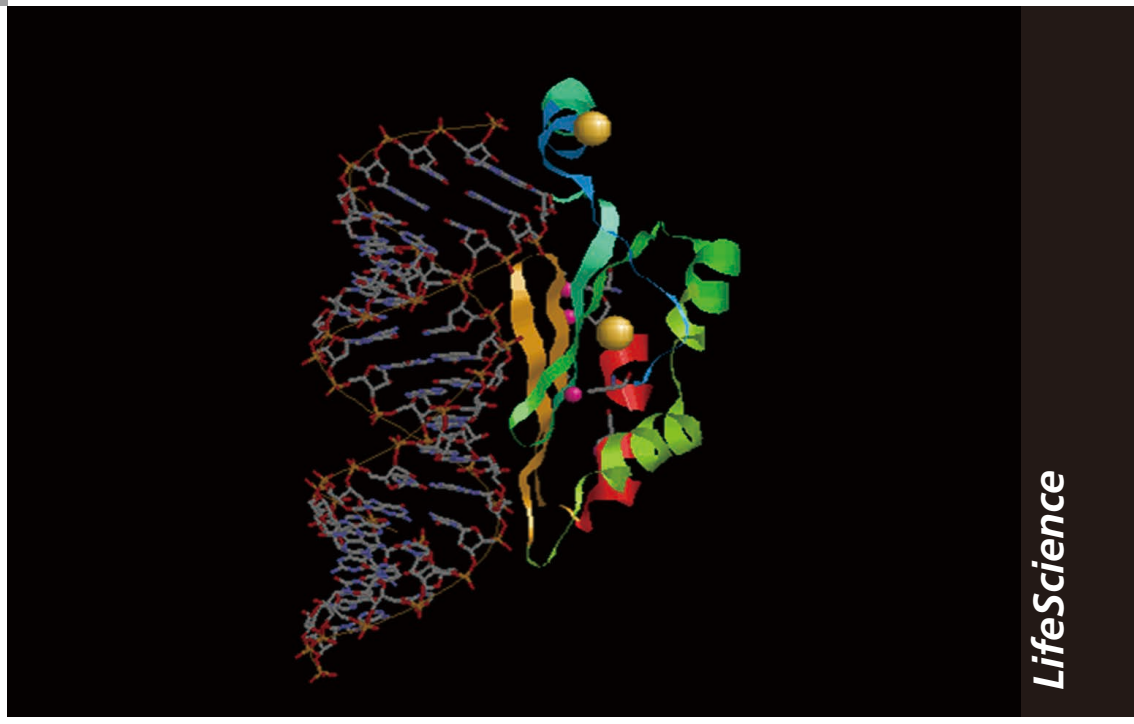
SHIMADZU
Excellence in Science

Application Note

No. 12

Primary Structural Analysis of Proteins / Peptides – Application of Protein Sequencer –

Yasushi SAKAMOTO[*1], Tomoko KURIKI[*2]



LifeScience

1. Introduction

Understanding of pathological states and life phenomena has accelerated dramatically with recent widespread advances in the field of proteomics, which encompasses the exhaustive comparison and identification of expressed proteins in normal and disease-state tissues and cells. The scope of this investigation includes pathology, differences in type, genetic variation and cell abnormalities across a range of scientific disciplines. For example, investigation of the causes of disease and treatment methods, and the search for specific biomarkers associated with those diseases in the field of proteomics has grown rapidly.

The word "proteomics" was coined in the early 1990s to indicate the entire set of proteins expressed in a specific cell type or organism associated with that genome.

In general, there is very little in the way of change in the genome in a given type of organism, but the change in the proteome between organs and cells can be vast. In order for an organism to make its genetic information actually function, the process of gene expression is required. Proteins are extremely important tool molecules which elicit the expression of a wide range of gene functions. Proteome analysis is an indispensible tool for elucidating these mechanisms of interaction between the genome and proteome, and hence is referred to as the field of proteomics. However, viewed historically, proteome analysis has proved to be much more difficult compared to genome analysis.

*1: Central Research Center, Faculty of Medicine, Saitama Medical University

*2: Kyoto ADC, Analytical Applications Dept., Analytical & Measuring Instruments Div., Shimadzu Corp.

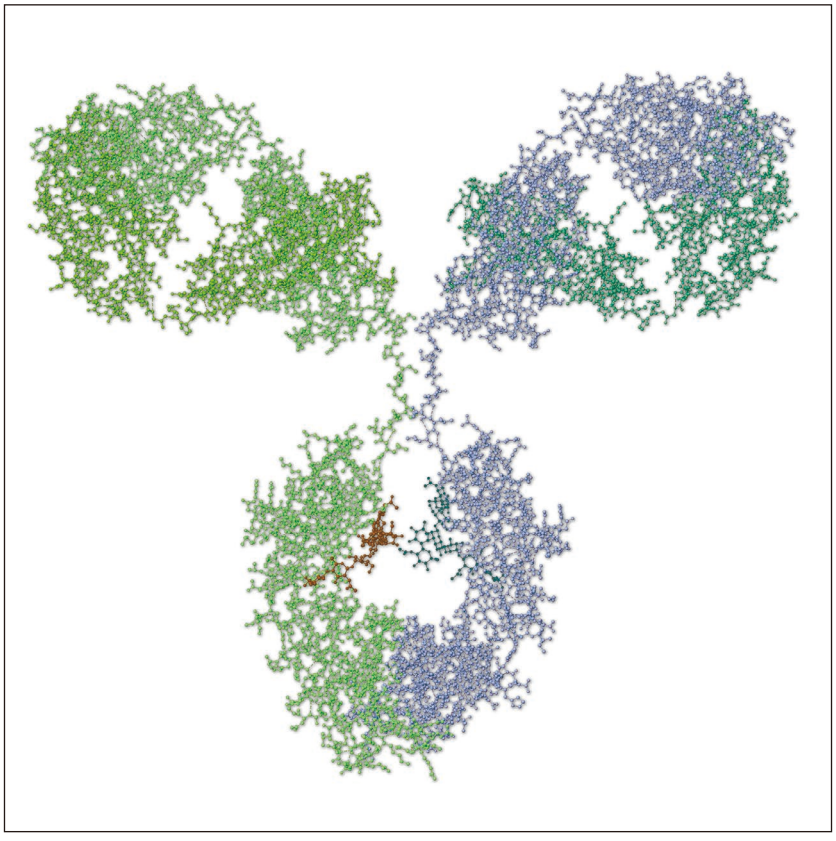
In genome analysis, a technology like Polymerase Chain Reaction (PCR) that amplifies a target molecule magnifies samples which are at trace levels. Many methods and techniques such as PCR have been developed to conduct such amplification and have now been incorporated into commercially-available kits. In addition, nucleic acids comprised of just four types of base compositions do not exhibit great changes in physical properties even in very large molecules, but the range of conditions under which they are handled is relatively narrow. In proteome analysis, since the quantity of proteins obtained from a just a small amount of sample is insufficient for analysis even if analysis accuracy and sensitivity are improved to their utmost limits, even higher sensitivity is required.

In addition, because proteins are basically composed of 20 kinds of amino acids, substitution of just one amino acid could have a profound effect on its physical properties, so a wide range of dissolution and separation conditions are used. Furthermore, biological reactions referred to as post translational modifications (PTMs) including terminus amino acid modifications, as well as such changes as bonding of sugar chains (glycosylation), lipids, phosphates (phosphorylation), sulfates (sulfation), methyl groups (methylation), etc., can occur non-uniformly within the same protein molecule. When separation technology improves, the collection of multiple molecules that are different, even of the same protein, will become possible, so situations requiring even higher sensitivity will become necessary. Analysis of these modification sites is critical for determining the functions of proteins.

There are two main types of instruments used for "primary structural analysis of proteins," the central technology of proteome analysis. One of these is the protein sequencer, and the other is a type of mass spectrometer referred to as a Matrix-Assisted Laser Desorption Ionization-Time-of-Flight Mass Spectrometer (MALDI-TOF-MS), which has greatly extended the range of protein analysis. From the standpoint of throughput, the mass spectrometer is superior to the protein sequencer, but for identification of proteins not included in a database, or verification of amino acid sequences (especially, unknown sequences that are not registered in a database), the protein sequencer has the edge.

The success of proteome analysis is also tied to "bio-pharmaceuticals," such as antibody drugs. Primary structural analysis of the proteins, the central technology of proteome analysis, is used for research and development and quality control of antibody drugs, etc.

Here we explain the features of these two types of instruments, the protein sequencer and mass spectrometer, which are used for the primary structural analysis of proteins.



2. Protein Sequencer

The protein sequencer is an instrument that was developed by Pehr Victor Edman in 1950 with the intention of automating the Edman reaction, a degradation system consisting of the following 3 steps.

1: Coupling
PITC (Phenylisothiocyanate) is linked to the N-terminus amino group of an isolated, purified sample protein (including peptides) in a reaction tank (reactor) under alkaline conditions.

2: Cleavage
This N-terminus amino acid with its linked PITC is cleaved under acidic conditions, and is extracted as an anilinothiazolinone (ATZ) derivative in the organic layer.

3: Conversion
The organic solvent is volatilized, and the extracted ATZ derivative is further derivatized in acidic aqueous solution to form the more stable phenylthiohydantoin (PTH)-amino acid derivative (Fig. 1, Fig. 2).

In step 1, the protein remaining in the reactor is reacted with PITC, and by repeating the above steps, the N-terminus amino acid PTH derivative is obtained sequentially, cycle by cycle. Then, using chromatography to conduct identification, the amino acid sequence of the protein can be elucidated.

The instrument that completely automates this series of operations consisting of chemical reaction, extraction and identification is the protein sequencer. In particular, with respect to coupling of step 1 above, by binding the sample peptides with the polybrene-treated glass filter and/or PVDF membrane, and by using nitrogen carrier gas to deliver the PITC and triethylamine coupling reagents, as well as the trifluoroacetic acid (cleavage reagent) to the reactor, this instrument which suppresses sample loss to the minimum is referred to as a gas phase protein sequencer. The final step of analysis is identification of PTH amino acids by high-performance liquid chromatography (HPLC). This sequencer fully automates the chemistry of Edman degradation following sample preparation.

