

How to install the PME5 sequence databases in MASCOT

In this document:

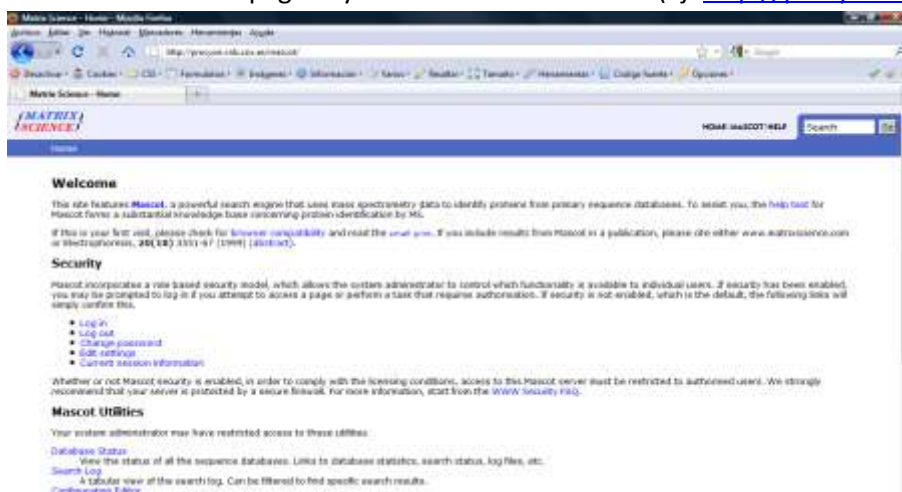
- [Copy the fasta files in your local MASCOT server](#)
- [Install PME5 database in your local MASCOT server](#)
- [Install PME5 Decoy database in your local MASCOT server](#)

Copy the fasta files in your local MASCOT server

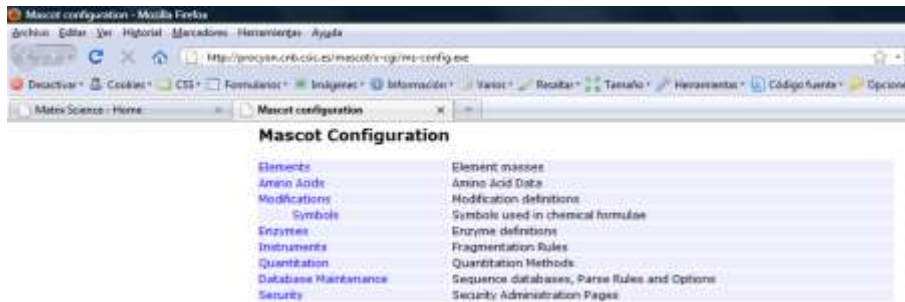
1. Go to sequence folder on your MASCOT server (ej: E:\MASCOT\sequence)
(note that your local server installation can probably be located in other folder or unit)
2. Create two folders, one for each database:
 - a) E:\MASCOT\sequence\PME5
 - b) E:\MASCOT\sequence\PME5_Decoy
3. Go to each folder and create a folder called "current":
 - a) E:\MASCOT\sequence\PME5\current
 - b) E:\MASCOT\sequence\PME5_Decoy\current
4. Copy the fasta file PME5_1.0.fasta to E:\MASCOT\sequence\PME5\current
5. Copy the file PME5_Decoy_1.0.fasta to E:\MASCOT\sequence\PME5_Decoy\current

Install PME5 database in your local MASCOT server

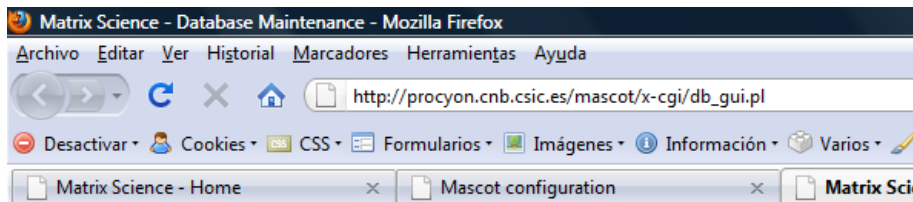
1. Go to the main web page in your local MASCOT server (ej: <http://procyon.cnb.csic.es/mascot/>)



