UNIVERSITÄT
DUISBURG
ESSEN

Label-free and site-specific detection of protein recognition by supramolecular ligands in solution using ultraviolet resonance Raman spectroscopy

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften – Dr. rer. nat. –

vorgelegt von

Banafshe Zakeri

Institut für Physikalische Chemie
der
Universität Duisburg-Essen

Essen, 2018

Die vorliegende Arbeit entstand in der Zeit vom February 2015 bis December 2017 im Fachbereich Chemie der Universität Duisburg-Essen, am Institut für Physikalische Chemie unter der Leitung von Herrn Prof. Dr. Sebastian Schlücker.

1. Gutachter: Prof. Dr. Sebastian Schlücker
2. Gutachter: Prof. Dr. Carsten Schmuck

Vorsitzende: Prof. Dr. Maik Walpuski

Datum der disputation: 21.09.2018

"The important thing is not to stop questioning."
Albert Einstein

Thesis Abstract

In this thesis, ultraviolet resonance Raman (UVRR) spectroscopy was successfully applied for label-free monitoring of molecular recognition of two proteins dermcidin and leucine zipper by multivalent supramolecular ligands. The ligand molecules contain guanidinium carbonyl pyrrole (GCP), a binding pocket which can be selectively enhanced by UV excitation wavelength. Therefore, the binding between two large systems can be monitored by probing this motif. Preliminary pH and concentration-dependent UVRR experiments were performed in order to find the optimum pH and concentration of ligand, respectively. We performed the main binding studies with a constant ligand concentration to keep the UVRR spectrum of chromophore as a signature for evaluating the spectral changes upon protein addition. In addition to GCP-based supramolecular ligands and with the purpose of extending the UVRR binding studies to a different class of supramolecular ligands, we performed UVRR experiments to evaluate the binding between molecular tweezers and a tripeptide containing lysine moiety. The initial results are promising for future binding studies with molecular tweezers.

Two multivariate data analysis methods, NMF and MCR-ALS, were applied for analyzing the pH dependent and UVRR binding studies, with the main advantages of using the whole available spectral information and without any *a priori* information of the system. The two methods were used for the analysis of the binding study with a three armed GCP ligand and the protein leucine zipper in order to determine the concentration profiles and the spectral contributions from both free and complexed ligand. Data analysis with MCR/ALS was performed with the assumption of multisteps equilibria and the qualitative analysis of the calculated concentration profiles resulted in the specification of an average binding constant and also two stepwise binding constants using a defined sigmoid function and multi-steps equilibria equations, respectively.

Zusammenfassung der Dissertation

In dieser Doktorarbeit wurde die UV-Resonanz-Raman (UVRR)-Spektroskopie als markierungsfreie Technik für die Verfolgung der Bindung von supramolekularen multivalenten Liganden mit zwei verschiedenen Proteinen, Dermcidin und Leucin-Zipper, erfolgreich eingesetzt. Die verwendeten multivalenten Ligandenmoleküle besitzen eine Guanidiniocarbonylpyrrol-basierte Bindetasche, deren Raman-Siganl selektiv mit UV-Anregung verstärkt werden kann. Folglich kann die Bindung zwischen supramolekularem Ligand und Protein mittels UVRR-Spektroskopie über die selektive Erfassung des GCP-Motivs verfolgt werden. Vorläufige pH- und Konzentrations-abhängige UVRR-Experimente wurden durchgeführt, um den optimalen pH-Bereich und die optimale Ligandenkonzentration zu ermitteln. Die UVRR-Bindungsstudien wurden mit konstanter Ligandenkonzentration durchgeführt, damit das UVRR-Spektrum des Chromophors als eine Signatur für die Auswertung der spektralen Veränderungen nach Zugabe des Proteins verwendet werden kann. Zusätzlich zu den GCP-basierten supramolekularen Liganden und der Absicht, die UVRR-Bindungsstudien für andere Klassen an supramolekularen Liganden zu erweitern, wurden auch UVRR-Experimente mit molekularen Pinzetten und einem Tripeptid mit einer Lysin-Einheit durchgeführt. Die ersten Resultate sind vielversprechend für weiterführende Bindungsstudien an molekularen Pinzetten.

Zwei multivariate Verfahren, NMF und MCR-ALS, wurden zur Analyse der pH-abhängigen UVRR-Experimente und UVRR-Bindungsstudien verwendet. Die oben genannten Methoden erlauben es, die komplette verfügbare spektrale Information zu verwenden ohne a priori Wissen über das zugrundeliegende System. Beide Verfahren wurden in den Bindungsstudien eines drei-armigen GCP-basierten Liganden mit dem Protein Leucin-Zipper für die Bestimmung des Konzentrationsprofils und der Reinspektren der beiden Komponenten ("freie" versus "komplexierte" GCP-Einheit) eingesetzt. Die Datenanalyse wurde unter Annahme eines mehrstufigen Gleichgewichts durchgeführt. Eine qualitative Analyse der berechneten Konzentrationsprofile erlaubt die Bestimmung einer mittleren Bindungskonstante mittels einer definierten Sigmoidfunktion und zweier stufenweiser Bindungskonstanten mit Hilfe mehrschrittiger Gleichgewichtsgleichungen.

Contents

Chapter 1

Introduction

Molecular recognition in biological systems, which relies on the existence of specific attractive interactions between two partner molecules [1], has opened up an interesting area of collaboration between biochemists and other research specialists, such as spectroscopists, for the benefit of fundamental research and drug discovery. Since the Nobel Prize was awarded to the scientists for the development of supramolecular chemistry[1], designing artificial host molecules for specific interactions with proteins has greatly grown. The synthesized molecules are nowadays applied by biologists to biological systems for protein recognition and functional modulation. For optimizing the interactions, an approximation of affinity contributions of attractive interactions as well as the knowledge about the interaction geometries are required. This information can be gleaned from X-ray crystallography and NMR spectroscopy complemented by computational studies. In addition, vibrational spectroscopy combined with multivariate analysis methods can accompany other techniques for a deeper understanding of protein-protein interactions and, as an ultimate goal, to optimize the specific interactions. One good example of such a collaboration between different research groups is CRC[2] "supramolecular chemistry on proteins" (https://www.uni-due.de/crc1093/) and in this thesis we describe a part of this collaboration including establishing UVRR spectroscopy as a new method in this context.

Development of charged host molecules is still the subject of many research field . In supramolecular chemistry, since both factors of selectivity and strength are required for molecular recognition, therefore a successful host molecule exhibits a strong affinity for one particular guest molecule and a much lower affinity for others. The selectivity itself is governed by an enormous number of factors such as complementarity, solvation, size and shape effects, which makes it very complicated to

[1] Lehn, Pederson and Cram

[2] Collaborative Research Center 1093

design a highly selective synthetic host for a given charged residue [2]. Moreover, due to some of the intrinsic properties of anions, the application of anion host is made much more difficult than cation receptors. For example, many anions exists in a relatively narrow pH window and receptor needs to be fully protonated in the pH region in which the anion is present in desired form. Also, anion hosts must compete more effectively with the surrounding medium because anions have high free energy of solvation [2–4]. Guanidinium carbonyl pyrrole (GCP) is an anion host introduced by Schmuck and co-workers for a specific binding with carboxylate anions. While a large number of host systems can function in hydrophobic solvents, a well designed combination of electrostatic interaction and hydrogen bonding in GCP motif can provide the necessary binding energy even in polar solvents [5–7]. The polarity of surrounding solvent which weaken the strength of hydrogen bond and electrostatic interaction is a big challenge also for cation receptors. Molecular tweezers developed by Klärner and Schrader is a receptor for cationic residues lysine and arginine which successfully overcame the polarity of solvent in the physiological condition by a special combination of hydrophobic and electrostatic interactions [8]. These phosphate tweezers are made with anionic phosphate or phosphonate groups on the central benzene ring feature, an electron-rich and torus-shaped cavity. By a unique threading mechanism, the positively charged side chains of these amino acids are drawn into the tweezers' tailored cavity and locked by the salt bridge formation with phosphate/phosphonate [9, 10]. Both receptors are used for protein recognition in the context of CRC as two different classes of artificial supramolecular ligand.

From the early step of developing GCP-based host molecule for peptide recognition [11], UVRR spectroscopy was promised to be a good partner technique for monitoring molecular recognition of GCP motif [12]. A series of experiments was performed by UVRR spectroscopy in order to evaluate the complexation between a small GCP receptor and a tetrapeptide [13–15]. The results, analyzed by computational chemistry and multivariate data analysis, provided fundamental principles of binding studies between two partner molecules by UVRR spectroscopy [16]. Since UVRR spectroscopy was rarely used for binding studies, it was an important step for establishing a technique which monitor the vibrational change of the GCP motif to quantify its molecular recognition. Along with the development of GCP tailor made supramolecular ligands, the motivation of this work was demonstrating the potential of UVRR spectroscopy for binding studies in supramolecular chemistry, explaining the whole procedure from establishment of new setup to experimental part and data analysis. In addition, the first UVRR spectrum of molecular

tweezers is presented here.

The organization of the thesis is divided into two parts. The first reviews the whole project by providing the knowledge which is needed to bridge chemistry and spectroscopy. The second part deals with the experimental aspects of UVRR spectroscopy and the potential of multivariate data analysis methods for describing the behavior of the system when no *a priori* information is available about the procedure of binding. The first part is composed of chapters 2 and 3. In addition to the theoretical background of UVRR spectroscopy and multivariate data analysis, chapter 2 also contains a very brief history about the evolution of artificial supramolecular ligand for molecular recognition three of which will be discussed later on in more detail in chapter 3 as the main motivations for this project. The second part of thesis, including chapters 4, 5 and 6 is the main body of this project in which the experimental and analytical aspects of binding studies are described. Partner molecules including ligand and receptors as well as optical spectroscopy and data analysis methods applied in this thesis are mentioned in chapter 4. The experiments are explained in more detail in the first part of chapter 5. Then, the experimental spectra are analyzed by data analysis methods qualitatively and quantitatively. As well as these main features, the nonlinearity of Raman intensity versus concentration was practically used to determine the optimum range of concentration for our binding studies. The chapter also consists of molecular docking performed in computational biochemistry group (CRC subproject A8). Chapter 5 is closed with a section discussing the experimental results of molecular tweezers. Chapter 6 starts with the comparison between the used molecules for binding studies, two proteins on one hand and two ligand molecules on the other hand, and also two multivariate data analysis methods. These materials and methods were compared from the view of final results of our binding studies in order to conclude the main aspects and have an outlook over the whole project. Finally, the summary of the thesis is presented in chapter 7.

Generally, different chemical, experimental and analytical principles and techniques are brought together in order to solve a problem in supramolecular chemistry (binding studies of molecular recognition). However, it should be reminded that there are always other possibilities of more complicated methods and techniques which would result in more certain analysis of binding studies, e.g., the information from the crystal structure of the complex could be used in final analysis if it was available. However, it can be said that the guidance of Albert Einstein to "make everything as simple as possible but not simpler" was followed in our work.

Chapter 2

Background information and basic concepts

2.1 Supramolecular ligands for peptide recognition

Molecular recognition is a very fundamental process. It may be said that without molecular recognition, there would be no life on this world. Enzymes, antibodies, membranes, and their receptors, carriers, and channels all use the phenomenon of molecular recognition. According to the concept of "lock and key" proposed by Fischer in 1894, molecular recognition images the complementarity of a molecular receptor (lock) and the substrate (key) that is recognized to give a defined receptor-substrate complex. Then, molecular recognition represented a new area in chemistry called "supramolecular chemistry", the chemistry beyond the molecule, where non-covalent bonds and special fit between molecular individuals, which form a specific host-guest complex, are in the foreground [1]. Designing artificial host molecules is an important part of molecular recognition, the main goal of which is the achievement of selective binding with a given target. In particular, supramolecular host molecules for peptide recognition are powerful tools for investigating the principles of protein-protein interactions (PPIs). In this way, the fundamental principles of molecular recognition in biological processes would be well understood [17]. This context has been the central point by which people in CRC started their collaboration.

Quite often protein function and signaling is localized on the hot spots which involve only a few amino acids or sequences of amino acids. Therefore, artificial binders specific for single amino acids, represent a promising tool for protein targeting with potential applications in mechanistic elucidation, diagnostics and disease-modifying therapy [18]. One effective class of host molecule for the selective binding of carboxylate

is guanidinium carbonyl pyrrole (GCP) which was firstly designed by Schmuck and coworkers to improve the binding affinity of guanidinium cation by addition of hydrogen bonding donors. As shown in Figure 2.1, the

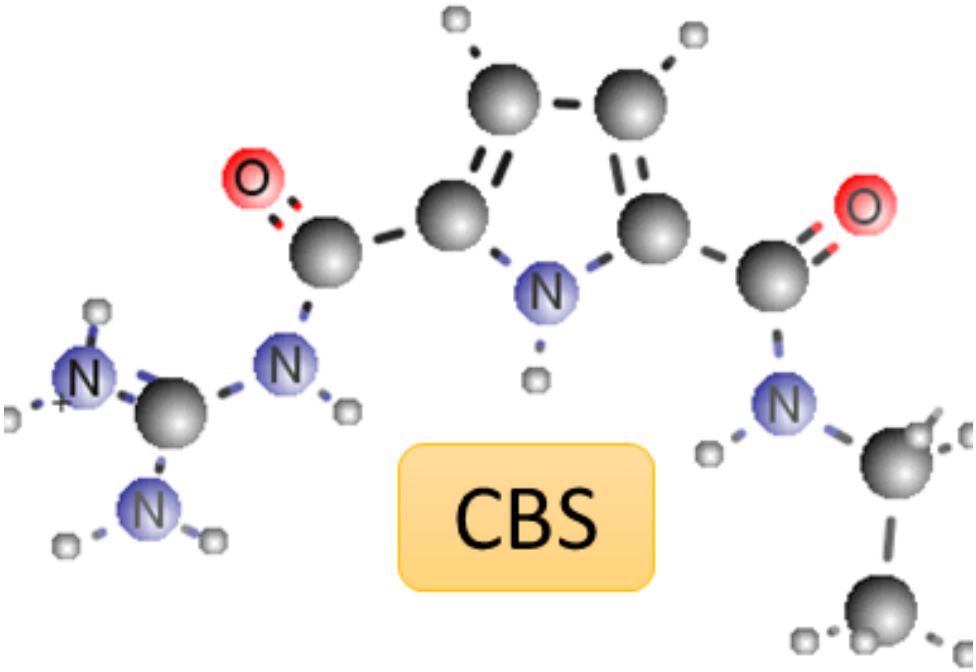


Figure 2.1: The GCP cation; a tailor-made carboxylate binding site.

GCP motif can adapt conformation with all NHs pointing inwards forming the carboxylate binding site (CBS). The polarity of the solvent decreases the mutual repulsion between NHs helping this suitable conformation of the GCP motif for the binding [5]. However, according to molecular dynamics, the molecule is non planar since the pyrrole ring is tilted relative to the plane of the guanidinium cation. Therefore, upon complexation with carboxylate, GCP has to undergo unfavorable conformational changes [19]. The selectivity of the recognition process was another challenging task to achieve. In peptide recognition, Schmuck and co-workers could achieve this selectivity by appropriate selection of additional amino-acids attached to GCP motif which can selectively bind to the side chains of target peptides [20]. They used this principle for designing host molecules for a selective binding to the C-termini of Amyloid [21] or in a sequence-dependent binding of dipeptides [22]. Along with the development of GCP-based receptors for accomplishing the molecular recognition with high affinity and selectivity in physiological condition, the characteristic modeling of GCP moiety revealed some special properties of this motif. For example, an anion switchable self-aggregation of GCP in DMSO was presented [23], or it was shown that this small and flexible host molecule can present a remarkable cooperativity [24]. All of this information, achieved in different research works, together with applying the dynamic combinatorial library

[25] provided a valuable base knowledge, which was then used for extending the approach from peptide recognition to protein recognition.
Molecular tweezers, introduced by Klärner and Schradar [9], is another promising supramolecular ligand in CRC for protein recognition. Comparing to GCP, molecular tweezers has a very different structure, though it also uses a combination of noncovalent interactions for stable host-guest complexes with various guest molecules containing either lysine or arginine moieties. As shown in Figure 2.2, molecular tweezers features an electron-rich and torus-shaped cavity with anionic groups (e.g., phosphate) on the central benzene ring. The anionic groups are responsible for water solubility of molecular tweezers and also essential for locking ammonium or guanidinuim cations of basic amino acids by an internal ion pair. The inclusion of guest molecule into the cavity is guided by π-cation stabilization and the substantial dispersive forces [10, 26]. Understanding

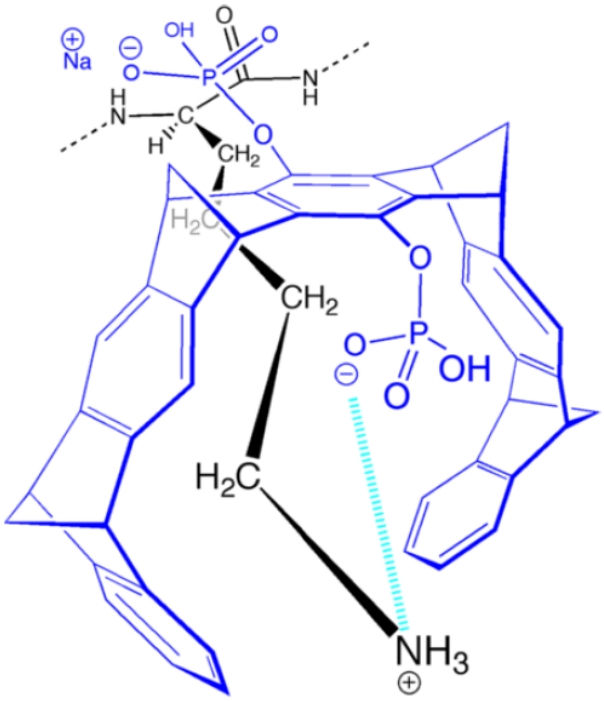


Figure 2.2: Structure of phosphate tweezers and schematic representation of its interaction with Lysine, [27].

the contribution of each interaction was demanding for development of molecular tweezers into more powerful and selective receptors. Some efforts were made by comparative studies of different anions [28] and linkers [29]. The latter caused a new approach to design unsymmetrical tweezers which carry additional recognition site and hence become peptide- or protein-specific. In general, these studies provided deep insight for understanding the exciting influence molecular tweezers have on the aggregation of proteins inhibition [26, 30, 31], modulating the protein-protein interactions [30] and the activity of enzymes [32].

2.2 UV Resonance Raman spectroscopy

Raman spectroscopy is a fundamental form of molecular spectroscopy which has been gradually developed in instrumentation from its discovery by Raman in 1928 [33]. Raman scattering is used to obtain information about the structure and properties of the molecules from their vibrational transitions [34]. Since the physiological functions of biological macromolecules are determined by the structural organization, the elucidation of the structure-function relationship in macromolecules is a challenging task which is usually beyond the resolution of the classical methods, i.e., X-ray crystallography and NMR spectroscopy. For instance, hydrogen bonding or electrostatic intermolecular interactions, which are essential for biochemical and biophysical processes can only be assumed but not determined by X-ray crystallography. NMR spectroscopy technique can be an alternative, but size limitations impose severe constraints. The current and future contributions of vibrational spectroscopy to this field can overcome these limitations [35].
The property involved in Raman spectroscopy is the change in the polarizability of the molecule with respect to its vibrational motion. When a molecule is placed in an electric field (laser beam), it suffers distortion since the positively charged nuclei are attracted toward the positive pole. This charged separation produces an induced dipole moment and the radiation emitted by this induced dipole moment contains the observed Raman scattering. The light scattered by the induced dipole of the molecule consists of both Rayleigh scattering and Raman scattering. Rayleigh scattering corresponds to the light scattered at the frequency of the incident radiation, whereas the Raman radiation is shifted in frequency, and hence energy, from the frequency of the incident radiation by the vibrational energy that is gained or lost in the molecule. The polarizability is a tensor with two Cartesian components; one is associated with the incident photon and the other with the scattered photon. The two photons are connected by a single quantum mechanical process that makes Rayleigh and Raman scattering different from the two one-photon event of absorption followed by emission. Though Raman scattered radiation is weak, using the resonance Raman effect, it is possible to selectively enhance vibrations of a particular chromophoric group in the molecule. It occurs when the exciting line is chosen so that its energy intercepts the manifold of an electronic excited state [36, 37]. Mathematically, Raman scattering can be described by either classical theory or according to quantum mechanics. Here, we briefly mention both aspects.

Classical theory of Raman scattering

According to classical theory, Raman scattering can be explained as follows: The electric field strength (E) of the electromagnetic wave (laser beam) fluctuate with time (t) as shown by the following equation,

$$E = E_0 cos(2\pi \upsilon_0 t) \tag{2.1}$$

where E_0 is the vibrational amplitude and ν_0 is the frequency of the laser. If a diatomic molecule is irradiated by this light, an electric dipole moment P is induced:

$$P = \alpha E = \alpha E_0 cos(2\pi \upsilon_0 t) \tag{2.2}$$

Here, α is the proportionality constant and is called polarizability. If the molecule is vibrating with the frequency ν_m, the nuclear displacement q is written

$$q = q_0 cos(2\pi \upsilon_m t) \tag{2.3}$$

where q_0 is the vibrational amplitude. For a small amplitude of vibration, α is approximately a linear function of q. Thus, we can write

$$\alpha = \alpha_0 + (\partial\alpha/\partial q)_0 q_0 \tag{2.4}$$

Here, α_0 is the polarizability at the equilibrium position and $(\partial\alpha/\partial q)_0$ is the rate of change of α with respect to the change in q, evaluated at the equilibrium position. Combining the last three equations, we obtain

$$P = \alpha_0 E_0 cos(2\pi\nu_0 t) + \frac{1}{2}(\partial\alpha/\partial q)_0 q_0 E_0 [cos2\pi(\nu_0 + \nu_m)t + cos2\pi(\nu_0 - \nu_m)t] \tag{2.5}$$

According to classical theory, the first term represents an oscillating dipole that radiates light of frequency ν_0 (Rayleigh scattering), while the second term corresponds to the Raman scattering of frequency $\nu_0 + \nu_m$ (anti-Stokes) and ν_0 - ν_m (Stokes). If $(\partial\alpha/\partial q)_0$ is zero, the vibration is not Raman-active. Namely, to be Raman-active, the rate of change of polarizability (α) with the vibration must not be zero.
In actual molecule, however, such a simple relationship between dipole moment and electric field of laser radiation (P = αE) does not hold since both P and E are vectors consisting of three components in the x, y and z direction, so that Equation 2.2 can be rewritten as follows

$$\begin{bmatrix} P_x \\ P_y \\ p_z \end{bmatrix} = \begin{bmatrix} \alpha_{xx} & \alpha_{xy} & \alpha_{xz} \\ \alpha_{yx} & \alpha_{yy} & \alpha_{yz} \\ \alpha_{zx} & \alpha_{zy} & \alpha_{zz} \end{bmatrix} \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix} \tag{2.6}$$

The first matrix on the right-hand side is called polarizability tensor. In normal Raman scattering, this tensor is symmetric: $\alpha_{xy} = \alpha_{yx}$, $\alpha_{xz} = \alpha_{zx}$ and $\alpha_{yz} = \alpha_{zy}$. The vibration is Raman-active if one of these components of the polarizability tensor is changed during the vibration [36].