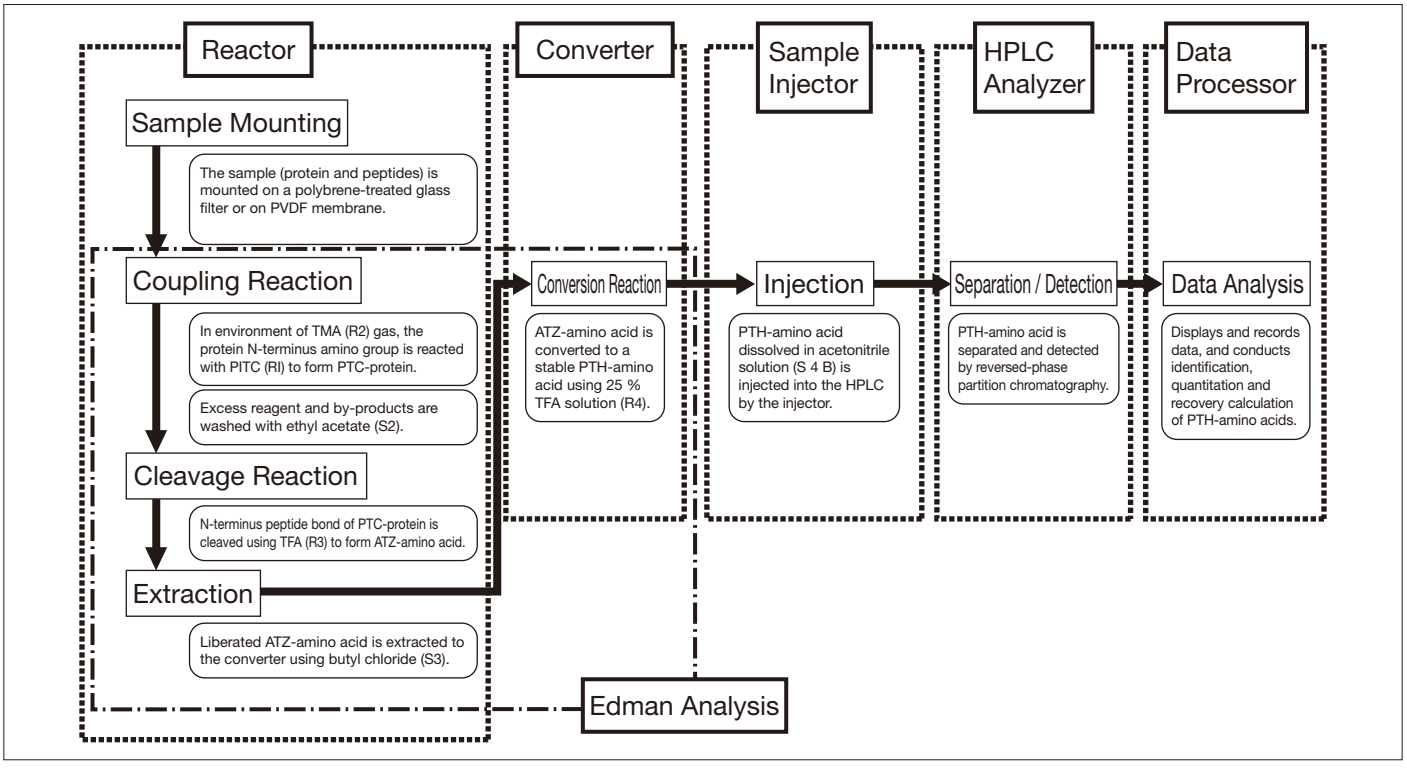


Fig. 1 Amino Acid Sequence Analysis Flow Chart

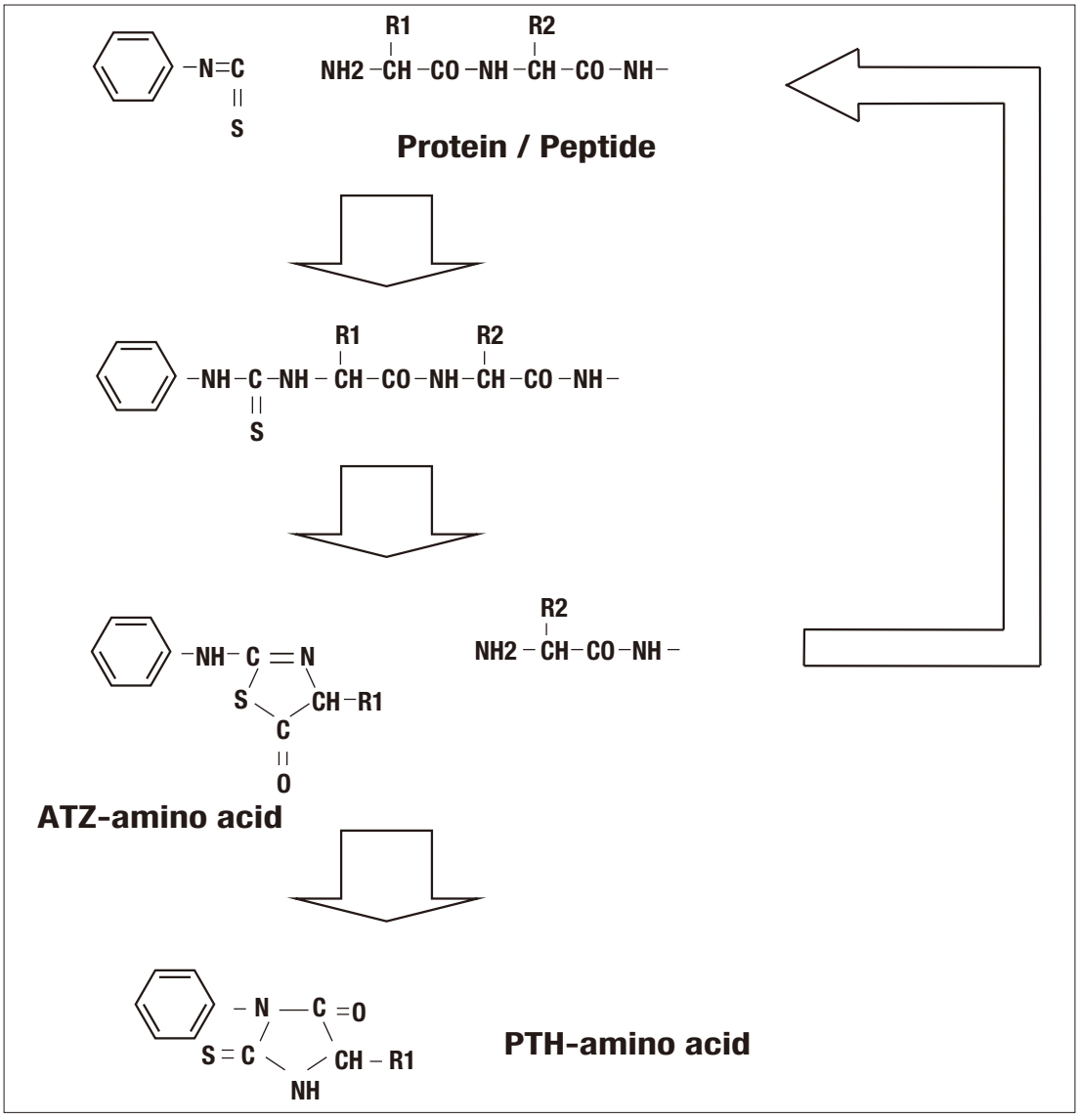


Fig. 2 Scheme of Edman Reaction

In order to determine an amino acid sequence using the Edman reaction, a highly purified protein sample at the pmol level is required, although the operation itself is extremely simple. Once again, sample preparation is extremely critical before cycling on the sequencer. For example, following electrophoresis, the separated protein is electrically transferred (electroblotted) onto PVDF membrane and stained, and the stained protein spot is then cut out from the membrane and placed directly in the reactor. Subsequent analysis can be performed completely automatically.

Listed below are the salient features of the protein sequencer. Except in the field of basic research, these clearly indicate how indispensible this instrument is in quality management of protein or bio-pharmaceuticals, which requires a high degree of reliability.

- Provides highly reliable amino acid sequencing.
- Proteins can be analyzed as is.
- Ile and Leu, with exactly the same mass number, can be differentiated.
- The existence and position of the S-S bond can be determined.
- Operation is extremely easy.

It is an especially difficult task to distinguish between Ile and Leu, with their identical mass numbers, even using current mass spectrometry (MS) technology. As an example, we analyzed the amino acid sequence [MAPRGFSCLLLLTGEIDLPVKRRA] of the synthetic analog (Gly^{14}-Humanin) of the peptide Humanin. This peptide contains one Ile residue (Ile^{16}) and 5 Leu residues ($Leu^{9,10,11,12,18}$), but since the HPLC retention times of PTH-Ile and PTH-Leu are different (PTH-Ile retention time is smaller), they are easily distinguished (Fig. 3A, B, C). In addition, the complete amino acid sequence, including Cys^{8}, could be determined, so detection and identification of the C-terminus amino acid Ala^{24} was also possible (Fig. 3D).