2. Click on Configuration Editor



3. Click on Database Maintenance



Mascot Database Maintenance: Edit Database Definitions

Help Window

Select: MSDB

Name: MSDB Active Inactive

Path: E:/MASCOT/sequence/MSDB/current/MSDB_*.fasta

AA NA Mem map Mem lock

Threads: 8 Local ref file

Taxonomy: MSDB REF

Rule to parse accession string from Fasta file:
Rule 17: ">\ ([^]*)"

Rule to parse description string from Fasta file:
Rule 18: ">[^]* \\.*)"

Rule to parse accession string from local reference file:
Rule 19: ">[A-Z] [0-9]; \ ([^]*) []*"

Source and parse rule for full text report (optional):

Host: localhost Port: 80

Path: C:/INETPUB/MASCOT/x-cgi/ms-getseq.exe MSDB #ACCESSION# a
Rule 9: "\.*)"

4. We have to add a new database definition, so click on New definition button and fill de fields as follows:

- Name: PME5
- Path: E:\MASCOT\sequence\PME5\current*.fasta (or the same folder created in subsection 3a plus "\.fasta").
- Local ref file: No selected
- Taxonomy: --- None ---
- Rule to parse accession string from Fasta file: You have to select a rule like this:
 - ">.\ ([^]*)"
 - If you don't have this rule in your rule list:

- Click on "Edit parse rules" button
 - Click on "Add new rule" button
 - Paste ">..\|^[^]*\" in the last rule field, the one that is empty.
 - Click on "Return to database definitions" button.
- Rule to parse description string from Fasta file: You have to select a rule like this: ">[^]* \\.*)\"
 - If you don't have this rule in your rule list, proceed as in previous section to add this new rule.
 - Rule to parse accession string from local reference file: ---no local reference file---
 - Host: www.expasy.org
 - Port:80
 - Path: /cgi-bin/get-sprot-raw.pl?#ACCESSION#
 - Rule: "\\.*)\" (proceed as in previous sections if you don't have this rule in your rule list).

Select	PME5	
Name	PME5	Active <input checked="" type="radio"/> Inactive <input type="radio"/>
Path	E:/MASCOT/sequence/PME5/current/*.fasta	
	AA <input checked="" type="radio"/> NA <input type="radio"/>	Mem map <input checked="" type="checkbox"/> Mem lock <input type="checkbox"/>
	Threads 8	Local ref file <input type="checkbox"/>
Taxonomy	--- None ---	
Rule to parse accession string from Fasta file:		
	Rule 45: ">..\ ^[^]*\"	
Rule to parse description string from Fasta file:		
	Rule 46: ">[^]* \\.*)\"	
Rule to parse accession string from local reference file:		
	--- no local reference file ---	
Source and parse rule for full text report (optional):		
Host	www.expasy.org	Port 80
Path	/cgi-bin/get-sprot-raw.pl?#ACCESSION#	
	Rule 23: "\\.*)\"	
Test this definition		Delete this definition
New definition		Edit parse rules
		Edit options

No changes are written to mascot.dat until you choose:

- 5. Click on "Test this definition" to test the parse rules. You have to obtain a result like this:

Mascot Database Maintenance

Testing Database Definition PME5

Testing entries at beginning and end of E:\MASCOT\sequence\PME5\current\PME5_1.0.fasta:

Accession	Description
P05100	DNA-3-methyladenine glycosylase 1 OS=Escherichia coli (strain K12) GN=tag PE=1 SV=1
P04395	DNA-3-methyladenine glycosylase 2 OS=Escherichia coli (strain K12) GN=alkA PE=1 SV=1
P37754	6-phosphogluconate dehydrogenase, decarboxylating OS=Escherichia coli GN=gdh PE=3 SV=1
P00350	6-phosphogluconate dehydrogenase, decarboxylating OS=Escherichia coli (strain K12) GN=gdh PE=1 SV=2
A72822	6-phosphogluconolactonase OS=Escherichia coli O139:H28 (strain E24377A / ETEC) GN=pgl PE=3 SV=1
P0ACS1	Zinc uptake regulation protein OS=Escherichia coli (strain K12) GN=zur PE=1 SV=1
P00004	Cytochrome c OS=Equus caballus GN=CYCS PE=1 SV=2
P68082	Myoglobin OS=Equus caballus GN=MB PE=1 SV=2
P00083	Fructose-bisphosphate aldolase A OS=Cryptosporidium parvum GN=ALDOA PE=1 SV=2
P02769	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Example of full text report:

```
ID   SMS1_ECOLI              Reviewed:      187 AB.
AC   P05100; Q9HFL0;
DT   13-NOV-1987, integrated into UniProtKB/TrEMBL.
DT   13-NOV-1987, sequence version 1.
DT   15-DEC-2008, entry version 87.
DE   RefName: Full=OHA-3-methyladenine glycosylase 1/
DE       EC=1.1.1.201
DE   AltName: Full=OHA-3-methyladenine glycosylase 1/
DE       AltName: Full=3-methyladenine-OHA glycosylase 1, constitutive/
```

6. Click on “Return to database definitions”. If you have obtained some error, or the Accession column has captured any other data, please review the parse rules again. It is important that all laboratories capture the Accession of each protein as you can see in previous figure (P05100, P04395, ...).
7. Click on “APPLY” button to save changes.
8. You will obtain this screen if all is OK:

Mascot Database Maintenance

Changes have been saved to	mascot.dat
The previous contents of mascot.dat are in	mascot.dat.160
To make further changes, do not press the browser back button, Follow	This link
Display	Database Status

9. Click on Database Status to see how the database is being installed automatically by MASCOT:

```
Name       = PME5                      Family    = E:\MASCOT\sequence\PME5\current\*.fasta
Filename   = PME5_1.0.fasta             Pathname  = E:\MASCOT\sequence\PME5\current\PME5_1.0.fasta
Status     = In use                    Statistics
State Time = Wed Dec 16 12:16:46 # searches = 0
Mem mapped = YES Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = 8 Current = YES
```

Install PME5_Decoy database in your local MASCOT server

1. Follow the steps 1, 2 and 3 from the previous section “Install PME5 database in your local MASCOT server”
2. We have to add a new database definition, so click on New definition button and fill the fields as follows:
 - o Name: PME5_Decoy
 - o Path: E:\MASCOT\sequence\PME5_Decoy\current*.fasta (or the same folder created in subsection 3b plus “*.fasta” from previous section).
 - o Local ref file: No selected
 - o Taxonomy: --- None ---

- Rule to parse accession string from Fasta file: You have to select a rule like this: ">..\|^[^|]*\"
 - If you don't have this rule in your rule list:
 - Click on "Edit parse rules" button
 - Click on "Add new rule" button
 - Paste ">..\|^[^|]*\" in the last rule field, the one that is empty.
 - Click on "Return to database definitions" button.
- Rule to parse description string from Fasta file: You have to select a rule like this: ">[^]* \\.*\\"
 - If you don't have this rule in your rule list, proceed as in previous section to add this new rule.
- Rule to parse accession string from local reference file: ---no local reference file---
- Host: www.expasy.org
- Port:80
- Path: /cgi-bin/get-sprot-raw.pl?#ACCESSION#
- Rule: "\\.*\\" (proceed as in previous sections if you don't have this rule in your rule list).