Quantum mechanical theory of Raman scattering

In quantum mechanics, the vibration of a diatomic molecule can be treated as a motion of a single particle having mass μ. So, the quantum-mechanical frequency of the vibration is exactly the same as the classical frequency, $v_0 = \frac{1}{2\pi}\sqrt{\frac{K}{\mu}}$ where K is the force constant. However, the quantum vibrational energy is quantized by the vibrational quantum number (ν=0, 1, 2, ...) as follows

$$E_\nu = hv_0(\nu + \frac{1}{2}) \tag{2.7}$$

Therefore, quantum mechanically, the energy can change only in units of $h\nu$ and there is a zero point energy of $E = \frac{1}{2}hv_0$ for the lowest energy state ($\nu = 0$), while the energy of a such a vibrator change continuously in classical mechanics and it is zero when q is zero. According to quantum mechanics, the change in the polarizability before and after transition is determined by the integrals

$$[\alpha_{xx}]_{\nu',\nu''} = \int \psi^*_{\nu'}(Q_a)\alpha_{xx}\psi_{\nu''}(Q_a)dQ_a \tag{2.8}$$

where α_{xx} is one component of the polarizability tensor in Equation 2.6, $\nu^{'}$ and $\nu^{''}$ are respectively the vibrational quantum numbers before and after the transition, $\psi_{\nu'}$ and $\psi_{\nu''}$ are vibrational wavefunctions and Q_a is the normal coordinate of the normal mode, a. If one of six integrals calculated for polarizability tensor components of α_{xx}, α_{yy}, α_{zz}, α_{xy}, α_{xz} and α_{yz} is nonzero, this vibration is Raman-active. If all the integrals are zero, the vibration is Raman-inactive.
Resonance Raman (RR) scattering occurs when the sample is irradiated with an exciting line whose energy corresponds to that of an electronic transition of a particular chromophoric group in the molecule. Under these conditions, the intensities of Raman bands originating from this chromophore are selectively enhanced by a factor of 10^3 to 10^5. This selectivity is important for identifying vibrations of this particular chromophore. Theoretically, the intensity of a Raman band observed at $\nu_0 - \nu_{mn}$ is given by

$$I_{mn} = constant \cdot I_0 \cdot (v_0 - v_{mn})^4 \sum_{\rho\sigma} \mid (\alpha_{\rho\sigma})_{mn} \mid^2 \tag{2.9}$$

Here, I_0 is the intensity of incident laser beam with the frequency ν_0, m and n denote the initial and final states, respectively, of the electronic ground state and $(\alpha_{\rho\sigma})_{mn}$ represents the change in polarizability α caused by the transition of $m \rightarrow e \rightarrow n$ and ρ and σ are x, y and z components of the

polarizability tensor with the following formulation

$$(\alpha_{\rho\sigma})_{mn} = \frac{1}{h}\sum_{e}\left(\frac{M_{me}M_{en}}{v_{em} - v_0 + i\Gamma_e} + \frac{M_{me}M_{en}}{v_{en} + v_0 + i\Gamma_e}\right) \tag{2.10}$$

where e is the excited electronic state, v_{em} and v_{en} are the frequencies corresponding to the energy differences between the states subscribed and h is Planck's constant. Γ_e is proportional to the band width of the eth state, hence, inversely proportional to its lifetime. M_{me}, etc., are the electric transition moments, such as

$$M_{me} = \int \psi_m^* \mu_\sigma \psi_e d\tau \tag{2.11}$$

Here, ψ_m and ψ_e are total wavefunctions of the m and e states, respectively, and μ_σ is the σ component of the electric dipole moment. The summation in Equation 2.11 is over all excited electronic states, e, of the molecule. The first of the two terms is called resonance term because iunder resonance condition, as v_0 approaches v_{em} , this term becomes dominant. In normal Raman scattering, v_0 is chosen so that $v_0 \ll v_{em}$, it means the energy of the incident beam is much smaller than that of an electronic transition. Under these condition, the Raman intensity is proportional to $(v_0 - v_{mn})^4$ [37].
In biochemistry, the intrinsic sensitivity and selectivity of UVRR method can provide a spectroscopic signature of the a special chromophoric group of the molecule which is changed upon various perturbations like pH, temperature or adding a potential binding partner. These variations can be elucidated qualitatively by a comparative analysis of the Raman spectra and the results of the molecular modeling (DFT calculation) based on assigning the different vibrational mode. Nonetheless, like any other spectroscopic and evaluating techniques, the combination of UVRR spectroscopy with an efficient chemometric analysis is the crucial final step for quantitative analysis of the spectral variations. Especially, in the presence of multiple components in the data set, e.g., pH studies of multi-protic samples or binding studies of multivalent ligand and/or protein, when little or no information about the system composition and the spectra of individual components is in hand, an efficient multivariate spectral analysis can quantitatively isolate the sources of variations (e.g., concentration of different species) in an experimental data set.

2.3 Data analysis of binding studies

Representation or transformation of the multivariate data for extracting their hidden information is a long-standing problem in statistics and

analytical areas. Assuming that the data consists of r variables while the number of observations is t, a general formulation of the problem is finding a function from an n-dimensional space to an r-dimensional space such that the transformed variables give information on the data that is otherwise hidden in the large data set. That is, the transformed variables should be the underlying factors or components that describe the essential structure of the data. In most cases only linear functions are considered because the interpretation of the representation is simpler. So, each observation can be expressed as a linear combination of the observed variables,

$$v_i(t) = \sum_j w_{ij} h_j(t), for i = 1, ..., n, j = 1, ..., r \tag{2.12}$$

where w_{ij} are coefficients which define the representation and t is the observables t=1,...,T (e.g., wavenumbers upon which the spectra are recorded). The problem can be rephrased as the problem of determining the coefficients w_{ij} which is the basis of the most multivariate data analysis methods [38].

There are several analytical methods frequently applied in the qualitative and/or quantitative analysis of experimental data acquired by different instrumental techniques. Structure of the data and complexity of the problem define which method is more favorable. Typically, the data are obtained by spectrometric techniques when spectra are recorded in one direction (e.g., wavenumber) while the system under study is subject to a modulation in a second direction (e.g., caused by time, pH or protein titration in the case of binding study). So, the data matrix is an instrument response in two directions and including the information of a mixture and a standard sample, which contains simply the pure analyte. As a result, the concentration of the analyte in the mixture can be determined in the presence of unknown species. Multivariate data analysis are second-order methods which are broadly used especially for analysing the spectra of mixtures.

Generally, the multivariate data analyses are *soft-modeling* approaches which describe processes without using a chemical model linked to them. It identifies a model from numerical and statistical analysis of a data set for the isolation of the variation sources. In contrast, the process can be designated by *hard-modeling* if an accurate picture of the chemical process and an analytical or numerical solution to its mathematical model are available [39–41]. In practice, however, there is not always a previously assumed physical or chemical model for describing the system under study. For evaluating the result of binding studies, two multivariate data analysis soft-modeling methods, NMF and MCR-ALS were implemented in this thesis. The theoretical aspects of both methods are available elsewhere [42, 43]. Here, we briefly mention the basic principals relevant

to the application of binding studies.

2.3.1 Multivariate data analysis: NMF and MCR-ALS

Non-negative matrix factorization (NMF) is a resolution method which decomposes mathematically a global mixed instrumental response into the pure contributions of the components in the system. Given a set of multivariate n-dimensional data vectors, the vectors are placed in columns of a $n \times t$ matrix V where n is the number of examples in the data set. This matrix is then approximately factorized into an $n \times r$ matrix W and an $r \times t$ matrix H, assuming that there are r hidden components constructing each example in the data matrix V. The algorithms written for NMF are based on iterative updates of W and H so that at each iteration of the algorithm, the new value of H or W is found by multiplying the current value by some factors so that the quality of the approximation $V \simeq WH$ can be improved. This quality is quantified by cost functions which can be constructed using some measure of distance between two non-negative matrices V and WH. One useful measure is simply the square of the Euclidean distance between V and WH,

$$C = \| V - WH \|^2 \tag{2.13}$$

which should be minimized with respect to W and H, subject to the constraints W,H ≥ 0 [42]. For the optimum strategy of minimization, the following "multiplicative update rules", proposed by Lee and Seung, have been found as a good agreement between speed and ease of implementation [44].

$$H = H\frac{W^T V}{W^T W H} \quad , \quad W = W\frac{V H^T}{W H H^T} \tag{2.14}$$

In the previous binding studies [13,14], this strategy was used to write an algorithm for NMF in order to analyze the UVRR spectra of GCP based ligand and its mixture with a tetrapeptide, which allowed for recovering the concentration profiles and the spectra of two components occured in the experimental condition. A predefined number of iteration (100000) was used as a stopping criterion.

Multivariate curve resolution based on alternating least square (MCR-ALS) is another soft-modeling technique which has been applied successfully for analyzing the experimental data of second-order techniques such as UV-vis absorption spectroscopy [41], deep-UVRR [45] and FT-IR absorption [46] spectroscopy. In all of these applications, the goal was the resolution of the original data into the pure signals and concentration profiles of the species. Generally, in MCR methods the

bilinear decomposition of the experimental data matrix is usually noted using the following equation:

$$D = CS^T + E = D^* + E \tag{2.15}$$

where D, C and S have features in common with the introduced V, W and H in NMF. Similarly, given the data matrix D, the number of chemical components or species (N) causing the observed data variance should be specified. Then, the concentration profiles of these species in the matrix C and the pure response or spectra profiles of these species in the matrix S^T are determined. Equation 2.15 is solved in an alternating least-squares (ALS) optimization by minimizing the residual matrix E. The two alternating steps in the iterative optimization are

$$S^T = C^+ D^* \tag{2.16}$$

$$C = D^* (S^T)^+ \tag{2.17}$$

where $(C)^+$ and $(S^T)^+$ are pseudoinverses of C and S^T, respectively. D^* is the PCA-reproduced data matrix for the number of modeled components. Principle component analysis (PCA) algorithm computes an orthogonal decomposition of the correlation matrix produced by the original data matrix [47]. Working with D^* calculated by PCA, instead of the original data matrix, minimizes the rotational ambiguities which is described by the following equation

$$D = C_1 S_1^T = (C_1 T)(T^T S_1) = C_2 S_2^T \tag{2.18}$$

In this equation, by linear combination of the initial solutions C_1 and S_1^T using a nonsingular matrix T, a new set of solutions C_2 and S_2^T can be obtained, and both solutions fit the data matrix D well. PCA provides unique solutions as initial data matrix to be processed by ALS [48]. The theoretical principles of PCA is like singular value decomposition (SVD) which is briefly described in the next section. The stopping criterion for ALS optimization is the threshold value which is defined for the relative difference in lack of fit (LOF) between consecutive iterations. The lack of fit (LOF) is defined as

$$LOF = \sqrt{\frac{\sum_{ij} e_{ij}^2}{\sum_{ij} d_{ij}^2}} \tag{2.19}$$

where d_{ij} is the original element in the data set and e_{ij} is its related residual [39].

Although the aim of soft-modeling approaches is to propose a reliable model from the analysis of the empirical data, without applying a prior model,

practically this is difficult to achieve because of the rotational and intensity ambiguities due to the pure mathematical solution. Applying appropriate constraints, explored from the known physical and chemical properties of the system under study, removes or minimizes these ambiguities. The intensity ambiguity can be described from Equation 2.18. For n*th* particular component, the concentration profile C_n can be arbitrarily increased in an unlimited way by multiplying it by an arbitrary scalar number m if at the same time its spectrum S_n^T is decreased by the same amount by dividing it by the same number m. Properly using the normalization and closure constraints can limit the size of C_n or S_n^T profiles. Normalization is a mathematical constraint and can be implemented in different ways; either by using the norm of the species spectra by forcing it to be equal to one, i.e. $\|S_k^T\| = 1$, or by constraining the signal height or maximum intensity of a profile to be equal to a constant value. The closure constraint is different since it is not merely a mathematical constraint, rather it is based on the known information about the species. The closure is implemented when the total concentration of the species in the considered spectra is constant ($\sum_{k=1}^{n} C_{i,k}(T) = TOT$, if spectra are placed in the row of the original data matrix). The most commonly used constraint in curve resolution is non-negativity since physical concentrations can be only positive or zero ($C \geq 0$), and in many spectroscopies, spectral values can also be only positive or zero ($S^T \geq 0$). However, in the case of multivariate data analysis of the second-derivative spectra, the non-negativity constraint should be only applied on the concentration profile. Selectivity and local rank are the constraints which refer to the fact that in certain windows or regions of the data matrix D a particular species is known to exist while other species are known not to exist. If there is such an information about the system, it is extremely useful to use them for a partial or even total elimination of rotational ambiguities [43]. In general, every multivariate data analysis method has the capability to be included by all of mentioned constraints depending on the need for the system under study, however, the flexibility of applying these constraints during the process of MCR-ALS makes it easy to be adapted by different experimental techniques and chemical conditions.

2.3.2 Data treatment

The preprocessing the original experimental data is a crucial step before starting the main analytical process, since even the most powerful analysis methods can not provide the full potential results without a suitable primary data. For this regard, applying basic preprocessing such as "smoothing" and "baseline correction" on the original spectra for removing the unnecessary information of the noise and background

are necessary for useful data analysis. Moreover, depending on the task, some other preprocessing methods such as data transformation and scaling are suggested for simplifying data analysis. However, sometimes in multivariate data analysis, it is not enough to have the preprocessed data for starting the procedure. For example, MCR-ALS requires initial estimation of either spectra or concentration profile of the modeled components. Also, in data set produced from one titration with more than one equilibrium system, the number of components to be modeled with MCR-ALS should be determined.
Therefore, in general, we can divide data treatment into two sections. One is associated to the processing functions which transform the raw data to the more suitable data which are going to be analyzed. The second part comprises the techniques which do not modify the data, as preprocessing does e.g, smoothing and baseline correction, but rather they are kind of pre-analysis by which the important information for the main multivariate data analysis is explored. As it is described in the following, the algorithms which can be used for this initial estimation are also from the category of multivariate data analysis. Though a tremendous number of functions and methods are available for both purposes, the methods mentioned here are the ones which were implemented in this thesis.

Preprocessing

Centering and scaling refer to column-wise manipulation of the data matrix, assuming that the columns are occupied by different experimental data (e.g., the spectra of different mixtures), with the purpose that all columns have mean zero (centering) and the same variance (scaling). If $\overline{x}_j$ be the mean and s_j the standard deviation of a variable j, then

$$x_{ij}(mean-centered) = x_{ij}(original) - \overline{x}_j \tag{2.20}$$

$$x_{ij}(autoscaled) = \frac{x_{ij}(original) - \overline{x}_j}{s_j} \tag{2.21}$$

After mean-centering, the variable have a mean of zero, the data shifted by $\overline{x}_j$ and the center of the data becomes the new origin. Mean centering simplifies many methods in multivariate data analysis. Variance scaling standardizes each variable j by its standard deviation s_j. Usually it is combined with mean centering and is then called autoscaling. Autoscaled data have a mean of zero and a variance (or standard deviation) of one, thereby giving all variables an equal statistical weight [49].
The typical method used for data smoothing is Savitzky-Golay smoothing filters based on least squares polynomial approximation. We used the commercial available algorithm in MATLAB for this purpose. The mathematical basics of this method are described in the paper published

by Savizky and Golay in 1964 [50]. After smoothing, the spectra should be treated with an appropriate method for removing background. The baseline in the original Raman spectra stem from fluorescence either originated from optical setup (specially for any spectroscopic methodology which utilizes a high numerical aperture collecting optics [51]) or generated by the chemical or biological samples. Polynomial curve-fitting is a usual software method implemented for fluorescence reduction by simply fitting a polynomial curve to the raw spectrum in a least square manner. However, since it is based on minimizing the difference between the fit and the measured spectrum, which includes both the fluorescence background and Raman bands, it is not always efficient especially for more complicated background curvature (e.g., for the spectrum which has different curvature in various wavelength ranges). Thus, a modified polynomial fitting method was introduced. The basis for this method is also a least-square-based polynomial curve-fitting function. However, this function is modified such that all data points in the generated curve that have an intensity value higher than their respective pixel value in the input spectrum are automatically reassigned to the original intensity. This process (curve fitting and subsequent reassignment) is repeated typically for a defined number of iterations, gradually eliminating the Raman peaks from the underlying baseline fluorescence [52]. This method was then improved by introducing a stopping criterion. This value, which can be adjusted by the user, takes into account the contribution of noise which could have adverse effect [53].

Initial estimations

Decomposition of the experimental data can be performed by random initial matrices, known as "blind source separation", means either W or H in Equation 2.13 can be randomly selected. However, if a unique and meaningful response is desired, these estimations are better to be performed based on the evolution of the system rather than a random estimation. One example is the previous binding study [13, 14] performed by NMF in which the spectra at the first and the last point of titration were selected as an initial estimation of the components spectra. This estimation was based on the known evolution of the chemical system, because at the first point there was just one component and it was assumed that in the last point of titration the solution and therefore the spectrum is dominated by the second component (complexed ligand). But what if the system under study becomes more complex such that its evolution can not be easily interpreted? Assuming a chemical system with more than two different species involved in the equilibrium, estimating the number of components is essential for a good resolution of experimental data. However, it is

not always an easy task, especially when no selective observation variable (e.g., wavenumbers in the experimental spectra) exist for the species of interest and the data corresponding to each component are overlapped. For this purpose, the dimension reduction strategy is initially used to reveal the change in the system under study and solve the Equation 2.12. The methods working with this principle determine the coefficients w_{ij} by limiting the number of components v_i so that they contain as much information on the data as possible. This family of techniques are called principle component analysis or factor analysis [54].

Principle component analysis (PCA) can be considered as the mother of all methods in multivariate data analysis. The idea in PCA is finding the axis along which the variance of the data are maximized. Put another way, the data can best be viewed as lying along this axis. They are corresponding to the eigenvectors of VV^T and V^TV, where V is the given data matrix. The high-dimensional data can then be projected to the most important axes corresponding to the largest eigenvalues [49]. In MCR-ALS, applying PCA at the start of the procedure decreased the rotational ambiguity described by Equation 2.18. It is applied to the original data matrix V with a similar decomposition equation as stated in Equation 2.15 where C should be an orthogonal matrix whose columns are the eigenvectors of DD^T, scaled with the corresponding eigenvalues of DD^T (σ_{ii}^2) and S^T is an orthonormal matrix whose rows are eigenvectors of D^TD [38].

Evolving factor analysis (EFA) is a local rank exploratory method based on PCA which detects and localizes the selective zones with a number of compounds smaller than the total rank. By providing knowledge about the evolving and sequential nature of the pure contributions of the components in a step-wise manner, it monitors the chemical processes. EFA performs subsequent PCA runs, by using a singular value decomposition (SVD), on windows gradually enlarged by addition of a row in the process direction. This analysis is performed by building the windows from the beginning to the end of the process (forward EFA) and in the opposite direction (backward EFA). The forward and backward EFA plots are built by representing the singular values of each PCA analysis versus the process variable (e.g., the protein concentration) related to the last row included in the window analyzed. The lines connecting all the analogous singular values, i.e., all the first singular values, the second singular values and ith singular values and so on, indicate the evolution of the singular values along the process. The visualisation of these results gives a local rank map of the data set, i.e., a representation of how many components are simultaneously present in the different zones of the data set [55, 56].

A possible alternative to EFA for the initial observation of the change in the system is pure variable methods, which extract the spectra of individual components by exploring the highest intensity variations in the data set. To

understand the principle underlying the purest variable methods, one can consider a set of Raman spectra of the multicomponent mixture. Assume that the Raman spectrum of some individual component consists of several bands centered at certain wavenumbers. An increase in concentration of this component will result in simultaneous rise of Raman intensities at the peak wavenumbers, and vice versa. The band with the highest intensity variation is assumed to be contributed by a simple individual component. In fact, if the intensity at some wavenumber is a sum of contributions from various components then the variation in the total intensity at this wavenumber will be relatively small. Therefore, the pure variable methods identify the least correlated wavenumbers which have the largest intensity variations. Such wavenumbers are called the purest variables [57]. SIMPLISMA (Simple-to-use interactive self-modeling mixture analysis) is a pure variable based method. It exhibits a high relative standard deviation (coefficient of variation), which is the basis for the way Simplisma determines its pure variables [58].
Generally, in addition to the main multivariate data analysis methods, NMF and MCR-ALS, we applied EFA and SIMPLISMA for providing the initial information for the transformation of the system under study. In particular, SIMPLISMA was applied for the initial estimation of the spectra of the different protic components in pH-dependent UVRR study in subsection 5.2.2. We implemented EFA to initially estimate the number of involved components and their concentration profiles in binding studies with the assumption of two components (section 5.4.3). All of the mentioned methods are described practically and in more details during the qualitative analysis of the experimental results.

Chapter 3

Motivation and objectives

Binding to protein is thought to involve a combination of electrostatic and aromatic or hydrophobic interactions, as most proteins present a distinct and unique hydrophobic and charged residues on their surface. However, reaching both aspects of high affinity and selectivity in protein recognition is more challenging than peptide recognition. For example, since one specific residue on protein surface can possess different conformations and directions, the direction and complementarity of hydrogen bonds are hard to be well designed in the structure of ligand for recognition of this specific residues. On the other side, comparing to the development of traditional enzyme inhibitors which bind in deep pockets at active sites of the protein, supramolecular ligands need to bind in shallow grooves or large pores on the protein surface which is even more challenging [59]. These are the main challenges in supramolecular chemistry which are tried to be overcome in CRC. Along with the traditional methods such as ITC and fluorescence spectroscopy for quantifying the binding between host and guest molecules, UVRR spectroscopy was going to be established as a complementary technique. The main motivation and objectives were originated from the progress which has been made for designing the supramolecular ligand for protein recognition. Mainly for two reasons, Raman spectroscopy is ideal for studying biochemical systems; first because water is a weak Raman scatterer, it does not interfere with Raman spectra of the solutes in aqueous solution, secondly by taking the advantage of resonance Raman (RR) scattering, it is possible to selectively enhance a particular chromophoric vibrations using a small quantity of samples. Moreover, since the diameter of laser beam is normally 1-2mm, only a small sample area is needed to obtain Raman spectra. However, Raman spectroscopy has some disadvantages. The laser source which is needed to observe weak Raman scattering may cause local heating and/or photodecomposition. Especially, in resonance Raman studies where the

laser frequency is tuned in the absorption band of the molecule. The Raman spectra of some compounds also is hindered by fluorescence [36].

3.1 From small peptide to protein recognition

In the biochemical context, Raman spectroscopy is concerned primarily with the vibrational energy level transitions of the molecule. In other words, by monitoring the inelastically scattered photons we can probe molecular vibrations, and the Raman spectrum is a vibrational spectrum of a molecule [60]. The selectivity and sensitivity of RR spectroscopy were mostly applied for elucidating the proteins secondary structure and peptide/protein dynamics and mostly in the range of deep UV [61–63]. These advantages are also useful for binding studies between host and guest molecules by enhancing the vibrations of the chromophoric group in one of counterpart molecules (host and guest) which is responsible for the target recognition. Despite this fact, RR spectroscopy has been rarely used for binding studies between host and guest molecules [64]. In this regard, a ligand-based UVRR technique was introduced for the first time in order to evaluate the binding event between two molecules by enhancing the vibration of one chromophore involving in the binding [13, 14]. In this manner, the GCP motif, designed for the carboxylate recognition, was selected as the chromophoric group enhanced by UV excitation for the proof of concept. The concentration of ligand (host molecule) including GCP had to be constant for monitoring the spectral change upon addition of binding partner. As a result, the name "ligand-based" was devoted to the technique based on this manner of titration. The interfere of binding partner signature in the UVRR spectra was neglected since Raman scattering of binding site (GCP) was intensified upon resonance condition. Therefore, UVRR spectroscopy was applied for characterizing a receptor molecule comprising a GCP motif and a tripeptide (Figure 3.1, up), regarding determination of its pK_a value by a UVRR pH titration experiment [13], and determination of its association constant with a tetrapeptide (Figure 3.1, bottom) by a UVRR binding study [14]. During these continuous studies, and even comparing with other vibrational technique like FT-IR absorption spectroscopy [15], it was shown that UVRR spectroscopy combined with multivariate data analysis methods is capable to quantify the host-guest interactions with a high sensitivity and selectivity of GCP motif used in the host molecules.

