Journal of Proteomics & Bioinformatics - Open Access

Research Article JPB/Vol.1/April 2008

Automated Calculation of Unique Peptide Sequences for Unambiguous Identification of Highly Homologous Proteins by Mass Spectrometry

Michael Kohl[#][*], Gorden Redlich[#], Martin Eisenacher, Anke Schnabel, Helmut E. Meyer, Katrin Marcus and Christian Stephan

Medizinisches Proteom-Center, Ruhr-Universität Bochum, Bochum, Germany

*Corresponding author: Michael Kohl, Medizinisches Proteom-Center, ZKF E.2.051, Ruhr-Universitaet Bochum, Bochum, Universitaetsstr. 150, D-44801 Bochum, Germany
Tel: +49-234-32-29275; Fax: +49-234-32-14554; E-mail: Michael.Kohl@rub.de

[#] Both authors contributed equally to this work

Received April 03, 2008; Accepted April 09, 2008; Published April 20, 2008

Citation: Michael K, Gorden R, Martin E, Anke S, Helmut EM, et al. (2008) Automated Calculation of Unique Peptide Sequences for Unambiguous Identification of Highly Homologous Proteins by Mass Spectrometry. J Proteomics Bioinform 1: 006-010. doi:10.4172/jpb.1000003

Abstract

In regular proteomics approaches proteases are used to digest the proteome into a set of peptides. Unfortunately, determining proteins on the basis of peptides implies some uncertainty since a peptide may be part of different proteins. Therefore, a targeted detection of unique peptides of particular proteins is a promising task for an unambiguous identification of a specific protein. Here we present a software solution that offers the possibility for a highly efficient and simple detection of such unique peptides. In a first step a SQL-based database of theoretically digested peptides from a given FASTA file formatted protein database is generated by choosing a protease. In a second step, *in silico* generated peptides from a pre-defined protein sequence are compared to this peptide database in order to identify unique peptides. Amongst others, possible applications are identification of proteins when only sparse peptide information is available or advanced proteomics techniques that require information about the uniqueness of peptides such as Multiple Reaction Monitoring (MRM).

Keywords: Multiple reaction monitoring ; Unique peptides; In silico digest; SQL database; Java

Abbreviations

AUC: Area Under the Curve; **AQUA:** Absolute Quantification; **DBMS:** Database Management System; **IPI:** International Protein Index; **JDK:** Java Development Kit; **JVM:** Java Virtual Machine; **MPC:** Medizinisches Proteom-Center, **MRM:** Multiple Reaction Monitoring; **MS:** Mass Spectrometry; **UPF:** Unique Peptide Finder

Introduction

Proteomics is a powerful methodology to investigate protein expression in cells, tissues, organs or whole organisms. One fundamental idea of proteomic approaches is the expression analysis of thousands of proteins at the same time. Proteome analysis more and more appears into the spotlight of classical fundamental as well as clinical research. Differential quantitative proteome analysis allows direct comparison of proteomes of different cellular states whereas descriptive qualitative approaches provide an insight in the protein composition of a given cell, organelle, tissue etc. Protein identification usually is done by mass spectrometry (MS). MS can be either performed in the form of whole-protein analysis (“top-down” approach) or by investigation of enzymatically produced peptides (“bottom-up” approach).

In bottom-up proteomics approaches trypsin or other proteases are used to digest the proteins into a set of peptides prior to their mass spectrometric analysis (Kocher and Superti-Furga, 2007). Substantial advantages when working with peptides instead of proteins such as superior solubility of peptides in a wide variety of solvents as well as their lower nonspecific adsorption to surfaces make the bottom-up approach very attractive for comprehensive proteome analysis. On the other hand, after protein digestion the number of molecular species increases dramatically, which complicates the analysis (Lohaus et al., 2007). Unfortunately, all information about a particular intact protein is lost. Therefore, determination of proteins on the basis of detected peptides implies some uncertainty since a peptide may be part of different proteins.

The detection of unique peptides is especially important, when doing targeted proteomics. In this (special) case the proteins that shall be analyzed by mass spectrometry are already known before the analysis.

