Application Note

Mid-Infrared-based quantitation of highly modified proteins, labeled peptides and peptidomimetics

Introduction

Mid-infrared (MIR) spectroscopy exploits the fact that all molecules absorb specific frequencies of electromagnetic radiation and that the absorbance pattern is characteristic to the chemical composition of the analyzed molecule. The primary structure of all proteins and peptides is defined by a chain of amino acids linked covalently via peptide (amide) bonds. The MIR absorption pattern of an amide bond consists of nine bands. Among these, Amide I (1600 – 1690 cm^{-1}) and Amide II (1480 – 1575 cm^{-1}) have been identified as the most representative vibration modes[1,2]. In the past, analysis of the Amide I and II absorption regions has been severely hindered in aqueous solutions due to spectral interference of a strong water absorption band (at 1500 – 1700 cm^{-1})[4]. Application of biological material directly onto a polytetrafluoroethylene (PTFE) membrane that is spectrally transparent above 1300 cm^{-1} has enabled quantitative and qualitative analysis of a dried analyte without any additional sample manipulation, using the Direct Detect® spectrometer (EMD Millipore, Cat. No. DDHW00010-WW).

EMD Millipore is a division of Merck KGaA, Darmstadt, Germany

Enabling MIR-based quantification

MIR spectrometry is well established for the analysis of proteins and peptides, delivering valuable information of primarily qualitative and structural nature[2,3,5]. The unique design of the Direct Detect® assay-free sample card (EMD Millipore Cat. No. DDAC00010-GR) simplifies sample application and contains the sample within the focused MIR beam. Precise overlap of the MIR beam and a sample spot leads to highly reproducible analysis and permits accurate quantification. However, calibration is still required.

The MIR-based analysis is dependent on the number of amide bonds present in the protein or peptide. Review of the human protein database (Uniprot Release 2012_10) showed that many protein properties, like length and mass, the number of specific amino acids and even the predicted extinction coefficients at 280 nm, vary widely from protein to protein (Table 1). At the same time, the average mass per residue is highly consistent, despite the fact that individual amino acid contributions range from 57 Da for glycine to 186 Da for tryptophan. Additionally, since the number of amide bonds in a protein or peptide is only one less than the number of residues, the mass per amide bond is also very consistent, even for short proteins and peptides. The mass per amide bond calculated even when including the side chain amides of asparagine and glutamine residues present in natural sequences remains very consistent throughout the sequences reported in the database.

	Length	MW [Da]	Average MW/ AA [Da]	nGln	nAsn	EC [AU/(M cm)]	EC [AU mL/(mg cm)]	Amide Bonds	Mass / Amide [Da]
Average	558	62164.2	111.5	27	20	59310	1.01	604	103.6
STDEV	603	66776	3.4	31	24	65792	0.51	653	3.4
Max.	34350	3816037	138.3	942	1111	3991820	5.26	36402	166.9
Min.	4	500.6	82.6	0	0	0	0	3	65
CV	108%	107%	3%	118%	121%	111%	51%	108%	3%

Table 1. Highly consistent average mass per amino acid and mass per amide bond across all proteins, as determined by surveying the Human Protein Database (Uniprot Release 2012_10). The number of amide bonds was determined by adding the number of asparagines and glutamines to the total number of amino acids and subtracting one. The average mass per amide was determined by dividing the molecular weight by the number of amides. EC=extinction coefficient.

Consequently, it can be expected that the MIR absorbance (on a mass, not a molar basis) of proteins varying in length and sequence would be very consistent and highly comparable. Also, any protein or peptide for which the average residue mass does not deviate significantly from the typical ~112 Da per residue and ~104 Da per amide bond can serve as a reliable standard for calibration.

