European Directorate for the Quality of Medicines & HealthCare | Direction européenne de la qualité du médicament & soins de santé

Funded by the European Union and the Council of Europe

EUROPEAN UNION

COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Implemented by the Council of Europe

General European OMCL Network (GEON)
QUALITY MANAGEMENT DOCUMENT

PA/PH/OMCL (15) 04 2R

INTERPRETATION OF SCREENING RESULTS FOR UNKNOWN PEPTIDES AND PROTEINS BY MASS SPECTROMETRY BASED METHODS

Full document title and reference	Interpretation of Screening Results for Unknown Peptides and Proteins by Mass Spectrometry Based Methods PA/PH/OMCL (15) 04 2R
Document type	Recommendation Document
Legislative basis	Council Directive 2001/83/EC and 2001/82/EC, as amended
Date of first adoption	25 April 2016
Date of original entry into force	2 May 2016
Date of entry into force of revised document	N/A
Previous titles/other references / last valid version	N/A
Custodian Organisation	The present document was elaborated by the OMCL Network / EDQM of the Council of Europe
Concerned Network	GEON

1. Introduction

Recent advances in genomics, recombinant DNA technologies and peptide synthesis have led to an increased development of protein and peptide therapeutics. Unfortunately this goes hand in hand with a growing market of counterfeit and illegal biopharmaceuticals, including complex protein and peptide mixtures even of animal origin. These counterfeits and illegal protein and peptide substances could imply severe health threats as has been demonstrated by numerous case reports. One of the difficulties encountered by laboratories responsible for the analysis of these substances is the lack of reference standards for many of these samples. In order to sidestep these limitations we could envisage the use of mass spectrometry, analytical libraries and bioinformatics tools. The output from these Mass Spectrometry (MS)-based methods using database searches is expressed in statistical probabilities and therefore, interpretation of the results may not be straightforward. Thus, there is a growing demand to harmonise the interpretation of screening results of these MS-based methods and database searches. Identification is not always possible only by a MS-based screening approach but could require additional methods.

On this document we propose a workflow to acquire and evaluate analytical data on peptides and proteins obtained by MS and data base searches. Whilst proteins may be primarily analysed by bottom-up MS, top-down protein profiling (based on the molecular mass) and top-down sequencing, the identification of peptides is based on a system of Identification Points (IP) as recommended for sports drug testing applications [1].

2. Approaches of the proposed identification

2.1 Peptides

2.1.1 Reference standard available

The reliability of compound identification very much depends on the mass spectrometer in question and the detection mode of the MS method. The identification is based on the charge state of the precursor ion, and the simultaneous or subsequent generation fragment ions (typically b and/or y) from that precursor ion. Additionally, a reference standard should be analysed by the same MS methodology used for analysis of the suspected sample.

The scoring system for identification of peptides, based on Thevis et al. [1] is as follows:

- In low-resolution MS (FWHM < 10 000 [2-3]), 1 IP is earned for the molecular weight (Mw) identification and 1.5 IPs are earned per matching fragment ion.
- In high-resolution MS (FWHM ≥ 10 000 [2-3]), 2 IPs are earned for the Mw identification and 2.5 IPs are earned for one MS^n matching fragment ion.
- A minimum 5 IPs, including 2 fragment ions, is required for compound identification, for larger peptides higher IPs may be necessary.

Mass spec	IP
Low-resolution MS precursor ion/ molecular weight determination	1
Low-resolution MS^n fragment ion	1.5
High-resolution/high-accuracy MS precursor ion/ molecular weight determination	2.0
High-resolution/high-accuracy MS^n fragment ion	2.5

Table 1: IP scores given for precursor and fragment ions matching the ions observed in the reference standard

2.1.2 Reference standard not available

If no standard is available, information on the identity of the peptide can be obtained with a combination of approaches. After acquiring adequate analytical MS data of the intact or digested peptide [5, 8], the data is compared to that obtained from the literature [4-11], data bases or synthetic peptides (= peptide that is custom made to confirm the identity of the suspected peptide (natural or from a different origin) in the sample).

2.2 Proteins

The amino acid sequences of many biopharmaceuticals are engineered and differ in part from those of naturally existing proteins. In such cases, identification of biopharmaceuticals by comparisons against data bases containing natural proteins are of limited use, and special data bases containing the sequences of biopharmaceuticals are necessary. Preferentially, biological data bases, where the sequences of biopharmaceuticals are added ('spiked'), should be used in order to facilitate searches against naturally existing and engineered proteins.

For the identification of proteins, bottom-up and top-down approaches are applied. In addition, *de novo* sequencing and orthogonal biochemical, immunological or physiochemical strategies may be needed to confirm identity.

2.2.1. Bottom-up MS

The bottom-up approach involves digestion of the sample with a proteolytic enzyme (such as trypsin) or a chemical treatment that cleaves the protein at specific sites to create a complex peptide mixture. The digest is then analysed on LC-MS.