Studying the recognition of small peptide fragments by GCP and demonstrating its efficiency for C-terminal and carboxylate recognition in aqueous solution, reviewed in section 2.1, provided an outline for

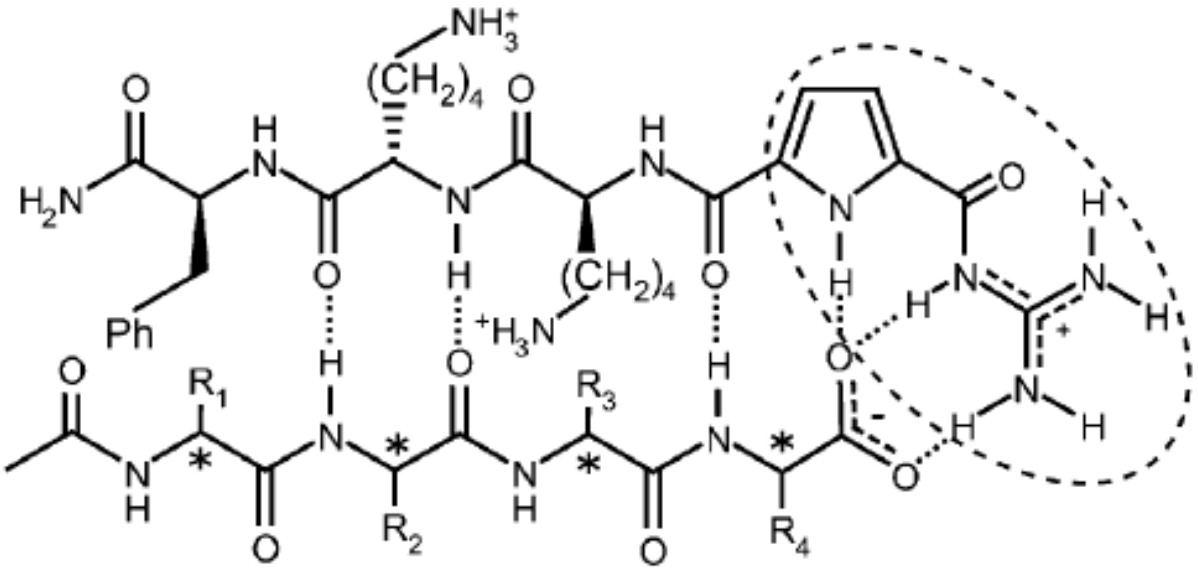


Figure 3.1: The structural schematic of a tetrapeptide recognition by a KKF-GCP receptor, from reference [15].

extending this approach from peptide- to protein-recognition. Especially, development of supramolecular ligand was demanded in the context of CRC for supramolecular chemistry on proteins. In particular, protein surface binding approach has been promising with development of supramolecular and multivalent ligands including GCP motif. As an example, Schmuck and co-workers identified four-armed ligands derived from lysine trimers as scaffolds which bind with nanomolar affinity to the entrance of the central pore of β-Tryptase (a tetramer protein) which is lined by anionic amino acids [59, 65]. Additionally, some efforts have been also doing for designing a binder ligand for protein 14-3-3 [66, 67] within the CRC.

Jumping to higher level of molecular recognition by GCP-based host molecules stimulated the idea of establishing UVRR spectroscopy for protein recognition. Therefore, encouraged by previous promising results from UVRR binding studies [13, 14] and in parallel with extended approach from small peptides to protein recognition by supramolecular multivalent ligands, we were interested to see whether the ligand-based UVRR technique is still sensitive enough to monitor the formation of the supramolecular complex. By developing a new set-up facilitated by a compact spectrometer, with lower focal length than the previous double monochromator spectrometer (with lower throughput and higher resolution) used in the previous binding studies [16], we expected to have higher signal. However, although we planned to monitor the spectral changes of ligand in constant concentration upon protein addition, it was limited in its applicability for different proteins recognition because of fluorescence. As the main drawback of UVRR spectroscopy especially for biochemical applications, ultraviolet fluorescence is contributed from

three aromatic amino acids phenylalanine, tyrosine and tryptophane. The absorption and emission spectra of these amino acids are shown in Figure 3.2. Phenylalanine displays the shortest absorption and emission wavelengths. It displays a structured emission with a maximum near 282 nm. The emission of tyrosine in water occurs at 303 nm and is relatively insensitive to solvent polarity. The emission maximum of tryptophan in water occurs near 350 nm and is highly dependent upon polarity and/or local environment. Protein fluorescence is generally excited at the absorption maximum near 280 nm or at longer wavelengths due to tyrosin and tryptophan while phenylalanine is mostly excited at lower wavelength [68, 69]. Therefore, phenylalanin was expected to be excited by the excitation wavelength of our UVRR experiments (266 nm). Accordingly, the proteins β-Tryptase and 14-3-3, which were under binding investigation with GCP-based supramolecular ligands in CRC, could not be used for UVRR binding studies. Therefore, we had to deal with the interference of Raman spectra and fluorescence stem from phenylalanine included in the structure of either protein or ligand. The usual techniques for partially or completely rejecting the fluorescence background require modification of the existing instrumentation to implement. Moreover, they are usually costly and complex in source and/or detection [70, 71].

Beside the hardware involved techniques, there are also mathematical techniques for fluorescence subtraction [52, 53]. In order to keep the simplicity of instrumentation and proving the concept of monitoring molecular recognition in supramolecular chemistry by UVRR spectroscopy, we implemented one of these algorithms for removing the fluorescence background in our UVRR spectra. But this technique can not be used for the spectra hindered strongly by the fluorescence background. Therefore, the protein receptors with minimum number of aromatic amino acids were required for binding experiments with supramolecular ligand. We found these properties in dermcidin-1L, a human antibiotic peptide (PDB: 2KSG) and leucine zipper (PKG I_α), a DNA protein binding (PDB: 4R4L), displayed respectively in Figure 3.3 and Figure 3.4. Dermcidin consists of four α-helices and has 47 amino acids with no aromatic amino acids while one of 49 amino acids in the sequence of leucine zipper is phenylalanine. As can be seen in the figures, leucine zipper is a dimer composed of two α-helices and several turns. With these two proteins, we examined the protein recognition of supramolecular ligands by UVRR spectroscopy.

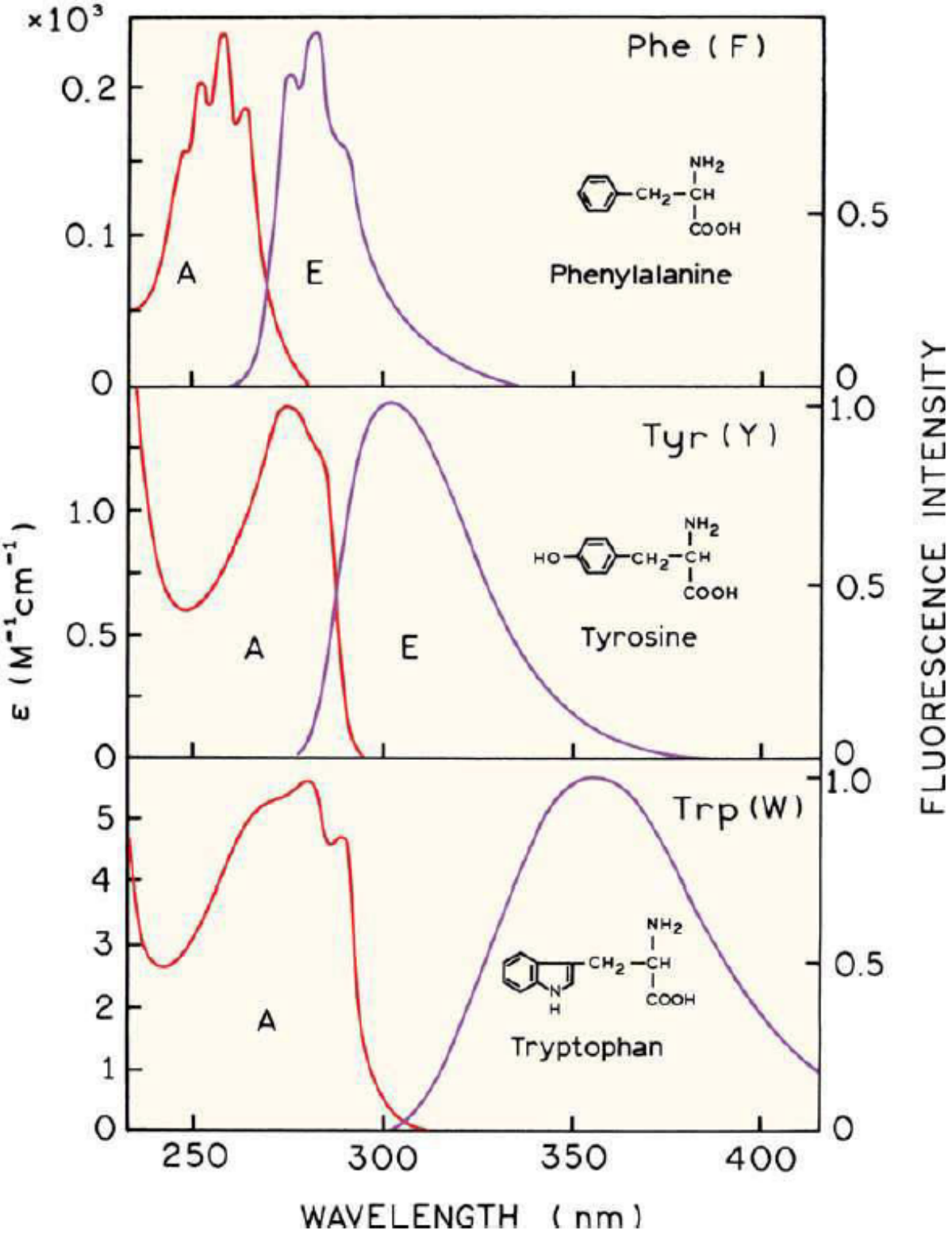


Figure 3.2: Absorption (A) and emission (E) spectra of the aromatic amino acids in aqueous solution with pH 7 [68].

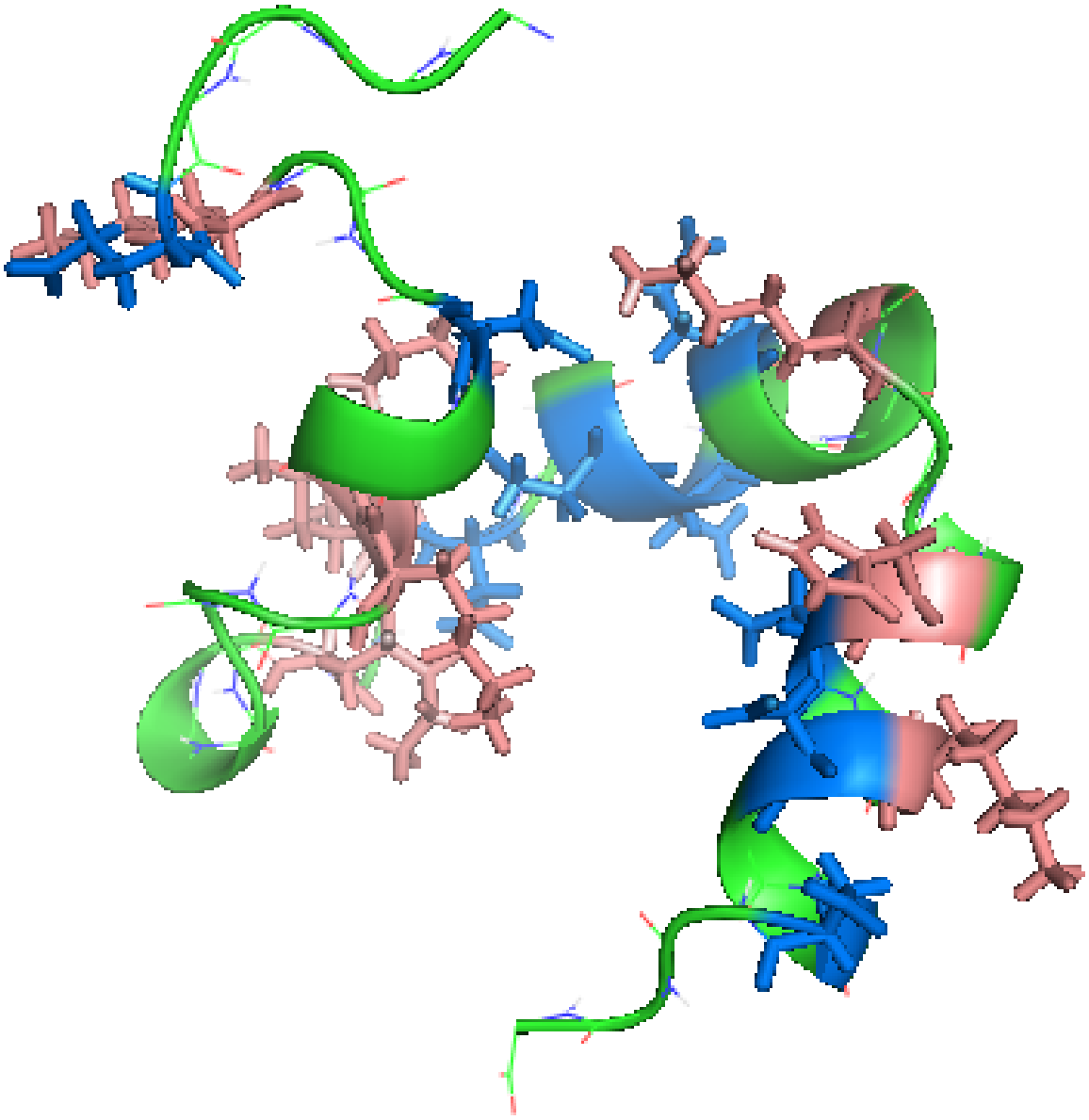


Figure 3.3: The structure of the protein dermcidin with negatively charged residues in blue and positively charged residues in pink.

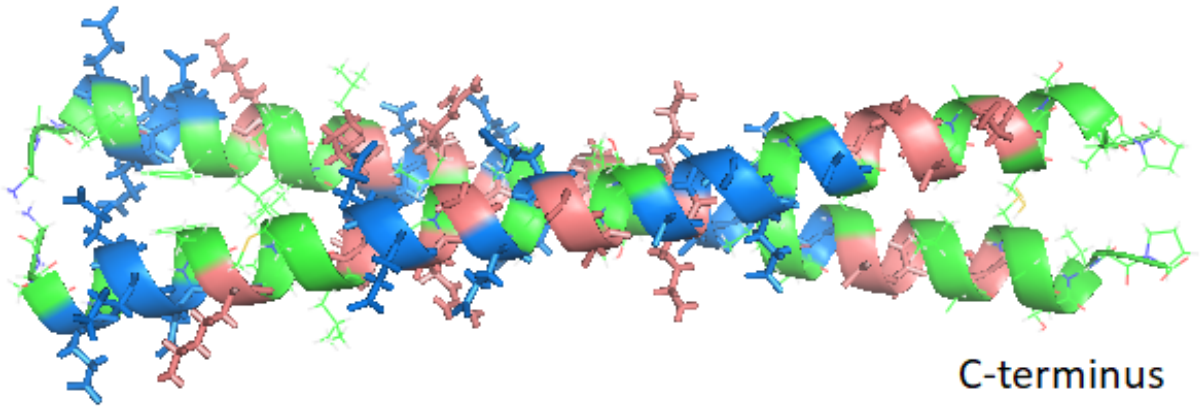


Figure 3.4: The structure of the leucine zipper homodimer with negatively charged residues in blue and positively charged residues in pink.

3.2 From mono- to multi-valent ligands

For protein recognition with high affinity and specificity, the ligand molecules not only need to involve a combination of electrostatic and hydrophobic interactions, but also require a careful design regarding their sequence and conformation if selective binding is demanding. Selective host molecules for peptides could be achieved by using molecular modeling followed by the rational and design-based methods [72]. However, this approach could not be used in designing more challenging host molecules for the bigger and more complex target substrates. Since the larger the substrate or host becomes, the more difficult the design gets. In this regard, the dynamic combinatorial method was demonstrated as a more useful technique for identifying the host molecules with high affinity in supramolecular chemistry [25].
Multivalency is a very important concept in supramolecular chemistry which is often described as the "Gulliver effect" and can ensure stable complexation. A multitude of individually small interactions act in concert to achieve strong binding, similar to what happened to the giant Gulliver in the country Lilliput in Jonathan Swift's famous novel [7]. Assuming that the ligands with more GCPs and/or more positively charged residues might be more affine or more selective, different multi-armed supramolecular ligands, depending on the task, were designed and synthesized for targeting different proteins. The multivalent ligand using aldehydes as scaffold can be mentioned as an example of multivalency, which were used by dynamic combinatorial method to discover potent inhibitors of β- tryptase. Shown in Figure 3.5 is the structure of the rigid central aromatic scaffold used for a ligand with three cationic peptide arms (R: Lys_X-Phe-Lys_Y, where X,Y= GCP or H) [73]. The composition of tripeptide part containing GCP motif within the arms was investigated before [65] to identify the most important position of the variable tripeptide for the protein inhibition. The aldehyde scaffold with arms ending in two positively charged groups, a Lysine and Guanidinocarbonylpyrrole (GCP) was also successfully applied for stabilizing the interaction between 14-3-3ζ and its effectors c-Raf or Tau [67].
Another possibility for designing multiarmed ligands is when different arms are connected through flexible linker instead of a rigid scaffold. Shown in Figure 3.6 is a structural schematic of a three-armed GCP ligand in which the GCP binding motifs are connected through a flexible linker to a scaffold. It is supposed that such a ligand with more than one GCP-group connected by a flexible linker are able to bind to anionic groups on different adjacent monomers. We used this multi-armed GCP ligand for binding studies with the protein leucine zipper. For further optimizing the multiarmed ligands, the optimal linker length and the number of arms as well as suitable scaffold

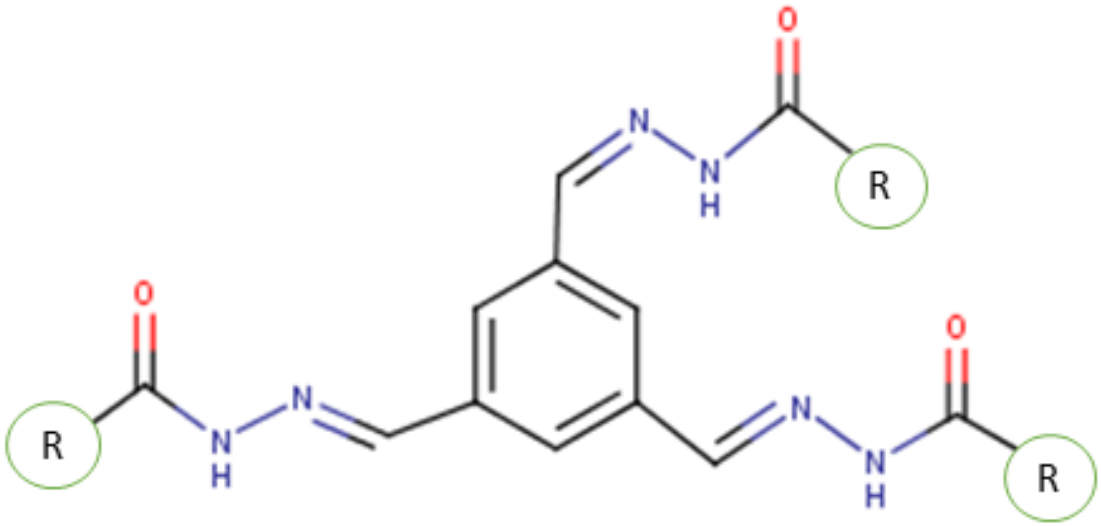


Figure 3.5: A structural schematic of a scaffold used in multi-armed GCP ligands.

should be optimized. For example, in a previous study on bis-cations [74] with a simple primary ammonium cation attached via flexible linkers to a GCP binding site as depicted in Figure 3.7, the effect of the linker length on the affinity of the complex with N-acetyl amino acid carboxylates was investigated. It was shown that with increasing linker length, complex stability first increases until an optimum is reached for n=4 and then

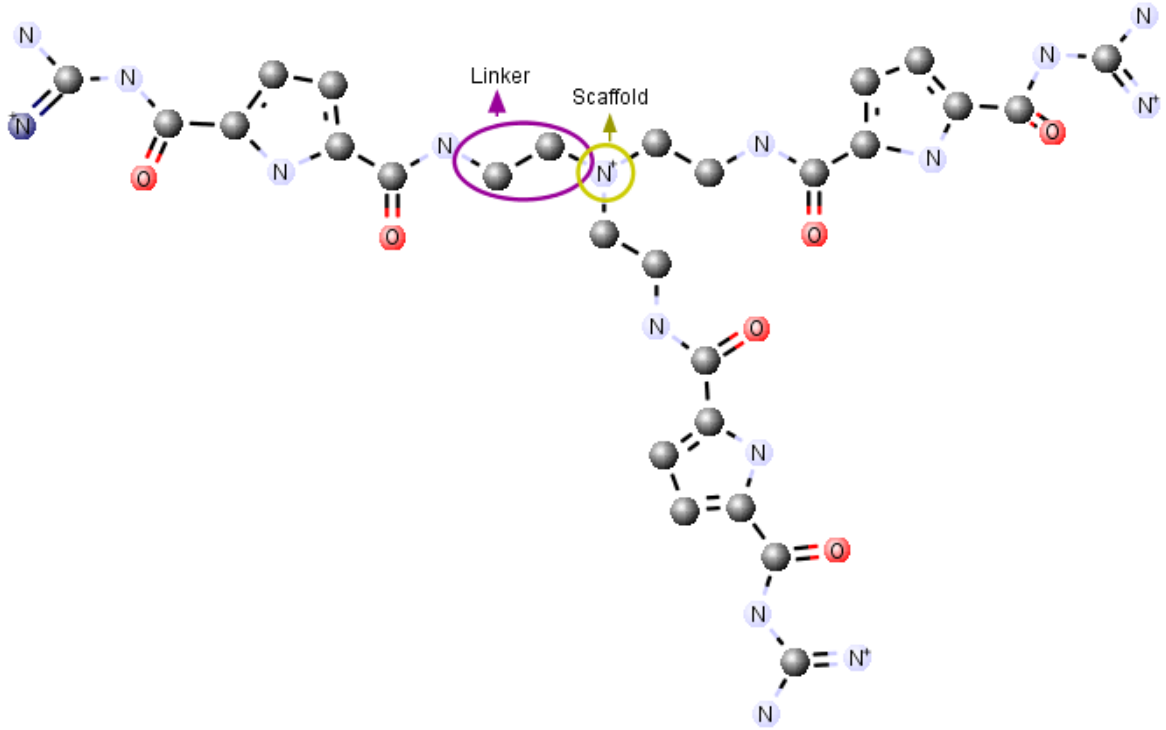


Figure 3.6: A structural schematic of a multi-armed GCP ligand

decreases again. For the multiarmed ligand, the optimum linker length can

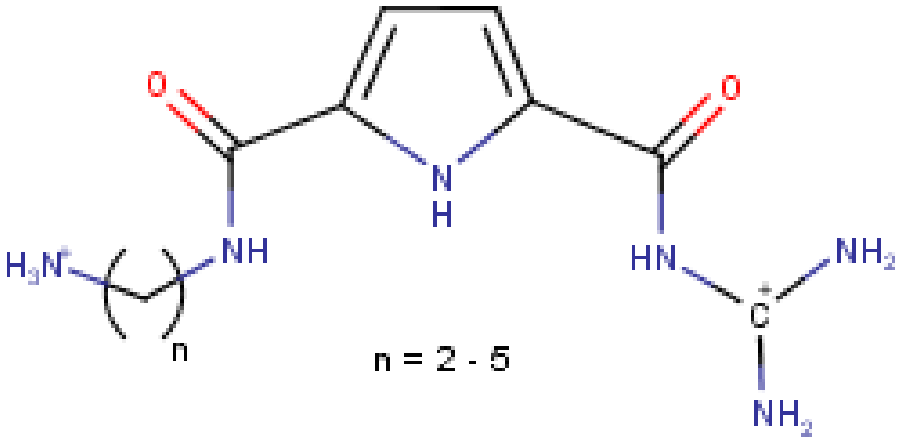


Figure 3.7: A GCP motif attached to an ammonium cation via flexible linkers of varying length, from reference [74].

be changed if two adjacent arms are correlated in targeting one carboxylate, while it also depends on the structure and sequence of the target.
Designing multivalent ligand and tailor-made binding site for protein recognition was another interesting motivation for the establishing UVRR technique for binding studies in supramolecular chemistry since, as a methodology partner, it was important to be involved in this transition from mono- to multi-valent supramolecular ligand. This contribution even gets more attention when we use UVRR as a ligand-based technique, because as a result of its sensitivity to the vibrations of GCP motif, it could be possible to evaluate the contribution of each motif or different arms (if GCP motifs are distributed between arms). However, since the UVRR spectra of GCP motifs are totally overlapped, resolving the contribution of each motif from the UVRR spectra is not possible without labeling or using isotopic atoms. Therefore, we could not evaluate the binding properties of each arm at this stage. However, the concept of multivalency is applied in the analysis of spectral results in subsection 5.4.4.

3.3 From GCP-based ligands to molecular tweezers

Being a part of collaboration in CRC with the available different classes of supramolecular ligands, designed and synthesized by organic chemistry groups, stimulated the idea of studying the molecular recognition of different categories of supramolecular ligands by UVRR technique. Therefore, in addition to our main binding studies using GCP-based supramolecular ligands, molecular tweezers were selected to

be monitored spectroscopically by UVRR for further investigation of molecular recognition. The special characteristics of this supramolecular ligand is not only in using the combination of hydrophobic and electrostatic interactions as a water-soluble host, but also in its "process-specifity" which open the door for biological applications [27, 75–77]. These remarkable biological effects inspired the design of new molecular tweezers to reach site specificity in protein recognition. One example from the new generation of molecular tweezers is shown in Figure 3.8 (left). Compared to symmetrical tweezers in which the same linkers are used for locking the amino acid lysine in the cavity, the asymmetrical tweezers utilizes one different linker for a specific interaction with the protein, e.g., through click reaction, while the presence of one phosphate linker is necessary for maintaining the water solubility of the molecule [28]. Such a development in the structure of molecular tweezers demands new techniques which are literally capable of evaluating the contribution of each interaction by probing the chromophoric groups mainly involved in the related interaction. With this ultimate goal, we took the first step to examine the UVRR spectra of molecular tweezers.