MIR-based analysis using the Direct Detect® spectrometer starts with a measurement of Amide I band absorption. Absorption strength is compared to the standard curve and converted to a concentration that is displayed in mg/mL. Bovine serum albumin (BSA) is accepted and widely used as a standard in protein concentration determinations. With 583 amino acids, a mass of 66,433 Da, 14 asparagines and 20 glutamines, the protein is very typical of an average human protein (Table 1). The average residue mass for BSA is slightly below 114 Da with average mass per amide bond approaching 108 Da. Both values are also close to the average obtained for proteins in the Human Protein Database. If BSA is used for calibration, the concentration values reported by the Direct Detect® software are accurate for all protein and peptides with average mass per residue comparable to BSA.

However, MIR-based quantitation using BSA calibration will not be accurate when applied to analysis of highly modified proteins, labeled peptides and peptidomimetics, where average residue mass values vary significantly from 114 Da. In the case where the modification is known, a manual recalculation of the results will permit accurate estimation of sample concentration. The following examples illustrate accurate MIR-based quantitation of five types of proteins and peptides whose residue masses are not comparable to that of the average human protein.

Example 1. MIR-based quantification of unnatural sequences

MIR-based quantification of peptides with sequences derived from naturally occurring proteins can be achieved with the general method relying on comparison to BSA. However, unnatural sequences like poly-glycine or poly-tryptophan cannot be quantified accurately using BSA as a standard. For example, a 20-mer composed exclusively of glycine residues is represented by average mass per amide bond of only 61 Da, which is only about half of the average observed for human database and BSA. If BSA is used as the standard for quantification of this peptide, the result will be grossly overestimated. A simple adjustment can be used to correct concentration. The BSA-based result should be multiplied by the analyte's real MW per amide (61) and divided by the BSA MW per amide (108).

Sequence	MW (Da)	MW/residue (Da)	MW/amide (Da)
GGGGGGGGGG GGGGGGGGGG	1159.00	57.95	61.00

correct concentration = BSA-based reading * (current "MW per amide"/BSA "MW per amide")

BSA-based reading	Recalculation		Correct concentration
0.500 mg/mL	0.5 * (61/108)	=	0.282 mg/mL
1.000 mg/mL	1 * (61/108)	=	0.565 mg/mL
2.000 mg/mL	2 * (61/108)	=	1.130 mg/mL

Similar to poly-glycine, a 20-mer composed solely of tryptophans will not be quantified correctly if BSA is used as a standard. The average residue mass of a poly-tryptophan peptide approaches 187 Da and its average mass per amide bond increases to 197 Da. Again, comparison to BSA or any other protein from human database will not deliver accurate results – the peptide concentration will be underestimated.

Sequence	MW (Da)	MW/residue (Da)	MW/amide (Da)
WWWWWWWWWW WWWWWWWWWW	3742.20	187.11	196.96

Recalculation using the above equation will allow proper correction:

BSA-based reading	Recalculation		Correct concentration
0.500 mg/mL	0.5 * (197/108)	=	0.912 mg/mL
1.000 mg/mL	1 * (197/108)	=	1.824 mg/mL
2.000 mg/mL	2 * (197/108)	=	3.648 mg/mL

Example 2. MIR-based quantification of highly modified proteins: glycosylation

Human type I interferon α receptor 1 (IFNAR1) is a 530 amino acid protein with a sequence that contains 35 asparagines and 23 glutamines. The average residue mass calculated for INFNAR1 is 115 Da and average mass per amide bond is 104 Da, confirming that the protein can be quantified using BSA as a standard. However, INFAR1 in its native form is heavily glycosylated, resulting in a molecular weight of 120 – 130 kDa[6]. The majority of INFAR1 glycosylation is reported to be introduced without formation of additional amide bonds. Therefore, the resulting average mass per amide bond for glycosylated protein increases to around 215 Da. Consequently, analysis of the heavily glycosylated IFNAR1 using BSA as a standard will underestimate its true concentration. Since glycosylated IFNAR1 contains almost the same amount of amide bonds as BSA but is twice as heavy, the BSA-based quantification results need to be doubled in order to be accurate. Again, one must multiply by the analyte's real MW per amide (215) and divide by the BSA MW per amide (108).