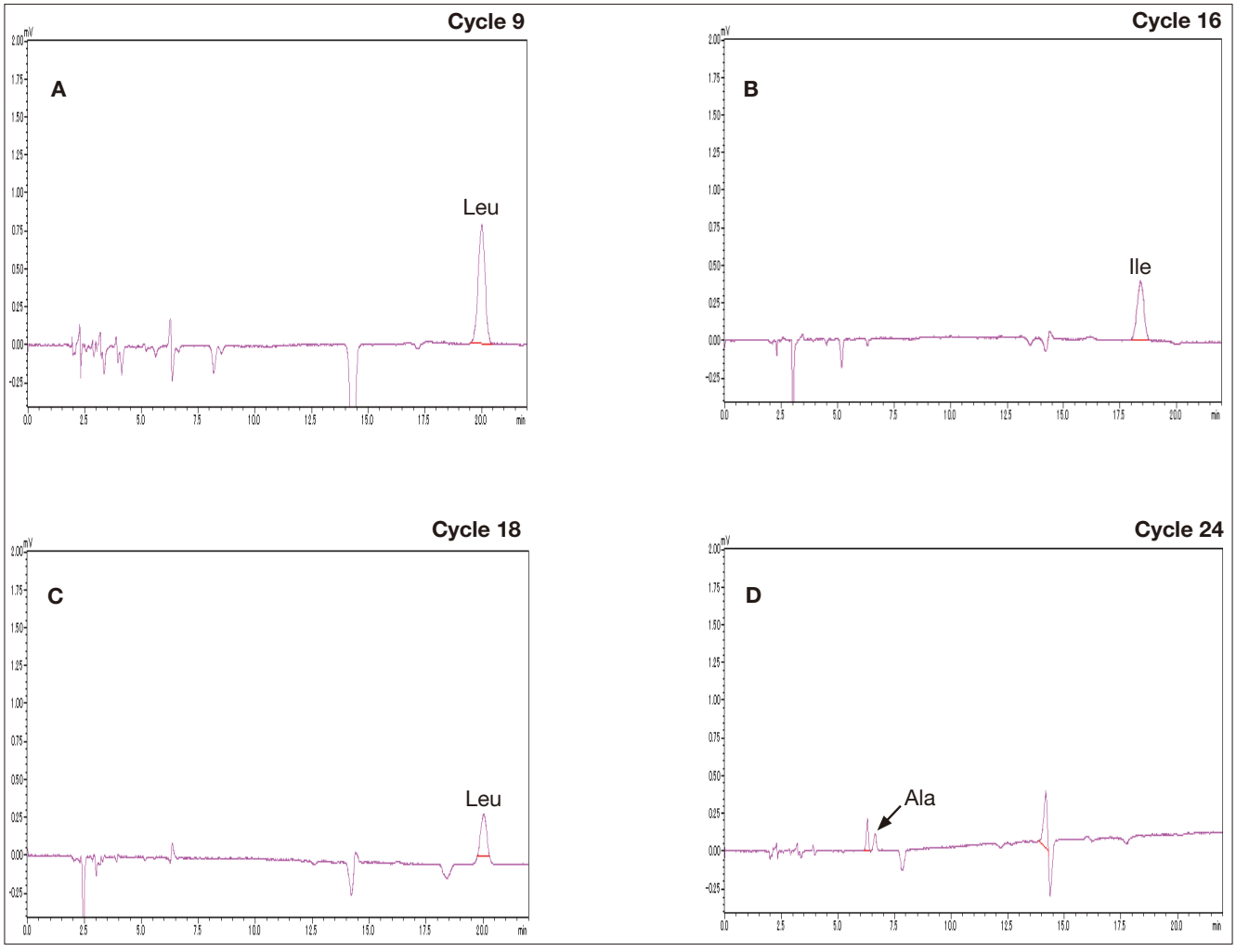


Fig. 3 Analysis of Gly14-Humanin Using Protein Sequencer: Leu9(A), Ile16(B), Leu18(C), Ala24 (D)
(Displayed chromatograms are the results after subtraction from the background of previous chromatograms, thereby selecting the most prominent amino acid per cycle.)

On the other hand, there are also many problems to be resolved with protein sequencers.
For example, if the N-terminus of the protein is blocked (acetylized, acylated, etc.), Edman degradation does not occur, preventing the acquisition of amino acid sequence information. In general, N-terminus blockage occurs in about 70 % of mammalian proteins. Since amino acid sequence information can also be obtained from proteins in which the N-terminus is blocked, the removal of the N-terminus block is being actively researched. Furthermore, sensitivity is inferior compared to that of MS, and more sample is required for sequence analysis.

Protein identification using a protein database is the predominant trend in present proteome analysis. Here we describe how it relates to amino acid sequencing (amino acid residue number when the sequence is known).

3. Partial Amino Acid Sequence Information and the Protein Database

A search of the PIR (Protein Information Resource) protein database for an N-terminus amino sequence IVVYTDK of "some protein" that occurs in the myelin sheath of a peripheral nerve returns hits for 2 types of the same protein.

This means that the protein can be identified even from a rather uncharacteristic 7-residue amino acid sequence. A search using the first 6 residues including IVVYTD returned as many as180 candidates or "hits." Even in this case, if the type is limited to human or rat, the protein candidates are limited to 1 type, and when mouse is specified, 2 protein candidates are returned. When the search is limited to the organ from which the protein is extracted, there is, as might be expected, only 1 candidate. Using up to 5 residues including IVVYT, there are 1210 candidates. Even in this case, if the type is limited according to organ, the protein candidates are limited to 1 for human, rat, or mouse. Furthermore, assuming that MS analysis is used, considering the lack of confidence in the first Ile, even if a search is conducted for VVYTD, the number of candidates increases to about 1800, but the result is the same after narrowing the search criteria. If a search is conducted using LVVYT with the judgment that Ile is indeed Leu, about 90 candidates are returned, but of course, all of these are irrelevant. In this case, mammalian protein was not included.

Therefore, a target protein can probably be identified if the target protein has been properly isolated and purified and is unique to a specific organ, and if a confirmed 5-residue amino acid sequence is known. If however, the protein is subjected to limited digestion using a protease, and the digested peptides are isolated, and further, 2 or more constituent amino acid sequences can be obtained, the certainty can be raised even more. Realistically, protein identification using a database is probably possible if several amino acid sequences of a 5 to 7-residue amino acid sequence per each protein can be determined.

Protein identification is widely conducted using a combination of MS and database search, as discussed below, but if a 5 to 7-residue amino acid sequence can be confirmed, protein identification is possible using a database in the same way.

4. Mass Spectrometer (Mass Spectrometry: MS)

The type of analysis most generally used for protein identification now is MS, which can be broadly categorized into MALDI-MS and LC-MS which is combined with HPLC. MS is used as a tool for conducting Peptide MS Fingerprinting (PMF) and MS/MS ion search, as described below. In these techniques, the protein is subjected to limited digestion using limiting degradative enzymes such as trypsin and lysyl endopeptidase to obtain peptide fragments. Next, the peptide fragment masses obtained by MS, or the peptide fragment partial sequence information obtained using MS/MS are matched against an existing database. In this matching against the database, the results are returned with the higher percentage matching rates at the top of the candidate list. In addition, other information related to the protein, including molecular weight, isoelectric point, type of organism, site of expression, etc., can be used to narrow the range of relevant proteins. Conducting matching with this existing database makes possible high-throughput protein identification.

On the other hand, to determine the amino acid sequence of a peptide (or protein) of unknown sequence, a gate that can selectively control the passage of a ions having only a certain mass number is installed in the instrument to allow selection of a specific peptide. Then, by analyzing the naturally dissociated ions, Post Source Decay (PSD) analysis which can be used to obtain amino acid sequence information, and further, advanced instruments like MS/MS analytical instruments consisting of tandemly arranged MS units, and Quadrupole Ion Trap (QIT)-MS with a built-in ion trap, etc., can be used similarly as in PSD analysis to guide a specifically selected peptide ion into a collision site (collision chamber, ion trap, etc.) using inert gas, to form secondary collision induced dissociation (CID) ions. This allows for the prediction of amino acid sequences (de novo sequencing). Although, this method is actively being researched, reliability is still inferior compared to determination of amino acid sequences by the protein sequencer. Also, from the standpoint of differentiating between Ile and Leu, the fact that their mass numbers are the same makes determination problematic due to this inherent theoretical difficulty. (See Table 1 Immonium Ion.)

The MS/MS spectrum obtained by QIT-TOF MS analysis of Humanin is shown in Fig. 4.
Since this is the Gly replacement for the 14th Ser residue, it was basically assumed that there would be no relevant information in the database. However, when the search criteria were entered into Mascot (search engine which uses MS data to identify proteins from amino acid sequence databases), one of the items that had been registered in the database was returned as a candidate. Table 1 shows the actual fragment ions that were observed that reflect this amino acid sequence. The analysis was easily conducted because it was already registered in the database, however, we went as far as conducted MS/MS/MS analysis, and even small fragment ions were investigated. The intensities of the fragment ions reflecting the existence of the Leu core portion from Leu^9 – Leu^{12} in addition to Pro were extremely small, so if the information hadn't been registered in the database, reliable determination of the entire amino acid sequence using only this spectrum would have been difficult.

Thus, even with primary structural analysis of proteins using MS, the reliance on de novo sequencing without using a database, and such problems as distinguishing between Ile and Leu still remain, indicating that there is no universal instrument for sequencing.

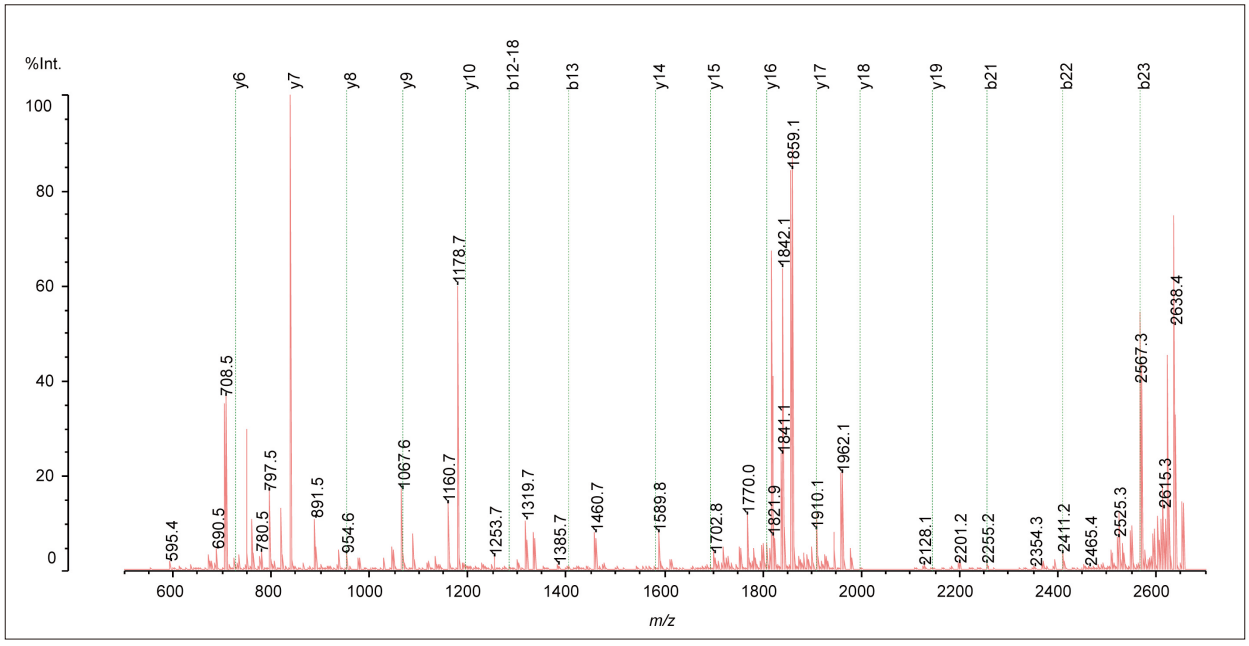


Fig. 4 MS/MS Spectrum Using QIT-TOF-MS

Table 1 MS/MS Spectrum Using QIT-TOF-MS

#	Immom.	a	a*	A^0	b	b*	B^0	Seq.	y	y*	y^0	#
1	104.0528							M				24
2	44.0495							A	2525.53		2507.55	23
3	70.0651							P				22
4	129.1135							R	2351.34			21
5	30.0338							G	2201.37	2183.36	2183.36	20
6	120.0808							F		2167.30	2167.30	19
7	60.0444							S		1979.26	1979.26	18
8	76.0215	822.57	805.54	804.54				C	1910.23	1892.21	1892.21	17
9	86.0964							L				16
10	86.0964							L	1694.08	1677.09	1675.10	15
11	86.0964				1189.72			L		1561.94	1561.94	14
12	86.0964				1302.79			L				13
13	74.0600				1403.84	1385.81	1385.81	T		1336.83	1336.83	12
14	30.0338				1460.86	1442.86	1442.86	G	1253.78			11
15	102.0550				1589.91	1571.91	1571.91	E		1179.40	1178.77	10
16	86.0964				1702.07	1686.05	1684.10	I	1067.72			9
17	88.0393				1818.05	1801.463	1800.07	D	954.63	937.60	936.63	8
18	86.0964							L	839.60	822.57		7
19	70.0651							P				6
20	72.0808				2167.30			V				5
21	101.1073				2255.21			K				4
22	129.1135				2411.45	2394.43	2393.47	R				3
23	129.1135				2567.57	2549.57	2549.57	R				2
24	44.0495							A				1

