Journal of Pharmaceutical and Biomedical Analysis 53 (2010) 29–36

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

Infrared spectroscopy for biopharmaceutical protein analysis

Peter Christian Groß, Michael Zeppezauer*

Symbiotec GmbH, Science Park Saar 1, D-66123 Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 24 October 2009
Accepted 8 March 2010
Available online 12 March 2010

Keywords:
IR
Spectroscopy
Histone H1
Validation
Quantitative analysis

ABSTRACT

A mid-infrared transmission spectroscopic method with the possibility of high sample throughput was developed and validated on the basis of GMP requirements for protein therapeutics. In comparison with the method predominantly used, i.e. the surface-sensitive attenuated total reflection (ATR), we have shown that transmission measurements in solution possess several advantages, such as quantitative analysis on the basis of Lambert–Beer's law, determination of identity and purity based on specifically developed identity criteria, characterization of protein structure and structural changes including aggregation in solution and study of ligand binding to the protein. The usefulness of this method is exemplified by the characterization of the drug substance of ONCOHIST® (recombinant human histone H1.3 and bis-Met-histone H1.3) and its interaction with phosphate ions. Our conclusion is that transmission mid-infrared spectroscopy is a powerful tool for protein analysis in biotechnology and supplements the current analytical techniques for biopharmaceutical quality control of therapeutic proteins.

© 2010 Published by Elsevier B.V.

1. Introduction

To ensure the quality and safety of a biopharmaceutical it is necessary to analyse the finished product by different methods for identity, purity and quantity. Especially during the development of new biopharmaceuticals, the objective is to collect a maximum of information, covering all accessible structural and process parameters. Currently most biopharmaceutical protein characterizations rely on different HPLC techniques, electrophoresis, and spectroscopic techniques [1]. Secondary/tertiary structure analysis is routinely done by X-ray, CD, fluorescence, NMR and FTIR spectroscopy. In particular the latter is applicable to samples in solution and in solid state, can be applied in numerous modes and has a very high sensitivity. Besides the possibility to study a protein's secondary structure with FTIR, its potential for detecting impurities is also high, for example the detection of purification-related buffer residues, detergents or protein impurities. In case of measuring protein solutions in transmission, the possibility for *direct* protein quantification exists by using Lambert–Beer's law. Protein quantification can cause problems with *indirect* colorimetric or spectrophotometric assays or when working with small proteins or peptides with fewer or no aromatic side chains and consequently a small UV signal at ~280 nm.

Application of mid-infrared spectroscopy in the field of biotechnology has mainly been limited to the analysis of protein secondary structure. Different approaches exist in the field, which all rely on the fine structure analysis of the protein's peptide bond vibrations, namely the Amide-I, -II and -III absorption bands [2–6]. The Amide-I band of a protein, which is most widely used for secondary structure analysis, is quite dependent on the protein's overall conformation. Numerous amino acid side chains also absorb in the same range and disturb the analysis [7]. The absorption bands of the different secondary structure elements are very broad and overlap to a great extent, which further complicates the evaluation. The assignment of a defined IR region to a single secondary structure element is a simplification, which is not applicable in every case. Even α-helices of different length can have various band positions [8]. Therefore the theoretical basis for secondary structure analysis with IR is not yet fully understood and suffers from different shortcomings. But the monitoring of the change of secondary structure, without exact interpretation, between samples of different production stages, during stability studies or for the reason of batch comparison is a very useful application of IR spectroscopy, disregarding the mentioned theoretical problems.

The study presented below explores the potential of IR spectroscopy for biopharmaceutical quality control by measuring *transmission* spectra of proteins in aqueous buffers. Due to the high absorptions of water at the bands of interest for a protein (e.g. 1650 cm^{-1}, Amide-I) many authors have preferred the surface-sensitive method ATR (attenuated total reflection spectroscopy) in order to avoid elaborate and time-consuming H/D exchange or the complications by using a very thin liquid layer cell (5–10 μm). However the transmission mode has distinct advantages compared to surface-sensitive analytical methods like ATR: in ATR spectroscopy the IR signal is mostly determined by the first few protein layers on

* Corresponding author. Tel.: +49 681 959 14 430; fax: +49 681 959 14 431.
E-mail address: m.zeppezauer@symbiotec.biz (M. Zeppezauer).

0731-7085/$ – see front matter © 2010 Published by Elsevier B.V.
doi:10.1016/j.jpba.2010.03.009

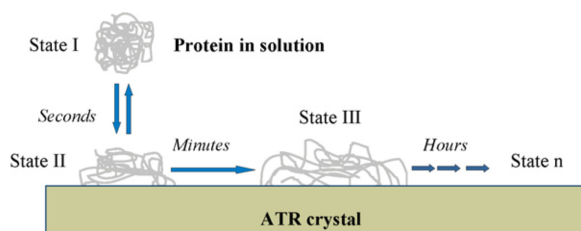

Fig. 1. Model of the irreversible protein adsorption onto an ATR crystal for dissolved proteins [9]. After the protein adsorbs onto the solid state it normally changes its secondary structure during long lasting transformation processes.

the interface between crystal surface and solution. In surface biochemistry the rule of thumb is, that 'every protein adsorbs on every surface' [9]. Considering the size and amphiphilic nature of proteins this is not surprising. In most cases the protein adsorption is irreversible and associated with extensive and long lasting (in the range of hours) protein secondary structure changes (Fig. 1) [9]. These processes can have a large impact on the ATR spectrum of a protein, resulting in erroneous conclusions. For example the secondary structure analysis can be disturbed, because the measured IR spectrum corresponds mainly to the adsorbed species with modified conformation and 'permanently changing' secondary structure.

