Analytica Chimica Acta 892 (2015) 59–68

Contents lists available at ScienceDirect

Analytica Chimica Acta

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

"Parallel factor analysis of multi-excitation ultraviolet resonance Raman spectra for protein secondary structure determination"

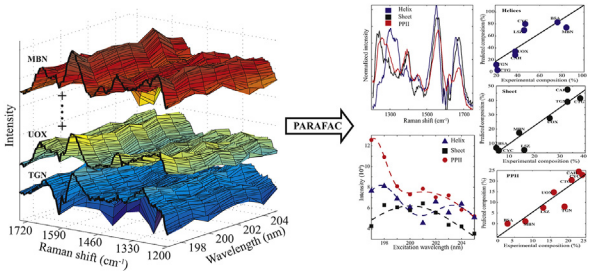
Olayinka O. Oshokoya, Renee D. JiJi*

University of Missouri, Department of Chemistry, 601 S. College Avenue, Columbia, MO 65211, USA

HIGHLIGHTS

- PARAFAC analysis of multi-excitation ultraviolet resonance Raman (ME-UVRR) spectra.
- *A priori* resolution of pure secondary structure Raman spectral profiles.
- Correlation of compositional profiles with distribution of dihedral angles.
- Improved quantification of PPII-type structure.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 19 June 2015
Accepted 11 August 2015
Available online 28 August 2015

Keywords:
Parallel factor analysis
Ultraviolet resonance Raman
Protein secondary structure
Multiway
Trilinear

ABSTRACT

Protein secondary structural analysis is important for understanding the relationship between protein structure and function, or more importantly how changes in structure relate to loss of function. The structurally sensitive protein vibrational modes (amide I, II, III and S) in deep-ultraviolet resonance Raman (DUVRR) spectra resulting from the backbone C—O and N—H vibrations make DUVRR a potentially powerful tool for studying secondary structure changes. Experimental studies reveal that the position and intensity of the four amide modes in DUVRR spectra of proteins are largely correlated with the varying fractions of α-helix, β-sheet and disordered structural content of proteins. Employing multivariate calibration methods and DUVRR spectra of globular proteins with varying structural compositions, the secondary structure of a protein with unknown structure can be predicted. A disadvantage of multivariate calibration methods is the requirement of known concentration or spectral profiles. Second-order curve resolution methods, such as parallel factor analysis (PARAFAC), do not have such a requirement due to the "second-order advantage." An exceptional feature of DUVRR spectroscopy is that DUVRR spectra are linearly dependent on both excitation wavelength and secondary structure composition. Thus, higher order data can be created by combining protein DUVRR spectra of several proteins collected at multiple excitation wavelengths to give multi-excitation ultraviolet resonance Raman data (ME-UVRR). PARAFAC has been used to analyze ME-UVRR data of nine proteins to resolve the pure spectral, excitation and compositional profiles. A three factor model with non-negativity constraints produced three unique factors that were correlated with the relative abundance of helical, β-sheet and poly-proline II dihedral angles. This is the first empirical evidence that the typically resolved "disordered" spectrum represents the better defined poly-proline II type structure.

* Corresponding author.
E-mail address: jijir@missouri.edu (R.D. JiJi).

http://dx.doi.org/10.1016/j.aca.2015.08.035
0003-2670/

1. Introduction

The relationship between protein structure and function have made determination and monitoring of protein secondary structure an area of great importance in biochemical and biophysical research. This increased interest in resolving and quantifying protein secondary structure content also stems from the observation that secondary structure changes without changes in the primary structure are involved in some protein based diseases [1–5].

Protein structure has four levels and a change at any level may result in changes in protein function. The primary structure of a protein is the amino acid sequence while the secondary structure refers to the structures within the protein that are defined by the phi (φ) and psi (ψ) dihedral angles of the amide backbone (Fig. 1). The three dimensional arrangement of these secondary structure motifs is the tertiary structure and finally the arrangement of protein tertiary subunits to each other in larger complexes that function as a single unit is the quaternary structure [6,7]. Because of the scope of this article, discussion would be limited to secondary structure.

Protein secondary structure is determined by the presence of specific hydrogen bonds as well as the set of dihedral angles (φ, ψ) which define spatial orientation of the peptide backbone. When these backbone dihedral angles have repeating values, the peptide forms regular secondary structure. Of the different secondary structures, the helical [8] and β-sheet [8,9] structures are the most common with dihedral angles corresponding to [α- (−60°, 45°), 3_{10}- (−49°, −26°) and π- (−57°, −69°)] and [parallel (−139°, 135°) and anti-parallel (−119°, 113)] respectively. Another category in protein secondary structure is the unfolded conformation, in which each amino acid in the backbone randomly samples all sterically-allowed φ and ψ angles [10]. Recent evidence [11] suggests that these seemingly disordered proteins could have local regions of order that adopt turn-like conformations or those of poly-proline II [12] (PPII, a left handed 3_1-helix formed by trans-L-poly-proline). Turns in a protein [13] (where the protein reverses its general direction) are non-repetitive and occur over only a few residues but are abundant in protein structures and have conformation flexibilities which are much greater than helices and β-sheets.