As an example, multiple reaction monitoring (MRM) turned out to be a fast and efficient technique, which allows quantifying proteins either relatively or absolutely by means of triggering a set of specific peptides (Janecki et al., 2007; Le Blanc et al., 2003; Mayya et al., 2006; Wolf-Yadlin et al., 2007). Analysis of very similar proteins usually deals with only a small set of unique peptides that needs to be calculated for unambiguous identification of these proteins.

Manually this can for example be done by theoretic digestion of the sequences of interest and blasting of all resulting peptides (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi). However, performing BLAST searches manually in order to identify unique peptides is both tedious and time-consuming.

Nowadays sample preparation, nanoHPLC, mass spectrometry and protein identification in modern proteomics can be automated (Alterovitz et al., 2006) , but to our knowledge no software is available for automated exact calculation of unique peptides for unambiguous identification of proteins. Here we report the construction of a software tool, which is called Unique Peptide Finder (UPF). UPF is used to calculate sets of unique peptides from protein sequences. Furthermore, UPF provides possibilities to calculate the frequency of peptides and to retrieve information from which proteins a particular peptide can be derived.

The automation of this process requires the theoretic digestion of sequences of interest and a comparison with a predigested protein sequence database avoiding the discussed problems caused by using the BLAST function against a non-digested sequence database.

The software will help on the interpretation of regular proteomics data and the generation of peptide lists for targeted proteomics especially when it comes to a desired differentiation between very similar proteins.

Material and Methods

Software Requirements

The developed tool is a Java™ -based software solution (JDK 6, Sun Microsystems Inc., Santa Clara, CA, USA). Therefore, the user needs to install the Java Virtual Machine (JVM) on his computer. Since Java™ provides platform independent technology, UPF can be used on various computer systems (e.g. Microsoft™ Windows™ or Linux operating systems). In order to identify unique peptides with respect to a given set of proteins, it is necessary to generate databases containing all peptides from a whole set of proteins. Therefore, the UPF software uses the relational database model for storing peptide information obtained from predigested protein sequences. Users need to have access to an adequate database management system (DBMS). During the development of the software, Microsoft™ SQL Server 2000 (Microsoft™ Corp., Redmond, WA, USA) was used as database management system. However, for providing the use of UPF free of charge, the open source database software MySQL can be utilized as well. Adequate database drivers are supplied with UPF for both DBMSs.

Conceptual Design of the UPF Software

In general the software readily generates an exact list of peptide related information. Counting the peptides from an *in silico*-digested protein of interest in a pre-digested protein database results in a frequency being of particular importance. This holds true, because then it is quite simple to detect peptides that are unique for a particular protein when considering the frequency information.

The presented software consists of two modules:

Module 1 - The first module is used to generate databases, which contain comprehensive information about peptides obtained from enzymatic digestion of a set of proteins. An important objective during the development was to provide easy access to current and common protein databases. Currently, designated input can be IPI and UniProtKB/Swiss-Prot databases in FASTA file format. Current IPI databases are available from ftp://ftp.ebi.ac.uk/pub/databases/IPI/current/ and the UniProtKB/Swiss-Prot database is available from ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/knowledgebase/complete/. Processing of the NCBInr database will be integrated into UPF in the near future.

For each considered enzymatic digestion a particular peptide database has to be created. Each of those databases consists of two tables. The first table stores information retrieved from the headers of the protein entries of the FASTA file (e.g. alias accession numbers, taxonomic identification, etc.) Great importance is attached on retrieving a maximum of information. The second table stores sequences and masses of the peptides obtained from an *in silico* digestion. However, no masses are calculated for the peptides when their sequences contain any wildcards (symbols B, X and Z).

The second table is linked to the first table. Therefore, starting from a specific peptide sequence ‘A’, which occurs in the second table, it is possible to track any protein from the first table that contains peptide sequence ‘A’.

This database generating tool (Module 1) is a command line tool to allow integration into scheduled batch processes. Hence, tool generated databases can be easily synchronized with current protein databases, that are available from the internet.

Module 2 - The major task of the second module is an automatic generation of database queries and the presentation of the results. Users simply need to provide a protein or peptide sequence of interest as input for module 2. An *in silico* digest is performed in order to get all possible peptides from this input sequence. Then, frequencies of occurrence are computed for this set of peptides with respect to the database that was generated by module 1. In order to allow intuitive applicability, the second module was designed with an easy to use graphical user interface (Figures. 1, 2).

