Methods 52 (2010) 23–37

Contents lists available at ScienceDirect

Methods

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

Quantitative methods for structural characterization of proteins based on deep UV resonance Raman spectroscopy

Victor A. Shashilov [a], Vitali Sikirzhytski [b], Ludmila A. Popova [b], Igor K. Lednev [b,*]

[a] Aegis Analytical Corporation, 1380 Forest Park Circle, Suite 200, Lafayette, CO 80026, USA
[b] Department of Chemistry, University at Albany, SUNY, 1400 Washington Ave., Albany, NY 12222, USA

ARTICLE INFO

Article history:
Accepted 19 May 2010
Available online 24 May 2010

Keywords:
Raman spectroscopy
Bayesian statistics
Partial least squares
Least squares support vector machines
Protein structure
Protein aggregation
Hydrogen–deuterium exchange

ABSTRACT

Here we report on novel quantitative approaches for protein structural characterization using deep UV resonance Raman (DUVRR) spectroscopy. Specifically, we propose a new method combining hydrogen–deuterium (HD) exchange and Bayesian source separation for extracting the DUVRR signatures of various structural elements of aggregated proteins including the cross-β core and unordered parts of amyloid fibrils. The proposed method is demonstrated using the set of DUVRR spectra of hen egg white lysozyme acquired at various stages of HD exchange. Prior information about the concentration matrix and the spectral features of the individual components was incorporated into the Bayesian equation to eliminate the ill-conditioning of the problem caused by 100% correlation of the concentration profiles of protonated and deuterated species. Secondary structure fractions obtained by partial least squares (PLS) and least squares support vector machines (LS-SVMs) were used as the initial guess for the Bayesian source separation. Advantages of the PLS and LS-SVMs methods over the classical least squares calibration (CLSC) are discussed and illustrated using the DUVRR data of the prion protein in its native and aggregated forms.

1. Introduction

1.1. Deep UV resonance Raman spectroscopy and protein structure characterization

Understanding the relationship between protein sequence and its structure and function is one of the central problems of structural biology in the post genome era [1]. Protein folding, misfolding and aggregation, amyloid fibril formation in particular, are of special interest due to the direct relevance to devastating neurodegenerative diseases [2,3]. Amyloid fibrils are straight, unbranched protein aggregates about 100 Å in diameter which are non-crystalline and insoluble in water. Despite a great interest of scientific community to amyloid fibrils, their molecular structure remains elusive [4]. This is because the classical tools of structural biology, conventional X-ray crystallography and solution nuclear magnetic resonance (NMR), exhibit serious limitations when applied to non-crystalline and insoluble amyloid fibrils [5]. Solid state NMR [6] and small-angle X-ray scattering [7,8] have been recently developed for structural characterization of amyloid fibrils, but the routine application of these techniques is yet to be established. X-ray diffraction provides atomic-level resolution for short (5–7 amino acid residues) peptide microcrystals [9,10], but not for fibrils formed from full-length proteins. Electron diffraction and scanning probe microscopy are valuable tools for fibril morphology studies, but these methods are not sensitive to the molecular structure.

Raman spectroscopy, deep UV resonance Raman (DUVRR) spectroscopy in particular, has been proven to be a valuable tool for protein structural characterization [11–16]. The Raman signature of the amide chromophore is sensitive to ψ and φ Ramachandran dihedral angles and thus provides direct quantitative information about the secondary structure of proteins [12]. Vibrational transitions within aromatic amino acid side chains are susceptible to the local environment and report on the protein tertiary structure [17]. In addition, the application of Raman spectroscopy is not limited because of the sample phase and the protein aggregation state. Specifically, the applicability of Raman spectroscopy to gelatinous and highly optically opaque samples make this method a uniquely valuable tool for amyloid fibril structural characterization [17–19].

1.2. DUVRR spectroscopy and statistical methods

Several research groups including Asher in Pittsburgh and Spiro in Princeton worked on the development of DUVRR basis spectra for secondary structure characterization of proteins [20,21]. A Raman spectroscopic signature of a pure secondary structural element in a globular protein can rarely be obtained since it is

* Corresponding author.
E-mail address: lednev@albany.edu (I.K. Lednev).

1046-2023/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2010.05.004

typically found in a combination with other structures. DUVRR spectra of complex protein systems could be decomposed into the spectra of individual structural elements by using chemometrics or source separation statistical methods [22]. The estimation of the protein secondary structure composition from DUVRR spectra requires the application of multivariate calibration methods. In this manuscript we propose a new Bayesian source separation algorithm for extracting the spectra of individual structural components of complex aggregated proteins, cross-β core and unordered parts of amyloid fibrils in particular. The proposed method utilizes hydrogen–deuterium (HD) exchange to reveal various structural elements of amyloid fibrils due to different solvent accessibility. Gradual HD exchange result in 100% correlation of the concentration profiles of protonated and deuterated species. To eliminate this ill-conditioning of the problem prior information about the concentration matrix and the spectral features of the individual components incorporated into the Bayesian equation. Specifically, the secondary structure composition obtained by partial least squares (PLS) and least squares support vector machines (LS-SVMs) provided the initials estimate for the Bayessian source separation. DUVRR spectroscopy combined with PLS and LS-SVMs methods were applied for structural characterization of prion protein (PrP) in its native and aggregated forms.