Traditional methods for protein secondary structure quantification such as x-ray crystallography [14,15], nuclear magnetic resonance [16,17], and circular dichroism [18,19] are now complimented by vibrational methods like IR, conventional and resonance Raman [20–27]. Deep-ultraviolet resonance Raman (DUVRR) particularly, has proven useful for quantification due to the structural sensitivity of the observed backbone amide modes. Some advantages of DUVRR include increased signal intensity (10^2–10^6 times) versus conventional Raman spectroscopy, minimal contribution from solvent water bands, elimination of background fluorescence and selective enhancement of the peptide backbone modes derived from the various vibrations of the backbone amide group (−CO−NH−) as a result of the π_2 to π_3^* dipole-allowed transition [26,28–32]. DUVRR sensitivity to protein secondary structure is observed in the shifting and intensity changes of the four observable amide modes [33–38]. The position and intensity of the four amide modes: amide I, II, III and S are dependent upon the secondary structure of the protein with their relative contributions being proportional to the relative amount of each secondary structure conformation. Similarly, each secondary structure type exhibits distinct absorption profiles leading to a compositional dependence in resonance enhancement versus excitation wavelength [29]. Therefore, the position and intensity of amide modes change with varying excitation wavelengths and secondary structure composition [39].

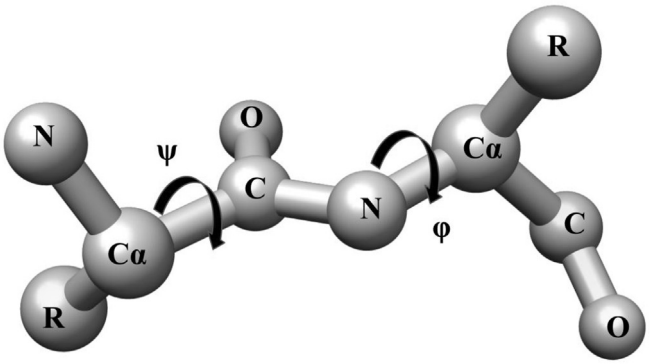


Fig. 1. Peptide backbone of protein showing phi (φ) and psi (ψ) dihedral angle.

Initially, quantification studies of protein secondary structure employing DUVRR focused on univariate calibration methods of single amide modes [26], but have evolved to include multivariate calibration and multivariate curve resolution methods and other advanced statistical analyses of all observable amide modes [22,35,36,40–45]. In these methods, first order data are decomposed using bilinear models showing that individual protein spectra, **x**, are a linear combination of the different underlying pure secondary structure motifs, **s**, and their respective fractional amounts, *c* (Equation (1));

$$\mathbf{x} = c_\alpha \mathbf{s}_\alpha + c_\beta \mathbf{s}_\beta + \cdots \quad (1)$$

where α designates α-helical and β designates β-sheet related variables. Other structures that have been defined include α-L, turns, random coil or disordered and 3_{10} helices.

Spectra from different proteins with different secondary structure compositions can be combined into a matrix and decomposed to give a matrix of known secondary structural content of each protein and a matrix of underlying pure secondary. The pure underlying secondary structure spectra may then be calculated via multivariate least square regression which in turn can be used to determine the structural content of a new protein.

The "known" concentration profiles are typically determined from proteins whose atomic level structure has been characterized by X-ray crystallography or NMR spectroscopy and published in the protein database, PDB (www.rcsb.org) [46]. However, the conditions under which the structures were determined do not necessarily represent the conditions under which the spectroscopic measurements are being conducted. This is especially true of crystal structures, which are static in nature. "Known" spectral profiles can be estimated from model polypeptides, such as poly-L-lysine or poly-L-glutamic acid, that adopt "pure" secondary structures with varying environments [40]. Given that these predetermined or "known" inputs are not taken from the proteins themselves in their relevant environments, they can themselves introduce errors or biases into the multivariate analysis.

Trilinear methods are not prone to the rotational ambiguity of bilinear methods and reduce or eliminate the need for good initial estimates of either the **C** or **S** matrices. Multi-excitation UVRR (ME-UVRR) spectra of multiple proteins results in trilinear data that can be modeled with higher order algorithms such as parallel factor analysis (PARAFAC). This is particularly important for proteins whose disordered regions are not well characterized by NMR or X-ray crystallography but appear to have increasingly important functional significance [47–51]. Using a combination of ME-UVRR of nine globular proteins and PARAFAC, the pure spectral profiles of folded, helical and β-sheet structures are resolved along with the spectral profiles corresponding to the disordered portions of the