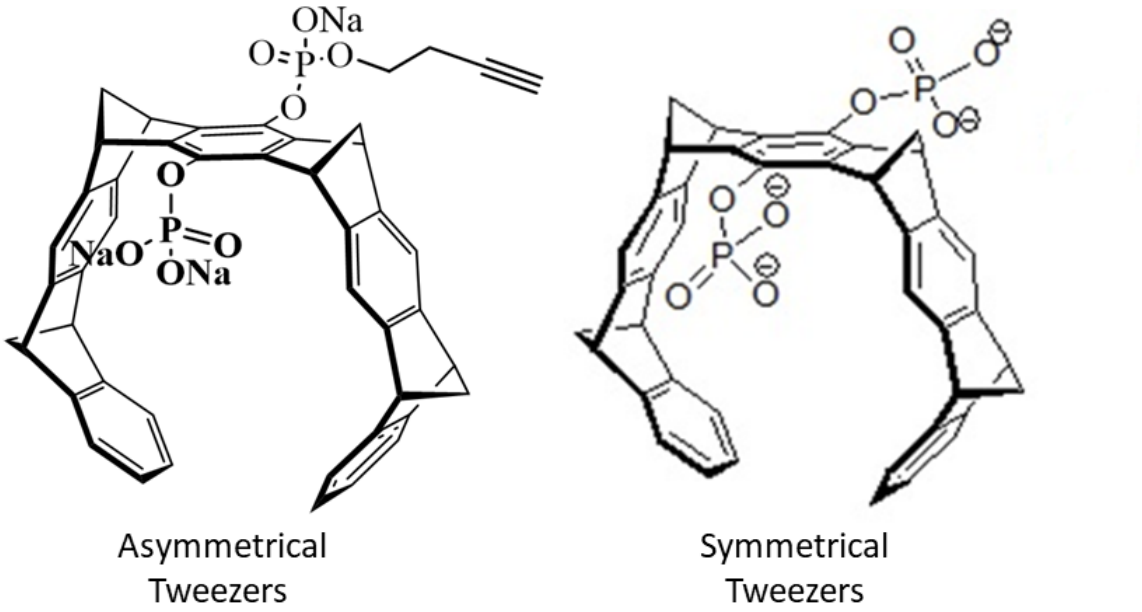


Figure 3.8: Chemical structure of an asymmetrical (left) and a symmetrical (right) molecular tweezers.

The significant point in using UVRR technique for binding studies in supramolecular chemistry is to enhance the vibrations of the chromophore involved in the main binding interactions. For example, in monitoring molecular recognition of GCP-based ligand by UVRR, since the GCP motif contributes to both electrostatic and hydrogen bonds, the observed spectral changes upon binding by probing GCP can show the whole process

of complexation by using one single laser wavelength to enhance this particular chromophore. The situation, however, is different for binding studies of molecular tweezers because, even in symmetrical tweezers, hosting is organized by two different parts: the cavity and the anionic groups attached to the tweezers skeleton. The vibrational modes of these parts can not be enhanced by using one excitation wavelength. Clearly, in the UV range, the vibrations of the cavity might be dominant. For directly measuring the electrostatic interaction between the anionic groups of tweezers and the guest cationic groups, the excitation wavelength should be selected in resonance with these groups. While the net result of all contributions decides about the stability of the complex, the evaluation of each contribution was not reported. Therefore, we came up with the idea of knowing the strength of each interaction contribution which caused the starting this collaboration, though it seemed a far and ultimate goal.

Chapter 4

Materials and methods

4.1 Supramolecular ligands

By applying UVRR spectroscopy as a ligand-based technique for binding studies in supramolecular chemistry, we investigated two different classes of artificial ligand. One was the ligand including GCP motif as the carboxylate binder, successfully used for binding studies with peptides. Keeping on the previous binding studies with GCP, we focused on supramolecular multivalent GCP-based ligand molecules for probing their interaction with protein. In general term, these ligand were used for proving the concept of supramolecular binding studies by UVRR with an emphasis on two concepts of multivalency and protein recognition. On the other hand, molecular tweezers, a different class of selective binder to lysine and arginine, was utilized for our UVRR experiments.
Three multivalent supramolecular GCP-based ligands were used for UVRR experiments: Li-40, shown in Figure 4.1, was mostly used for initial experiments in order to find the optimum concentration and pK_a value, while compound **1** (Figure 4.2) and compound **2** (Figure 4.3) were

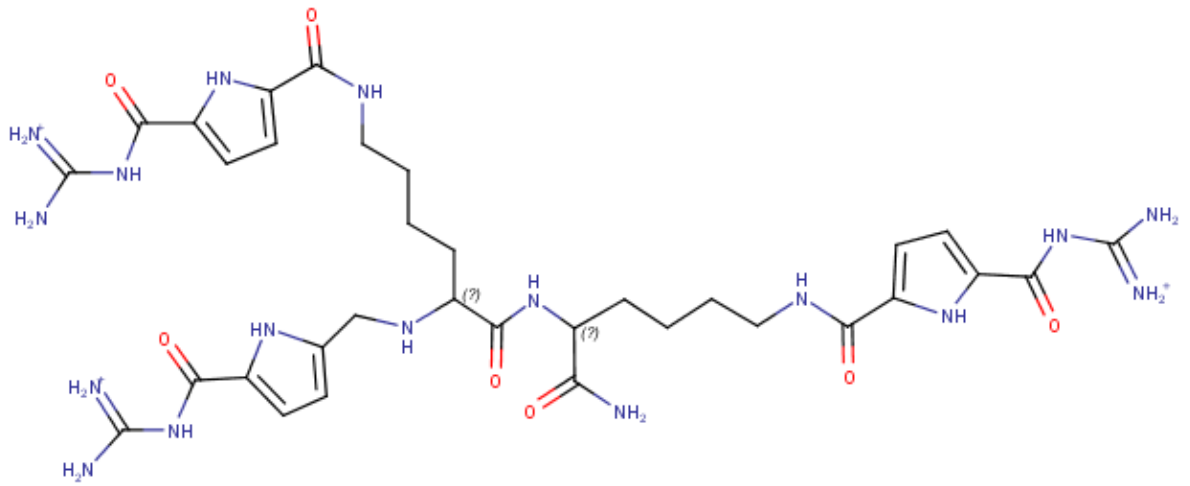


Figure 4.1: The chemical structure of Li-40

applied for binding studies with the proteins dermcidin and leucine zipper, respectively. All of these molecules are multi-armed ligand equipped with

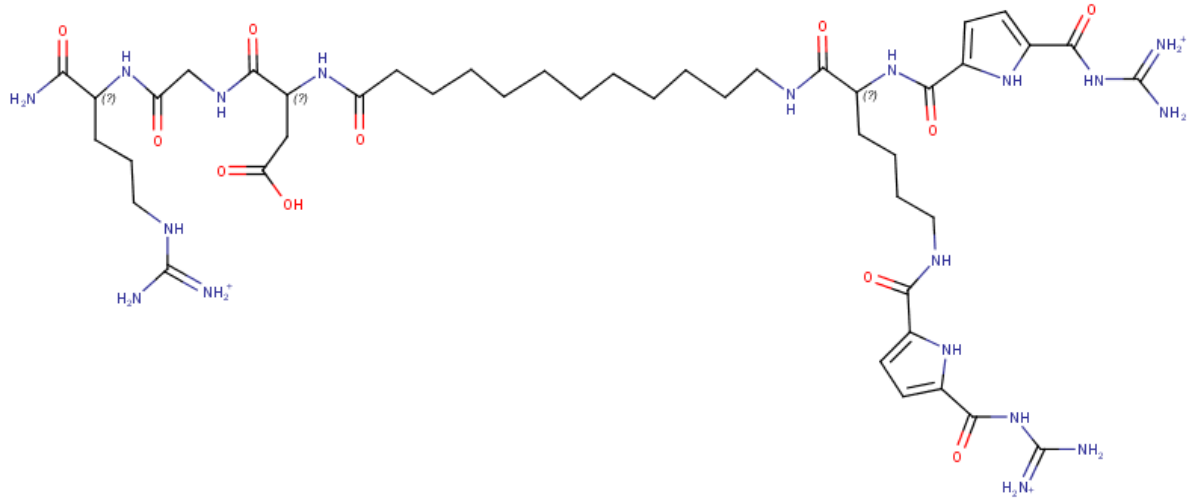


Figure 4.2: The chemical structure of compound **1**

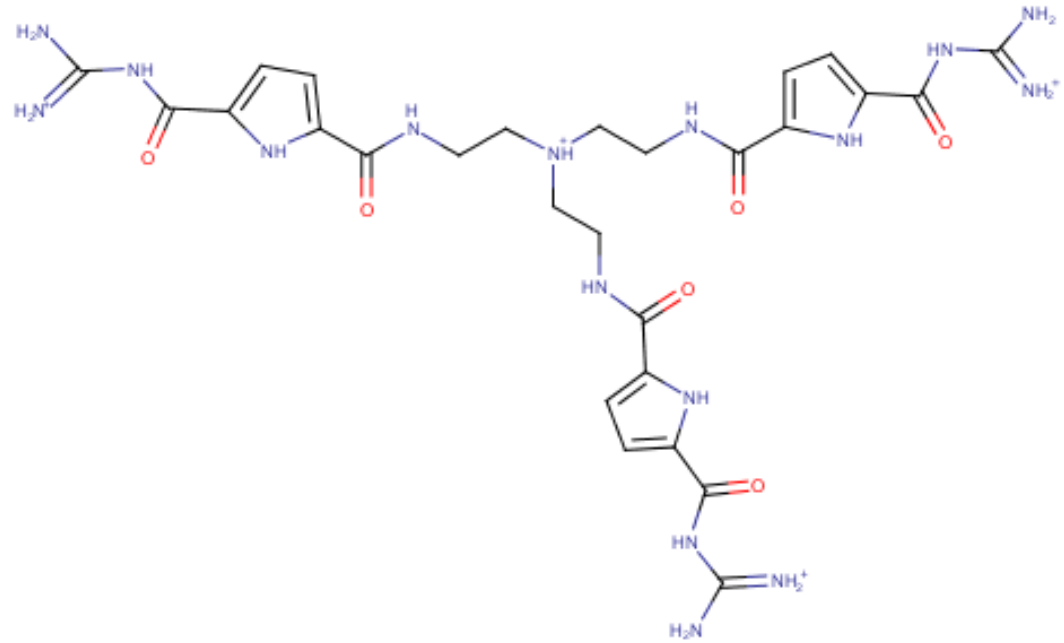


Figure 4.3: The chemical structure of compound **2**

GCP motif for the recognition of carboxylate group included in peptide or protein. Compound **1** contains a carboxyl group and compound **2** has an extra possible positive charge on the scaffold. The ligands are also different regarding the length of their linker. The minimized energy conformations of tri-armed ligand (compound **2**), calculated by Molecular Dynamic toolbar in MarvinSketch depicted in Figure 4.4, interestingly shows that the GCP motif maintains its special bent shape in all conformations.

Phosphate molecular tweezers (CLR01) shown at the right of Figure 3.8,

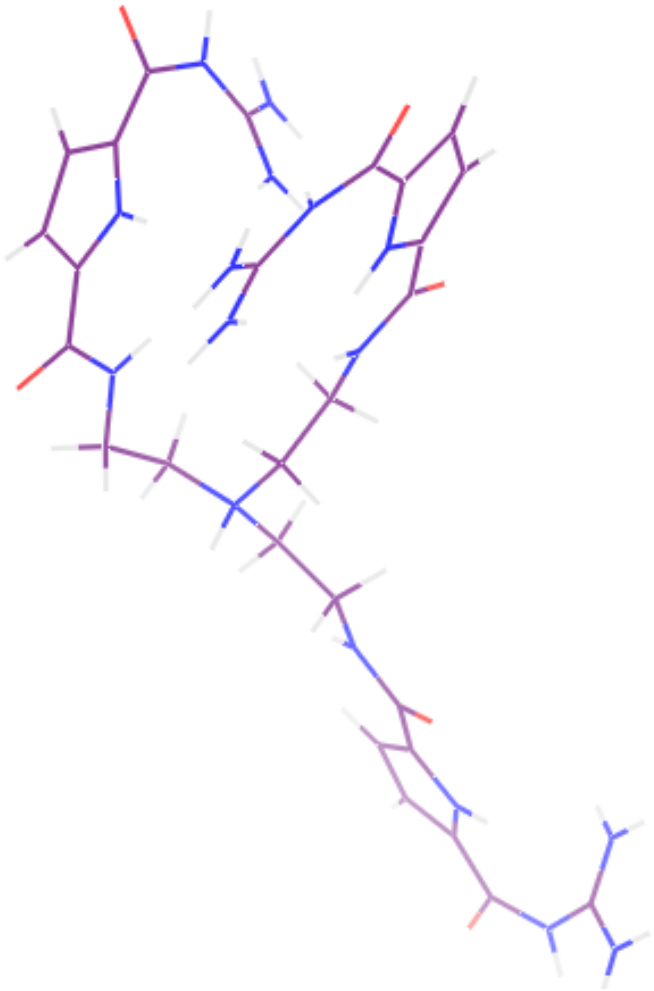


Figure 4.4: The minimized energy conformation of three armed GCP ligand in Figure 4.3.

on the other hand, has a high potential to interfere with biologically relevant peptides or proteins in which lysine or arginine residues play crucial roles [76, 78]. The first UVRR spectra of phosphate molecular tweezers was examined as a starting point for forecasting an experimental plan for the future binding studies of molecular tweezers with UVRR spectroscopy.
Except Li-40, which was synthesized by Mao Li, all syntheses and subsequent purification steps of GCP-based ligands were carried out by Alba Gigante Martinez (Organic chemistry, Prof. Carsten Schmuck; CRC1093, subproject A1). The molecular tweezers was synthesized by Andrea Sowislok (Organic chemistry, Prof. Thomas Schrader; CRC1093, subproject A3).
For pH-dependent UVRR studies, the sample Li-40 was magnetically stirred and titrated from pH 3 to 9 with NaOH and HCl which were used as concentrated as possible to keep the effect of sample dilution negligible. For the UVRR binding studies, constant concentrated solutions of ligands were used in order to record the vibrational change of GCP motifs upon

protein addition during UVRR titration. In one preliminary experiment for stoichiometry estimation, this titration was done in a sequential manner by a continuous addition of protein to the constant concentrated ligand. But in the main binding studies experiments, we prepared separate solutions of ligand-receptor mixtures with constant concentration of ligand and changed the concentration of receptor in different equivalents relative to the neat ligand. Providing separate mixture solutions has two major advantages: it is easier and more accurate to adjust the pH value for every mixture, secondly the probability of photo-damaging the sample by the laser radiation is decreased compared to the sequential titration because of the reduced exposure time. The concentration of ligands was kept constant in all experiments for binding studies, since the GCP moieties of the ligand are selectively enhanced upon UV excitation and therefore dominate the Raman spectra of the mixtures. The pH for the solutions was carefully adjusted to 6.00± 0.05 using the pH electrode (Metrohm, 877 Titrino plus).

4.2 Protein receptors

In the first step of developing UVRR technique for binding studies of supramolecular GCP-based ligands and proteins, we searched for the receptors with minimum number of aromatic amino acids and high numbers of carboxylate anion side chains in their structure. The mentioned criteria were necessary for acquiring good UVRR spectra without interfering with UV-excited fluorescence in order to go one step further and study the binding between multi-armed ligands and bigger molecules than tetrapeptides. In this regard, two protein receptors dermcidin and leucine zipper were ordered from Centic Biotech company to be synthesised and purified. Their characterization is available in protein data bank (PDB) account (leucine zipper: 4R4L, dermcidin: 2KSG). Our main binding study was performed between leucine zipper and compound **2** (Figure 4.3). Therefore, we summarize here the most important structural and biological characteristics of this protein derived from its crystal structure [79]. The protein structural specifications will be presented in detail in chapter 6 for interpreting the experimental results.
Leucine zipper is from the *α-helical coiled-coils* family including two α-helices that intertwine, creating rope-like structures. Generally, coiled coil structures have a characteristic seven residues repeat, $(a.b.c.d.e.f.g)_n$ with hydrophobic residues at positions a and d and polar residues generally elsewhere. The hydrophobic residues at "a" and "d" positions pack together in a **"knobs-into-hole"** fashion to minimize their interaction with water molecules. This hydrophobic effect is the main driving force for association and folding of coiled coils. In addition,

some parallel and dimeric coiled-coils like leucine zippers exhibit strong preference for specific amino acids at the hydrophobic a and d positions of the heptad repeat. Unlike other coiled-coils, leucine zippers contain leucine in $\sim$ 80 percent of all d positions [80].
Biologically, leucine zippers play an important role in many biological functions. Leucine zippers are included in the N-terminal regulatory domain of cyclic GMP-dependent protein kinase (PKG) which is a serin/threonine kinase, regulating smooth muscle tone, vasorelaxation, erectile function, platelet aggregation, circadian rhythm, and gastrointestinal secretion. Domain organization of PKG is depicted in Figure 4.5 [79]. Three types of PKG are found in mammalian cells. PKG I_α and I_β are splice variants of the type I gene and are found mostly in the cytosol, whereas PKG II is transcribed from a separate gene and is associated with the membrane via a myristoylated glycine at the N terminus [81]. The three leucine zipper domains share little sequence similarities, except for leucine or isoleucine residues at a and d position of the heptad repeat that form the dimer interface. The leucine zipper domain mediates homo dimerization and cellular targeting of PKG [79]. For example, it was shown that $PKGI_\alpha$ is targeted to the smooth muscle cell contractile apparatus by a leucine zipper interaction with the myosin-binding subunit (MBS) of myosin phosphatase. Uncoupling of the $PKGI_\alpha$-MBS interaction prevents cGMP-dependent dephosphorylation of myosin light chain, demonstrating that this interaction and therefore leucine zipper domain is essential to the regulation of vascular smooth muscle cell tone [82]. Hence, studying the binding possibilities of these leucine zipper domains would help to understand the underlying principles of related protein-protein interactions. Thus, in our binding studies, we focused on leucine zipper domain of $PKGI_\alpha$ and the name leucine

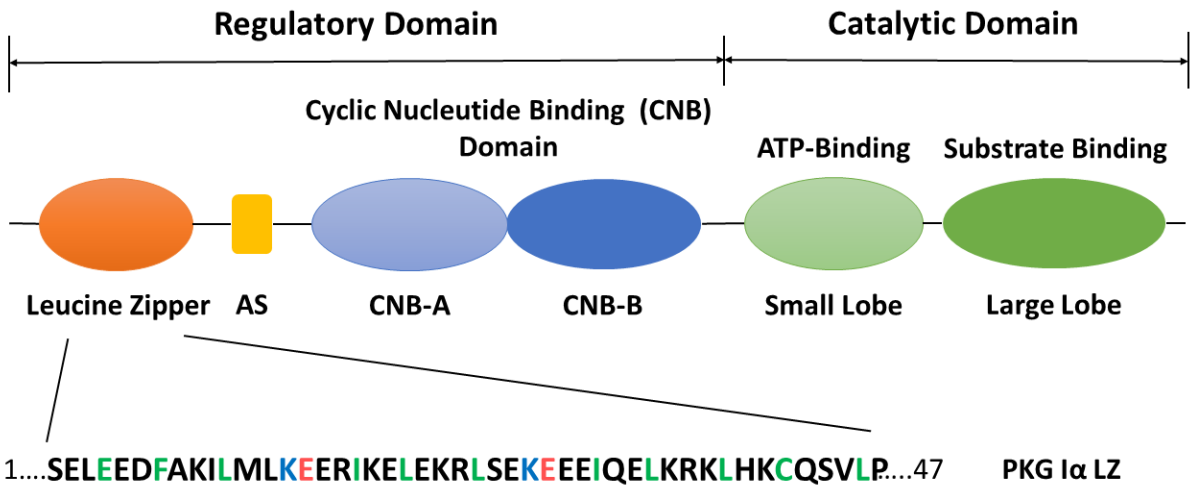


Figure 4.5: Domain organization of PKG with the sequence of its leucine zipper domain, adopted from [79].

zipper will be a statement of this domain with the sequence displayed in Figure 4.5.

4.3 Optical spectroscopy

Generally, we applied two spectroscopic techniques during our research studies for detection of the protein recognition by supramolecular ligand. UV-vis absorption spectroscopy was mostly used for preliminary experiments and preparation of the experimental condition. We used a 2-channel UV-vis spectrometer (JAS.CO, Spectrophotometer V-630), in which the radiation of same frequency and intensity is simultaneously passed through both channels. One channel is the radiation path through the reference cell containing only the solvent (water in our case), while in the second channel the radiation is transmitted through the sample. Radiation across the UV range from 200 nm to 350/400 nm is scanned over a period of approximately 30 seconds. The absorption spectra are then recorded by a multichannel array detector by comparing the difference between the intensity of the radiation passing through the sample and the reference cell. The sample was the prepared solution of ligand in different concentrations for measuring the wavelength-dependent absorptivity coefficients of ligand required for concentration-dependent UVRR (see subsection 5.2.1), while the absorption spectra of the mixtures (ligand and protein) were recorded for estimating the stoichiometry in section 5.4.1. Moreover, we used UV-vis absorption spectroscopy for checking the absorption band of the compound (ligand and protein). This is the basic application of UV-vis absorption spectroscopy in complementary with the UV resonance Raman (UVRR) spectroscopy. As far as the molecule has the absorption band within the UV range, UV light can excite the chromophore to the higher energy level. Then, UVRR spectroscopy can be used for monitoring the molecular vibrations of this particular chromophore.

Resonance Raman scattering was performed using a 266 nm cw laser excitation (CryLaS, FQCW 266) illuminated to the sample in a custom-made rotating quartz cuvette (company Hellma). The rotating quartz cuvette was used in order to minimize the photo damaging of the sample. The scattered light was collected in a 90° geometry set-up and focused on the entrance slit of a double monochromator with a focal length of 50 cm and a 2400 rules/mm grating (Acton, SpectraPro 500i). The monochromator was equipped with a LN2-cooled UV-enhanced CCD (Princeton Instruments, model 2KBUV) for detection of Raman scattering. The CCD camera and spectrometer was connected to the host computer on which the *LightField* software was installed for monitoring

the spectra and controlling the parameters of the experiment. Building the set-up with the mentioned components as well as the calibrating of the spectrometer is described in section 5.1.
All spectra were recorded between 900 and 1800 cm^{-1}. We tried to keep the total measurement time up to 30 min for all experiments. It was the optimum illumination time for having the maximum signal to noise ratio. However, we changed the ratio between the exposure time and accumulation based on the structure and the intrinsic fluorescence of the sample under experiment because increasing the exposure time for a sample with fluorescence could saturate the CCD camera. For checking the sample damage by UV radiation, the spectrum after the first accumulation was saved to be compared with the final spectrum.