Re-calculation examples:

BSA-based reading	Recalculation		Correct concentration
0.500 mg/mL	0.5 * (215/108)	=	0.995 mg/mL
1.000 mg/mL	1 * (215/108)	=	1.991 mg/mL
2.000 mg/mL	2 * (215/108)	=	3.981 mg/mL

Example 3. MIR-based quantification of modified proteins and peptides: phosphorylation

Protein phosphorylation represents an important regulatory mechanism in eukaryotic cells. For example, a tyrosine-based activation motif (ITAM) is an intrinsic part of T-cell signaling pathway[8], and ITAM-containing sequences are often investigated in both unmodified and phosphorylated forms[9]. The concentration of the test peptides can be estimated by MIR-based analysis.

However, if BSA is used as a standard in the analysis of both the unmodified and phosphorylated forms, the accuracy of concentration estimation will be compromised. A simple adjustment will correct the error. For example, the twenty amino acid long fragment of human T-cell zeta receptor chain contains three residues that can be phosphorylated (two tyrosines and one serine).

T-cell receptor zeta chain	MW (Da)	MW/residue (Da)	MW/amide (Da)
(106) QEGL**pY**NELQKDKMAEA**pYpS**EI (125)	2599.57	129.98	123.79

Estimating the concentration of the fully phosphorylated version (124 Da per amide bond) using BSA as a standard (108 Da per amide bond) will introduce approximately 15% error. This can be corrected in a single-step manual adjustment:

BSA-based reading	Recalculation		Correct concentration
0.5 mg/mL	0.5 * (124/108)	=	0.574 mg/mL
1.0 mg/mL	1 * (124/108)	=	1.148 mg/mL
2.0 mg/mL	2 * (124/108)	=	2.296 mg/mL

Example 4. MIR-based quantification of modified proteins and peptides: biotinylation

Biotin-Avidin affinity is commonly used as a detection system and biotin-modified sequences are frequently used in scientific investigations. Again, MIR-based quantification method can be found useful in estimating the concentration of biotinylated peptides. While attachment of a single biotin (via amide bond) will not have any significant effect on average mass per amide, addition of multiple biotin molecules will most likely cause a meaningful increase. Therefore, the quantification of some biotinylated materials by MIR-based method using BSA as a standard might result in underestimation of the concentration. The fragment of T-cell zeta receptor analyzed in example 3 contains two lysines, so it can be modified by the addition of three biotin molecules:

T-cell receptor zeta chain	MW (Da)	MW/residue (Da)	MW/amide (Da)
(106) **bQ**EGLYNELQ**bK**D**bK**MAEAYSEI (125)	3037.57	151.88	126.57

Estimating the concentration of triply biotinylated T-cell receptor zeta chain fragment (127 Da per amide bond) using BSA as a standard (108 Da per amide bond) will introduce approximately 18% error, which again can be corrected in a single-step manual adjustment:

BSA-based reading	Recalculation		Correct concentration
0.5 mg/mL	0.5 * (127/108)	=	0.588 mg/mL
1.0 mg/mL	1 * (127/108)	=	1.176 mg/mL
2.0 mg/mL	2 * (127/108)	=	2.352 mg/mL

Example 5. MIR-based quantification of peptidomimetics

The use of peptidomimetics as antimicrobial agents, immunosuppressors and enzyme or signaling pathway inhibitors is well known. Most peptidomimetics studies require accurate concentration estimation. The example described below demonstrates how MIR-based analysis can be applied in order to accurately estimate the concentration of any peptidomimetic.