Select	--- New Database ---	
Name	PME5_Decoy	Active <input checked="" type="radio"/> Inactive <input type="radio"/>
Path	E:/MASCOT/sequence/PME5_Decoy/current/*.fasta	
	AA <input checked="" type="radio"/> NA <input type="radio"/>	Mem map <input checked="" type="checkbox"/> Mem lock <input type="checkbox"/>
	Threads 8	Local ref file <input type="checkbox"/>
Taxonomy	--- None ---	
Rule to parse accession string from Fasta file:		
	Rule 45: ">..\ ^[^]*\"	
Rule to parse description string from Fasta file:		
	Rule 46: ">[^]* \\.*\\"	
Rule to parse accession string from local reference file:		
	--- no local reference file ---	
Source and parse rule for full text report (optional):		
Host	www.expasy.org	Port 80
Path	/cgi-bin/get-sprot-raw.pl?#ACCESSION#	
	Rule 23: "\\.*\\"	
	<input type="button" value="Test this definition"/>	<input type="button" value="Delete this definition"/>
	<input type="button" value="New definition"/>	<input type="button" value="Edit parse rules"/>
	<input type="button" value="Edit options"/>	
No changes are written to mascot.dat until you choose: <input type="button" value="APPLY"/>		

3. Click on "Test this definition" to test the parse rules. You have to obtain a result like this:



Mascot Database Maintenance

Testing Database Definition PMES_Decoy

Testing entries at beginning and end of E:/MASCOT/sequence/PMES_Decoy/current/PMES_Decoy_1.0.fasta:

Accession	Description
P05100	DNA-3-methyladenine glycosylase 1 OS=Escherichia coli (strain K12) GN=tag PE=1 SV=1
rndP05100	DNA-3-methyladenine glycosylase 1 OS=Escherichia coli (strain K12) GN=tag PE=1 SV=1
P04395	DNA-3-methyladenine glycosylase 2 OS=Escherichia coli (strain K12) GN=alkA PE=1 SV=1
rndP04395	DNA-3-methyladenine glycosylase 2 OS=Escherichia coli (strain K12) GN=alkA PE=1 SV=1
P37754	6-phosphogluconate dehydrogenase, decarboxylating OS=Escherichia coli GN=gnd PE=3 SV=1
rndP08082	Myoglobin OS=Equus caballus GN=MB PE=1 SV=2
P00883	Fructose-bisphosphate aldolase A OS=Oryctolagus cuniculus GN=ALDOA PE=1 SV=2
rndP00883	Fructose-bisphosphate aldolase A OS=Oryctolagus cuniculus GN=ALDOA PE=1 SV=2
P02769	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4
rndP02769	Serum albumin OS=Bos taurus GN=ALB PE=1 SV=4

Example of full text report:

```
ID: P05100, ECOLI Reviewed: 167 AB...
AC: P05100; Q2NTLO;
DT: 13-MAY-1991, integrated into UniProtKB/Swiss-Prot.
DT: 13-MAY-1991, sequence version 1.
DT: 15-DEC-2009, entry version 87.
DE: RecName: Full=DNA-3-methyladenine glycosylase 1;
DE: EC=3.2.2.20;
DE: AltName: Full=DNA-3-methyladenine glycosylase 1;
DE: AltName: Full=3-methyladenine-DNA glycosylase 1, constitutive;
```

4. Click on "Return to database definitions". If you have obtained some error, or the Accession column has captured any other data, please review the parse rules again. It is important that all laboratories capture the Accession of each protein as you can see in previous figure (P05100, rndP05100, ...).
5. Click on "APPLY" button to save changes.
6. You will obtain this screen if all is OK:



Mascot Database Maintenance

Changes have been saved to [mascot.dat](#)
 The previous contents of mascot.dat are in [mascot.dat.161](#)
 To make further changes, do not press the browser back button, Follow [This link](#)
[Display Database Status](#)

7. Click on Database Status to see how the database is being installed automatically by MASCOT:

```
Name = PMES_Decoy Family = E:/MASCOT/sequence/PMES_Decoy/current/*.fasta
Filename = PMES_Decoy_1.0.fasta Pathname = E:/MASCOT/sequence/PMES_Decoy/current/PMES_Decoy_1.0.fasta
Status = Creating decompressed files 28% complete
State Time = Wed Dec 16 12:26:14 # searches = 0
Mem mapped = NO Request to mem map = YES Request unmap = NO Mem locked = NO
Number of threads = 8 Current = NO
```