During ATR analysis of different proteins, we could confirm the model of irreversible protein adsorption on the surface of ATR crystals exposed to protein solutions (unpublished own results).

Besides avoiding problems with protein secondary structure changes in transmission analysis compared to ATR, the second advantage of transmission measurements is the possibility of *direct* protein quantification on the basis of Lambert–Beer's Law; whereas no strict correlating law exists in case of ATR. The third advantage of transmission spectroscopy is a practical one: the lack of disturbing protein-surface interactions allows a high sample throughput without meticulous cleaning of the cell after each measurement.

A fundamental question by application of IR spectroscopy to biotechnology is, how the huge information, contained in a protein IR spectrum, can be properly extracted. Every protein spectrum is dominated by broad and intensive peptide bond vibrations of the backbone and numerous side chain vibrations. The polymeric nature of proteins built up from amino acids leads to a large overlap und summation of closely related IR absorption bands. Therefore, at first glance one could expect that the IR spectra of different proteins are similar to the extent that the identification of a particular protein by this method would seem very difficult if not impossible. One goal of this study was to explore the potential of transmission IR spectroscopy to contribute to the identification of a protein by analysing its individual infrared spectrum.

The subject of the present study was the drug substance of ONCOHIST®, a protein drug developed for the treatment of malignant diseases. This drug consists of an aqueous solution of recombinant human histone H1.3 and recombinant human N-bis-Met-H1.3. A first clinical trial Phase I/II with relapsed or refractory acute-myeloic-leukaemia (AML) patients has shown, that the protein is very well tolerated without side-effects and first signs of efficacy were observed [10].

2. Experimental

2.1. Spectra acquisition

All spectra were measured with a Bruker Tensor 27 FTIR spectrometer (Karlsruhe, Germany) with the Software Opus 5.5 and a HgCdTe detector. A resolution of 4 cm^{-1} and 25 scans per measurement were taken in the range between 3000 cm^{-1} and 1000 cm^{-1}. The flow-through liquid transmission IR cell was equipped with CaF_2 windows of 4 mm thickness, had an optical path length of 7.4 μm and an aperture of 5 mm (Microbiolytics GmbH, Freiburg Germany). The whole cell was temperature controlled, all measurements were done at 25 °C. For every measured sample, a spectrum of the pure buffer, in which the sample was dissolved, was measured as reference or background single channel spectrum for calculating the absorption spectrum of the sample.

2.2. Samples

Throughout the study the drug substance ONCOHIST® was used as solution in 0.9% NaCl for infusion (Ph. Eur.) at different concentrations, exceptions are mentioned separately. All samples were measured in duplicate while the instrument was flushed with gaseous nitrogen at a flow-rate of approximately 160 l/h. The sample volume necessary for one measurement was 20 μl. For the study of protein specificity 13 arbitrarily selected, proteins were measured in 0.9% NaCl at a concentration of 10.0 mg/ml: BSA (Sigma, A-9085), Anti-human IgG (Beckman Coulter, PNIM 0837), Chymotrypsin (Calbiochem, 230834), Cytochrome *c* (Merck, 1.24804), Bovine Insulin Chain A (Sigma, I-2254), Bovine Insulin Chain B (Sigma, I-6380), β-Lactoglobulin (Sigma, L-4520), Lysozyme (Sigma, L-6876), α-2-Macroglobulin (Dako Cytomation, A0033), Papain (Merck, 1.07144.0025), Transferrin (Behringwerke, 1227/34), Trypsin-Inhibitor (Serva, 37329), Urease (Boehringer Mannheim 1070225).

2.3. Spectra evaluation

If necessary a buffer correction before analysis was done by adjusting a defined Amide-I to Amide-II ratio. This ratio was quite sensitive to small differences between sample buffer and the corresponding background buffer. The correction was done by defining the Amide-I/Amide-II ratio of two ONCOHIST® standard spectra as reference values and by adjusting all other spectra between these two given limits by subtraction or addition of the background buffer spectrum. This kind of buffer correction is only applicable for very similar protein samples as it is the case in this study.

2.3.1. Comparison of second derivative spectra

Second derivative spectra are useful for resolution enhancement and determining band positions or shoulders. The software OPUS 5.5 offers the possibility for automated comparison of second derivative spectra by a mathematical algorithm. The results are quoted as a coefficient of correlation in percentage between 0% and 100%, where 100% means identical spectra. The comparison was always done at a fixed ONCOHIST® concentration of 10.0 mg/ml between the test and the standard spectrum in the range between 3000 cm^{-1} and 1000 cm^{-1}. As standard spectrum we defined the spectrum of an ONCOHIST® batch of the highest attained quality.

2.3.2. Secondary structure analysis

The prediction of secondary structure elements α-helix and β-sheet were done with the methods supplied by the Confocheck™ system (Bruker Karlsruhe, Germany). These methods calculate the secondary structure with a multivariate partial-least-squares algorithm (PLS) on the basis of a calibration data set of 45 different proteins.

2.3.3. Difference spectra

A very sensitive method to compare two spectra is the generation of a difference spectrum between them with preceding normalization of spectra. To this end is reasonable to define an ONCOHIST® standard spectrum and subtract it from the spectrum of interest. This new spectrum is then termed the difference spectrum of the sample. Afterwards the difference spectra of various

Table 1
Results for identity criteria after the examination of 16 ONCOHIST® IR standard spectra with a concentration of 10.0 mg/ml measured on two different days.