4.4 Data processing and analysis

In order to analyze the underlying principles of the complex formation, e.g., the contribution and the spectra of different species, UVRR spectroscopy needs to be combined with chemometric methods. In general, multivariate data analysis of the spectral data matrix containing rows of UVRR spectra per mixture allows for the quantification of analyte (ligand) in the presence of spectrally unknown and silent interference (protein). This feature can be achieved by simultaneous analysis of only one pure ligand used as calibration standard with different mixtures of ligand and protein (see section 2.3 for more detail).
In this regard, we utilized non-negative matrix factorization (NMF) and multivariate curve resolution-alternative least square (MCR-ALS) for a comparative analysis of UVRR spectroscopic data. While NMF was used for data analysis in the previous UVRR binding studies, we additionally applied MCR-ALS which has been used specifically for diverse series of titration recorded by different experimental methods. Even though two approaches present the same information, i.e., the concentration profile and the spectra of the components, MCR-ALS is a flexible method for applying the known chemical and mathematical information about the data set through the use of constraints. We used this advantage for interpreting the multiple equilibrium binding with some assumption about the hidden components during the analysis. The flexibility of using constraints during the calculation of MCR-ALS also allowed us to apply it for the second derivative spectra, in which the constraint of non-negativity was applied just for the concentration profile rather than for both concentrations and spectra. This possibility is not available with NMF since it can be applied only for non-negative data. The basic principles of methods have been reviewed in section 2.3 and all applied

MATLAB codes are freely available[1]. Here, only the procedure and the sequence by which they are applied in this thesis are mentioned.
Both methods are applied on the treated Raman spectra. Data pre-treatment including baseline correction and smoothing are necessary for further analysis by any chemometric methods. In MCR-ALS algorithm, since the experimental data matrix D is reproduced by PCA for the number of modeled components, the scaling of the data (e.g., autoscaling for treating the self-absorption) is not crucial for further analysis. However, for acquiring a quantitative information from NMF it is important to perform the scaling on experimental data (see Figures 5.24 and 5.26 for a comparison between two methods on non-scaled and auto-scaled experimental data, in respect). Shown in Figure 4.6 and Figure 4.7 are the flowcharts for the procedure of NMF and MCR-ALS, respectively. Both methods need an initial estimation of either spectra (S-type) or concentration profile (C-type which referred as the contribution of the species) for all modeled components and also the number of components in the case of multicomponent analysis. In addition, applying different constraints stem from the chemical conditions of the problem improve the results. The constraints are applied in the form of a penalty function during the ALS fit which is one of the most outstanding features of the method and has allowed the adaptation of the MCR-ALS to many different chemical problems.

All the algorithms used for NMF were the same as the written algorithm for previous data analysis. In addition we used the function *nnmf* in MATLAB for an initial estimation of component spectra and number of species observed in recorded spectra in multi-component analysis. We used this function because it is completely a blind source NMF algorithm and needs no initial assumption about the system of interest. The initial estimation can be also done by finding the pure variables during the titration. For MCR-ALS calculations, the algorithms implemented by Tauler and co-workers were performed using MATLAB routines.

Data pre-treatment

The raw UVRR spectra were smoothed by a Savitzky–Golay filter (2nd order polynomial, 21 points). Baseline correction and smoothing were applied to prepare the Raman spectra for further analysis. Since an appropriate choice of baseline correction methods could significantly improve subsequent analysis, we examined different baseline correction methods. In particular, the baseline contribution from phenylalanine fluorescence included in the structure of protein leucine zipper have

[1] http://www.cid.csic.es/homes/rtaqam/tmp/WEB_MCR/down_mcrt.html

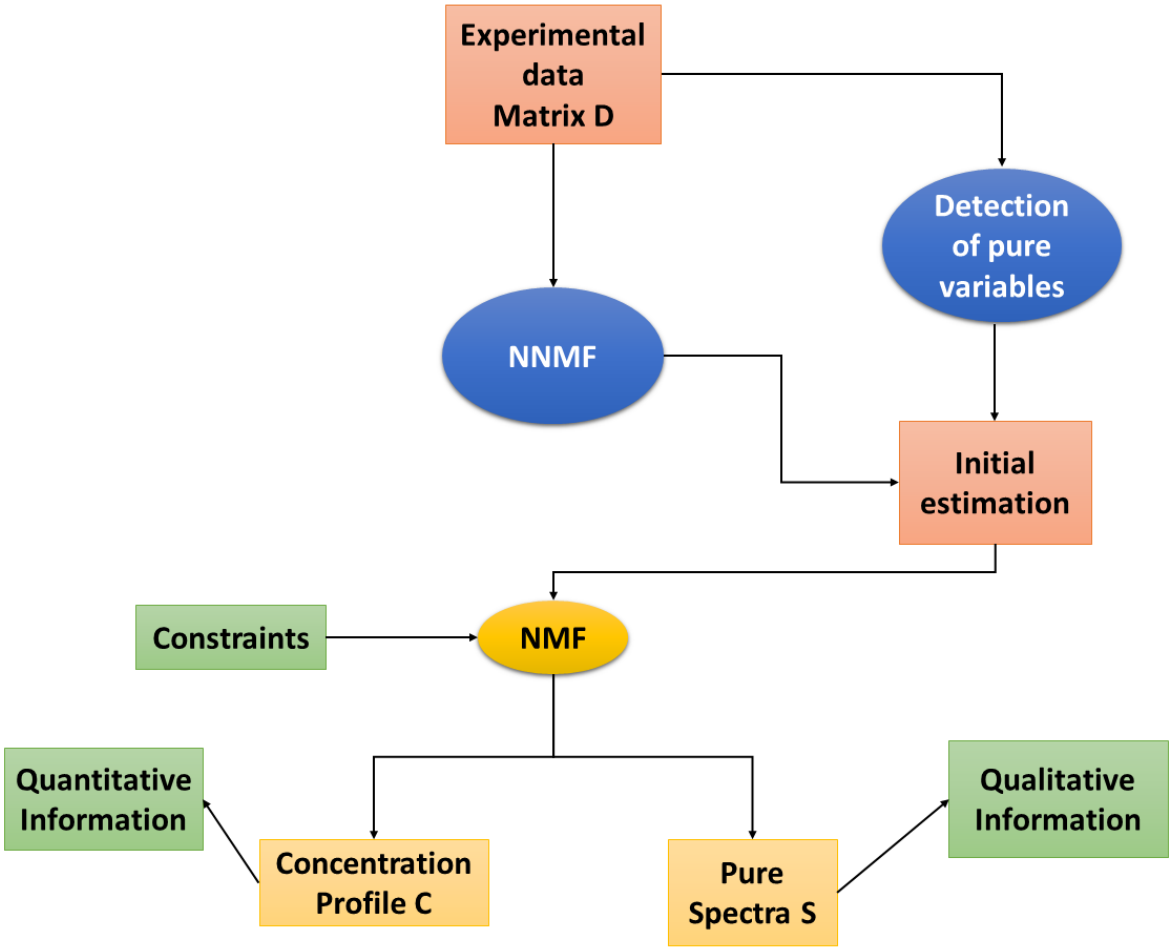


Figure 4.6: The flowchart of preparation steps for NMF and its final results.

different curvatures. These curvatures are dependent on the used protein concentration which was probably correlated with the protein self-absorption effect. Therefore, applying a method which could suitably match the baseline with different curvatures in various mixture spectra was important for the pre-treatment of the data. Although the polynomial curve fitting is used widely for removing fluorescence background, this technique traditionally relies on user-selected spectral locations for fitting the base to these points. It is time-consuming because the user must process each spectrum individually to identify the spectral regions which are non-Raman active. To address this limitation, the modified polynomial method was developed for fluorescence subtraction specially from biological Raman spectra [52, 53]. This method smooths the spectrum in such a way that Raman peaks are automatically eliminated. At the end of the process only the fluorescence baseline remains intact to be subtracted from the raw spectrum. The algorithm was written in MATLAB for which the minimal user intervention is required making the method less time-consuming than the other fluorescence background suppression. The baseline correction of one spectrum selected from binding study between compound **2** and

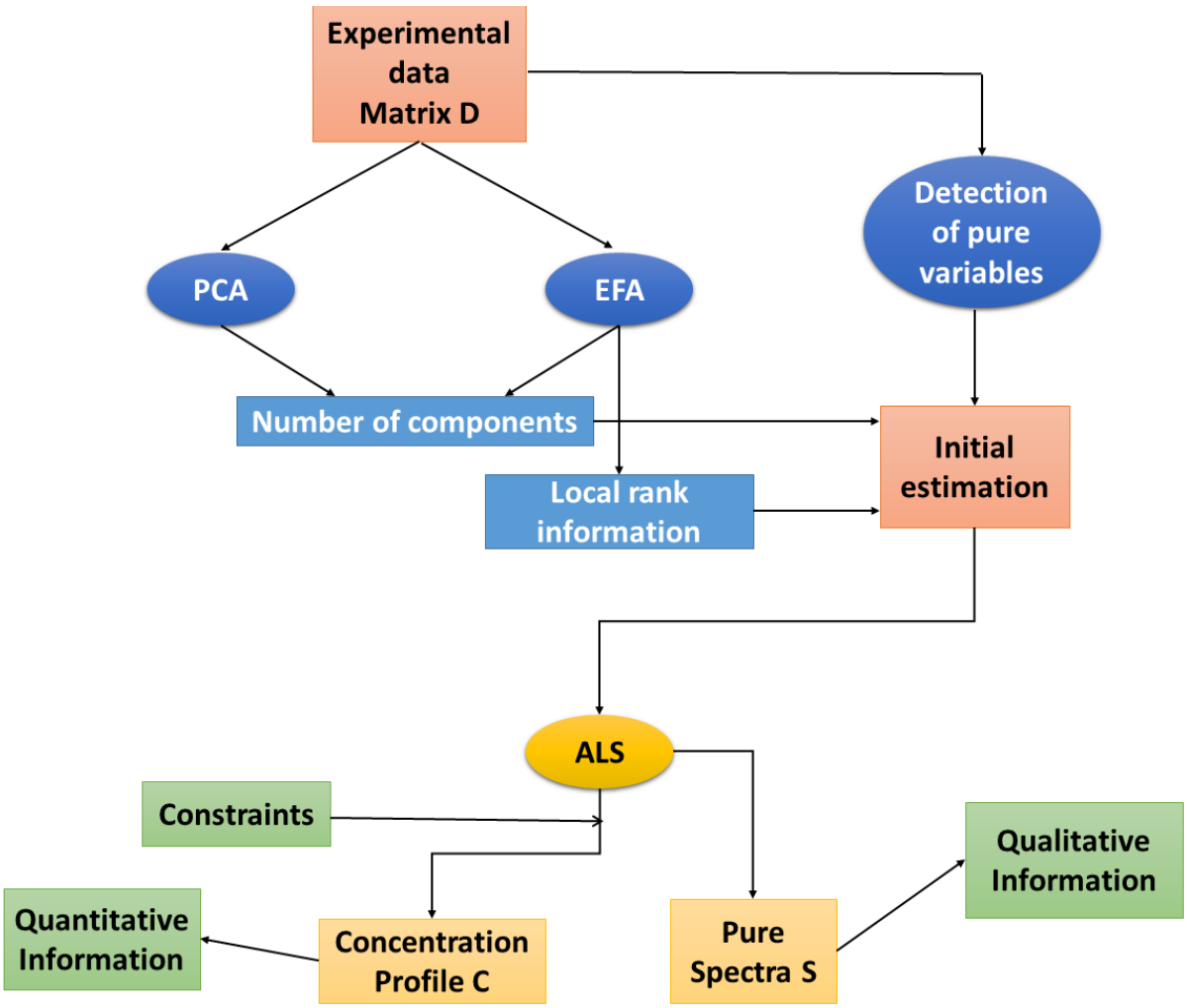


Figure 4.7: The flowchart of preparation steps for MCR-ALS and its final results.

leucine zipper, at the final point of titration, is displayed in Appendix A.1. The MATLAB code has been added to Appendix A.2. For auto-scaling, we used one band unaffected by titration, the corresponding normal mode of pyrrole hydrogens in the 2- and 3-position shown in Appendix A.1.3, as an internal standard for normalizing the UVRR intensities in order to correct the absorption and re-absorption upon protein addition. Mean centering was the last step of spectra preparation with which all spectra were normalized to the mean value of the Raman intensities of the neat ligand. As a result, the calculated concentration profile of complexed ligand can be expressed as the concentration ratio of complexed to the neat ligand with the value between 0 and 1.

Initial estimates

In the previous binding studies, the initial estimates were the component spectra including the spectrum of neat ligand for the "free" ligand while the spectrum at the last point of titration was considered for the "complexed"

component. Here different algorithms were utilized for this step. Especially for the case of multicomponent analysis, in which more than two spectra are needed for initial estimation, this step seems to be more essential. Evolving Factor Analysis (EFA) is a local rank exploratory method based on Principle Component Analysis (PCA) which can provide an initial guess for concentration profile with an initial estimation for the number of components in a multicomponent analysis. The components spectra can be estimated by selection of purest variables based on SIMPLISMA. Another simple alternative is running the NMF algorithm by initial random data for both concentration profile and component spectra. The results can be taken as the initial estimation of a constrained NMF or even for the initial estimation of either spectra or concentration for MCR-ALS.
All algorithms used for the initial estimation of either concentration profiles or spectra including *nnmf*, *simplisma* and *EFA* need a guess for the number of components which are going to be modeled. Only Evolving Factor Analysis (*EFA*) has the advantage to give an estimation for the number of components by visual interpretation of the process evolution curve acquired by singular value decomposition (*SVD*).

Chapter 5

Results and discussion

5.1 Setup for UVRR spectroscopy

Building a setup for guiding the light from the source to the sample and then to the spectrometer by using different optical components is the first step for any kind of spectroscopy. Especially for Raman spectroscopy, in which the intensity of scattered light is basically so weak, an efficient setup alignment in both parts of the focusing optics (from the source to the sample) and the collecting optics (from the sample to the spectrometer) is necessary in order to have a high signal to noise ratio.

Shown in Figure 5.1 is the schematic configuration of the UVRR set-up, in which the sample holder (quartz quvette) was aligned in a 90 degree scattering geometry. The laser beam diameter is 0.6 mm (± 20%). In some experiments, depending on the absorption of the sample, we need to illuminate a bigger area of the sample. For this purpose, a telescope (1) was used to increase the laser beam diameter by two times. After focusing the light into the sample by a set of mirrors and focusing lens (1 to 4 in Figure 5.2, f_1 =10 cm) and vertical alignment, the scattered light should be carefully collected to the entrance slit of the monochromator (6). The beam aperture Ω, defined as the active internal area (the illuminated area of grating) divided by the square of the focal length ($\omega = \frac{A}{f^2}$), was calculated to choose the right lenses for the collecting optics (5). Ω represents the normalized cone angle through the spectrometer and reflects the fact that the light density (irradiance) decreases with the square of the distance. If the Ω of the illumination and the spectrometer are identical, the losses are minimal and the spectrometric performance is the best. In contrast, f-number calculations only regard two linear dimensions, the focal length and one dimension of illuminated object ($f - number = \frac{r}{f}$) [83]. However, even with considering Ω instead of f-number, there is always some losses by reflection or absorption of the optical components used in the set-up which make the calculation complex. To compensate for the losses, the environmental circle of the grating (the green circle in Figure 5.3) was

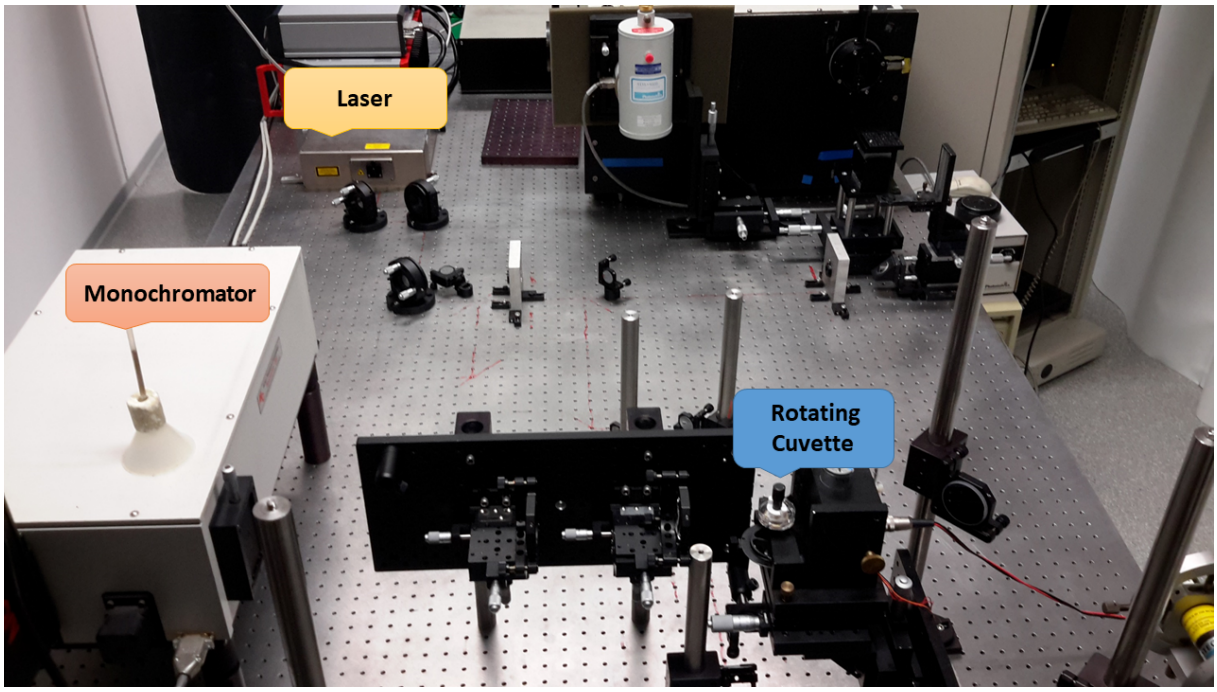


Figure 5.1: UVRR spectroscopy setup with 90° scattering geometry.

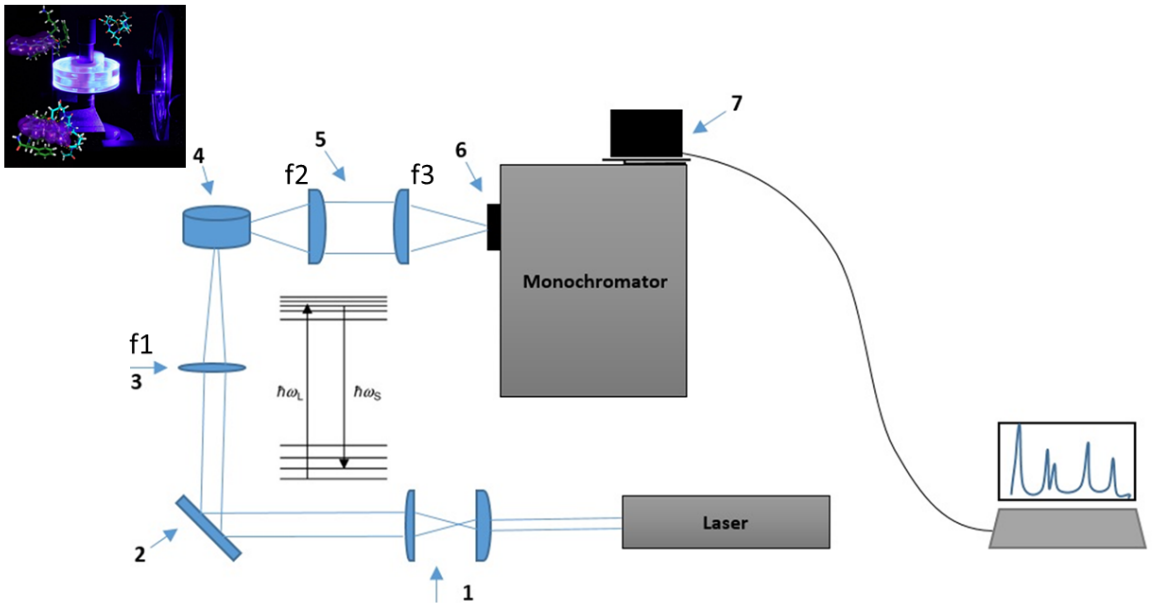


Figure 5.2: Schematic illustration of the setup with (1) telescope, (2) mirror, (3) focusing lens, (4) rotating cuvette for holding the sample, (5) collecting optics, (6) entrance slit and (7) CCD cooled camera.

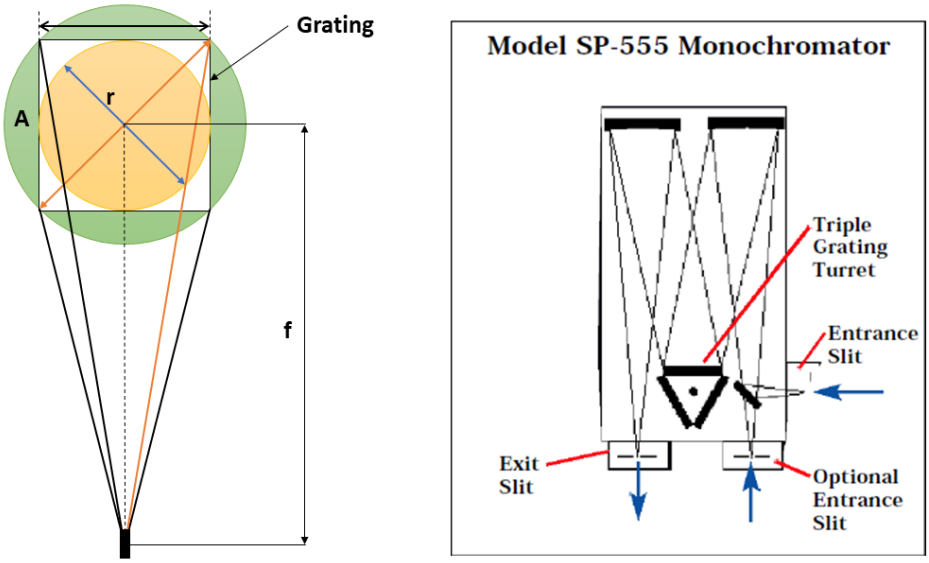


Figure 5.3: (left) Illustration of f-number versus ω, (right) the double monochromator and its components.

considered as the projected area from the entrance slit for calculating the Ω of the spectrometer. By matching this value with the Ω of the illumination, the focal length of the lenses for the collecting optic was calculated (f_2 =10 cm, f_3 =20 cm). The adjusting of the illuminations is more flexible with the arrangement of two lenses for collecting optics (telescope 5). While the focal length of second lens has to be selected to fulfill the equality of the Ω for the illumination and spectrometer, the focal length of the first lens can be small enough in order to collect more scattered light from the sample. An accurate focusing and alignment are important for a Raman spectroscopy set-up because the observed Raman intensity are dependent on them. However, even on a particular instrument and under apparently identical experimental condition, a significant variation of intensity may occur for a given sample from day to day. Therefore, a calibration is necessary for correcting the variation of the instrument response across a Raman spectrum. In fact, calibration for the spectroscopy is the process of preparing the software (e.g., LightField) to assign appropriate calibration values over the scanned range of an acquired spectrum. In this regard, we used a mercury lamp as a light source with calibration lines mounted in front of the entrance slit of the spectrometer and applied the standard calibration. A standard calibration is a broad calibration that precisely calibrates the movement of a spectrometer grating using the spectrometer stepper motor. Since this application of Raman spectroscopy does not require the evaluation of absolute intensity, a standard calibration was performed for wavenumber without calibrating the absolute intensity.

However, the reproducibility of the observed intensity has been always the main basic issue of the photometric accuracy of Raman spectroscopy. Therefore, we used cyclohexane for recording a reference Raman spectrum in order to check the variation of the instrumental set-up. Shown in Figure 5.4 is the UVRR spectrum of cyclohexane acquired in illumination time of 0.0001 seconds and 10 accumulations. This spectrum was used for correction of the instrument intensity response from time to time.

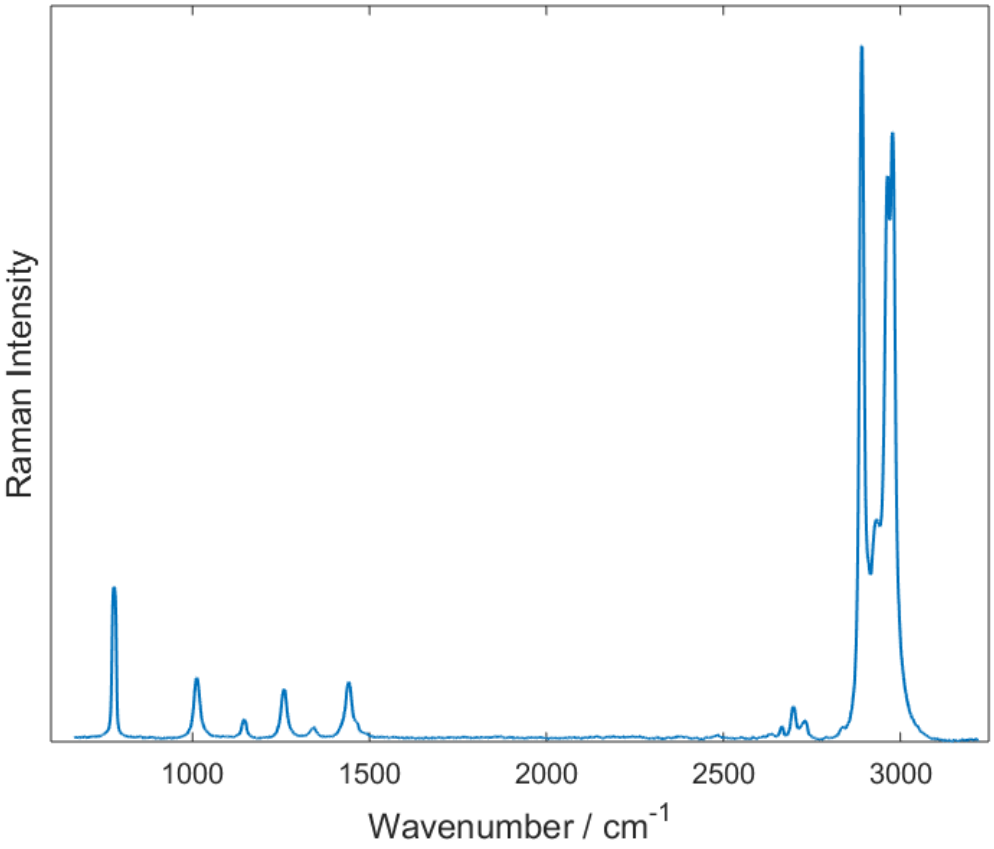


Figure 5.4: The UVRR spectrum of cyclohexane used as a reference spectrum for checking the response of the system.