5. Conclusion

Up to now, several kinds of reagents (Table 2) [1] have been developed to improve PITC with the aim of introducing fluorescence for the visualization of PTH-amino acids in the Edman degradation reaction. However, as a result, the method selected at the automation stage was the original Edman degradation reaction only. Therefore, there seems to be no doubt that future improvements to the reaction system will only result in greater complexity of the instrument as long as higher sensitivity cannot be achieved without changing almost all of the extraction conditions and reaction conditions. Ultraviolet absorption (in the vicinity of 269 nm) due to the benzene ring as the chromophore is currently being used for detection of the PTH-amino acids. Even if higher sensitivity becomes possible, identification accuracy of the amino acid may even decrease unless there is a reduction in the quantity of Edman degradation by-products that have a benzene ring. Other problems that can be raised are the need to improve the mechanical accuracy of the protein sequencer, reducing the amount of solvent required, and increasing the purity of the solvents themselves, but given the level of currently available technology, it is possible that these problems can be resolved.

In addition to the proteome analysis for in-depth comparison and identification of expressed proteins, more focused proteome analysis is expected to become quite common. Amino acid sequence analysis of the N-terminus and C-terminus of proteins is indispensable for researching the processes of splicing and selective splicing in transcription from the genome to mRNA, and the process of RNA editing, etc.

From the standpoint of throughput as measured by the number of identified proteins, MS has a slight edge when compared with the protein sequencer. However, as far as reliability of the determined sequence (especially when the unknown protein is not registered in the protein database), the protein sequencer has the edge.

Unlike genome analysis in which nucleic acid molecules having similar properties are the target of analysis, protein analysis involves a much wider array of amino acids having different properties. The quality of the research will depend on the effectiveness of future analytical instrumentation development.

Table 2 Previously Reported Isothiocyanate Reagents

1. Methylisothiocyanate (Vance and Feingold, 1970), Identification by mass spectrometry (Richards and Lovins, 1972).
2. Pentafluorophenylisothiocyanate (Lequin and Niall, 1972), Identification by gas-liquid chromatography, with electron-capture detection.
3. p-Phenylazophenylisothiocyanate (Deyl. 1970), Coloured products.
4. 4-N-Dimethylamino-1-naphthyl isothiocyanate (Ichikawa et al., 1970), Fluorescent thiocarbamyl derivatives.
5. p-Sulphophenyl isothiocyanate (Birr el al., 1970), Water-soluble reagent.
6. Diphenylindenonyl isothiocyanate (Ivanov and Mancheva. 1973, 1976), Coloured reagent: sensitive detection of products by colour and fluorescence on polyamide layers (Mancheva and Vladovska-Yukhnovska, 1978).
7. 4-N, N-Dimethylaminoazobenzene-4'-isothiocyanate (Chang et al.. 1976, 1978), Coloured reagent and products.
8. Fluorescein isothiocyanate (Muramoto et al., 1978), Fluorescent derivatives.
9. 3-Isothiocyanato-4-methoxy-4'-nitrostilbene (Nowak et al., 1977), Fluorescent reagent.

Appendix: Shimadzu PPSQ-50 Series Protein Sequencer

The Shimadzu PPSQ-50 Series Protein Sequencer carries out analysis of PTH-amino acids using high-performance liquid chromatography in isocratic mode. The instrument boasts various features, such as low running cost thanks to the reduced consumption of analytical reagents, easy operation, and compliance with FDA 21 CFR Part11. Here, we describe analysis examples of the PPSQ-50 Series applying these features.

Appendix-1. System Composition

The PPSQ-50 Series is comprised of the Edman degradation reactor component, the high-performance liquid chromatograph (HPLC) which handles analysis of the PTH-amino acids generated in the Edman degradation reaction, and a control / data processing component (personal computer). (The sample injector and analytical column are incorporated in the Edman reactor.) The PPSQ-51A and PPSQ-53A differ in the number of reactors making up the Edman reactor component. The 1-reactor type model is the PPSQ-51A, and the 3-reactor type model is the PPSQ-53A. The PPSQ-53A can continuously analyze amino acid sequences of samples one by one that are loaded into multiple reactors.

Edman analysis is repeated in the Edman reactor, cutting amino acids from proteins and peptides in order from the N-terminus to form stable PTH-amino acid derivatives. In addition, these are injected online to the HPLC, where analysis is conducted by UV detection. The HPLC data (chromatogram) is stored in the personal computer, where the obtained chromatogram is processed using the LabSolutions PPSQ software, and amino acid identification and sequence prediction are conducted using amino acid sequence prediction software. The waste liquid volume is small due to the instrument's environmentally friendly design.

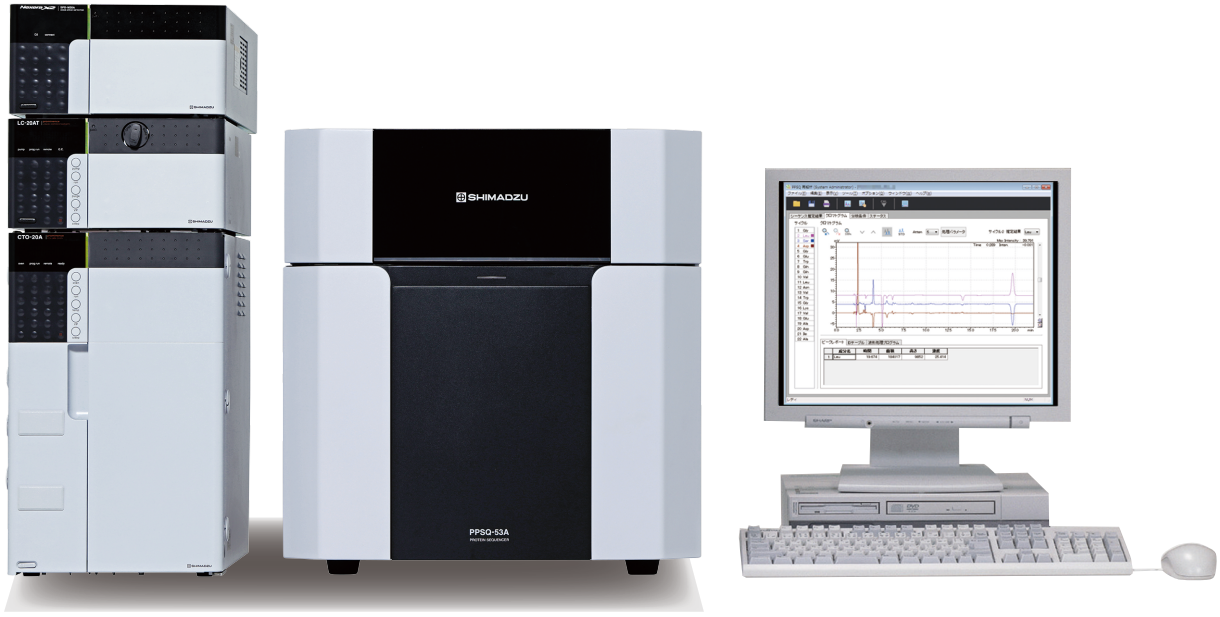


Fig. 5 PPSQ-50 Series

Fig. 6 PPSQ-51A Single Reactor

Fig. 7 PPSQ-53A Triple Reactor

Appendix-2. PTH-Amino Acid Identification Using Isocratic Mode

Fig. 8 shows an analysis example using 5 pmol of a PTH-amino acid standard mixture as the sample. In the analysis of the PTH-amino acids using the PPSQ-50 series in isocratic mode, it can be seen that the baseline is stable, and the analysis itself was completed in about 21 minutes. In addition, the elution time repeatability of all the PTH-amino acids was extremely high (Table 3), and correction of the PTH-amino acid elution times is unnecessary for each sequence analysis. Since a special eluent continuously circulates through the system, eluent adjustment that is dependent on operator experience (pH, ion intensity, organic solvent concentration, gradient conditions, etc.) is unnecessary. In addition, maintenance with respect to the eluent is easy and consumption volume is low, an important feature which translates into low running costs.

Table 3 Elution Time Repeatability of PTH-Amino Acids

PTH-Amino Acid	Elution Time		
	1st Day (min)	8th Day (min)	15th Day (min)
Asp	2.586	2.578	2.58
Glu	2.873	2.879	2.888
Asn	3.57	3.571	3.571
Gln	3.779	3.78	3.781
Ser	3.989	3.991	3.991
Thr	4.357	4.359	4.359
His	4.71	4.714	4.719
Gly	5.002	5.005	5.006
Ala	6.429	6.434	6.438
Tyr	6.79	6.797	6.803
Arg	8.205	8.189	8.188
Dth	9.18	9.19	9.2
Met	11.431	11.45	11.463
Val	11.821	11.84	11.853
Pro	12.296	12.314	12.327
Trp	14.475	14.502	14.523
Phe	16.722	16.754	16.779
Lys	17.28	17.317	17.346
Ile	18.945	17.977	18.001
Leu	19.447	19.487	19.514