No.	Evaluation	Identity criterion
1	Second derivative spectra comparison (SDSC)	Spectra similarity >95%
2	α-Helix content	28.3 ± 2.0%
3	β-Sheet content	22.3 ± 2.0%
4	Difference spectrum	Relative deviations <0.005 AU

samples can be compared among each other. The definition of the reference spectrum depends on the purpose of the IR measurement. Sometimes, e.g. in case of stability studies, it is useful to define the first measured spectrum of the test sample at time 0 as reference spectrum. In other cases it is better to use an ONCOHIST® standard spectrum measured immediately before or after the measurement of the test sample. All spectra were prepared prior to subtraction in the following way: offset correction to 1713 cm^{-1}, vector normalization between 1600 cm^{-1} and 1500 cm^{-1} and a second offset correction to 2000 cm^{-1}.

2.4. Linearity

To evaluate the linearity of the method ten ONCOHIST® samples with different concentrations were prepared by dilution with 0.9% NaCl of an ONCOHIST® standard sample with a protein concentration of 58.6 mg/ml. Each of the solutions was measured on two different days in order to check the linearity and its possible day-to-day fluctuations. The height of the Amide-I band at 1651 cm^{-1} of the obtained IR spectra was analysed with the so-called quant1 method of Opus 5.5 software which calculates a linear regression on the basis of least squares between the measured absorbance at this wave number and the protein concentration.

3. Results and discussion

3.1. Specificity

3.1.1. Development of identity criteria

Proof of identity for unknown proteins can be made by direct information gained from the protein IR spectrum and/or comparison of the spectrum with reference data. In this work four different evaluation strategies were chosen:

1. Comparison of the second derivative of the test spectrum with the second derivative of the reference spectrum by calculation of a spectra similarity index with the OPUS software ('*second derivative spectra comparison*' or SDSC)
2. Chemometric Amide-I band analysis and prediction of α-helix content
3. Chemometric Amide-I band analysis and prediction of β-sheet content
4. Subtraction of an appropriate standard spectrum from the test spectrum (difference spectrum)

To avoid effects resulting from different protein concentrations, the identity analysis was done by using only a single ONCOHIST® concentration of 10.0 mg/ml. Standard values for the four identity criteria were established by evaluation of the spectra of 16 ONCOHIST® standard samples from a very pure GMP batch, the results are summarized in Table 1.

3.1.2. Protein specificity

With the four identity criteria defined in Table 1 it is now possible to explore the specificity of the method. For this purpose 13 arbitrarily selected non-cognate proteins were measured and compared with ONCOHIST® by the SDSC method (identity criterion 1, Table 1). The results are summarized in Table 2. None of the measured proteins has an IR spectrum similar enough to ONCOHIST® to fulfil the SDSC identity criterion of >95% similarity. This shows that IR spectroscopy is well suited to discriminate between different proteins, and the comparison of their second derivative spectra is sufficient to distinguish them. The application of the three other identity criteria to analyse these spectra in more detail was therefore not deemed necessary.

Table 2
Comparison between second derivative spectra ONCOHIST® and 13 arbitrarily selected non-cognate proteins.

Protein	SDSC (%)
Transferrin	73
Lysozyme	71
Cytochrome *c*	66
Bovine serum albumin	64
Trypsin inhibitor	56
Chymotrypsin	49
β-Lactoglobulin A	39
Papain	39
Anti-human IgG antibody	34
Insulin chain A	34
Insulin chain B	23
α-2-Macroglobulin	20
Urease	20

In view of this very small fraction of all known proteins further examinations have to be done. However considering the fact, that all proteins are chemically very similar, built up by 20 amino acids, this result is surprising. We saw that the differences between the proteins are not only limited to the conformation-sensitive Amide-I region, but even more the whole spectrum between 3000 cm^{-1} and 1000 cm^{-1} differs from protein to protein. Therefore it would be very useful to compile a protein infrared spectra databank (PISD), even if one could expect the more spectra are collected in a PISD databank, the possibility to meet a very similar spectrum from a structurally unrelated protein would increase [11].

3.1.3. Impurities

Mid-infrared spectroscopy bears the potential to detect both organic and inorganic impurities. For biopharmaceuticals the process-related impurities like buffer residuals or detergents, protein impurities in form of host-cell-proteins (HCP) or fragments and modifications of the main product are of high relevance. Two examples are given below. Fig. 2A shows the consequence of a small change in the production process, leading to an ONCOHIST® drug substance with significant deviations from the baseline in its difference spectrum. This contamination was not seen during analysis by RP-HPLC-ESI-MS or SDS-PAGE. The position and identity of the bands suggests the presence of a fragment derived from DNA, possibly a nucleoside strongly adsorbed on the histone H1 molecule. Complete identification of this impurity was not pursued further.

Fig. 2 panel B illustrates a particularly interesting observation: During the development phase of the manufacturing process of ONCOHIST®, an attempt was made to treat ONCOHIST® with the detergent Triton X-100 while being adsorbed on a cation exchange resin. The objective was to remove bacterial endotoxins and other lipids from the protein before eluting the pure protein, a method frequently used in the purification of therapeutic antibodies. After this washing step, the protein was eluted and further purified with two chromatographic steps without Triton X-100. In spite of this, the detergent molecule remained strongly adsorbed on the protein until completion of the purification. This could only be detected by the strong Triton X-100 absorption bands at 1100 cm^{-1} in Fig. 2B. To our knowledge, no other method would have allowed this iden-

