il database MedLine, integrato in Entrez. Ancora ci sono

## La query form



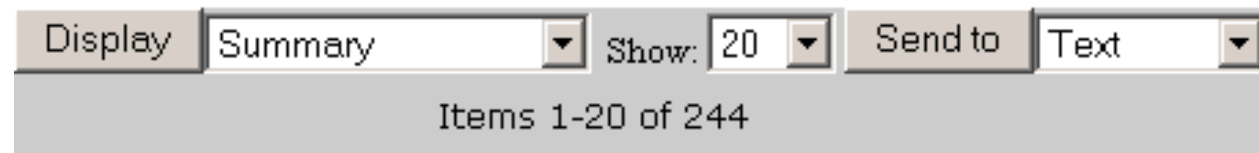
The screenshot shows the search interface for Entrez. At the top, there are tabs for 'Entrez', 'PubMed', 'Nucleotide', 'Protein', 'Genome', and 'Structure'. Below these tabs is a search bar with a dropdown menu set to 'Protein'. To the right of the dropdown is the word 'for' followed by a large empty text input field. Further right are two buttons: 'Go' and 'Clear'.

## La feature table



The screenshot shows a horizontal navigation bar with five buttons: 'Limits', 'Preview/Index', 'History', 'Clipboard', and 'Details'. The buttons are light blue with dark blue text.

## e la la results table



The screenshot shows the controls for the results table. It includes a 'Display' dropdown menu set to 'Summary', a 'Show:' dropdown menu set to '20', and a 'Send to' dropdown menu set to 'Text'. Below these controls, the text 'Items 1-20 of 244' is displayed.

## con il contatore dei risultati



The screenshot shows the page navigation bar. It includes a 'Page' dropdown menu set to '1', followed by the text 'of 13' and a 'Next' button.

# La pagina dei risultati

- 1: [BAC64697](#) BLink, Domains, Links  
putative acylphosphatase [Streptococcus pyogenes SSI-1]  
gi|28811768|dbj|BAC64697.1|[28811768]
  
- 2: [BAC59890](#) BLink, Domains, Links  
acylphosphatase [Vibrio parahaemolyticus]  
gi|28806618|dbj|BAC59890.1|[28806618]
  
- 3: [YP\\_004174](#) Links  
acylphosphatase [Thermus thermophilus HB27]  
gi|46198507|ref|YP\_004174.1|[46198507]
  
- 4: [NP\\_926939](#) BLink, Domains, Links  
probable acylphosphatase [Gloeobacter violaceus PCC 7421]  
gi|37523562|ref|NP\_926939.1|[37523562]
  
- 5: [NP\\_518340](#) BLink, Domains, Links  
PUTATIVE ACYLPHOSPHATASE PROTEIN [Ralstonia solanacearum GMI1000]  
gi|17544938|ref|NP\_518340.1|[17544938]

Si presenta come quella di PubMed, indicando anziche autori, titoli e summary, gli ID delle varie entries e una breve descrizione, con alcuni ID alternativi per quella entry.

Selezionando varie entries è possibile salvarle nella clipboard o su file in un formato richiesto, così da poterle riutilizzare.

Quello che manca ad Entrez rispetto ad SRS è l'integrazione dei programmi di analisi delle sequenze, che permettono di operare direttamente sui risultati senza il disagio del salvataggio su disco dei dati.

Anche il concetto di “progetto” che Entrez sviluppa con la Clipboard è diverso: mentre per SRS un progetto è un lavoro vero è proprio, per Entrez è come una rubrica in cui segnare le entry più interessanti.

In pratica, Entrez è un sistema di retrieval di informazioni potente e semplice da usare, ma non è ottimale per l'elaborazione. D'altra parte SRS ha i vantaggi di integrare molti programmi di analisi, ma il sistema di interrogazione è molto meno immediato.

Entrambi i sistemi di retrieval, comunque, permettono di ottenere gli stessi dati, visto che si basano su banche dati primarie e proteiche comuni, quindi hanno una grossa intersezione .

Cambiano invece i database accessori, rielaborazioni più o meno dirette dei database primari ma molto ben curati e aggiornati, quindi molto importanti per ottenere le informazioni desiderate.