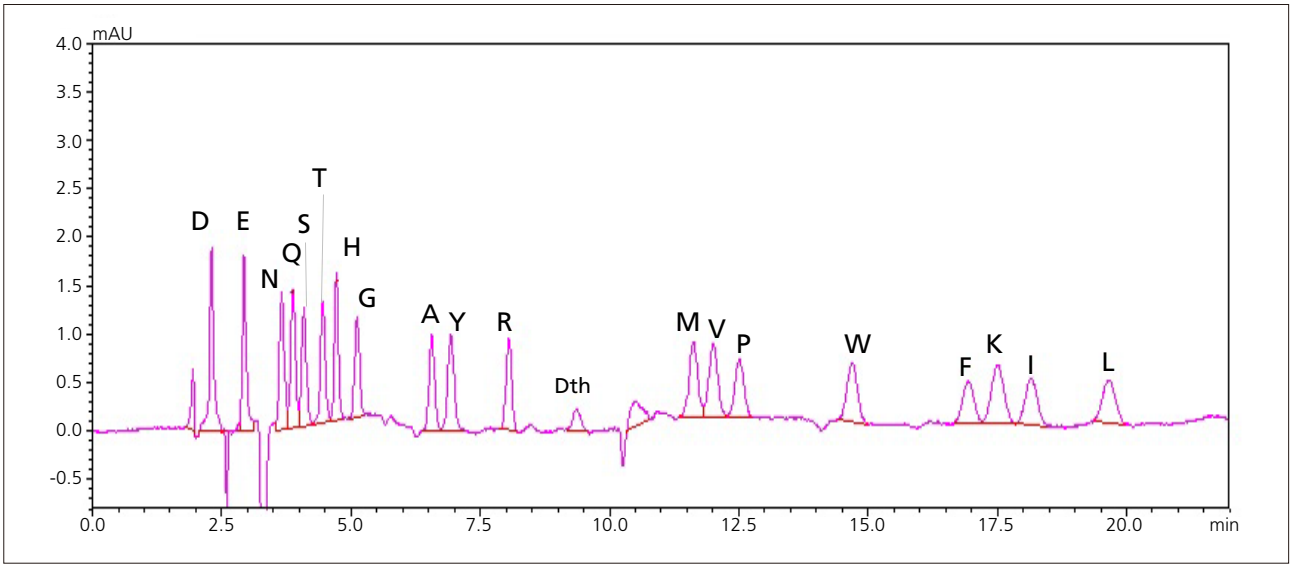


Fig. 8 Analysis of 5 pmol of a PTH-Amino Acid Standard Mixture

Appendix-3. Control and Data Processing Software

Control of the Edman reactor and HPLC analysis unit, as well as analysis of the acquired data are handled by the respective software application modules. This software is user friendly, providing excellent ease of operation.

Appendix-3-1. Control Software

The Edman reactor and HPLC analysis unit are controlled by the LabSolutions PPSQ software. Not only is this software used for operating the PPSQ-50 Series instrument, it also provides HPLC control, and displays chromatograms and the status of various components, thus allowing complete control of the PPSQ-50 Series from a single window.

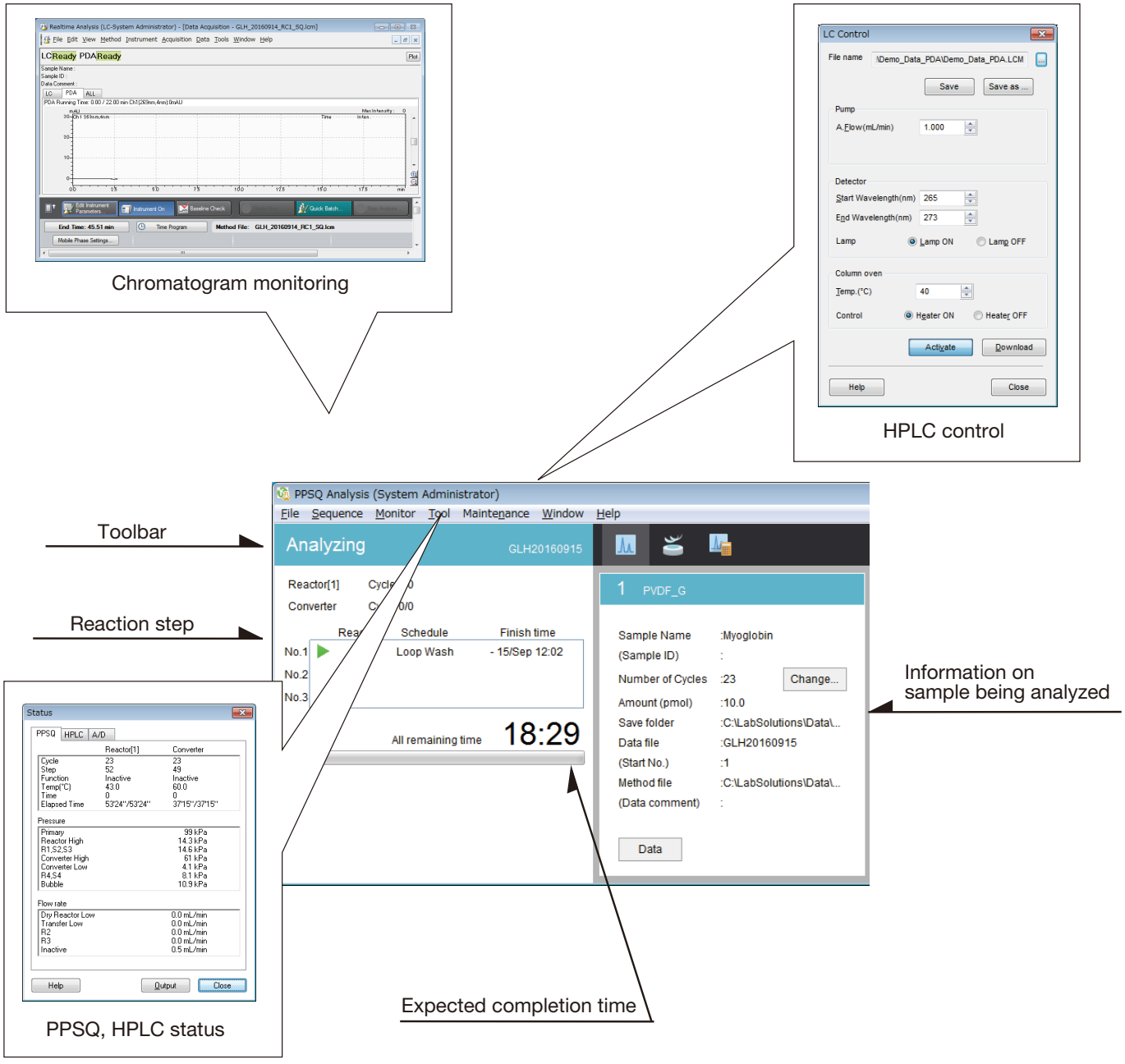


Fig. 9 LabSolutions PPSQ Window

Appendix-3-2. PPSQ Postrun Analysis Software

The PPSQ postrun analysis software is used to analyze the acquired chromatogram. This software automatically predicts the amino acid sequence, and is used to calculate the recovery, display the graph, as well as process the chromatograms. Chromatograms can also be reprocessed via the simple manual integration toolbar.

<Sequence Prediction>
Automatic prediction of the amino acid sequence is executed by comparing the chromatogram for which prediction will be conducted with the chromatogram acquired prior to this one, then determining the ratio of the peak heights of each amino acid, and finally applying our proprietary algorithm based on the differences. The displayed prediction results show the amino acids up to the 4th candidate, the degree of reliability of the 1st amino acid candidate (accuracy threshold), and the evaluation score (the value on which reliability is based). The display of amino acids up to the 4th candidate can be useful for deliberation in cases when multiple samples are mixed together or when the prediction reliability is low. In addition, the initial recovery as well as the repeat recovery calculated from selected amino acid yield can also be displayed in graph format for each cycle, annotated with the PTH-AA abbreviation.

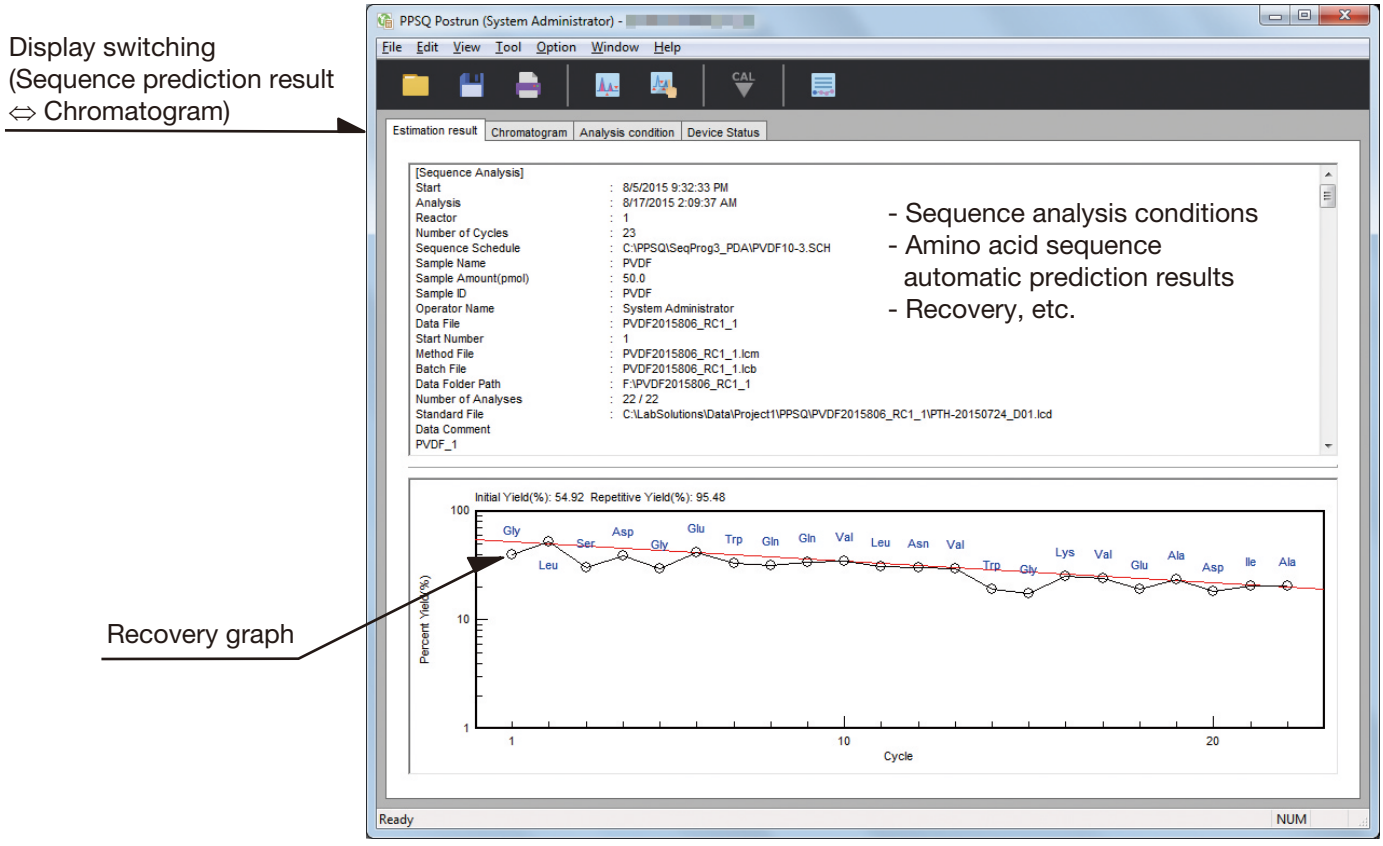


Fig. 10 PPSQ Postrun Analysis Window – Sequence Prediction Results –

<Chromatogram>
Reprocessing as well as display and printout of the chromatogram are conducted in this window.
Calibration of PTH-amino acids and peak integration of the chromatogram, which are necessary for data analysis, can be performed here. Also, since high repeatability of elution times is obtained using the isocratic mode for analysis of the PTH-amino acids, a subtraction chromatogram consisting of the current chromatogram minus that from the prior cycle can be generated. Displaying multiple chromatograms at the same time makes it easy to identify characteristic PTH-amino acid additions at each cycle. In addition, since the automatic prediction results obtained in sequence prediction are synchronized with the chromatograms, confirmation of the automatically predicted PTH-amino acids at each cycle is facilitated.