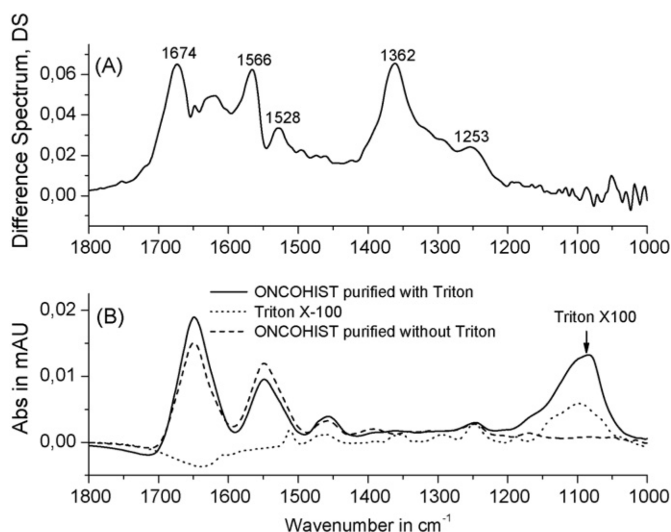

Fig. 2. (A) Difference spectrum of an ONCOHIST® GLP-batch, produced with small modifications in the purification process. The observed signals at the denoted positions are 10 times higher than the accepted identity criterion for ONCOHIST® (Table 1) and therefore significant. This difference spectrum could e.g. point to nucleoside adsorbed on the Histone H1. (B) IR spectra of an ONCOHIST® batch produced using Triton X-100 during purification process (—) compared to standard ONCOHIST® produced without Triton X-100 (—). The large peak at ~1100 cm^{-1} of ONCOHIST® purified with Triton X-100 originates obviously from Triton X-100 itself as proven by the IR spectrum of pure Triton X-100 (...). Therefore one can conclude, that Triton was not completely removed from the protein during the following steps of downstream processing.

tification in such a simple and speedy way. The IR method offers the potential also for a direct quantitation of the detergent in a protein solution, e.g. in the context of a cleaning validation.

To explore the potential of the method for the detection of protein impurities, mixtures of bovine serum albumin (BSA) and ONCOHIST® were taken as a model system. The aim of this work was to find the limit of detection (LOD) of BSA in an ONCOHIST® drug substance/product. First of all one has to identify significant differences between the pure spectra of both proteins. This is evident from Fig. 3, where at least eight differences could be detected. The requirements to detect BSA protein impurities in ONCOHIST® solutions with IR were therefore fulfilled. The results of the application of the first three identity criteria of Table 1 to the IR spectra of

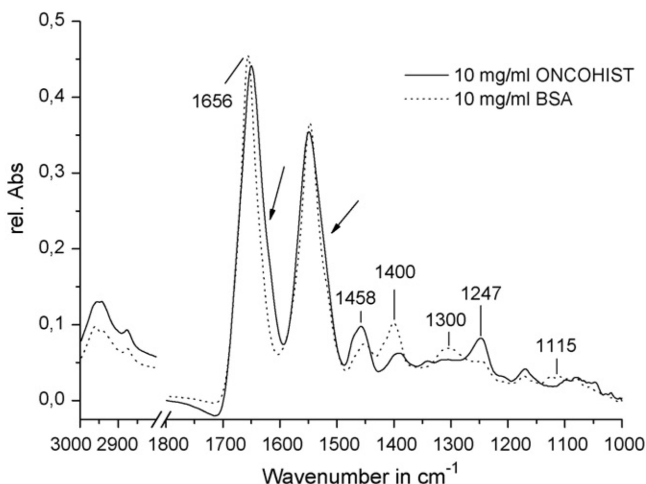

Fig. 3. Comparison between the FTIR spectrum of ONCOHIST® (—) and BSA (...). The spectra were vector normalized between 1600 cm^{-1} and 1500 cm^{-1} (Amide-II) and offset corrected to 2000 cm^{-1}. The differences between them are indicated by arrows or labelled with the wavenumber position in cm^{-1}.

Table 3
Detection of BSA impurities in ONCOHIST® solutions with the first three identity criteria SDSC, α-helix and β-sheet (total protein concentration 10.0 mg/ml for each).

BSA content (%) (w/w)	SDSC	α-Helix (%)	β-Sheet (%)
0	96.7	27.2	22.4
2	97.4	28.1	22.3
4	96.4	28.6	21.8
6	95.4	29.3	21.6
8	96.5	30.1	21.2
10	97.3	30.2	20.8
100	66.2	55.8	1.37

the different ONCOHIST®/BSA mixtures are summarized in Table 3. The SDSC criterion was not able to detect BSA up to 10% in the ONCOHIST®/BSA mixtures. The predictions for α-helix and β-sheet were showing a regular linear increase or decrease respectively, but all values in mixtures up to 10% BSA were within the defined identity criteria for ONCOHIST® (Table 1). The generation of difference spectra (not shown here) between the BSA/ONCOHIST® mixtures and a ONCOHIST® standard indicates at least two characteristic BSA marker bands at 1405 cm^{-1} and 1251 cm^{-1}, allowed the specific identification and even quantification of small BSA amounts in an ONCOHIST® solution. For example the peak height at 1405 cm^{-1} increases nearly linearly with rising BSA content in the range between 0 and 10% BSA ($y = 0.2857 + 0.6929x$; R 0.995). This result suggests, that the best method of analysing a small amount of protein impurity in a protein drug solution is the difference spectrum. Specific protein impurity marker bands in the difference spectra could then be used for quantification.

As it was shown the application of IR spectroscopy to analyse protein impurities, depends strongly on the protein impurity itself and its individual IR spectrum. The principal potential of IR spectroscopy to detect those protein impurities was not systematically examined until now. An amount of BSA as low as 2% in the ONCOHIST® solution is detectable by this approach. In this respect the IR method does not compete with more sensitive methods e.g. RP-HPLC, where protein impurities can be detected down to ~0.05%.