For identification using bottom-up MS, generally a minimum of two significant peptide matches is required. The matching peptides must contain sequences unique to the protein to validate identification. The significance of the hit (score) is defined by the programme or algorithm used and is based on the quality of the data (several algorithms are available and an example using Mascot is presented below as example 1). For unambiguous identification of a protein the two significant peptides have to be unique for the protein in question.

Example 1: Results LC-MS/MS analysis on the tryptic digest of unknown sample and analysis by MASCOT bio-informatics tools:

1) Protein scores:

The mascot search results gives a list (ordered by protein score) of possible protein hits. In the example a search was carried out against the Swisprot database and no species-related restrictions were used. Therefore the results reflect the possible protein and the possible species. However, care has to be taken with the species hit since homologues proteins may result in the same or similar protein score.

Mascot Search Results

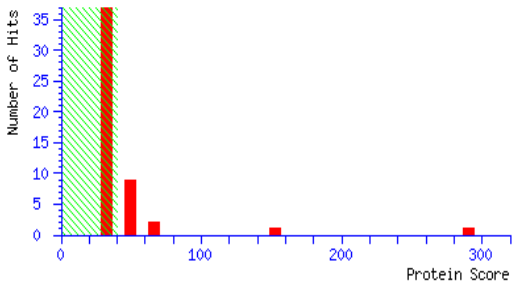
User :
Email :
Search title : sample cyprus
MS data file : DATA.TXT
Database : SwissProt 2012_03 (535248 sequences; 189901164 residues)
Timestamp : 10 Oct 2014 at 08:10:07 GMT
Protein hits :

TRYP_PIG	Trypsin OS=Sus scrofa PE=1 SV=1
CGHB_HUMAN	Choriogonadotropin subunit beta OS=Homo sapiens GN=CGB PE=1 SV=1
GLHA_HUMAN	Glycoprotein hormones alpha chain OS=Homo sapiens GN=CGA PE=1 SV=1
RS24_METM6	30S ribosomal protein S24e OS=Methanococcus maripaludis (strain C6 / ATCC BAA-1332) GN=rps24e PE=3 SV=1
IMDH2_DANRE	Inosine-5'-monophosphate dehydrogenase 2 OS=Danio rerio GN=impdh2 PE=3 SV=1
PYRB_PARD8	Aspartate carbamoyltransferase OS=Parabacteroides distasonis (strain ATCC 8503 / DSM 20701 / NCTC 11152) GN=pyrB PE=3 SV=1
DCR_ARATH	BAHD acyltransferase DCR OS=Arabidopsis thaliana GN=DCR PE=2 SV=1
QRIC2_HUMAN	Glutamine-rich protein 2 OS=Homo sapiens GN=QRICH2 PE=2 SV=1
PLC1_YEAST	1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase 1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) GN=PLC1 PE=1 SV=1
PS4_PINST	Putative LRR disease resistance protein/transmembrane receptor kinase PS4 (Fragment) OS=Pinus strobus PE=1 SV=1
CBPA1_RAT	Carboxypeptidase A1 OS=Rattus norvegicus GN=Cpa1 PE=2 SV=2
Y7955_DICDI	Putative uncharacterized protein DDB_G0288489 OS=Dictyostelium discoideum GN=DDB_G0288489 PE=4 SV=1
GCSP_GRAFK	Glycine dehydrogenase [decarboxylating] OS=Gramella forsetii (strain KT0803) GN=gcvP PE=3 SV=1
MRAY_BORHD	Phospho-N-acetylmuramoyl-pentapeptide-transferase OS=Borrelia hermsii (strain DAH) GN=mraY PE=3 SV=1
HIS1_BACHD	ATP phosphoribosyltransferase OS=Bacillus halodurans (strain ATCC BAA-125 / DSM 18197 / FERM 7344 / JCM 9153 / C-125) GN=hisG PE=3 SV=1
AGL12_ARATH	Agamous-like MADS-box protein AGL12 OS=Arabidopsis thaliana GN=AGL12 PE=2 SV=2
RPOB_BLOFL	DNA-directed RNA polymerase subunit beta OS=Blochmannia floridanus GN=rpoB PE=3 SV=1
AROE_ANADF	Shikimate dehydrogenase OS=Anaeromyxobacter sp. (strain Fw109-5) GN=aroE PE=3 SV=1
WDR1_CAEEL	Actin-interacting protein 1 OS=Caenorhabditis elegans GN=unc-78 PE=1 SV=1
SYFB_LACJO	Phenylalanine--tRNA ligase beta subunit OS=Lactobacillus johnsonii (strain CNCM I-12250 / La1 / NCC 533) GN=pheT PE=3 SV=1

Next, the mascot score histogram reflects the number of protein scores (x-axis) in function of the number of hits (y-axis). Briefly, all the protein hits that reside in the green region in the example probably reflect a random hit and all protein hits with a protein score > 40 is considered as a putative identification. The protein score in the result report from an MS/MS search is derived from the ions scores. The Ion Score or peptide score is a measure of how well the observed MS/MS spectrum matches to the stated peptide (observed an theoretical b and y ions). For more detailed information on the MASCOT data output we would refer to http://www.matrixscience.com.