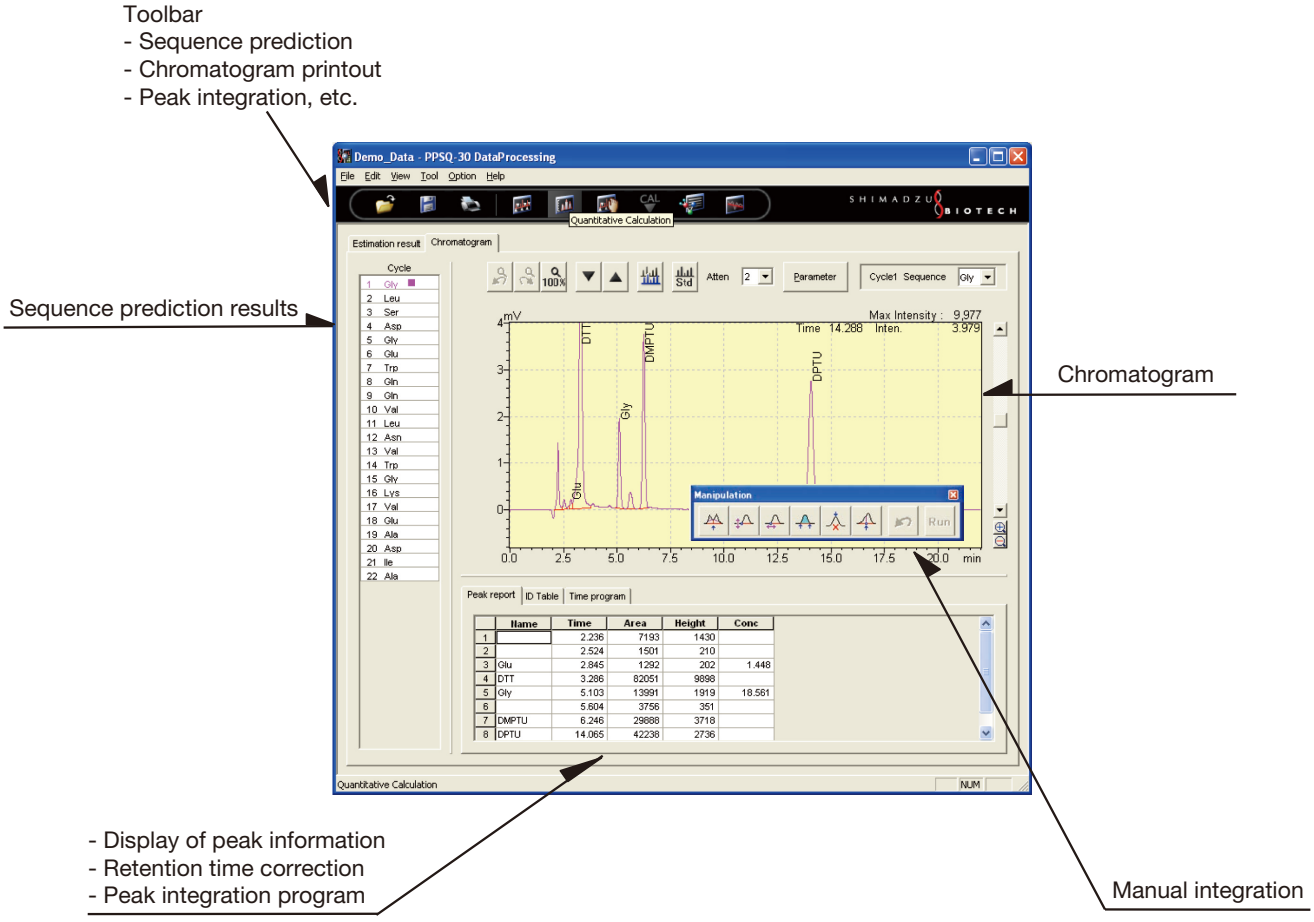


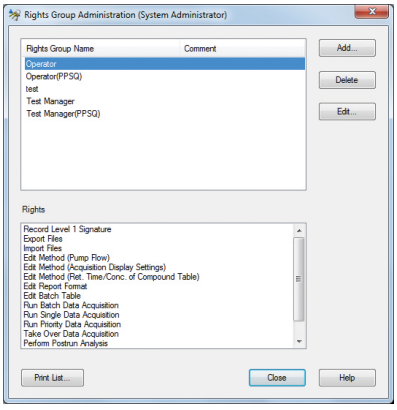
Fig. 11 PPSQ Postrun Analysis Window – Chromatogram –

Appendix-3-3. Compliance with FDA 21 CFR Part 11

LabSolutions PPSQ software provides compliance with FDA 21 CFR Part 11 guidelines and enables compliance with the security, user management, and audit trail requirements specified by FDA 21 CFR Part 11.

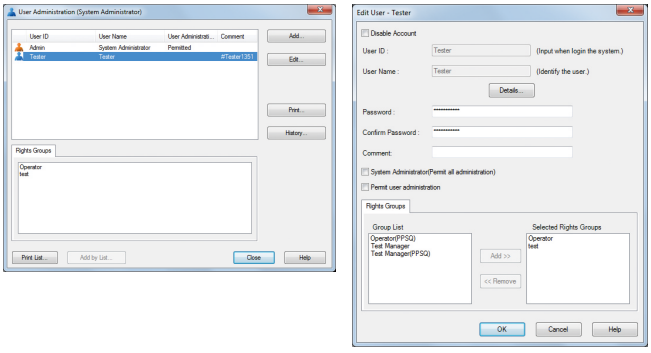
1 Security

Users are managed on the basis of groups, with each user being recognized by means of a username and password. Unique groups can be created by including persons having different access privileges. Clearly defining each user's access privileges prevents unauthorized changes to settings, instrument operation, and data access.



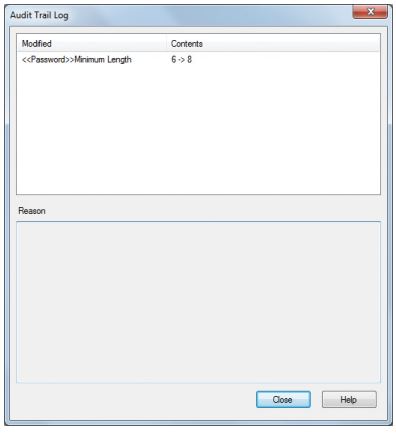
2 User Management

LabSolutions' user administration comprises the setting of rights groups and assignment of rights to users just as in Windows. Access rights required for each user can be set by assigning various levels of access to each user. This helps achieve effective user administration and more efficient laboratory operations.



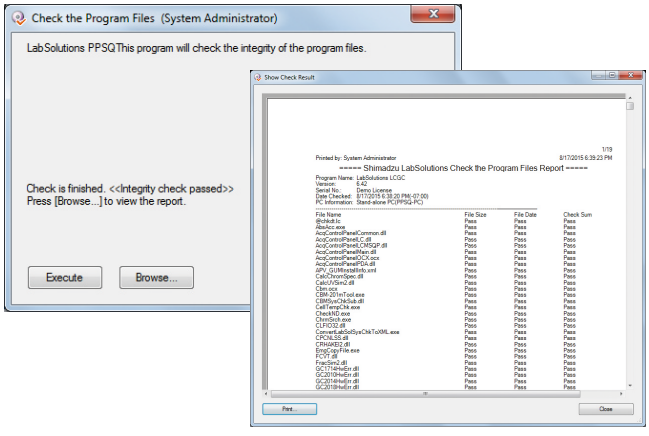
3 Audit Trail

Logins and logouts to the system, changes in users and groups, and the start and completion of acquisition, together with the username and the time, are all recorded. The recorded operational log can be registered in the database to provide traceability.



4 Software Validation

The integrity of the programs that comprise the system and the raw data acquired by instruments can be checked, ensuring the reliability of the system and data.
The results of these alteration checks can be managed as printouts.



Appendix-4. Sequence Analysis Example

The amino acid sequence analysis results obtained from analysis of 5 pmol Erythropoietin (CALBIOCHEM, cat#329871) are shown in Fig. 12. The raw chromatogram of the first cycle, and the subtraction chromatograms from the second and subsequent cycles are displayed. In the raw chromatogram, not only are PTH-amino acids at their characteristically increased numbers detected, but also detected are peaks originating from by-product substances generated at each cycle. However, by displaying the subtraction chromatograms, the by-product substances as well as the PTH-amino acids in the background that were detected in the prior cycle are cancelled, making it easy to pick out the amino acids that were added at each cycle (Fig. 13).