3.1.4. Protein aggregates

The occurrence of protein aggregates in pharmaceutical formulations of therapeutic proteins is a matter of great concern. To the extent possible their occurrence should be avoided since aggregates may add undesirable properties such as pyrogenicity to the drug. So far analytical size-exclusion-chromatography (SEC) is the method of choice to detect and quantify aggregates. FTIR spectroscopy in transmission mode for small and colloidal aggregates and ATR for insoluble aggregates are helpful complements to SEC. Well known is the so called 'β-aggregation', where different protein molecules are connected with each other via intermolecular β-sheets [2]. Those intermolecular β-sheets absorb in the typical β-sheet IR range between 1610–1640 cm^{-1} and 1670–1700 cm^{-1}. These bands are in most cases narrower than those for intramolecular β-sheets and can be identified accordingly [12].

The histone H1.3 forms aggregates in phosphate buffered saline solution (PBS) at concentrations higher than 25 mg/ml (unpublished results). To initiate a distinct aggregation an ONCOHIST® solution with 54.0 mg/ml in PBS (was prepared and afterwards directly diluted to 25.0 mg/ml. IR spectra of this solution were measured after different time intervals (Fig. 4). The evaluation of the three spectra using the three identity criteria is presented in Table 4. For the time 0 min the three identity criteria SDSC, α-helix and β-sheet corresponds to the standard values given in Table 1. This means at time 0 min, no detectable change or aggregation had been initiated yet. After standing for 120 min, the protein was in the process of conformation change, as observed by the increase of the

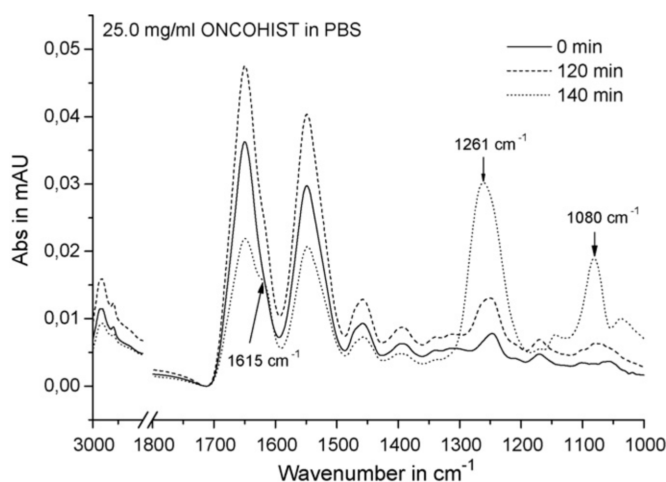

Fig. 4. Observing protein aggregation with FTIR spectroscopy *in situ*: IR transmission spectra of 25.0 mg/ml ONCOHIST® in PBS are shown at times: 0 min (—), 120 min (—) and 140 min (. . .) after dilution from a 54.0 mg/ml ONCOHIST® stock solution. The emerging shoulder at 1615 cm^{-1} in the 120 min spectrum indicates a so called β-aggregation. Furthermore in this spectrum the P═O asymmetric (1261 cm^{-1}) and symmetric (1080 cm^{-1}) stretching vibrations were increasing simultaneously during aggregation. This suggests the creation of specific phosphate binding sites in the protein molecule. The spectra were not normalized and represent the same protein solution, which was filled at time 0 min in the Aquaspec™ cell.

Amide-I band at 1650 cm^{-1} and two other bands at 1261 cm^{-1} and 1080 cm^{-1}. After 140 min, the process was still going on (140 min spectrum in Fig. 4) with now a decrease of the Amide-I band and an increase of the two bands at 1261 cm^{-1} and 1080 cm^{-1}.

The Amide-I band in Fig. 4 changes with time to a typical spectrum for a β-aggregation with a significant and narrow shoulder at 1615 cm^{-1}.

The prediction of β-sheet in the histone H1.3 molecule increased during aggregation from 22% to 31%, which confirms the hypothesis of β-aggregation.

3.1.5. Ligand binding

The interaction of phosphate with the histone H1.3 protein represents a good example for the usefulness of FTIR spectroscopy for the study of protein ligand interactions. Besides the 1615 cm^{-1} band, the simultaneous formation of the two other strong absorption bands at 1261 cm^{-1} and 1080 cm^{-1} in Fig. 4 is particularly noteworthy. Both bands are coinciding with the asymmetric and symmetric P═O stretching vibrations of the phosphate ion. The binding of the phosphate ion to the histone H1.3 molecule is probably an important feature of the interaction of DNA with this protein in the cell nucleus and has been shown to induce a change in the tertiary structure of the histone by circular dichroism [13]. The increase of the P═O double bond character during this binding process supports the suggestion of specific PO_4^{3-} binding sites in the histone molecule (Fig. 5).

The specific binding of the phosphate ion to the histone molecule leads to the formation of a defined P═O double bond, which could be then seen in the IR spectrum. This 'sharp' P═O double bond is different from the normal bonds of aqueous soluted phosphate ions at pH 7.5; the spectrum of this free soluted phosphate ions are eliminated by correcting for the buffer background. The question remains whether the P–O vibrations with more single bond character can be identified in the histone–phosphate complex. Since they would be expected in the range between 800 cm^{-1} and 900 cm^{-1}, they are not accessible with the CaF_2 windows of the used cuvette.

Table 4
Characterization of the time dependent spectra in Fig. 4 with the first three identity criteria.

No.	Time (min)	SDSC (%)	α-Helix (%)	β-Sheet (%)
1	0	97.8	27.8	22.4
2	120	96.2	27.3	25.2
3	140	58.2	22.5	31.1

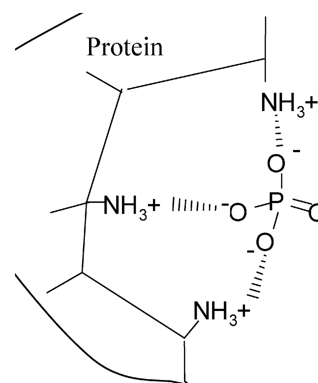

Fig. 5. Model for the increase of the P═O double bond character in phosphate ions by binding to a specific site in the ONCOHIST® molecule. In aqueous solution the P═O double bond of a dissociated phosphate ion is mesomericly distributed over the whole molecule. Spectroscopically it is therefore different from the P═O double bond of a phosphate ion bound to the histone H1. Three oxygen atoms bind in a fixed way to the protein, whereas the fourth oxygen is free with a localized double bond.