Mascot Score Histogram

Ions score is -10*Log(P), where P is the probability that the observed match is a random event.
Individual ions scores > 40 indicate identity or extensive homology (p<0.05).
Protein scores are derived from ions scores as a non-probabilistic basis for ranking protein hits.



However, one must verify if indeed significant non-redundant peptides are present. This info is given in peptide scores, which are based on observed and theoretical fragment ions (typically b and y ions) and are indicated by * in the summary report.

2) Peptide summary report for all predicted peptides for hit nr 2 (β-subunit of HcG):

	Query	Start – End	Observed	Mr(expt)	Mr(calc)	Delta	M	Score	Expect	Rank	U	Peptide
*	199	64 – 80	642.3570	1924.0492	1924.0717	-0.0225	0	49		1	U	R.VLQGVLPALPQVVCNYR.D + [-0.9840 at C-term]
*	22	95 – 102	417.6910	833.3674	833.4283	-0.0609	0	52		1	U	R.GVNPVVSY.A
	96	125 – 134	409.1730	1224.4972	1224.5193	-0.0222	0	29	0.94	1	U	K.DHPLTCDDPR.F
*	35	135 – 142	443.1500	884.2854	884.3876	-0.1021	0	51	0.0087	1	U	R.FQDSSSSK.A
*	68	143 – 153	553.2710	1104.5274	1104.5927	-0.0653	0	66	0.00028	1	U	K.APPPSLPSPSR.L

→ These scores indicate the presence of 4 significant non-redundant peptides

2.2.2. Top down MS

The top-down approach describes two different techniques on whole proteins. It is applied to "top-down profiling", which results in the determination of the mass of certain protein(s) (cfr. 2.1.1. High-resolution/high-accuracy MS precursor ion(s)). Further, "top-down sequencing" allows the determination of sequence information of a given protein.

For "top-down sequencing" ESI-ETD/PTR-IT or MALDI-TOF/TOF has emerged as powerful tools for the analysis of proteins. For a positive identification, the obtained results should cover at least 20 sequential amino acids with 100% identity, with the exception of prolines, and the sequence should be unique for the protein. It should be noted, that the information provided by "top-down sequencing" alone may not be adequate for identification, if sequence modifications outside the confirmed sequence are suspected.

3. Abbreviations

ESI-ETD: Electron spray ionisation-electron transfer dissociation

PTR-IT: proton transfer reaction-ion trap

MALDI-TOF: matrix assisted laser desorption ionisation- time of flight

4. References

[1] M. Thevis, W. Schänzer (2007). Current role of LC-MS(/MS) in doping control. Anal Bioanal Chem 388:1351-1358.

[2] 2002/657/EC: Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results (Text with EEA relevance) (notified under document number C(2002) 3044).

[3] M. Holcapek, R. Jirásko, M. Lísa (2012). Recent developments in liquid chromatography–mass spectrometry and related techniques. Journal of Chromatography A 1259: 3–15.

[4] A. Thomas, S. Höppner, H. Geyer, W. Schänzer, M. Petrou, D. Kwiatkowska, A. Pokrywka, M. Thevis (2011). Determination of growth hormone releasing peptides (GHRP) and their major metabolites in human urine for doping controls by means of liquid chromatography mass spectrometry. Anal Bioanal Chem 401:507-16.

[5] J. Henninge, M. Pepaj, I. Hullstein, P. Hemmersbach (2010). Identification of CJC-1295, a growth-hormone-releasing peptide, in an unknown pharmaceutical preparation. Drug Test Anal 2:647-650.

[6] T. Breindahl, M. Evans-Brown, B. Hindersson, J. McVeigh, M. Bellis, A. Stensballe, A. Kimergård (2014). Identification and characterization by LC-UV-MS/MS of melanotan II skin-tanning products sold illegally on the Internet. Drug Test Anal. Doi: 10.1002/dta.1655.

[7] M.C. Gaudiano, L. Valvo, A. Borioni (2014). Identification and quantification of the doping agent GHRP-2 in seized unlabelled vials by NMR and MS: a case-report. Drug Test Anal. 6:295-300.

[8] C. Vanhee, G. Moens, E. Deconcinck, J.O. De Beer (2014). Identification and characterization of peptide drugs in unknown pharmaceutical preparations seized by the Belgian authorities: case report on AOD9604. Drug Test Anal. 6:964-968.

[9] S. Esposito, K. Deventer, P. Van Eenoo (2012). Characterization and identification of a C-terminal amidated mechano growth factor (MGF) analogue in black market products. Rapid Commun Mass Spectrom. 26:686-692.

[10] C. Vanhee, G. Moens, E. Van Hoeck, E. Deconinck, J.O. De Beer (2015). Identification of the small research tetra peptide Epitalon, assumed to be a potential treatment for cancer, old age and Retinitis Pigmentosa in two illegal pharmaceutical preparations. Drug Test Anal. Doi 10.1002/dta.1771.

[11] C. Vanhee, S. Janvier, B. Desmedt, G. Moens, E. Deconinck, J.O. De Beer, P. Courselle (2015). Analysis of illegal peptide biopharmaceuticals frequently encountered by controlling agencies. Talanta 142: 1–10.