Further Notes

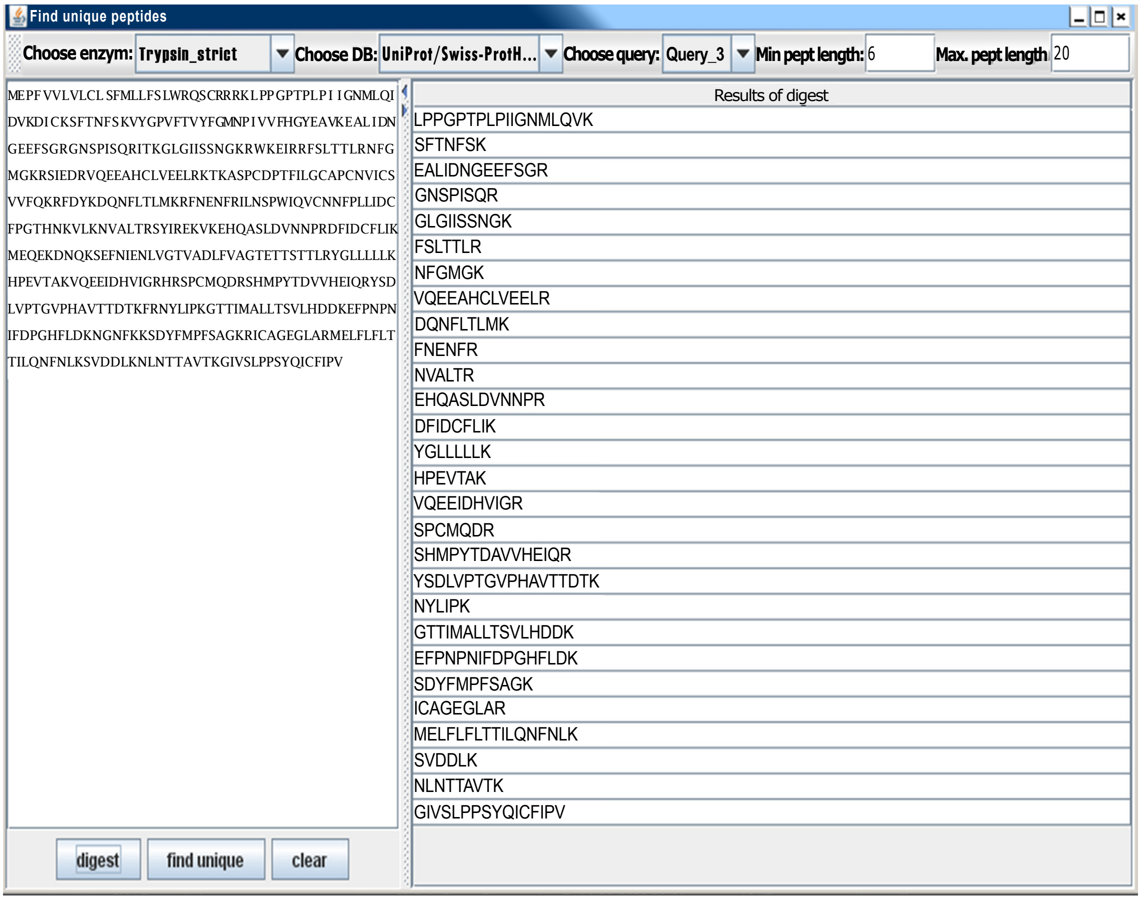
UPF is currently designed for usage of both IPI and the UniProtKB/Swiss-Prot database. Since the UniProtKB/Swiss-Prot database offers manual validated entries we recommend the use of this database. However, usage of other databases may be indispensible for finding proteins not included in the UniProtKB/Swiss-Prot. It is planned to provide opportunity for incorporating a wide range of different databases into future versions of the UPF software. Therefore, maximum freedom of decision regarding the choice of an appropriate database remains in the hands of the user.

The following proteins were used to evaluate the software (Swissprot accession in parentheses): CYP2C8 (P10632), CYP2C9 (P11712), CYP2C18 (P33260), CYP2C19 (P33261).