Our data show that MIR transmission spectroscopy can detect specific phosphate binding sites in proteins in general by observing the phosphate double bond vibrations. Besides the detection, the method bears furthermore the potential for quantitative binding studies and monitoring concomitant changes in the protein's secondary structure.

3.1.6. Detection of pH changes

The pH value of a biopharmaceutical formulation is a fundamental property, which has to be strictly controlled after the production and during stability studies. IR spectroscopy has the potential for the detection of pH fluctuations in the protein solution by using the protein itself as a pH indicator. Seven of the 20 normal amino acids possess side chains with ionisable groups; including the N- and C-terminus a protein can have up to nine different ionisable groups with pK values ranging from 3 to 13. These groups can be utilized as pH indicators in a protein IR spectrum. Eight of the 220 amino acids in histone H1.3 possess a carboxyl group, which corresponds to 3.6% of all amino acid residues; 66 lysine residues in histone H1.3 contain a basic amino group, which corresponds to 30% of all residues. Therefore in the case of ONCOHIST® the carboxylic and amino groups are of primary interest for monitoring pH fluctuations by IR spectroscopy.

Fig. 6 confirms that the decrease of pH from 8.7 to 2.6 changes the ONCOHIST® spectra in the expected manner: the three bands for the protonated form of the carboxyl group (–COOH) between 3000–2500 cm^{-1}, 1717–1680 cm^{-1} and at 1253 cm^{-1} are increasing by falling pH (Fig. 6). Simultaneously the two bands for the deprotonated form of the carboxyl group ($–COO^-$) between 1650–1550 cm^{-1} and at 1407 cm^{-1} are decreasing by falling pH (Fig. 6). Therefore it should be possible to define special pH marker bands for monitoring pH fluctuations in an ONCOHIST® solution. In Fig. 7 four of the most characteristic pH-dependent IR absorption bands are presented as a function of pH. All curves are showing a clear tendency of a regular change by pH variation.

By generating a suitable calibration data set it should be possible to measure the exact pH value of a protein solution in a

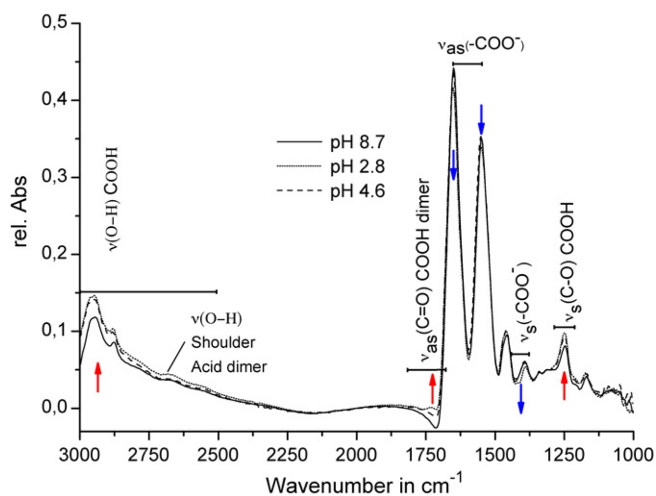

Fig. 6. FTIR spectra of 10 mg/ml ONCOHIST® at pH 8.7 (—), pH 4.6 (—) and pH 2.8 (. . .). All spectra were measured in 0.9% NaCl and normalized to the Amide-II band between 1600 cm^{-1} and 1500 cm^{-1} in the range between 3000 cm^{-1} and 1000 cm^{-1}.

volume as small as 20 μl. In case of limited protein amounts during process development and stability studies, this could be an interesting alternative to conventional pH measurement. The minimum amount of protein for pH measurement with such a method should be as small as ~20 μg.

3.2. Precision

3.2.1. Precision of ONCOHIST® assay

The applicability of the IR method for determining the concentration of ONCOHIST® in solution was tested by evaluating the repeatability for protein concentrations in the lower (1.0 mg/ml), middle (25.0 mg/ml) and upper (50.0 mg/ml) range. Every concentration was measured six times in succession. The protein concentration was determined on the basis of the Amide-I band height at 1651 cm^{-1}. The relative standard deviation (RSD) for the protein concentrations at 25 mg/ml and 50 mg/ml was always below 1%, which is comparable to e.g. a RP-HPLC-based protein assay. The higher RSD of 3.5% for the smaller concentrations at 1 mg/ml is also an acceptable value and still comparable to a RP-HPLC method. The intermediate precision was checked by measurements of an ONCOHIST® solution with a concentration of 7.4 mg/ml by two different persons on two different days. The determined ONCOHIST® concentrations lie between 7.3 mg/ml and 7.6 mg/ml, which means a maximal deviation of +2.7% between them.

The assay's precision of the IR method is high and in best accordance with GMP analytical requirements.

Table 5
Evaluation of 16 measurements of 10.0 mg/ml ONCOHIST® in 0.9% NaCl (measured on two different days by two different persons) with the identity criteria SDSC, α-helix and β-sheet prediction.