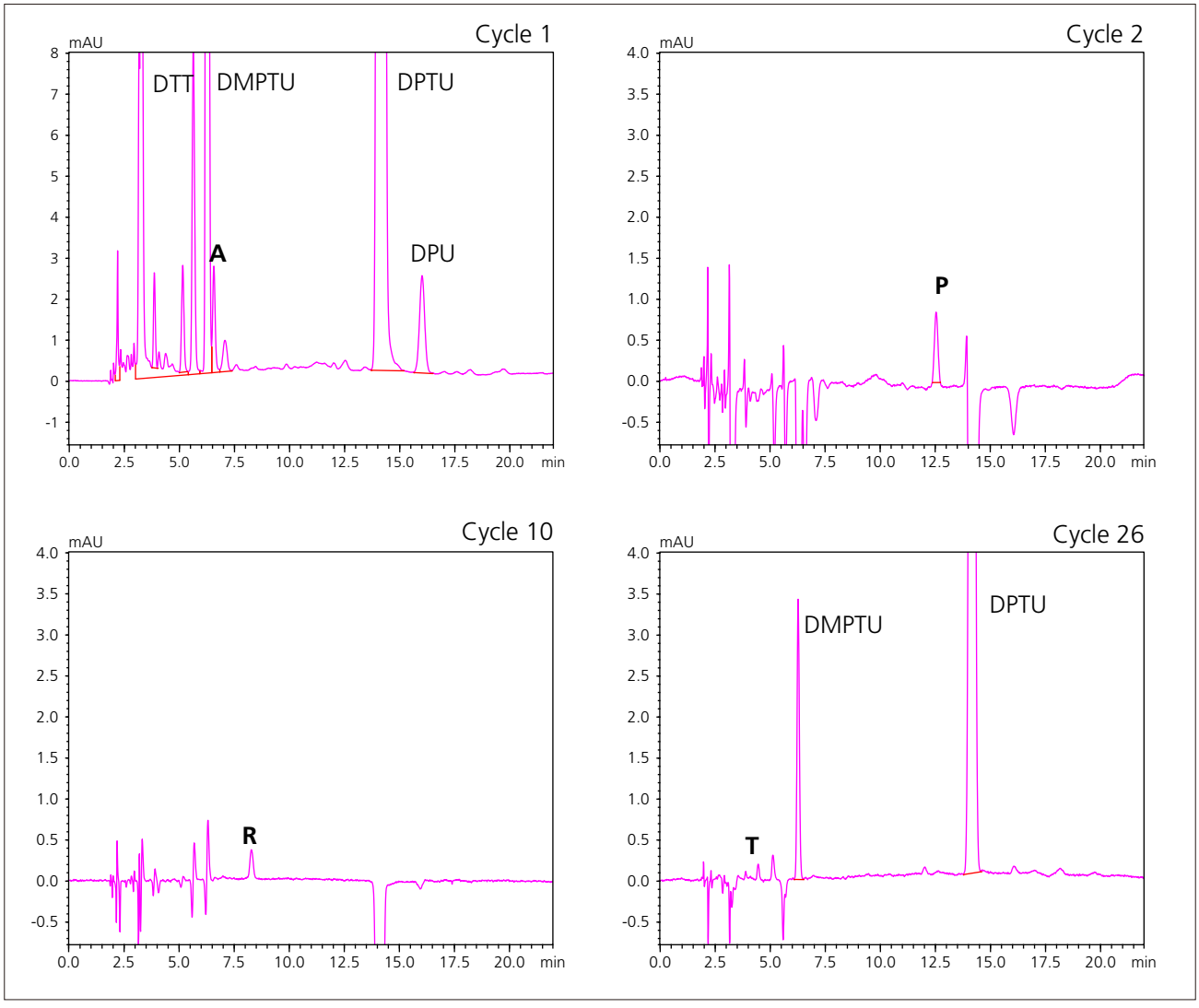


Fig. 12 Sequence Analysis of 5 pmol Erythropoietin

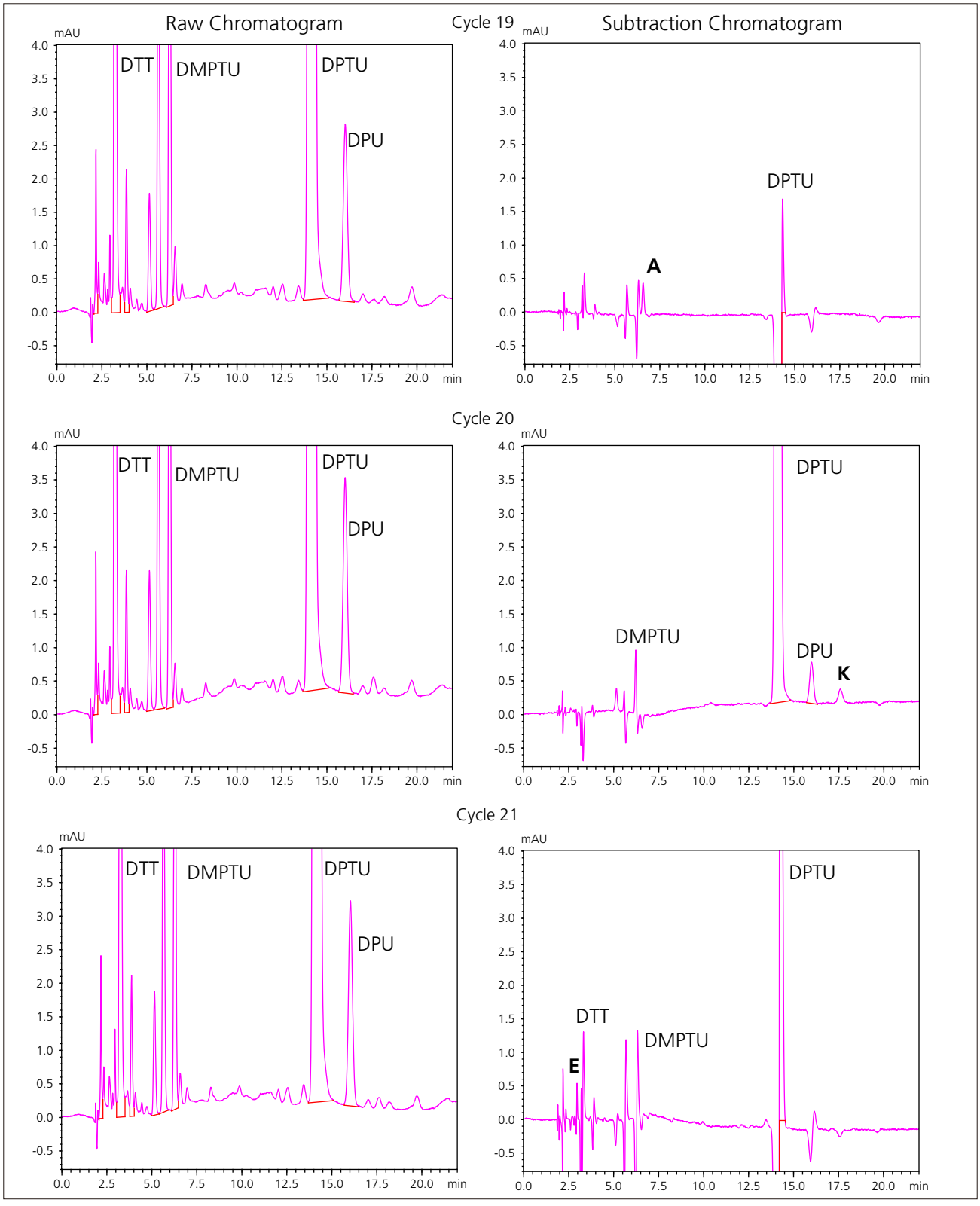


Fig. 13 Sequence Analysis of 5 pmol Erythropoietin (Raw Chromatogram and Subtraction Chromatograms)

Appendix-5. Identification of Cysteine by Pyridylethylation in Reactor

Cysteine residues typically form S-S bonds, so even if the half-cysteine residue of the N-terminus is converted to ATZ-half-cysteine due to Edman degradation, the half-cysteine and disulfide form a bond at the C-terminus, preventing elution and detection by HPLC. Therefore, identification in sequence analysis is generally conducted following reduction and alkylation. Using the PPSQ-50 Series, pyridylethylation (reduction-alkylation of Cys, Fig. 14) is first carried out in the reactor, and then continuous sequence analysis is conducted. Cysteine which has undergone pyridylethylation (PEC) is detected in the vicinity of 8.7 minutes. (Fig. 15) Since there is no overlapping of elution positions of the PTH-amino acids and Edman reaction by-products, identification can be carried out using the same HPLC conditions. Examples of oxytocin (human) analysis using this pyridylethylation program and using the typical Edman reaction program are shown side-by-side in Fig. 16 and 17. Thus, while there was no Cys cycle that could be found in analysis of intact cysteine, it is clear that identification is easy when detected as pyridylethylcysteine.