UPF can be downloaded as a zip formatted file from the website (software section) of the ‘Medizinisches Proteom-Center’ (MPC), Ruhr-Universitaet Bochum (http://www.medizinisches-proteom-center.de/). The zip file additionally includes a detailed user manual.

Results and Discussion

Mass spectrometry is the standard tool for protein identification, which is prevalently performed on peptide level after of a particular protein.

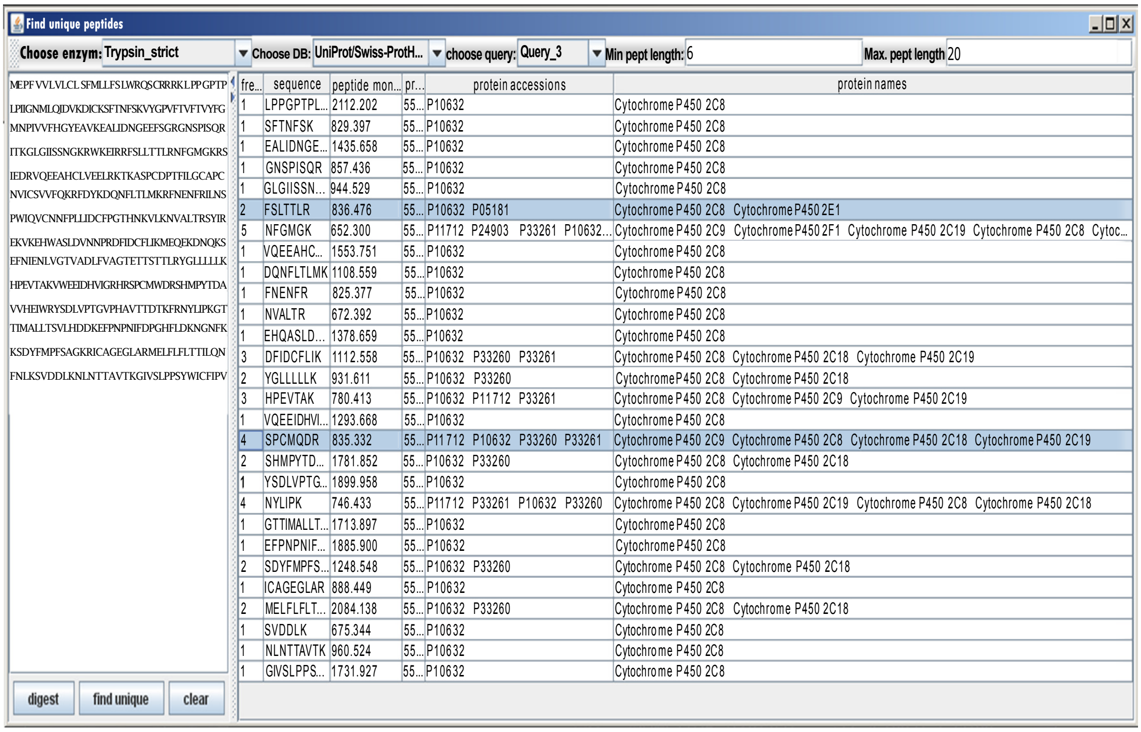


Choose enzym: Trypsin_strict Choose DB: UniProt/Swiss-ProtH... Choose query: Query_3 Min pept length: 6 Max. pept length 20

MEPF VVLVLCL SFMLLFS LWRQSCRRRKLPPGPTPLPI IGNMLQI
DVKDI CKSFTNFSKVYGPVFTVYFGMNPI VVFHGYEAVKEALIDN
GEEFSGRGNSPISQRITKGLGIISSNGKRWKEIRRFSLTTLRNFG
MGKRSIEDRVQEEAHCLVEELRKTKASPCDPTFILGCAPCNVICS
VVFQKRFDYKDQNFLTLMKRFNENFRILNSPWIQVCNNFPLLIDC
FPGTHNKVLKNVALTRSYIREKVKEHQASLDVNNPRDFIDCFLIK
MEQEKDNQKSEFNIENLVGTVADLFVAGTETTSTTLRYGLLLLLK
HPEVTAKVQEEIDHVIGRHRSPCMQDRSHMPYTDVVHEIQRYSD
LVPTGVPHAVTTDTKFRNYLIPKGTTIMALLTSVLHDDKEFPNPN
IFDPGHFLDKNGNFKKSDYFMPFSAGKRICAGEGLARMELFLFLT
TILQNFNLKSVDDLKNLNTTAVTKGIVSLPPSYQICFIPV