	SDSC	α-Helix	β-Sheet
Mean (%)	97.0	28.3	22.3
Standard deviation	0.8	0.3	0.2
RSD%	0.8	1.1	0.8
Minimum	95.5	27.7	22.1
Maximum	98.2	28.8	22.7

3.2.2. Precision of identity criteria

Repeatability and intermediate precision of the identity criteria were checked by measuring 16 IR spectra of 10 mg/ml ONCOHIST® in 0.9% NaCl by two different persons on two different days. Table 5 summarizes the results and shows, that the precision of identity criteria is high and free from day-to-day and person-to-person variations.

3.3. Linearity

Two independent calibration data sets were measured on two different days in the working range between 0.5 mg/ml and

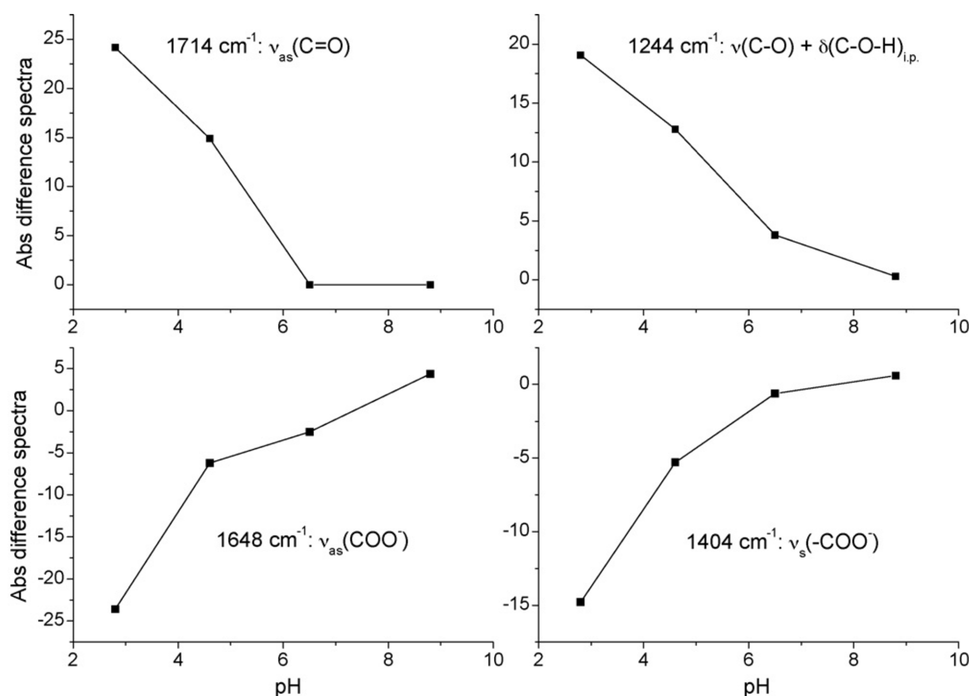

Fig. 7. The relative intensities of characteristic pH-dependent bands at 1714 cm^{-1}, 1244 cm^{-1}, 1648 cm^{-1} and 1404 cm^{-1} from the vector normalized IR difference spectra of ONCOHIST® solutions at pH 2.8, 4.6, 6.5 and 8.7 against an ONCOHIST® standard at pH 7.5 as subtrahend.

Table 6
Summary of information gain for biopharmaceutical quality control of MIR spectroscopy with the developed IR method.

No.	Information	Method	Remarks
1	Protein concentration	Amide-I band height	Specific method for quantifying proteins over a wide range of concentrations with high accuracy
2	Secondary structure	Chemometric Amide-I band analysis	Prediction of α-helix and β-sheet
3	Identity	Evaluation methods 1 to 4 (Table 1)	Secondary structure analysis and comparison with standard spectra
4	Purity	Evaluation methods 1–4	Secondary structure analysis and comparison with standard spectra
5	pH	Analysis of carbonic acid absorption bands	pH variation changes the vibrations in the carboxylic group
6	Aggregations	Monitoring the shoulder at 1615 cm^{-1}	β-aggregation deforms the Amide-I band with a significant shoulder at 1615 cm^{-1}
7	Phosphate binding	Monitoring ν_s(P=O) and ν_{as}(P=O) of the phosphate ion	In case of specific phosphate binding sites in the molecule the P=O double bond character in the phosphate ion changes, which is seen in IR spectrum

50.0 mg/ml ONCOHIST®. Each of the calibration data sets consisted of 20 independent values at 10 different concentrations distributed over the whole working range. Linear regression by the method of least squares resulted in two nearly identical linear equations ($y = 0.28 + 653.16x$ and $y = 0.11 + 649.69x$) with a correlation coefficient R of 0.9998 in both cases. The linearity is therefore in full accordance with GMP guidelines and furthermore free of day-to-day fluctuations.

3.4. Accuracy

The protein concentrations of two different GMP batches of the ONCOHIST® drug product were determined by quantitative amino acid analysis to be equal to 5.8 mg/ml and 6.7 mg/ml. The protein concentration of these two batches was determined with the FTIR transmission method to be equal to 6.1 mg/ml (+5.2%) and 6.6 mg/ml (−1.5%). These values for accuracy are in best accordance with usual biopharmaceutical requirements for protein quantity determination in a quality control lab.

3.5. Working range

The limit of detection (LOD) was determined by injection of 20 μl of 0.05 mg/ml ONCOHIST® (total protein 1 μg) in 0.9% NaCl two times and measuring the IR spectra. The mean value of the signal-to-noise ratio for these measurements was 8.7. According to the ICH-guideline Q2R1 a signal-to-noise ratio of 3 is sufficient for the LOD. The lower limit of quantitation (LLOQ) was determined by injection of 20 μl 0.5 mg/ml ONCOHIST® (total protein 10 μg) two times. The signal-to-noise ratio was determined to be 35.6; according to the ICH-guideline Q2R1 a value of 10 is sufficient for the LLOQ. The upper limit of quantitation was predominantly restricted by the viscosity of the protein solution, which must not to be too high for the reproducible and easy filling of the IR transmission cell. Own experiments have shown, that a protein concentration as high as 50.0 mg/ml is still possible to handle. Higher protein concentrations of ONCOHIST® have too high viscosities, which prolongs the filling procedure of the IR cuvette.