5.2 Preliminary studies

Transition from peptide recognition to protein recognition by multivalent supramolecular ligand instead of monovalent ligand requires the optimization of some experimental parameters such as concentration and pH. The three-armed GCP ligand (Li-40 shown in Figure 4.1) was utilized as a multivalent compound to find the optimum pH and concentration of the ligand.

5.2.1 Concentration-dependent UVRR spectroscopy

From two points of view, it is important to find the optimum range of ligand concentration for the binding studies experiments. First, though the concentration of ligand is kept constant in a ligand-based methodology (i.e., UVRR binding studies), this concentration defines the relative ratio of protein concentration for the titration. So, it is important to define a minimum range of concentration suitable for protein titration, which should be also high enough for detecting a good signal. The previous experiments for monitoring tetrapeptide recognition by UVRR were performed at the ligand concentration of 1 mM which is very high for protein titration. Secondly, Raman intensity versus concentration has a nonlinear dependence due to the sample self-absorption effect whereas linearity is a condition for applying multivariate data analysis methods as it was stated in Equation 2.12. Therefore, we need to work in a region of concentration in which the dependency of intensity versus concentration is linear. Consequently, we applied the self-absorption correction to investigate the nonlinearity of Raman intensity versus concentrations for three multivalent GCP-based ligands. From the resulting nonlinear intensity curve, minimum range of concentration with a linear approximation for each ligand can be found.
The well known Beer-Lambert law states that the intensity of a traveling light beam is exponentially attenuated as it transverse an absorbing material. Accordingly, the incident intensity I_0, after traveling the distance z through the absorbing sample, is exponentially attenuated to become I, as shown at the upper left of Figure 5.5,

$$I = I_0 e^{-\epsilon_\lambda c z} \tag{5.1}$$

where ϵ_λ is the molar absorptivity at the wavelength λ, and c is the concentration of the absorbing species [84]. In a 90 degree scattering geometry, the self-absorption attenuation of the light happens two times, as it is shown in the traveling path of the laser beam through the sample in a point scattering model in Figure 5.5. The light is attenuated first by the absorption of the sample in the rotating cuvette during the path length of l before reaching the scattered point (absorption), then the scattered light will be suffered again by the sample absorption from the scattered point to the cuvette wall (re-absorption through the path lengths of r). Therefore, the incident light is attenuated on its passage through the sample before reaching the spectrometer by the following function

$$I = I_0 e^{-kc(l\epsilon_i + r\epsilon_s)} \tag{5.2}$$

where ϵ_i is the molar absorptivity at the incident wavelength, ϵ_s molar absorptivity at the scattering wavelength, and $k \equiv loge = 2.303$.

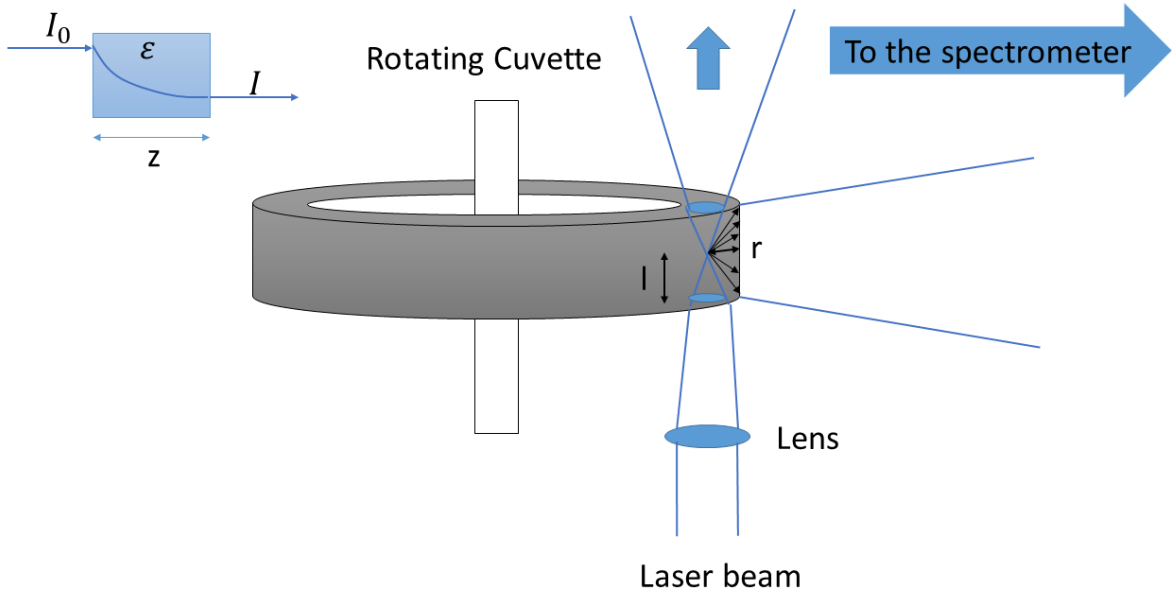


Figure 5.5: (left) Exponential attenuation of the light as it transverses an absorbing sample, (right) absorption and re-absorption (self-absorption) by the sample in a rotating cuvette and a 90° scattering geometry.

Combining with the scattering function $I_s = JIc$ where J is the molar scattering coefficient, we have the formulation for the observed Raman intensity vs concentration.

$$I_{obs} = I_0 Jce^{-kc(l\epsilon_i + r\epsilon_s)} \tag{5.3}$$

This equation has a maximum which gives the optimum concentration. It is more practical to normalize the function to the maximum observed intensity for avoiding the calculation of molar scattering coefficient (J) which results in a normalized curve

$$\frac{I}{I_{max}} = Ace^{(1-Ac)} \tag{5.4}$$

For calculating the coefficient $A = k(l\epsilon_i + r\epsilon_s)$, we need the molar absorptivity at the incident and scattered wavelength, which can be acquired by measuring the absorption spectrum of the sample. The half height of the cuvette containing the sample is considered as an approximation for the incident path length l. The exact measurement of the scattered path length r is quite complicated since it is an integral of all the possible routes through which the scattered light can travel from the scattering point to the cuvette wall [85]. A simple approximation is to assume an effective path length traversed by the scattered radiation under experimental conditions. This effective path length can be determined by measuring the relative intensity of two Raman bands as a function of the

sample concentration. With a known effective path length, one can use Equation 5.4 to construct the concentration-dependent Raman intensity curve for one Raman band, allowing the determination of optimum concentration [84]. We used this approximation method to anticipate the non-linearity of intensity versus concentration for the UVRR spectra of the ligands under our experimental condition.
First, the Raman spectra of Li-40 (see Figure 4.1) in different concentrations between 0 and 500 μM were recorded. Three selected measured spectra are shown in Figure 5.6 (left). While increasing the concentration from 75 to 200 μM caused more intense Raman bands in the entire spectrum (whole range of wavelengths), the intensity at 400 μM decreased dramatically and it illustrates the non-linearity of scattered Raman intensity of the entire spectrum versus concentration. Two Raman bands at 272.6 nm and 276.8 nm were chosen to estimate the effective scattering path length based on the equation $logR_2 = logR_1 - rc\Delta\epsilon$ [85], where R_1 and R_2 are the Raman intensity of two identified bands, r is the effective path length, c is the concentration and $\Delta\epsilon$ refers to the difference between the molar absorptivities of two wavelengths taken from UV-vis absorption spectrum (Figure 5.6, right). The fitted plot of this equation to the experimental values is shown in Figure 5.7 (left). The obtained straight line implies that the path length r can be effectively considered as a constant value under one experimental condition. Therefore, the calculated value of r (0.047 cm, calculated from the slope r$\Delta\varepsilon$) is used as the initial guess to fit the predicted formulation of normalized Raman intensity versus concentration in Equation 5.4 with the experimental data.

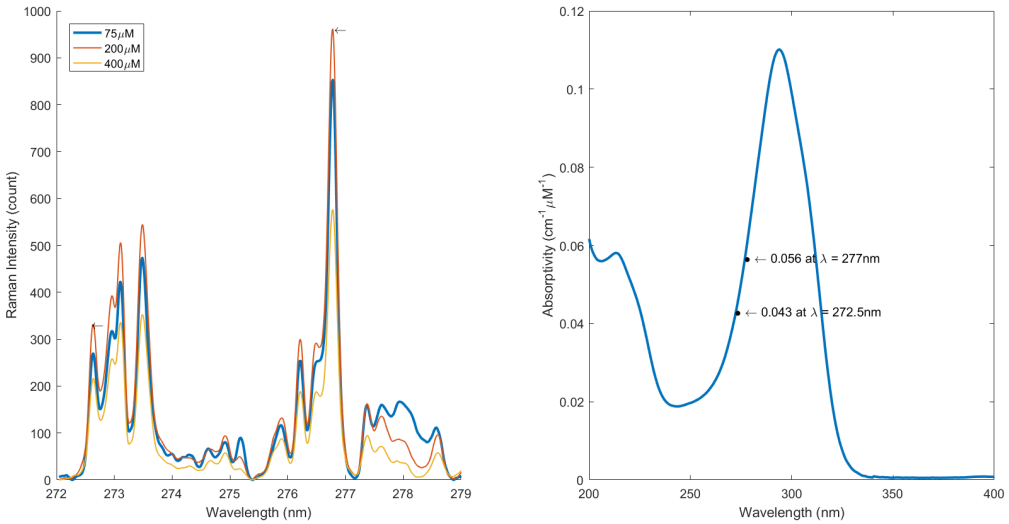


Figure 5.6: (left) concentration-dependent UVRR spectra of Li-40, (right) the molar absorptivity plotted against wavelength.

The Raman band at 276.8 nm was used for this purpose. Data points in Figure 5.7 (right) are Raman intensities normalized to the maximum

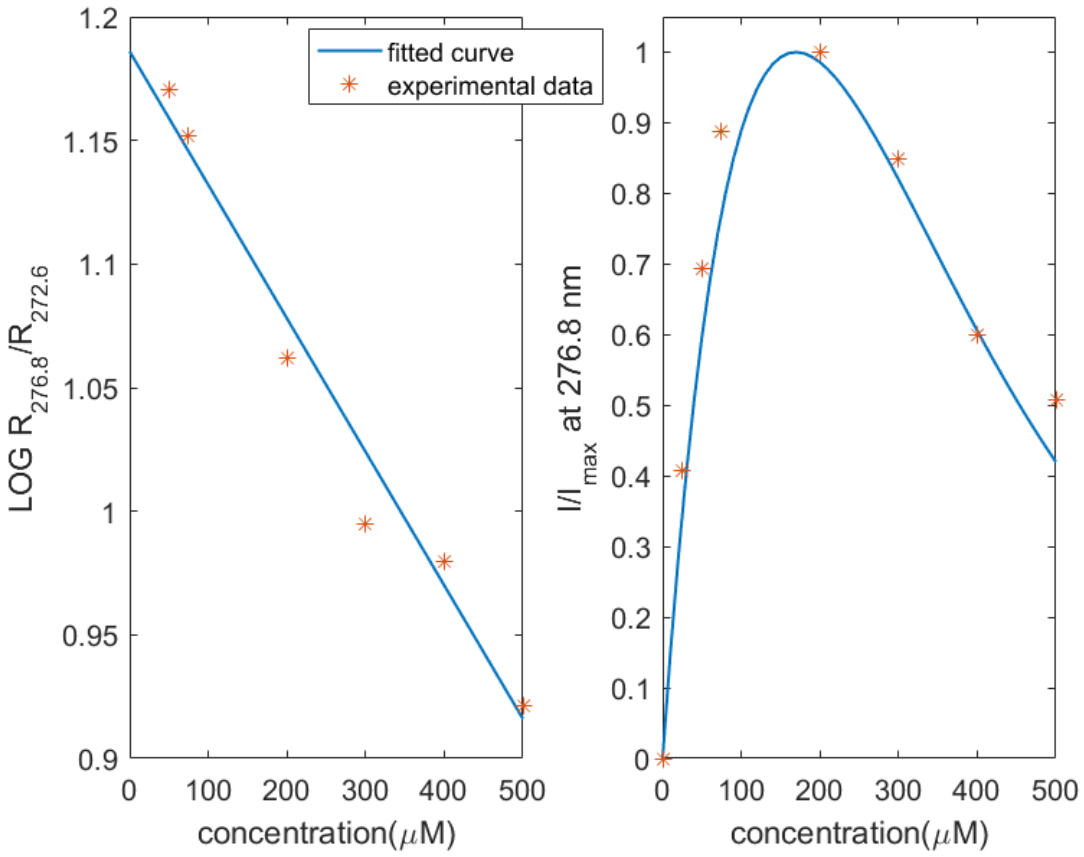


Figure 5.7: The nonlinear concentration-dependence of the Raman intensity for Li-40, (left) the logarithmic ratio of two selected Raman band intensities, the slope of the curve is an initial guess for the effective path length of scattered ligth, (right) measured Raman intensity at 278.8 nm, normalized to the maximum intensity.

intensity. The MATLAB function *fminsearch* has been implemented in a least square manner to evaluate the coefficient value (A in Equation 5.4) with the initial estimation of 0.047 cm for effective path length.
This numerical analysis was used for predicting the self-absorption function of compound **1** and compound **2**, measured in the same experimental geometry. In Figure 5.8, three different nonlinear plots of Raman intensity vs concentration are displayed, one of which (blue for Li-40) was acquired by using the experimental data. With considering the same experimental condition, the self-absorption effect of two other molecules was anticipated (yellow for compound **1** and red for compound **2**). Three ligands have different optimum concentration with maximum intensity which is obviously due to their different absorption coefficients. However, a same minimum range of concentration with a linear approximation can be assumed for three compounds ($0 < c < 100$ μM).

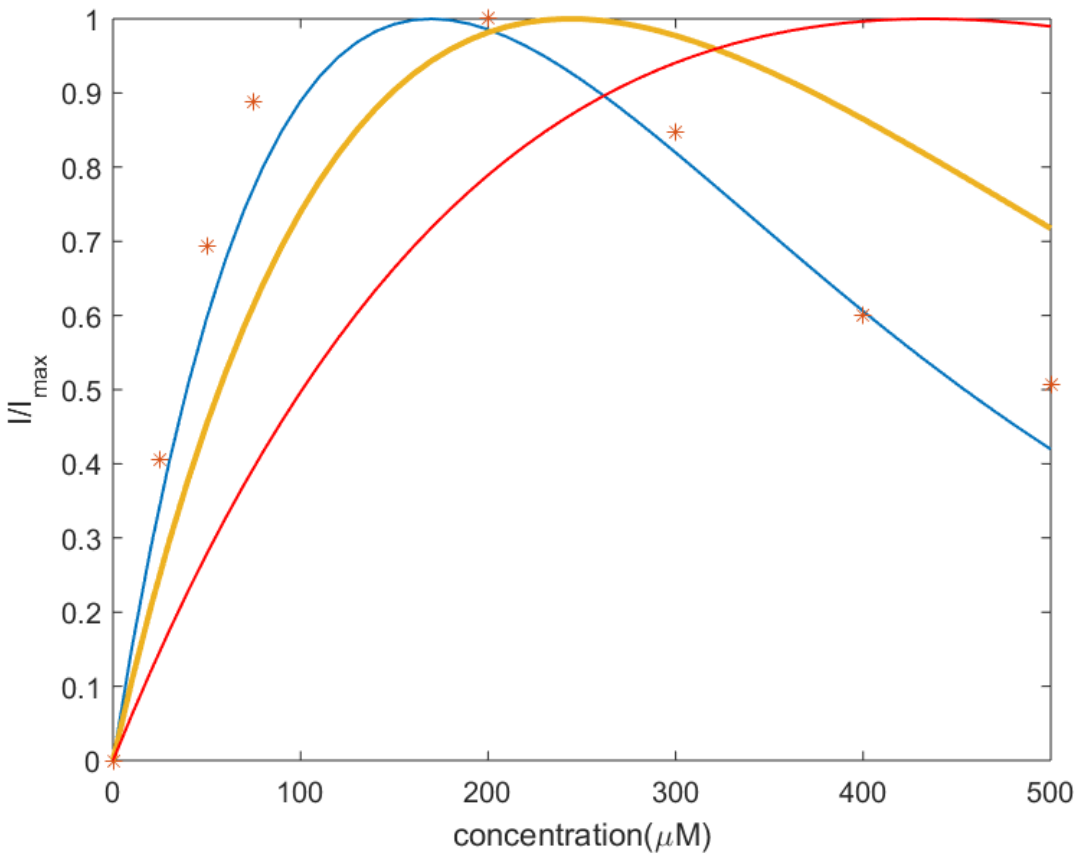


Figure 5.8: Normalized Raman intensity versus concentration for (blue) Li-40, (yellow) compound **2**, (red) compound **1**. The data points are the experimental data calculated for Li-40.

5.2.2 pH-dependent UVRR spectroscopy

The strong electrostatic interaction between the two oppositely charged binding partners, protonated GCP subunit (positively charged guanidinium cation) and deprotonated carboxylate, yield an efficient binding in molecular recognition of supramolecular ligand. Hence, the optimal pH for binding event requires a determination of pK_a values of the protonated ligand and deprotonated carboxylate of the protein receptor. The result of the previous site-specific pK_a value determination by UVRR [13, 86], with the contribution of two species of protonated and deprotonated GCP in acid/base equilibrium, was 6.5 for GCP motif. However, multivalent ligand molecules are in the category of multiprotic systems which can donate or accept more than one proton. Therefore, there are more than two species that should be taken into account in finding the optimal range of pH value for the dominating protonated form of GCP. In fact, the fully protonated multivalent ligand causes more efficient complexation rather than a partially protonated ligand. So, in the case study of multivalent ligand molecules with more than one GCP motif, there should be the characterization of more than two species, namely, the fully protonated, the intermediate and the fully deprotonated ligand (Figure 5.9). The number of the intermediate species (named as partially deprotonated ligand) depends on the number of GCP motifs included in the structure of the ligand.
Generally, the concentration profiles of the species involved in the acid-base equilibrium of an *n-protic* ligand (L) can be described as follows

$$[H_nL^{n+}] = c_L \frac{[H^+]^n}{[H^+]^n + K_1[H^+]^{n-1} + K_1K_2[H^+]^{n-2} + K_1K_2...K_n} \quad (5.5)$$

$$[H_{n-1}L^{(n-1)+}] = c_L \frac{K_1[H^+]^{n-1}}{[H^+]^n + K_1[H^+]^{n-1} + K_1K_2[H^+]^{n-2} + K_1K_2...K_n} \quad (5.6)$$

$$[L] = c_L \frac{K_1K_2...K_n}{[H^+]^n + K_1[H^+]^{n-1} + K_1K_2[H^+]^{n-2} + K_1K_2...K_n}. \quad (5.7)$$

In these equations, $[H_nL^{n+1}]$, $[H_{n-1}L^{(n-1)+}]$ and $[L]$ are the equilibrium concentrations of fully protonated, intermediate and fully deprotonated ligand, respectively. Obviously, the concentration profiles of all involved species are expressed as a function of the total concentration of ligand, c_L, and the suitable acidity constant, K_i. For estimation of these constants, a pH-dependent UVRR titration was performed for the tri-protic ligand Li-40 in order to find the range of pK_a for different acidic species. Some of the measured spectra in different pH are displayed in Figure 5.10.

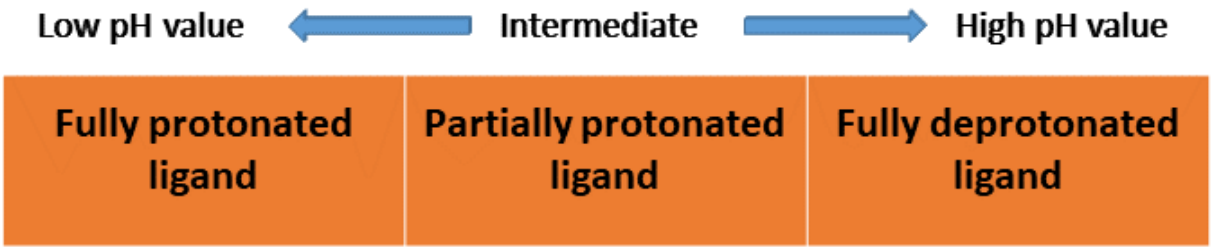


Figure 5.9: Species in acid-base equilibria of a multi-valent ligand.

In order to derive the concentration profiles of different species from the experimental spectra, we initially considered four species, as it is anticipated for a 3-protic compound. These are the maximum numbers of protic species which may exist at different pH values. Non-negative matrix factorization algorithm (*nnmf* in MATLAB) was applied to factorize the data matrix, including spectra at different pH between 3 and 10, into two matrices of lower rank which are the spectra of four species and their concentration profiles. The factorization is blind source because it uses an iterative method starting with random initial values for both factorized matrices. The UVRR spectra calculated for four probable species and their contribution are depicted in Figure 5.11. Among the four randomly calculated spectra, the spectra 2, 3 and 4 are somehow

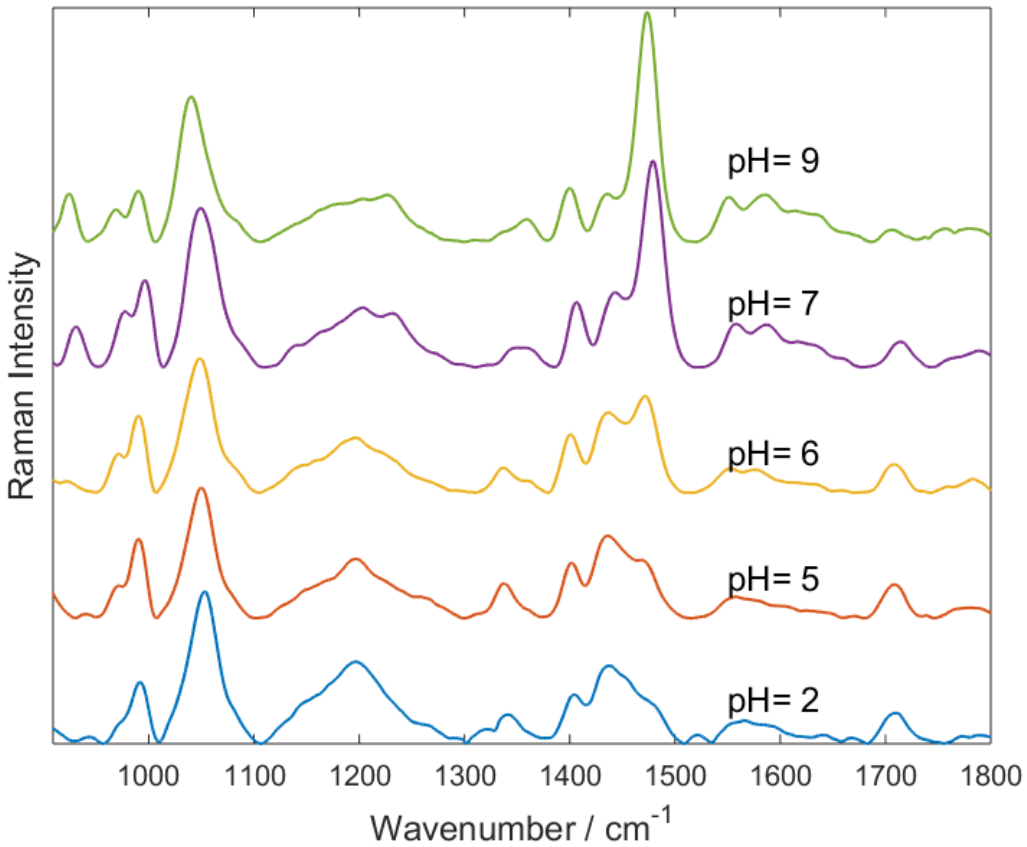


Figure 5.10: The pH-dependent UVRR spectra of three-armed ligand Li-40.