Results of digest
LPPGPTPLPIIGNMLQVK
SFTNFSK
EALIDNGEEFSGR
GNSPISQR
GLGIISSNGK
FSLTTLR
NFGMGK
VQEEAHCLVEELR
DQNFLTLMK
FNENFR
NVALTR
EHQASLDVNNPR
DFIDCFLIK
YGLLLLLK
HPEVTAK
VQEEIDHVIGR
SPCMQDR
SHMPYTDAVVHEIQR
YSDLVPTGVPHAVTTDTK
NYLIPK
GTTIMALLTSVLHDDK
EFPNPNIFDPGHFLDK
SDYFMPFSAGK
ICAGEGLAR
MELFLFLTTILQNFNLK
SVDDLK
NLNTTAVTK
GIVSLPPSYQICFIPV

Figure 1: Screenshot of the query tool. The text field on the left hand side shows the sequence of the human Cytochrome P450 (CYP) isoform 2C8. The results in the summary table show peptides obtained from the theoretical digest. A filter was applied in order to restrict the output to peptides with sequence lengths between 6 and 20 amino acids.



Choose enzym: Trypsin_strict Choose DB: UniProt/Swiss-ProtH... choose query: Query_3 Min pept length: 6 Max. pept length 20

MEPFVVLVLCLSFMLLFSLWRQSCRRRKLPPGPTP
LPIIGNMLQIDVKDICKSFTNFSKVYGPVFTVFTVYFG
MNPIVVFHGYEAVKEALIDNGEEFSGRGNSPISQR
ITKGLGIISSNGKRWKEIRRFSLLTTLRNFGMGKRS
IEDRVQEEAHCLVEELRKTKASPCDPTFILGCAPC
NVICSVVFQKRFDYKDQNFLTLMKRFNENFRILNS
PWIQVCNNFPLLIDCFPGTHNKVLKNVALTRSYIR
EKVKEHWASLDVNNPRDFIDCFLIKMEQEKDNQKS
EFNIENLVGTVADLFVAGTETTSTTLRYGLLLLLK
HPEVTAKVWEEIDHVIGRHRSPCMWDRSHMPYTDA
VVHEIWRYSDLVPTGVPHAVTTDTKFRNYLIPKGT
TIMALLTSVLHDDKEFPNPNIFDPGHFLDKNGNFK
KSDYFMPFSAGKRICAGEGLARMELFLFLTTILQN
FNLKSVDDLKNLNTTAVTKGIVSLPPSYWICFIPV