Until now, the accuracy and validity of calibration methods for predicting the protein secondary structure have been determined primarily by the quality and completeness of the set of basis Raman spectra [20,21]. The basis spectra have been obtained by using two alternative approaches: (i) measuring Raman spectra of model peptides or proteins with a predominant contribution from a single secondary structure [23] and (ii) calculating the basis spectra by using least squares regression (LSR) from the set of spectra recorded for proteins with known secondary structure [20,21].

Here we propose to use PLS and LS-SVMs multivariate calibration methods for protein secondary structure characterization that do not rely on the knowledge of basis spectra. Both algorithms were able to model non-linearity in spectral data and focus on relevant secondary structure features in DUVRR data disregarding uninformative yet intense spectral bands of aromatic amino acid side chains. The proposed approach is not limited to protein data, and can be readily extended to numerous quantitative Raman spectroscopic applications.

We have demonstrated previously [19] the application of various chemometric methods for quantitative characterization of protein structure evolution including the use of the alternating least squares (ALS) to model the kinetics of lysozyme structure transformations at the initial stages of fibrillation. All experimental Raman spectra of lysozyme recorded during the fibrillation lag phase were fitted with three pure component spectra, *i.e.* the spectra of nucleus β-sheet, partially unfolded intermediate calculated by independent component analysis (ICA), and the experimental spectrum of native lysozyme. A mixed soft–hard modeling approach [24] provided the refined DUVRR spectra of the β-sheet and partially unfolded intermediate, kinetic profiles for all three species and the characteristic times for each step of lysozyme structure transformation. We also applied ALS to fit the fluorescence, electrospray ionization mass spectrometry and DUVRR spectra of lysozyme at different stages of irreversible unfolding with two significant components [25]. The perfect two-component fit of all charge distribution envelopes of ESI–MS spectra served as a solid proof of the all-or-none mechanism of lysozyme partial unfolding. Shashilov et al. [19] reported on the first application of joint approximate diagonalization of Eigen-matrices (JADE) [26], second-order blind identification in the Fourier space data [27], and second-order nonstationary source separation [28] ICA algorithms for the analysis of protein structure evolution. The joint diagonalization ICA methods have been used to resolve sets of DUVRR spectra of lysozyme at the initial stages of fibrillation into pure spectra of native protein, partially unfolded intermediate and nucleus β-sheet [19]. The application of pure variable methods enabled us to resolve the spectra of newly formed fibrillar β-sheet and partially unfolded intermediates into separate components. To the best of our knowledge, that was the first report on the spectroscopic signature of the fibrillation nucleus [29].

1.3. Assignment of DUVRR spectra of proteins

Deep UV excitation (below 200 nm) resonantly enhances Raman scattering from the amide chromophore, a building block of the polypeptide backbone, exhibiting the $\pi\pi^*$ electronic transition at ~195 nm. As a result, a protein DUVRR spectrum is dominated by the amide bands with a small contribution of non-aromatic side chains. A high sensitivity of DUVRR spectra to the protein secondary structure is based on the complex nature of coupling between various amide vibrations. The Amide I mode consists of carbonyl C=O stretching [20,30], with smaller contributions from C–N stretching and N–H bending. Both Amide II and Amide III bands involve significant C–N stretching, N–H bending, and C–C stretching. The C_α–H bending vibrational mode involves C_α–H symmetric bending and C–C_α stretching [20,30]. Assignment of the major protein DUVRR bands are summarized in Table 1.

As seen from Table 1, Amide III and (C)–C_αH bending vibrational modes are the most sensitive Raman bands to the amide backbone conformation. Asher et al. [12] have recently demonstrated that this sensitivity results from the coupling of amide N–H and (C)C_αH bending motions. The magnitude of this coupling depends on the distance between N–H and C–H hydrogens which, in turn, depends on the geometry of peptide backbone characterized by the Ramachandran angle ψ. The vibrations are strongly mixed at ψ angles of ~120° corresponding to both the random coil and β-sheet conformations, but are almost completely uncoupled at ψ values approximately −60° for an α-helix conformation. In addition to a strong dependence on the ψ dihedral angle, the Amide III vibrational frequency shows a modest dependence on the ϕ dihedral angle [36,37]. Asher et al. [34] have developed a method for estimating the Ramachandran ψ-angular distributions from the Amide III deep UV Raman band shape.

The C_α–H bending mode is resonantly enhanced in DUVRR spectra of β-sheet and random coil and is absent in α-helix conformations. As shown by Asher et al. [12], upon excitation within the amide $\pi\pi^*$ transition, a strong enhancement of the C_α–H bending mode results from strong van der Waals interaction of (C)–C_α–H and N–H groups at ψ angles of ~120° corresponding to random coil and β-sheet-like conformations. The isolated C_α–H bending mode in the α-helical structure with the ψ angle of about −60° is not resonantly enhanced [12], and thus is not seen in the DUVRR spectrum of α-helix.

The shape and frequency (Table 1) of the Amide I band of α-helical, β-sheet and unordered structures are different and thereby the Amide I can be used for discriminating these structures. Specifically, an unordered protein gives rise to a broad Gaussian-shaped Amide I band while the rigid β-sheet structure produces a relatively narrow, Lorentzian-type Amide I bands [18,80]. Presumably, it is strong hydrogen bonding between C=O and N–H groups in α-helix-like structures that causes the downshift of the Amide I DUVRR band with the major contribution from the C=O stretching vibration. The Amide II Raman band is strongly enhanced in the case of unordered and β-sheet conformations [38] at deep UV excitation. The Amide II vibrational mode frequency shows almost no response to changes in the protein secondary structure and is responsive to the hydrogen bonding of –N–H group of the polypeptide backbone [39].