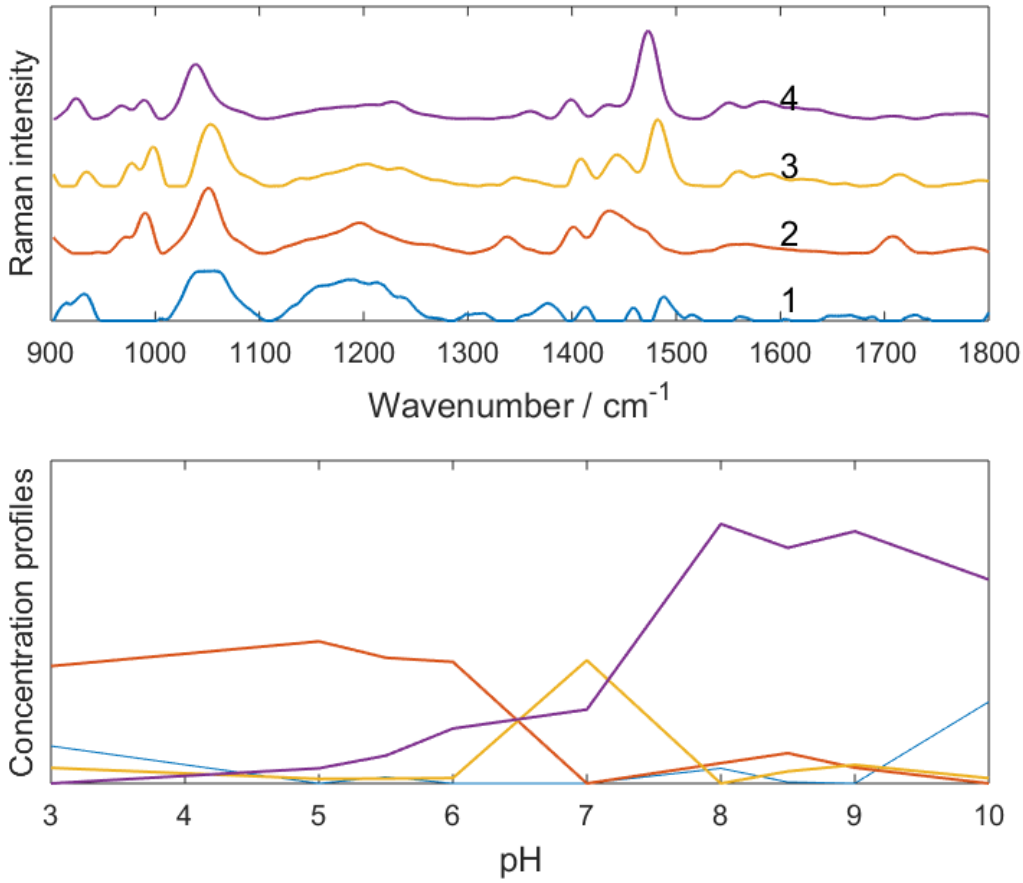


Figure 5.11: (top) The spectra and (bottom) the concentration profiles of four species calculated by unconstrained Non-Negative Matrix Factorization (NMF) from the UVRR spectra in Figure 5.10.

comparable with experimental data (with spectra at pH= 3, 7 and 10, respectively). This matching between the randomly calculated spectra and experimental data is supported by their calculated concentration profiles depicted at the bottom of Figure 5.11. The concentration of component 2 (red) decreases from a maximum value at lower pH to zero at higher pH value. This behavior resembles a protonated component. The calculated concentration of component 4 (violet) has a reversal behavior and resembles a neutral component, while the maximum contribution of component 3 (yellow) is between pH 6 and 8 and it is nearly zero in other pH values and, as a result, can be considered as the intermediate component. The spectrum of component 1 (blue) is not in accordance with experimentally recorded Raman bands of GCP motif. Therefore, the essence of only three principle species is assumed, indicating that two components have probably very close pK_a values and, hence, posses undistinguished spectra and concentration profiles.

In order to perform MCR-ALS on the pH-dependent spectra, to find spectroscopically distinguishable components with the assumption of three species, an initial estimation for the spectra of protonated, intermediate

and deprotonated species were needed. For this purpose, the spectra at pH=3, 7 and 9 were selected by *pure* algorithm as the initial guess. As mentioned in section 2.3.2, this algorithm finds the purest variables in a data set based on the SIMPLISMA method and is useful especially when more than two components are hidden in the data, because the presumption of the initial corresponding spectra is not simple in that case. Assuming that every component in the mixture under study has a variable (e.g., pH or wavenumbers), we applied the algorithm with pH as the variable and asked for three purest variables to be selected. The spectral set is consequently fitted with one, two and more spectra until the fitting residual shows high relative standard deviation. The found purest spectra by SIMPLISMA based on their standard deviation profiles are depicted in Appendix B.1. With these spectra as initial estimations, we applied MCR-ALS on the experimental data set. The **closure** constraint was applied for the optimization of ALS, because the sum of concentration contributions of three principle species is constant for each pH value. Figure 5.12 (top) shows the calculated concentration profiles of three species which identify how these principle species change over different

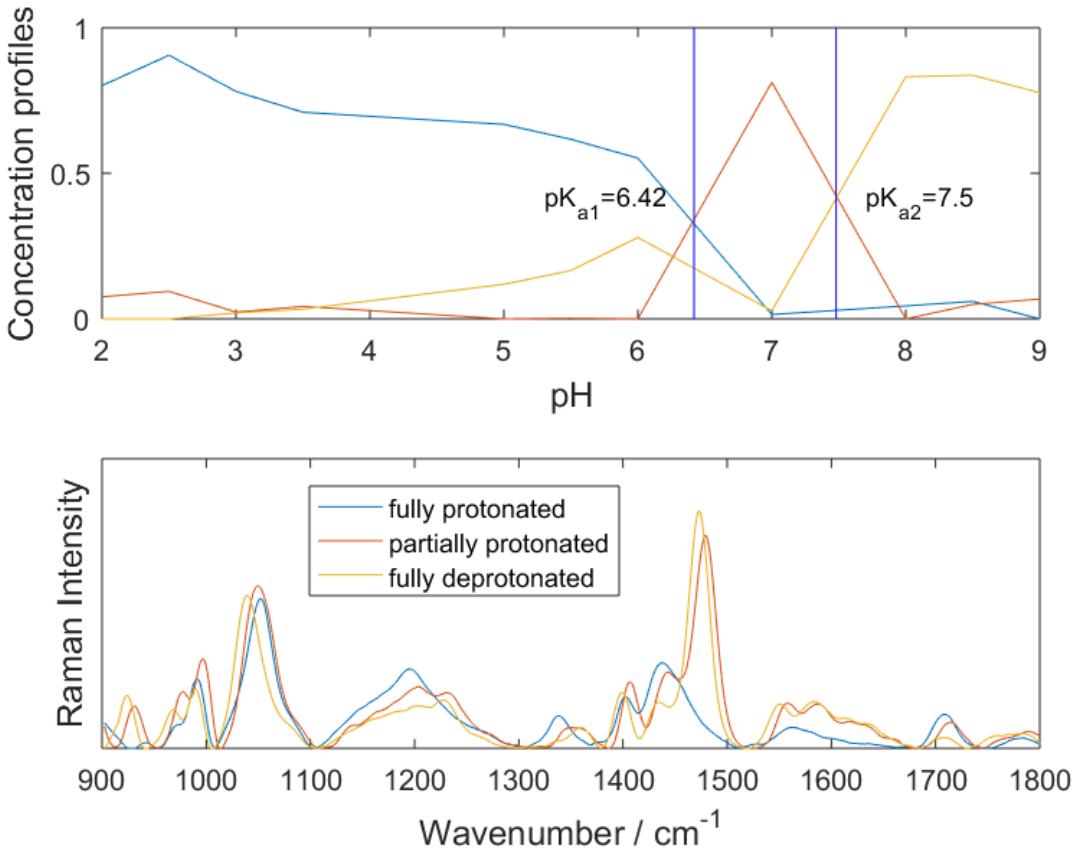


Figure 5.12: The calculation results of MCR-ALS for (top) the concentration profiles and (bottom) the spectra of three species. Lack of fit (LOF) = 8.2695 %.

pH values. At lower pH and mainly between pH=3 and 6, the protonated species predominates. Under pH between 6 and 8, most of the species are in the intermediate protonated states. Deprotonated species are found mostly at pH value higher than 8. Two pK_a values are identified at two equilibrium points.

5.3 Binding studies with the protein Dermcidin

In the sequence of protein dermcidin structure, there is one histidine which can be the origin of some enhanced Raman bands. The important issue, that should be taken into account in our binding evaluation, is the overlapping of these bands with that of GCP motif, especially in the region between 1400 and 1500 cm^{-1} where the changes due to binding are expected. As a result, the UVRR spectrum of dermcidin should also be treated and considered in the binding evaluation. Therefore,

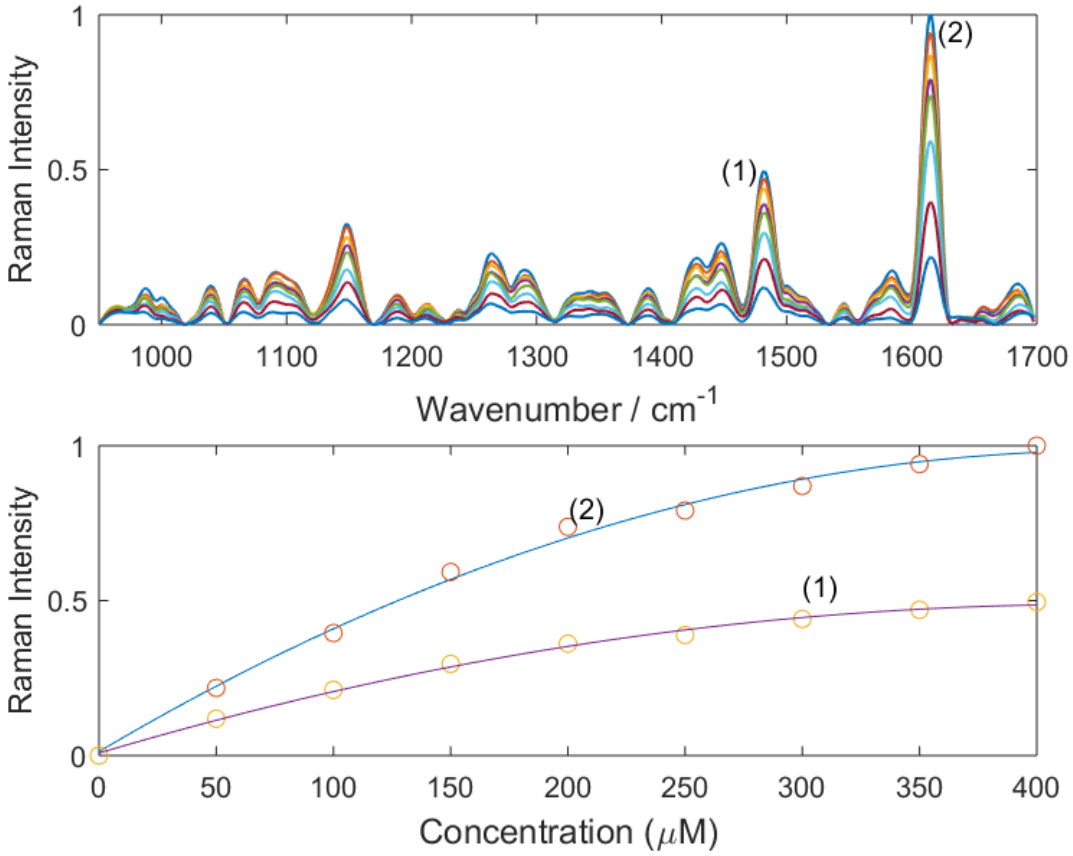


Figure 5.13: (top) The UVRR spectra of protein dermcidin in different concentration from 50 to 400 μM with the interval of 50 μM, (bottom) the nonlinear changes of two Raman marked bands versus concentration.

we performed a concentration-dependent UVRR experiment of protein

dermcidin to find the optimum region of concentration for further protein titration experiments. The spectra are shown in Figure 5.13 (top) and the general nonlinear behavior of two selected Raman bands intensity versus concentration is observed at the bottom of the figure. For accounting the

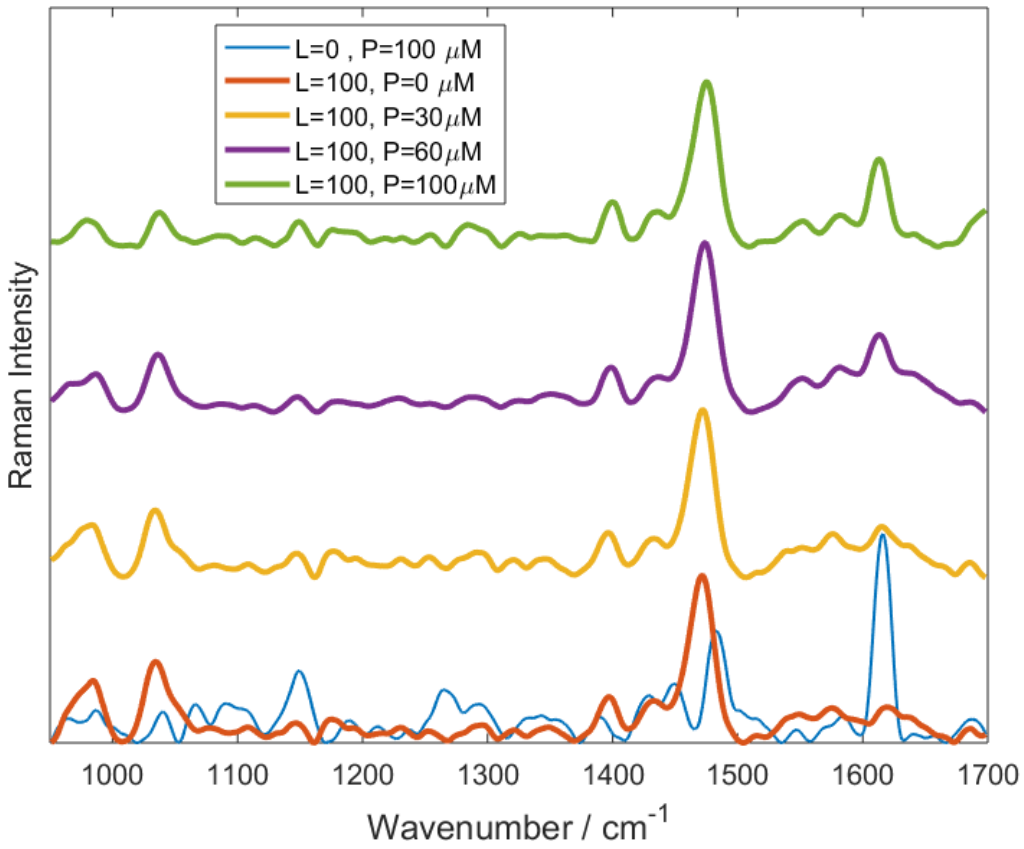


Figure 5.14: The spectrum of pure protein dermcidine (blue, bottom) and UVRR titration of compound **1** (constant concentration: 100 μM) with protein dermcidine from bottom (red) to top.

spectrum of dermcidin in binding studies by multivariate data analysis methods, the Raman intensity versus concentration should be linear. This is the criterion for using multivariate data analysis methods. The result from concentration-dependent experiment showed that the linearity of Raman intensity versus concentration is fulfilled in the range of 0 - 100 μM. So, we performed the UVRR titration experiment in this region.

For the UVRR titration experiment, the solutions were prepared with constant concentration of ligand (compound **1**) at 100 μM and pH 7.00 ($\pm$ 0.05) while the protein concentration was changed from 0 to 100 μM. With this compound, regardless of calculated pK_a value, we tried to do some experiments at neutral pH (pH 7). Four selected UVRR spectra acquired at different titration points are displayed in Figure 5.14. Additionally, the spectrum of neat protein was added (in blue) to show the interference of the Raman bands of the protein with ligand. Therefore,

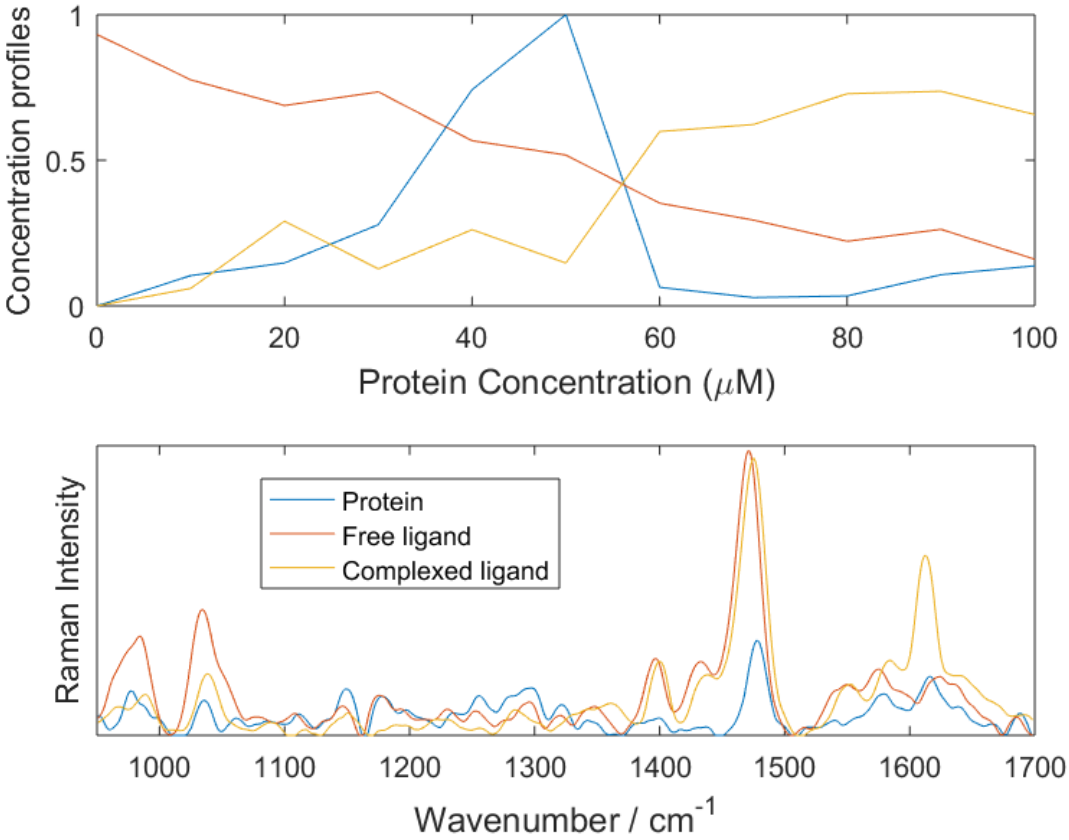


Figure 5.15: The calculation results of MCR-ALS for binding studies between compound **1** and protein dermcidin, (top) Concentration profiles and (bottom) UVRR spectra of three components. Lack of fit (LOF) = 8.4677 %.

the apparent unchanged intensity of Raman bands during the titration in this region can be a result of this interference. The Raman band at ca. 1614 cm^{-1} from protein is also clearly intensified during the titration. Since this band does not interfere with the spectrum of ligand, it can be used as an internal standard for the calculation of concentration profiles. Shown at the top of Figure 5.15 are the concentration profiles of three components: protein, "free" and "complexed" ligand which were calculated by ALS with two constraints of non-negativity and closure. The non-negativity constraint was applied for both concentration profile and spectra of species. We used the closure constraint for fulfillment of the concentration balance for both components, separately. In this experiment, since the concentration of ligand is constant, one closure constraint was used to keep a constant sum ratio for "free" and "complexed" ligand during the calculation. A separate constraint was applied for protein concentration to consider this point into calculation that ligand was titrated by protein up to one equivalent. The result of the calculated concentration profiles shows the complexation (shown in yellow) mainly starting from the protein

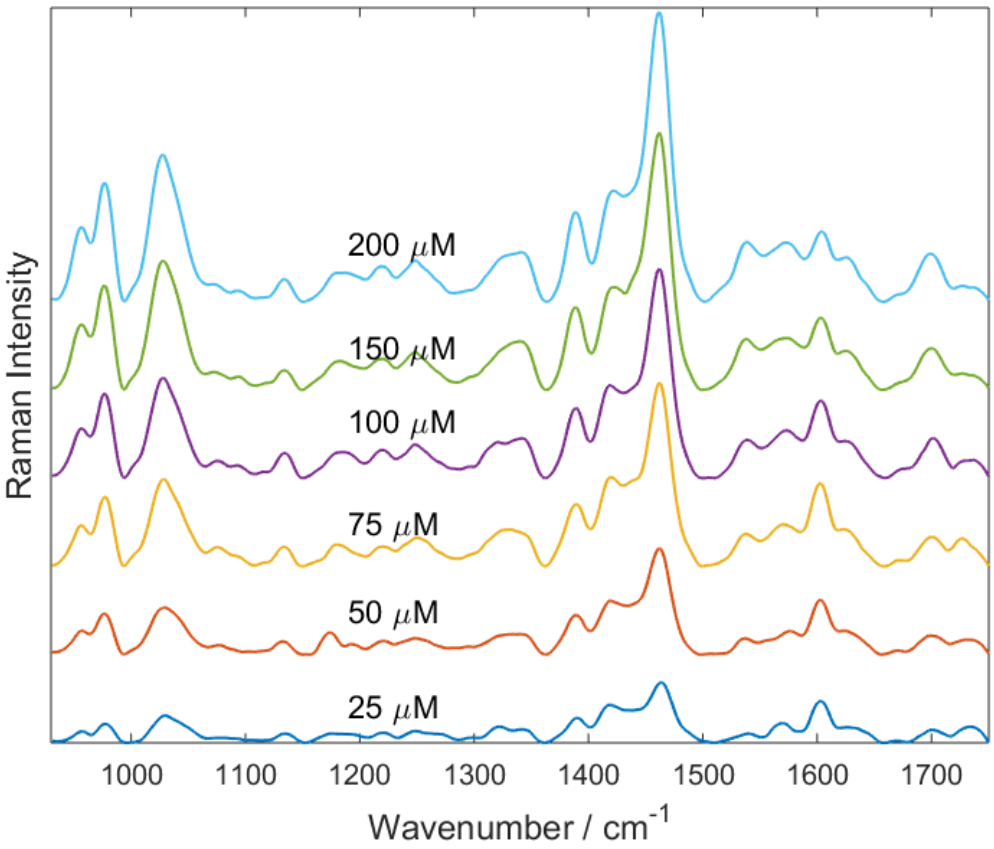


Figure 5.16: The UVRR titration of the protein dermcidin at constant concentration of 25 μM with increasing concentration of ligand (compound **1**) from 25 to 200 μM.

concentration at 50 μM where the contribution of protein (blue curve) significantly drops down. However, the change in the contribution of the complexed ligand is dramatically small so that its concentration slowly increases by addition of the protein concentration from 60 to 100 μM and even at the final point of the titration it does not reach the full potential of ligand concentration. This small change which is also clear by a comparison between the spectra of free and complexed ligand (Figure 5.15, bottom), could be due to either the weak binding or the low contribution of protonated species at pH 7. Therefore, a protein-based UVRR titration experiment at pH 6.5 was performed to reveal the main reason. The spectra are displayed in Figure 5.16. The protein concentration was kept constant in all mixture solutions at 25 μM and the concentration of ligand was changed from 1 to 8 equivalents of protein. Apparently, the change in the spectra is due to concentration increase of the ligand which dominantly overcomes the change caused by binding. Because of the ambiguity of the initial estimation of the complexed ligand spectrum, we performed non negative matrix factorization (NMF) initiating by random estimation of three components spectra (*nnmf* function in MATLAB) in order to

observe the general trend of the components concentration contribution during the titration. The results are displayed in Figure 5.17. The disadvantage of using NMF by random initial estimation is the uncertainty of the results regarding the attribution of the concentration profiles and spectra to each component. One solution is to use the information we have about the system for this attribution. For example, the blue curve can be easily attributed to the free ligand because it has a continuous increase in concentration and we expected such an increase due to the manner of our titration. Moreover, the corresponding spectrum also

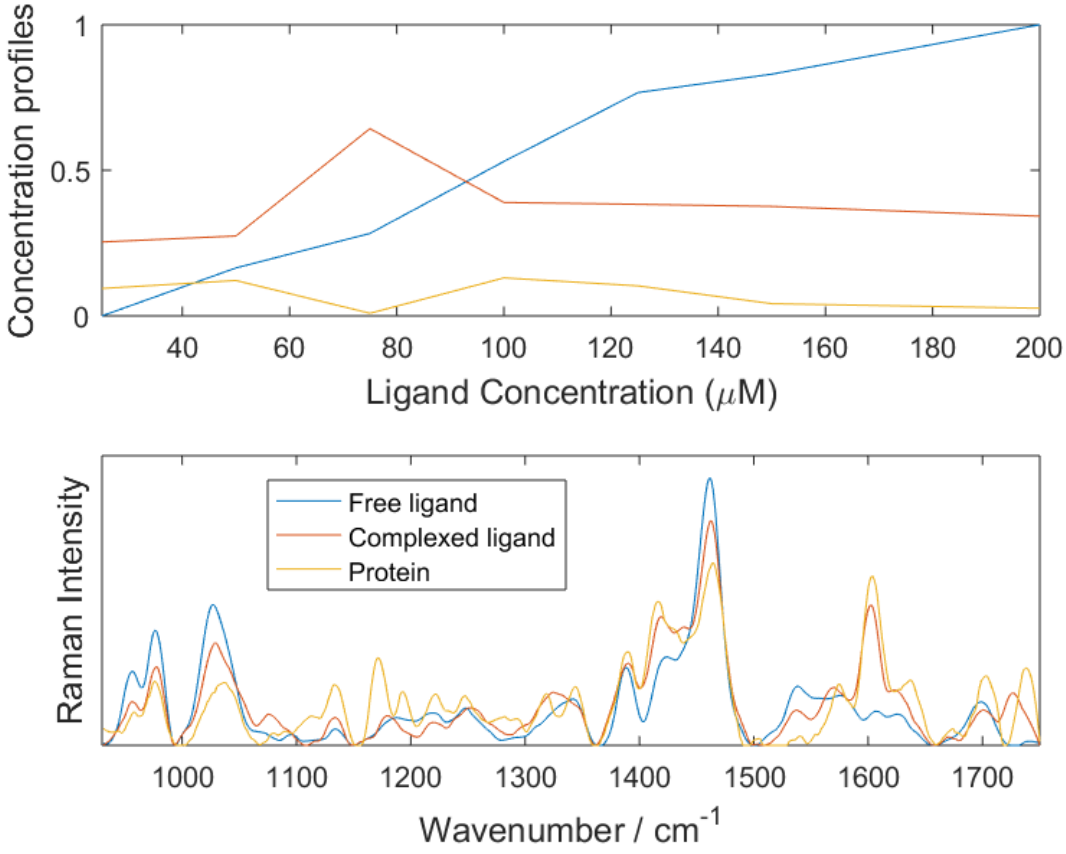


Figure 5.17: (top) The concentration profile and (bottom) the spectra of three distinguished species calculated by NMF from Figure 5.16.

belongs to neat ligand since there is no Raman band at ca. 1614 cm^{-1} which shows the presence of protein. We attributed the yellow curve to the protein because we do not expect an increase in the concentration of protein relative to its initial amount in a protein-based titration. The red concentration profile which is ascribed to the complexed ligand shows just a small change between 50 and 100 μM of the titration range. Consequently, the concentration of protein shows a reversal behavior in this range. In general, we could not extract new additional information about the binding by the ambiguous results of random NMF. However, with considering the concentration profiles in Figure 5.14, a weak binding is assumed between

compound **1** and protein dermcidin.