fre...	sequence	peptide mon...	pr...	protein accessions	protein names
1	LPPGPTPL...	2112.202	55...	P10632	Cytochrome P450 2C8
1	SFTNFSK	829.397	55...	P10632	Cytochrome P450 2C8
1	EALIDNGE...	1435.658	55...	P10632	Cytochrome P450 2C8
1	GNSPISQR	857.436	55...	P10632	Cytochrome P450 2C8
1	GLGIISSN...	944.529	55...	P10632	Cytochrome P450 2C8
2	FSLTTLR	836.476	55...	P10632 P05181	Cytochrome P450 2C8 Cytochrome P450 2E1
5	NFGMGK	652.300	55...	P11712 P24903 P33261 P10632...	Cytochrome P450 2C9 Cytochrome P450 2F1 Cytochrome P450 2C19 Cytochrome P450 2C8 Cytoc...
1	VQEEAHC...	1553.751	55...	P10632	Cytochrome P450 2C8
1	DQNFLTLMK	1108.559	55...	P10632	Cytochrome P450 2C8
1	FNENFR	825.377	55...	P10632	Cytochrome P450 2C8
1	NVALTR	672.392	55...	P10632	Cytochrome P450 2C8
1	EHQASLD...	1378.659	55...	P10632	Cytochrome P450 2C8
3	DFIDCFLIK	1112.558	55...	P10632 P33260 P33261	Cytochrome P450 2C8 Cytochrome P450 2C18 Cytochrome P450 2C19
2	YGLLLLLK	931.611	55...	P10632 P33260	Cytochrome P450 2C8 Cytochrome P450 2C18
3	HPEVTAK	780.413	55...	P10632 P11712 P33261	Cytochrome P450 2C8 Cytochrome P450 2C9 Cytochrome P450 2C19
1	VQEEIDHVI...	1293.668	55...	P10632	Cytochrome P450 2C8
4	SPCMQDR	835.332	55...	P11712 P10632 P33260 P33261	Cytochrome P450 2C9 Cytochrome P450 2C8 Cytochrome P450 2C18 Cytochrome P450 2C19
2	SHMPYTD...	1781.852	55...	P10632 P33260	Cytochrome P450 2C8 Cytochrome P450 2C18
1	YSDLVPTG...	1899.958	55...	P10632	Cytochrome P450 2C8
4	NYLIPK	746.433	55...	P11712 P33261 P10632 P33260	Cytochrome P450 2C8 Cytochrome P450 2C19 Cytochrome P450 2C8 Cytochrome P450 2C18
1	GTTIMALLT...	1713.897	55...	P10632	Cytochrome P450 2C8
1	EFPNPNIF...	1885.900	55...	P10632	Cytochrome P450 2C8
2	SDYFMPFS...	1248.548	55...	P10632 P33260	Cytochrome P450 2C8 Cytochrome P450 2C18
1	ICAGEGLAR	888.449	55...	P10632	Cytochrome P450 2C8
2	MELFLFLT...	2084.138	55...	P10632 P33260	Cytochrome P450 2C8 Cytochrome P450 2C18
1	SVDDLK	675.344	55...	P10632	Cytochrome P450 2C8
1	NLNTTAVTK	960.524	55...	P10632	Cytochrome P450 2C8
1	GIVSLPPS...	1731.927	55...	P10632	Cytochrome P450 2C8

Figure 2: Screenshot of the query tool showing the results of a database query. The text field on the left hand side shows the sequence of the human Cytochrome P450 (CYP) isoform 2C8. Results concerning both peptide and related protein information are given in the summary table on the right hand side. Thereby, the output is restricted to peptides with sequence lengths between 6 and 20 amino acids. The highlighted peptide information lines were further discussed in the text.

This is especially important, when very similar proteins with only few unique peptides - e.g. highly homolog protein isoforms - need to be differentiated by mass spectrometry.

Other important fields of application are functional proteomics approaches where the proteins of interest are already known before starting the experimental studies (targeted proteome studies).

One common procedure to find unique peptides for a protein of interest would be a protein BLAST search after performing a theoretical digest against a defined protein sequence database. Indeed, this approach occupies much time as every peptide has to be aligned and results have to be manually evaluated.

The aim of our work was to develop the UPF software solution as a tool that automates the detection of unique peptides.

In the next section both program features and the basic application of the UPF software will be briefly described and discussed.

Description of the UPF Software

Basic Features of the Command Line Tool (Module 1)

The command line tool (Module 1) can be used to generate up-to-date databases for storing peptide related information with respect to an *in silico* digest as part of a scheduled batch process.

Basically, peptide sequences are computed for each protein entry of the input FASTA file considering specific enzyme properties. Both monoisotopic and average masses are calculated for each peptide and additionally the protein mass is given**.**

Basic Features of the Query Tool (Module 2)

The query tool (Module 2) basically consists of two different areas and a tool bar (Figures 1, 2). Input sequences can be supplied by copy and paste on the left hand side. Sequence information must be given in one letter code. Both the enzyme and one of the databases generated by the first module can be chosen interactively from the tool bar.

Output is given in the summary table. After pressing the *digest* button in the panel on bottom left, the summary table shows all theoretic peptides of this protein that are generated specifically for the predefined protease (Figure 1). This list now serves for comparison with a predigested protein database, which was generated with module 1. An adequate database can be preselected in the tool bar. Furthermore, a peptide length restriction can be used for excluding peptides which might not be monitored with the applied experimental setup, e.g. very large or short peptides, respectively. Exclusion of very short peptides significantly reduces the time needed for the database search.