Ac-p-Tyr-Leu-Pro-Gln-Thr-Val-NH_2 has been proven as an effective inhibitor of signal transducer and activator of transcription 3 (Stat 3) protein[6]. The 840.9 Da hexamer contains an acetylated N-terminus, a phosphorylated tyrosine and amidated C-terminus. Consequently, the average residue mass is 140 Da while the average mass per amide bond is around 93 Da. Similar to the poly-glycine example, BSA-based quantification of the described peptidomimetic will deliver inflated concentration values. Simple adjustment for proper mass per amide corrects this:

BSA-based reading	Recalculation		Correct concentration
0.500 mg/mL	0.5 * (93/108)	=	0.431 mg/mL
1.000 mg/mL	1 * (93/108)	=	0.861 mg/mL
2.000 mg/mL	2 * (93/108)	=	1.722 mg/mL

Conclusions

MIR-based quantification delivers accurate results when applied to natural sequences. However, the use of bovine serum albumin as a standard will not be accurate in the analysis of sequences with average mass per amide bond differing significantly from the number observed for BSA, which is 108 Da per amide. The error associated with the measurement is easily correctable in all cases where average mass per amide bond of analyzed material is known. Simple, manual mathematical adjustment permits estimation of the accurate concentration.

Reference

1. T. Miyazawa and E.R. Blout. The infrared spectra of polypeptides in various conformations Amide I and II bands, J. Am. Chem. Soc. 1960; 83:712–719.
2. J. Kong and S. Yu, Fourier transform infrared spectroscopic analysis of protein secondary structure, Acta Biochim. Biophys. Sin. 2007; 39:549–559.
3. M. Jackson and H.H. Mantsch, The use and misuse of FTIR spectroscopy in the determination of protein structure, Critical Rev. Biochem. Mol. Biol. 1995; 30:95-120.
4. F. Cadet, S. Garrigues and M. de la Guardia, Quantitative analysis, infrared, Encyclopedia of Analytical Chemistry (2012) published online 15 September. DOI: 10.1002/9780470027318.a5610. pub2M.
5. K. Baginska et. al., Conformational studies of alanine-rich peptide using CD and FTIR spectroscopy, J. Pept. Sci. 2008; 14:283-289.
6. L.E. Ling et. al., Human Type I Interferon Receptor, IFNAR, is a heavily glycosylated 120-130 kD membrane protein, J. Interferon Cytokine Res. 1995; 15:55-61.
7. B. Debnath et. al., Small molecule inhibitors of signal transducer and activator of transcription 3 (Stat 3) protein, J. Med. Chem.2012; 55:6645-6668.
8. Isakov N., Role of immunoreceptor tyrosine-based activation motif in signal transduction from antigen and Fc receptors, Adv. Immunol. 1998; 69:183–247.
9. V. Chitu et. al., Comparative study on the effect of phosphorylated TCR ζ chain ITAM sequences on early activation events in Jurkat T cells, Peptides 2001; 22:1963-1971.

Ordering Information

Description	Qty	Catalogue No.
Direct Detect® Spectrometer and Starter Kit Includes:	1	DDHW00010-WW
Direct Detect® Spectrometer	1	
Universal Power Adapter	1	
Dell Latitude™ 2120 Netbook and power adapter	1	
Direct Detect® Software	1	
Netbook Stand	1	
Spotting Tray	1	
Ethernet Cable	1	
Direct Detect® Assay-free Cards (50/pk)	1	
Direct Detect® Assay-free Cards (50/pk)	1	DDAC00010-GR
Direct Detect® Assay-free Cards (50/pk)	4	DDAC00010-4P
Direct Detect® Assay-free Cards (50/pk)	8	DDAC00010-8P

To Place an Order or Receive Technical Assistance

In the U.S. and Canada, call toll-free 1-800-645-5476

For other countries across Europe and the world, please visit: **www.emdmillipore.com/offices**

For Technical Service, please visit: **www.emdmillipore.com/techservice**

Get Connected!

Join EMD Millipore Bioscience on your favorite social media outlet for the latest updates, news, products, innovations, and contests!

facebook.com/EMDMilliporeBioscience

twitter.com/EMDMilliporeBio

EMD MILLIPORE

www.emdmillipore.com/offices