The whole range of the IR method between 0.5 mg/ml and 50.0 mg/ml spans two orders of magnitude. For the determination of protein concentration this is an acceptable range. Thus transmission IR spectroscopy offers a direct means for quantification comparable to UV spectroscopy. But, since it is based on the protein's peptide bonds, the quantification is not dependent on the nature of the amino acid side chains like in UV.

3.6. Robustness

The identity of the buffer, in which the protein is dissolved and the background buffer for calculation of absorption spectra is a prerequisite for measuring IR spectra in solution. A problem in the measurement of IR spectra of aqueous protein solutions arises from the fact that, the main absorption bands of water and the peptide bonds overlap to a considerably extent and that the difference between the raw detector signal for the water bands and the peptide bonds is very small. Furthermore is the intensity of the water absorption at ~1650 cm^{-1} dependent on the nature of the buffer and its composition. We have therefore tested to what extent small changes in the background buffer composition would influence the protein IR spectrum in ONCOHIST® solutions. The isotonic saline solution for infusion with a sodium chloride concentration of 154 mM, which was used in most of the measurements in this study, was modified between 150 mM and 160 mM NaCl. This modified background buffers corresponded up to 4.4% (w/w) variation, which was assumed as a realistic buffer variation during a GMP production or during stability studies. The changed buffers were used as background buffers for the measurement of a 5.0 mg/ml ONCOHIST® solution in 154 mM NaCl. The measured spectra were buffer-corrected as described under Section 2.3 before the calculation of the protein concentration. The obtained results were 5.2 mg/ml ONCOHIST® for 150 mM NaCl as background buffer, 5.0 mg/ml ONCOHIST® for 154 mM NaCl as background buffer and 5.2 mg/ml ONCOHIST® for 160 mM NaCl as background buffer. The deviations were equal to 4% in both cases and prove the method to be robust enough for the compensation of minor buffer variations encountered during GMP production or product storage.

4. Conclusions

The objective of this study was to explore the usefulness of transmission IR spectroscopy of protein solutions in biopharmaceutical analytics in the framework of regulatory constraints.

One important aspect of this method is the multitude of parameters which can be determined and the information which can be extracted from them. Equally important is the fact, that the individual applications can easily be validated. As we have shown by the analysis of the drug substance of ONCOHIST®, the specificity, precision, linearity, working range, accuracy and robustness are all in accordance with GMP requirements. Table 6 is summarizing the type of information which become accessible and the principles and techniques used herewith.

Acknowledgements

Technical assistance of Bettina Breier and Andreas Wruck in measuring IR spectra is greatly acknowledged. Authors thank Dr. Grazyna Formicka-Zeppezauer for her valuable suggestions and correction of the manuscript. Part of this work was supported by the Bundesministerium für Bildung und Forschung of Germany (BMBF).

References

[1] R. Rodriguez, T. Wehr, S. Tuck, Analytical Techniques For Biopharmaceutical Development, Marcel Dekker, New York, 2007.
[2] B.R. Singh, Infrared Analysis Of Peptides And Proteins, ACS, Washington DC, 2000.
[3] D.M. Byler, H. Susi, Examination of the secondary structure of proteins by deconvolved FTIR spectra, Biopolymers 25 (1986) 469–487.
[4] D. Marsh, Quantitation of secondary structure in ATR infrared spectroscopy, Biophys. J. 77 (1999) 2630–2637.
[5] K.A. Oberg, J.M. Ruysschaert, E. Goormaghtigh, The optimization of protein secondary structure determination with infrared and circular dichroism spectra, Eur. J Biochem. 271 (2004) 2937–2948.
[6] E. Goormaghtigh, J.M. Ruysschaert, V. Raussens, Evaluation of the information content in infrared spectra for protein secondary structure determination, Biophys. J. 90 (2006) 2946–2957.
[7] K. Rahmelow, W. Hubner, T. Ackermann, Infrared absorbances of protein side chains, Anal. Biochem. 257 (1998) 1–11.
[8] S. Krimm, Interpreting infrared spectra of peptides and proteins, in: B.R. Singh (Ed.), Infrared Analysis of Peptides and Proteins, ACS, Washington DC, 2000, pp. 38–53.
[9] W. Norde, C.A. Haynes, Reversibility and the mechanism of protein adsorption, in: T.A. Horbett, J.L. Brash (Eds.), Proteins at Interfaces II, ACS, Washington DC, 1995, pp. 26–40.
[10] M. Zeppezauer, G. Formicka-Zeppezauer, P.C. Gross, H. Joernvall, S. Schubert, M. Pfreundschuh, A Phase I/II dose escalation trial of ONCOHIST® (Recombinant Human histone H1.3) in patients with relapsed or refractory acute myeloid leukemia, Blood 110 (2007) 279a–280a.
[11] J.A. Hering, P.R. Innocent, P.I. Haris, Towards developing a protein infrared spectra databank (PISD) for proteomics research, Proteomics 4 (2004) 2310–2319.
[12] A.L. Fink, Protein aggregation: folding aggregates, inclusion bodies and amyloid, Fold. Des. 3 (1998) R9–23.
[13] L. de Petrocellis, G. Quagliarotti, L. Tomei, G. Geraci, Structuring of H1 histone. Evidence of high-affinity binding sites for phosphate ions, Eur. J. Biochem. 156 (1986) 143–148.