After pressing the 'find unique' button a database query is performed and the results are displayed in the summary table. The frequency is always displayed in the frequency column showing the number of protein entries of the designated database in which the peptide sequence was found. Peptides with frequency values of one are unique with respect to the sequence given as user input. If the frequency is higher than one, the peptide is homolog in several database entries. The experimentalist of course has to evaluate if different database entries can be grouped to one protein class. Therefore additional information, which can be displayed in the summary table, should be considered. Beside the monoisotopic and average peptide mass and the corresponding protein mass, the accession number, the protein name and the peptide sequences (Figure 2) are given.

Results can be marked within the summary table and transferred into style sheet programs via the clipboard. Customizing the UPF software. In order to facilitate customizing both modules use configuration files. The configuration files were designed in the userfriendly Windows™ ini-file style, i.e. the files contain lines of key and values pairs, e.g. the line 'cleavageSites=KR' assigns the value 'KR' to the key 'cleavageSites'. This special key and value pair defines the amino acids where enzymatic cleavage occurs. The configuration files store properties necessary for establishing the database connection as well as specific parameters, which are needed to perform the theoretical digest, e.g. the cleavage sites of a particular proteases.

The UPF configuration files support tryptic digestion per default. However, the configuration files can easily be extended for the use of other proteases. Therefore, the user simply must attach another section and specify the given parameter values in order to add the functionality of any new enzymatic digest.

Regular expressions, which were also included in the configuration files, provide a succinct and flexible means for identifying patterns of characters and can therefore be used for parsing the header of IPI and UniProtKB/Swiss-Prot databases in order to obtain protein related information. From the supported databases (IPI and UniProtKB/Swiss-Prot) accession numbers, taxonomic identifiers and protein names are extracted. Additionally, the header of IPI databases contains cross references, which are stored in the generated database as well.

Application of UPF to a Concrete Question – CYP Subfamilies

To comment on the usability of UPF several isoforms from the human Cytochrome P450 (CYP) family have been selected and their sequences have been analyzed.

In human 57 CYP isoforms have been identified so far, which are classified on their sequence homology [Lewis, 2004]. CYPs are responsible for the oxidative metabolism of many xenobiotics as well as organic endogenous compounds. Especially members of the CYP3A* and CYP2C* families are highly homologous; CYP2C9 and CYP2C19 for example show a sequence homology of 91% (calculated by sequence alignment). Because no antibodies are available for the same purpose, mass spectrometry is the method of choice to differentiate between these protein isoforms.

In the following a subset of the UniProtKB/Swiss-Prot database, which considers only entries of human proteins, was used in order to evaluate the software. The sequence of CYP2C8 (left, Fig. 2) was *in silico* digested with trypsin (toolbar at the top, Fig. 2). With a peptide length restriction of six to twenty amino acids, eighteen unique peptides can be found for CYP2C8 (peptides showing a frequency of 1, Figure 2).

In case a peptide is homolog in another or several other proteins, the protein name information is very useful to evaluate the results in more detail; the peptide SPCMQDR for example is found to be present in all four CYP2C* isoforms (CYP2C8, CYP2C9, CYP2C18, and CYP2C19). Furthermore the peptide FSLTTLR can be used indeed to differentiate CYP2C8 from the other members of the CYP2C* subfamily, but it is homolog in another CYP isoform, namely CYP2E1. Finally the tool calculates eight unique peptides for CYP2C9, 19 unique peptides for CYP2C18 and 15 unique peptides for CYP2C19. These results show that even for highly homologous proteins in general enough theoretically generated unique peptides exist. Nevertheless, by disfavored digestion or ionization of some peptides or due to the low abundance of the protein in general proteins often get identified by only few peptides. The application of UPF tremendously reduces the time needed to argue about the uniqueness of those peptides to minutes in comparison to hours if a manual Blast search is done.

The application regarding to targeted proteomics is for instance the detection as well as the quantification of a protein by using multiple reaction monitoring (MRM), which is applied to the mea-

ISSN:0974-276X JPB, an open access journal

surement of specific peptides in complex mixtures (Kuhn et al., 2004). In the MRM approach the sensitive detection of precursor-to-product ion transition is a diagnostic trigger for the presence of a peptide. Because several triggering points are recorded in a certain time window, the AUC (area under the curve) can be used to relatively quantify the abundance of this peptide.