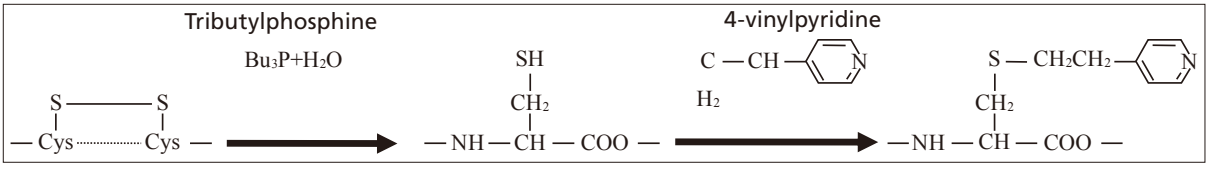


Fig. 14 Pyridylethylation Reaction Scheme

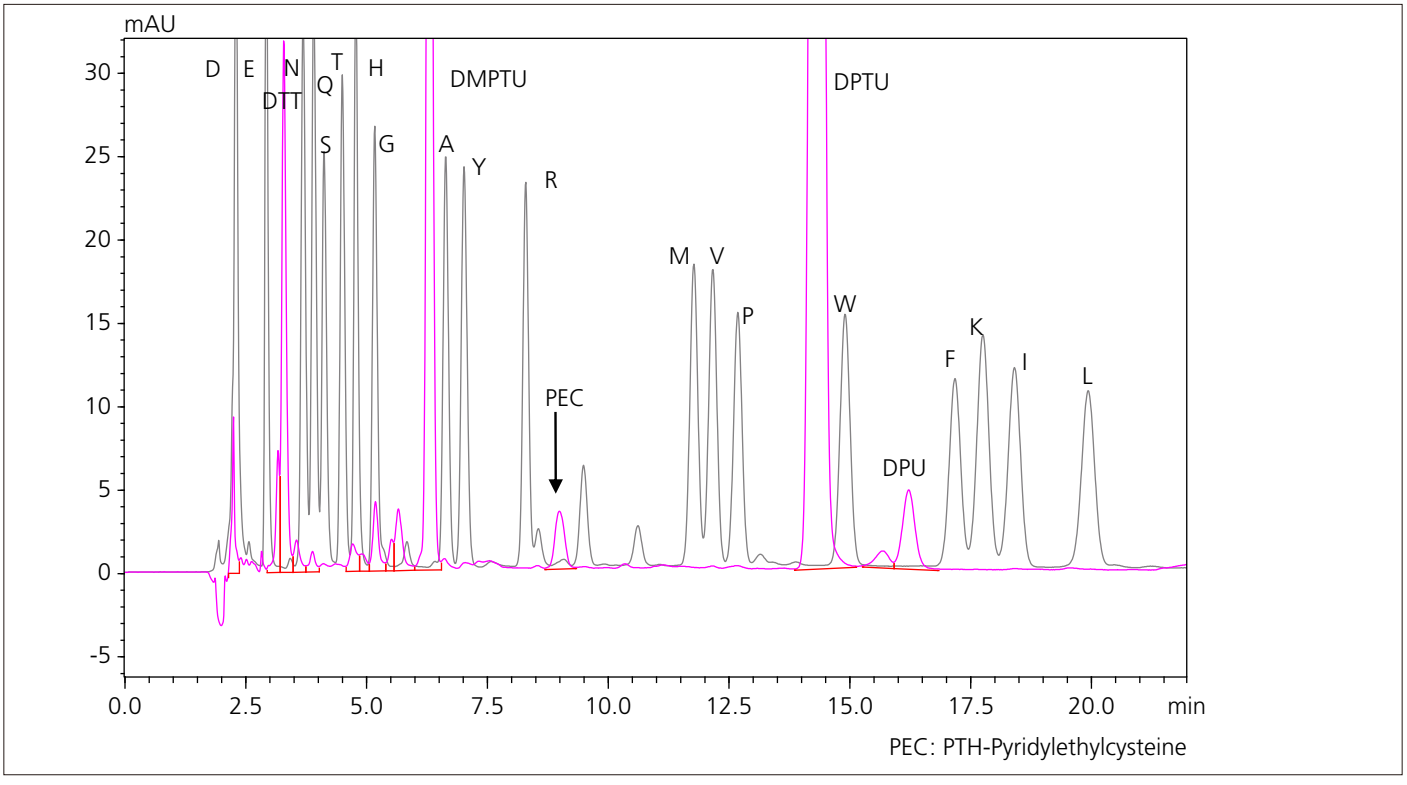


Fig. 15 PTH-Pyridylethylcysteine Elution Position

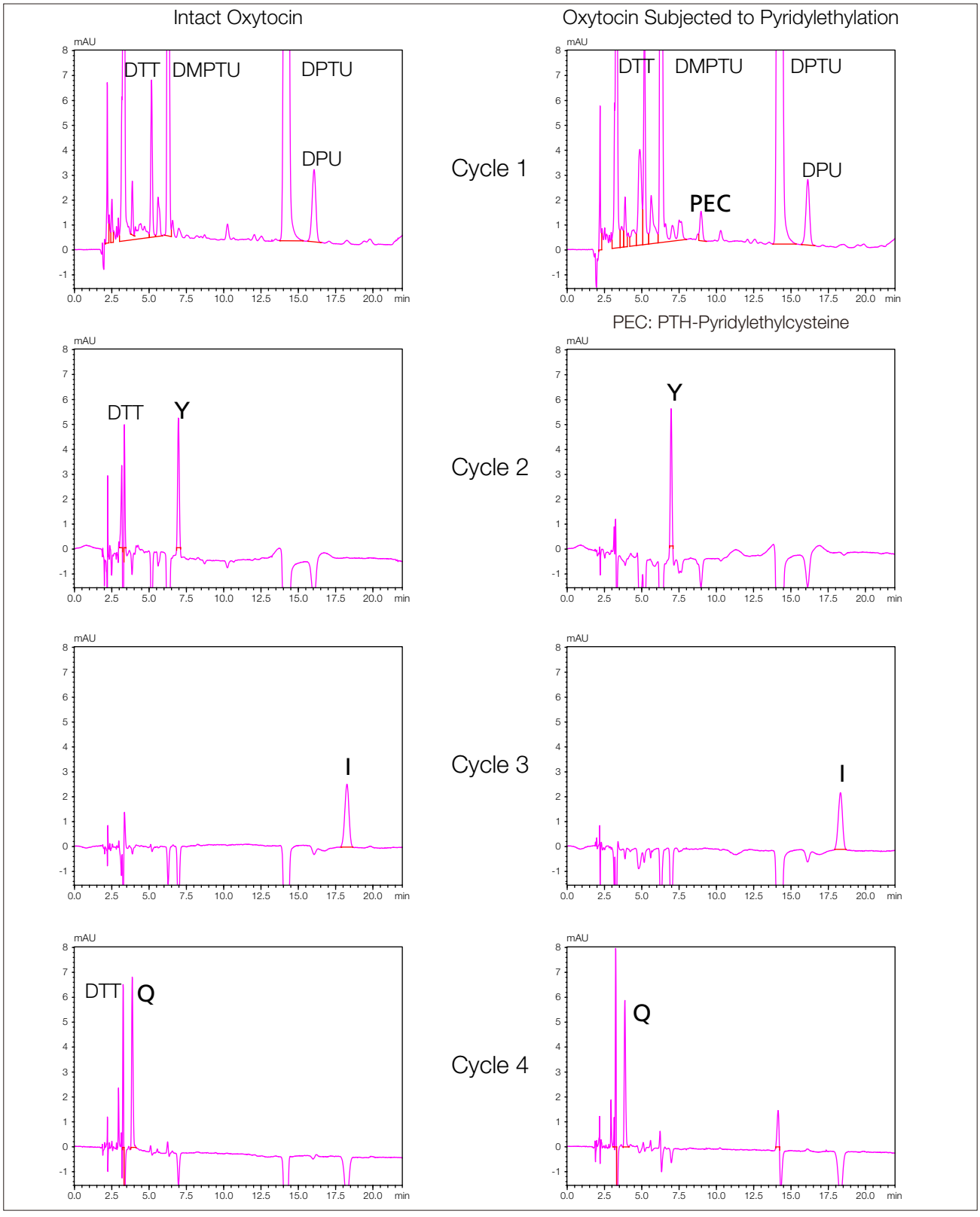


Fig. 16 Sequence Analysis of 10 pmol Oxytocin (Human)
(Subtraction chromatograms shown from 2nd cycle)

Oxytocin : Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH_2

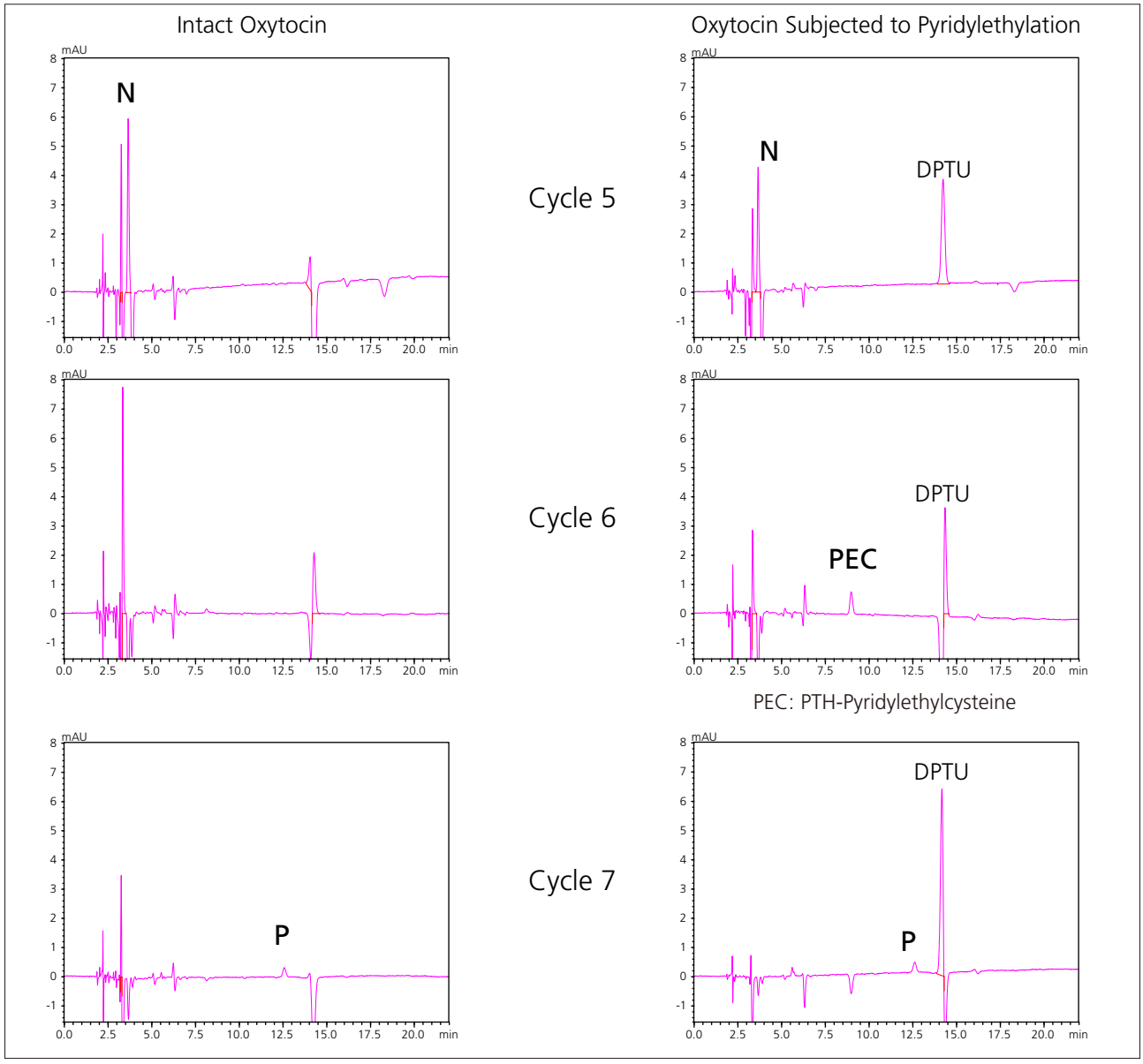


Fig. 17 Sequence Analysis of 10 pmol Oxytocin (Human)
(Subtraction chromatograms shown from 2nd cycle)

Oxytocin : Cys-Tyr-Ile-Gln-Asn-Cys-Pro-Leu-Gly-NH_2

References:
1. G. Allen, Laboratory Techniques in Biochemistry and Molecular Biology: Sequencing of Proteins and Peptides, North-Holland Publishing Company (1981)

*This document is based on information valid at the time of publication. It may be changed without notice.

First Edition: June, 2010
(Second Edition: November, 2016)

SHIMADZU

Shimadzu Corporation
www.shimadzu.com/an/

For Research Use Only. Not for use in diagnostic procedures.
This publication may contain references to products that are not available in your country. Please contact us to check the availability of these products in your country.

The content of this publication shall not be reproduced, altered or sold for any commercial purpose without the written approval of Shimadzu. Company names, products/service names and logos used in this publication are trademarks and trade names of Shimadzu Corporation, its subsidiaries or its affiliates, whether or not they are used with trademark symbol "TM" or "®".
Third-party trademarks and trade names may be used in this publication to refer to either the entities or their products/services, whether or not they are used with trademark symbol "TM" or "®".
Shimadzu disclaims any proprietary interest in trademarks and trade names other than its own.

The information contained herein is provided to you "as is" without warranty of any kind including without limitation warranties as to its accuracy or completeness. Shimadzu does not assume any responsibility or liability for any damage, whether direct or indirect, relating to the use of this publication. This publication is based upon the information available to Shimadzu on or before the date of publication, and subject to change without notice.

© Shimadzu Corporation, 2016

Related Solutions

- Life Science

- Price Inquiry
- Product Inquiry
- Technical Service / Support Inquiry
- Other Inquiry