In the AQUA (absolute quantification) strategy (Gerber et al., 2003; Barnidge et al., 2003), including MRM, a peptide can be used as a stoichiometric representative of the protein from which it is originated and related against a spiked stable isotope-labelled internal standard to calculate the absolute protein amount in a sample through the comparison of the AUCs. The uniqueness of a peptide ensures the specificity of the quantification approach for the targeted protein and therefore it is indispensable. Applying UPF for the search of such peptides will considerably facilitate and accelerate the workflow in the fields of targeted proteomics.

Future Prospects

Different features are planned to be integrated into the UPF software for the near future:

Efforts will be undertaken to provide UPF as a web service in the future. Then, several pre-digested databases will be hosted and maintained at the MPC.

Input is currently restricted to a single protein sequence. However, in order to ensure a more flexible usability the query module will be enhanced in order to accept many sequences or accession numbers as input. Moreover a batch processing should be integrated, i.e. multiple protein sequences or accession numbers should be accepted at the same time. Currently, results can be restricted to peptides with a particular sequence length. Because it is desirable to filter peptides with respect to their mass, such a selection will be integrated into UPF.

The configuration files will be extended for more predefined sections considering other proteases and databases in order to provide an easy access for users in most instances. Integration of the NCBInr database into UPF will obtain priority during the further work and may probably be accessible at the time point of publication.

Moreover, for enhancing adaptability a type of ‘SQL-Parser’ will be integrated into the software, which adopts the output of the query module to a specific SQL statement, which can be given by the user.

It is planned to take a predefined number of missed cleavage sites into account.

Conclusions

A software solution is presented for automatic calculation of unique peptide sequences. It was shown that identification of such peptides could be facilitated when using the software in comparison with a multitude of manually performed BLAST or alignment searches. Additionally, applying the software can be of value in particular when conducting larger MRM or AQUA experiments.

Acknowledgments

This work was funded by the HepatoSys network project of the German Ministry of Education and Science (BMBF, Hepatosys, grant 0313080J) and by ProDaC (European Commission project, 6th framework programme, project number LSHG-CT-2006-036814).

References

1. Alterovitz G, Liu J, Chow J, Ramoni MF (2006) Automation, parallelism, and robotics for proteomics. Proteomics 6: 4016-4022.» CrossRef » Pubmed » Google Scholar

2. Janecki DJ, Bemis KG, Tegeler TJ, Sanghani PC, Zhai L, et al. (2007) A multiple reaction monitoring method for absolute quantification of the human liver alcohol dehydrogenase ADH1C1 isoenzyme. Anal Biochem 369: 18-26. » CrossRef » Pubmed » Google Scholar

3. Kocher T, Superti-Furga G (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. Nat Methods 4: 807-815. » CrossRef » Pubmed » Google Scholar

4. Le Blanc JCY, Hager JW, Ilisiu AM, Hunter C, Zhong F, et al. (2003) Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications. Proteomics 3: 859-869. » CrossRef » Pubmed » Google Scholar

5. Lewis DFV (2004) 57 varieties: the human cytochromes P450. Pharmacogenomics 5: 305-318. » CrossRef » Pubmed » Google Scholar

6. Lohaus C, Nolte A, Bluggel M, Scheer C, Klose J, et al. (2007) Multidimensional chromatography: a powerful tool for the analysis of membrane proteins in mouse brain. J Proteome Res 6: 105-113. » CrossRef » Pubmed » Google Scholar

7. Mayya V, Rezual K, Wu LF, Fong MB, Han DK (2006) Absolute quantification of multisite phosphorylation by selective reaction monitoring mass spectrometry: determination of inhibitory phosphorylation status of cyclin-dependent kinases. Mol Cell Proteomics 5: 1146-1157.» CrossRef » Pubmed » Google Scholar

8. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20: 3551-3567. » CrossRef » Pubmed » Google Scholar

9. Sadygov RG, Cociorva D, Yates JR 3rd (2004) Large-scale database searching using tandem mass spectra: looking up the answer in the back of the book. Nat Methods 1: 195-202. » CrossRef » Pubmed » Google Scholar

10. Wolf-Yadlin A, Hautaniemi S, Lauffenburger DA, White FM (2007) Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. Proc Natl Acad Sci USA 104: 5860-5865. » CrossRef » Pubmed » Google Scholar