5.4 Binding studies with the protein Leucine Zipper

As mentioned before, leucine zipper with one phenylalanine shows auto-fluorescence upon 266 nm excitation and no detectable Raman bands. Therefore, it is not possible to do the concentration-dependent experiments for this protein. Although the lack of this experiment does not harm the final result since the protein spectrum is not considered in the multivariate data analysis, setting up a suitable range of concentration is necessary for doing a reliable experiment. For choosing the right concentration ratio of host/guest for the titration experiment, especially for the systems with more than one binding sites on the host and guest molecules, it is helpful to have at least some knowledge of what the stoichiometry of host and guest is. Therefore, we used UV-vis absorption and UVRR spectroscopy to initially estimate the stoichiometry or at least find a proper titration range for the main UVRR binding studies.

5.4.1 Estimation of stoichiometry

UV-vis absorption experiments

First, we used UV-vis absorption spectroscopy to monitor how the extinction of the supramolecular ligand changes upon addition of the protein Leucine Zipper. In Figure 5.18 the absorbance around 200 nm (amide backbone) shows a successive increase upon protein addition, but the changes of the ligand absorbance at 299 nm are small. A closer look in the region of ligand absorption, shown in the inset of Figure 5.18, reveals that the change after initial addition is more considerable compared to the changes which occur later at higher equivalents of protein. The only outcome of this experiment is an estimation of weak binding. Since the changes do not occur in a sequential manner, going to higher equivalents of protein will not give more information about binding.

In a second experiment with UV-vis absorption spectroscopy, we tried to determine the stoichiometry by using the Continuous Variation Method [87, 88]. For this purpose, different mixtures with constant sum of ligand and LZ concentration (100 μM) were prepared, while the host (ligand) and guest (protein) concentration have been changed from 0 to 100 μM. The recorded UV-vis absorption spectra of all mixtures are shown in Figure 5.19. Except the first and last spectrum, neat ligand and protein, respectively, the observed absorbance in each spectrum is the sum

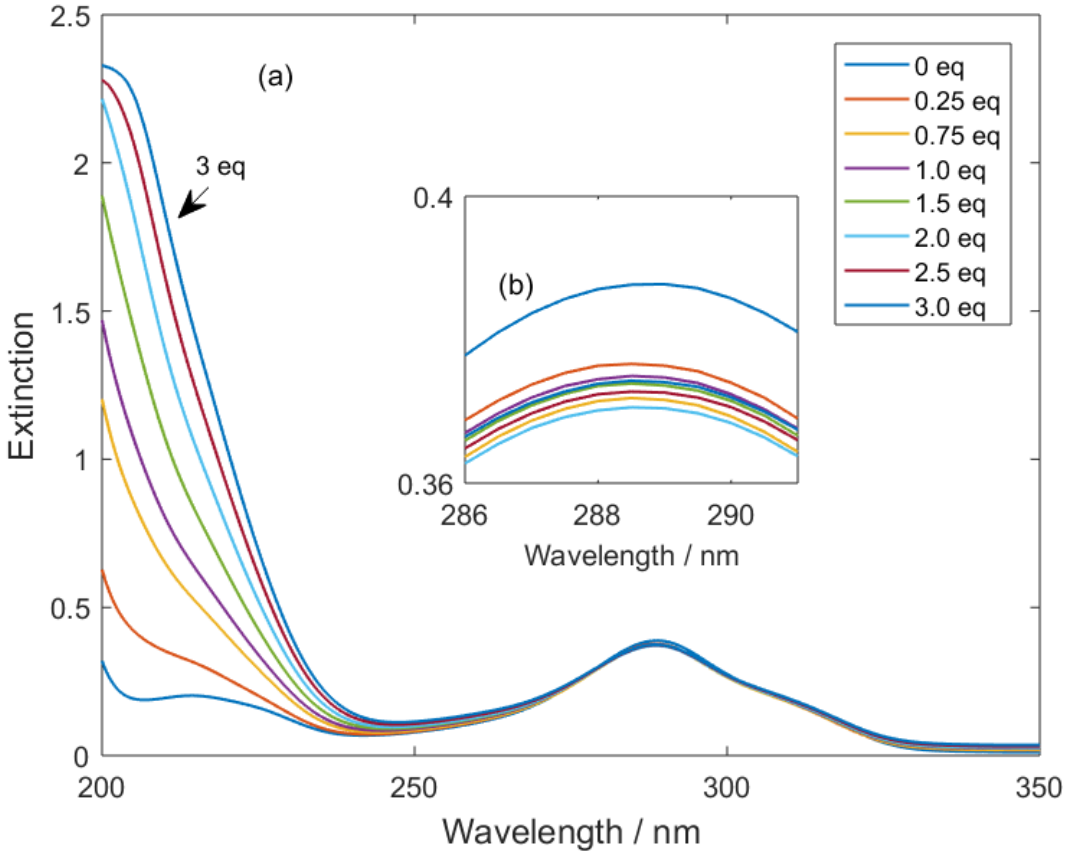


Figure 5.18: (a) UV-vis absorption spectra during the titration of supramolecular ligand (compound **2**, 10 μM) with the protein LZ (from 0.25 to 3 μM). (b) Inset: zoom into the region 286-295 nm.

of the absorbance of three components: free ligand (L), complex (C) and free protein (P), i.e., $A_{obs} = A_L + A_C + A_P$. The protein absorption at 264 nm is significantly weaker than the absorption by the ligand at 288 nm at the same concentration. Since the absorption spectrum of the complex is unknown and it overlaps with that of the free protein and the free ligand, univariate data analysis is not appropriate. Therefore, multivariate data analysis (MCR-ALS) was used to calculate the concentration profile of the complex as well as of the free protein and the free ligand in each UV-vis absorption spectrum. The results are displayed in Figure 5.20.

All spectra were pre-processed by centering and auto-scaling for baseline correction. Then, the MCR-ALS algorithm was applied to determine the contribution of each component in each spectrum. An initial guess for the spectrum of three components was needed for running the algorithm. The spectra of neat ligand and protein were used as initial guesses for two components. We applied the equation $A_{obs} - \epsilon_L[L] - \epsilon_P[P]$ for estimating the spectrum of the complex in which the absorptivity of ligand and protein, ϵ_L=0.0607 cm$^{-1}\mu$M^{-1} and ϵ_P = 0.0049 cm$^{-1}\mu$M^{-1}, were determined from other measurements using the pure ligand and

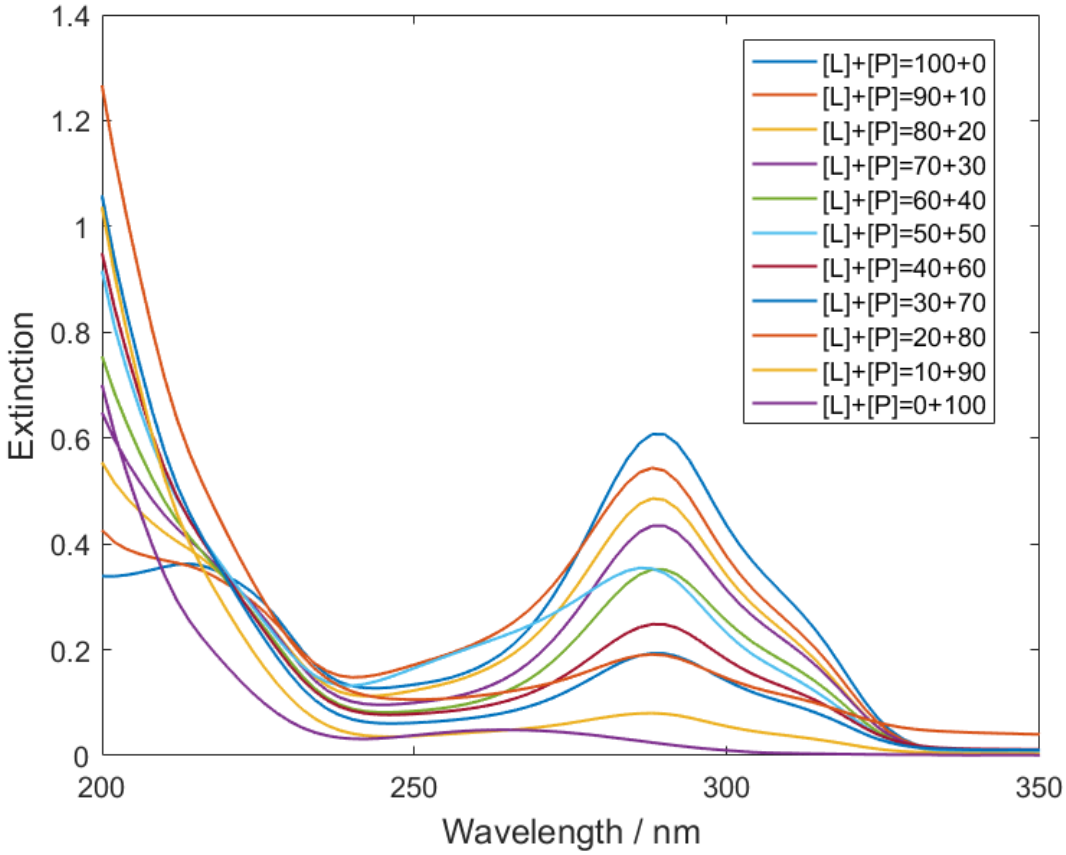


Figure 5.19: UV-vis absorption spectra of compound **1** [L] and Leucine Zipper [P] mixtures acquired via the continuous variation method

.

pure protein, respectively. These initial guesses for the spectra of three components can be seen at the bottom of Figure 5.20. Shown at the top of the figure are the calculated concentration profiles of the three components: neat ligand in blue, protein in yellow and complex in red. Derived from the calculated results, the maximum complex concentration occurs when the initial concentration of the neat ligand and leucine zipper are 20 and 80 μM, respectively. Knowing that the protein molecules have a dimeric structure, this ratio is not surprising, meaning that probably one ligand molecule can bind to two protein dimer molecules. However, one should keep in mind that the complex concentration was calculated without applying any constraint and prior information about the system except non-negativity and closure for the concentration amounts of components. So, the final results can not be counted as an accurate value.

UVRR spectroscopy

Since little information about the system has been exploited from the UV-vis absorption spectra, a protein-based UVRR titration experiment

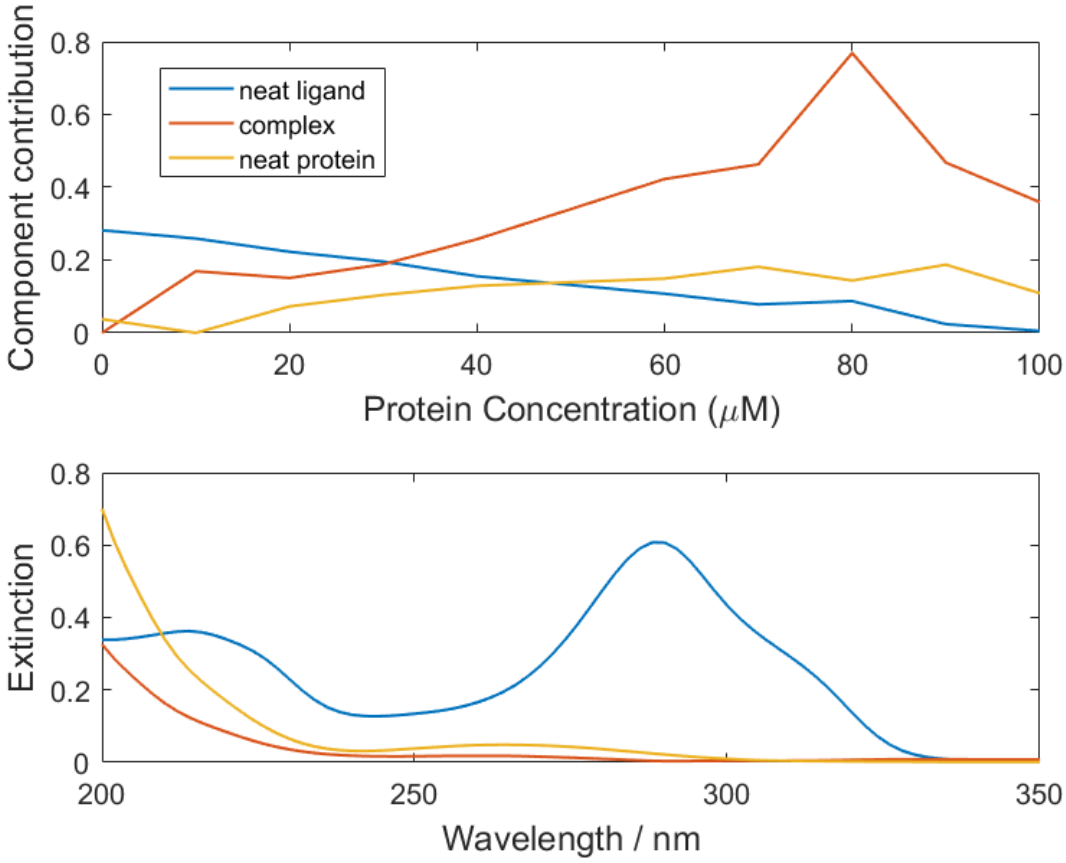


Figure 5.20: (top) Contributions of three components calculated by MCR-ALS from the spectra in Figure 5.19, (bottom) the UV-vis absorption spectra of three components. Lack of fit (LOF) $\simeq$ 3.11 %.

was performed in order to find the optimum concentration for protein titration. For this experiment, the concentration of the protein leucine zipper was kept constant at 50 μM and the UVRR spectra were recorded upon ligand addition from 20 to 100 μM. The continuous titration was performed from a concentrated 5 mM stock solution of ligand at pH 6. With this high concentration of ligand, the dilution due to addition of ligand during the titration was kept minimum so that the calculated error (the added volume to the general volume of the solution) for the last step of titration was 2%. Four UVRR spectra of this titration are displayed in Figure 5.21.

Based on the concentration-dependent function of Raman intensity which we calculated for the ligand, compound **1** in Figure 4.5, a continuous and nearly linear increase was expected for the Raman spectra over the whole measured wavelength for this range of concentrations from 20 to 100 μM. However, not all the Raman bands of the spectra in Figure 5.21 display this expected trend. The intensity of the Raman band at 1469 cm^{-1}, assigned to the bending mode of the guanidinium, shows a general decrease upon ligand addition. Specifically, when the ligand concentration

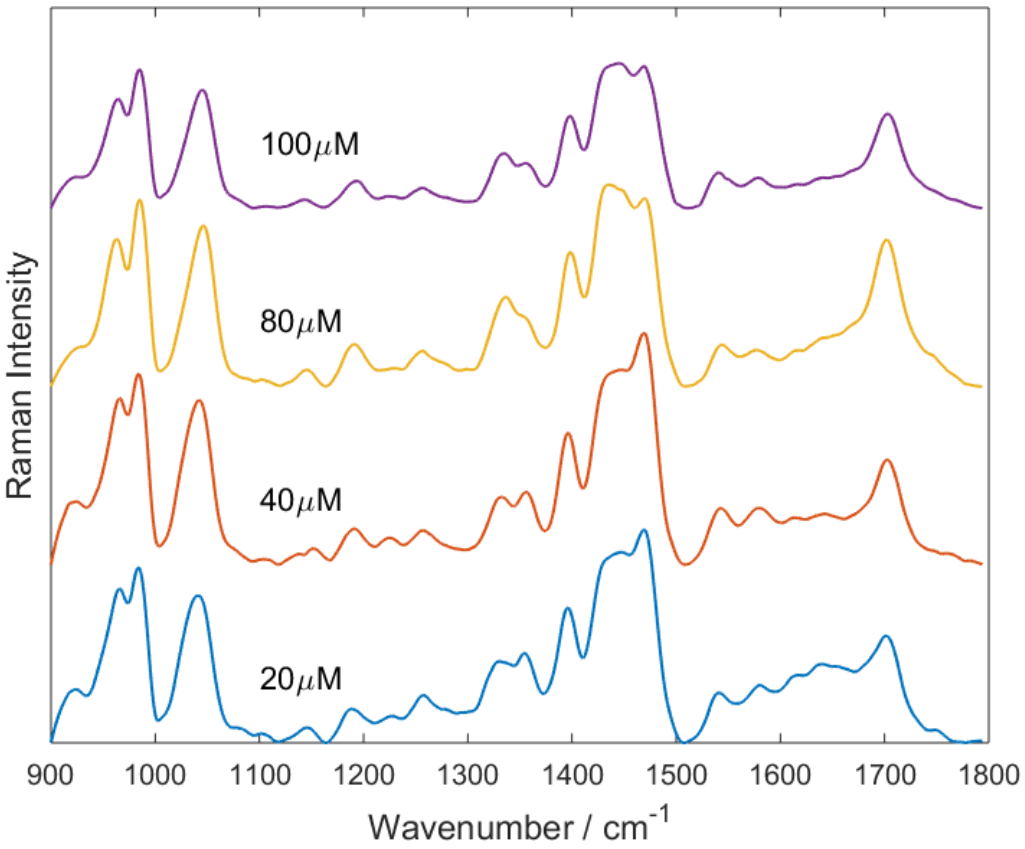


Figure 5.21: UVRR titration of the protein leucine zipper (constant concentration: 50 μM) with the supramolecular ligand.

exceeds the concentration of the protein, the Raman band decreases in intensity while we expected an increase due to increasing the concentration of ligand. This decrease is an obvious indicator of binding between two molecules. Only in the first step of titration from 20 to 40 μM, there is a slight increase which is due to the increase of ligand concentration by the factor of two. Among the four UVRR spectra, the most considerable change regarding both Raman marker bands at 1429 and 1467 cm^{-1} can be seen at a ligand concentration of 80 μM which shows a probable binding when the concentration of ligand is higher than that of protein. Nonetheless, the ratio of the two Raman bands decreases again at the ligand concentration of 100 μM. Moreover, the entire intensity of the last spectrum is weaker compared to the three UVRR spectra acquired at the beginning of the titration. We attribute this entire decrease to the photo-degradation of the sample due to the continuous UV illumination during the titration.
In conclusion, a general estimation of optimum concentration was provided by this quick UVRR titration experiments. The binding is expected to be more probable when the ligand concentration is higher than that of protein which makes much sense regarding the high number of protein binding

sites. However, the ratio (ligand concentration to protein concentration) in which both Raman marker bands are affected is less than 2. Based on this estimation, we set up the range of protein concentration in a ligand-based UVRR titration so that the titration range of protein covers both areas of the concentration less and more than the concentration of ligand.

5.4.2 UVRR binding study

In a ligand-based manner (constant ligand concentration), we performed a UVRR titration of the multi-armed ligand compound **1** with the protein leucine zipper at pH 6.0. The concentration of the multi-armed ligand was kept constant at 100 μM in all titration experiments since the GCP moieties of the ligand are selectively enhanced upon UV excitation and therefore dominate the Raman spectra of the mixtures. This ligand concentration falls in the range in which the concentration dependency of Raman intensity is approximately linear despite the general fact of its nonlinearity due to absorption and re-absorption (see Figure 5.8). As it was mentioned before in section 2.3, this linearity is a criterion for using multivariate data analysis methods. The concentration range for leucine zipper was selected based on the preliminary experiment, from 20 to 400 μM. Due to the UV-excited fluorescence from the aromatic amino acid (phenylalanine) included in the structure of protein, it was not possible to do the experiments with higher concentrations of leucine zipper.
The Raman spectra of different mixtures were acquired under the same experimental conditions: illuminated by 266 nm excitation wavelength and the power of 15 mW for 90 sec and 20 accumulations, total illumination time was 30 minutes. This combination was selected for obtaining a high signal to noise ratio. Each spectrum was recorded after the first accumulation (illumination time: 90 sec) and then was compared to the final spectrum (illuminated for 30 min). This is one way to monitor the Raman peaks during the total illumination time. Comparing two spectra of one sample with different illumination time, no change in the spectrum during the illumination means the sample was not photodamaged. The pH of the solution was carefully checked before and after every Raman experiment. Every spectrum was acquired two times and the mean value of two spectra was calculated for further analysis. The raw UVRR spectra were smoothed by Savitzky-Golay filter (sgolayfilt function in MATLAB) with a second polynomial order and frame width of 12. Baseline correction was performed by the modified polynomial algorithm (see section 4.4 for more detail) starting with different order polynomial fit. Major baseline changes were observed in the UVRR spectra during the titration at higher

protein concentrations. Therefore, different initial polynomial orders were used in order to suitably match with different curvatures in various mixture spectra.
The UVRR spectra of the neat ligand and of mixtures with an increasing concentration of leucine zipper are shown in Figure 5.22 (from bottom to top). The recorded small changes during the titration demonstrate the strong intrinsic sensitivity and selectivity of the Raman spectroscopic signature of the ligand. The highlighted region (1410-1480 cm^{-1}) represents the dominant spectral changes upon protein addition [14,15]. The behavior is the same as that of previous UVRR binding studies on monovalent GCP-based ligands and small tetrapeptides i.e., the decrease in the intensity ratio of the two Raman bands at ca. 1467 cm^{-1} and at ca. 1429 cm^{-1} upon protein addition. These specific bands are marked in green and red color in Figure 5.22. The results from density functional theory

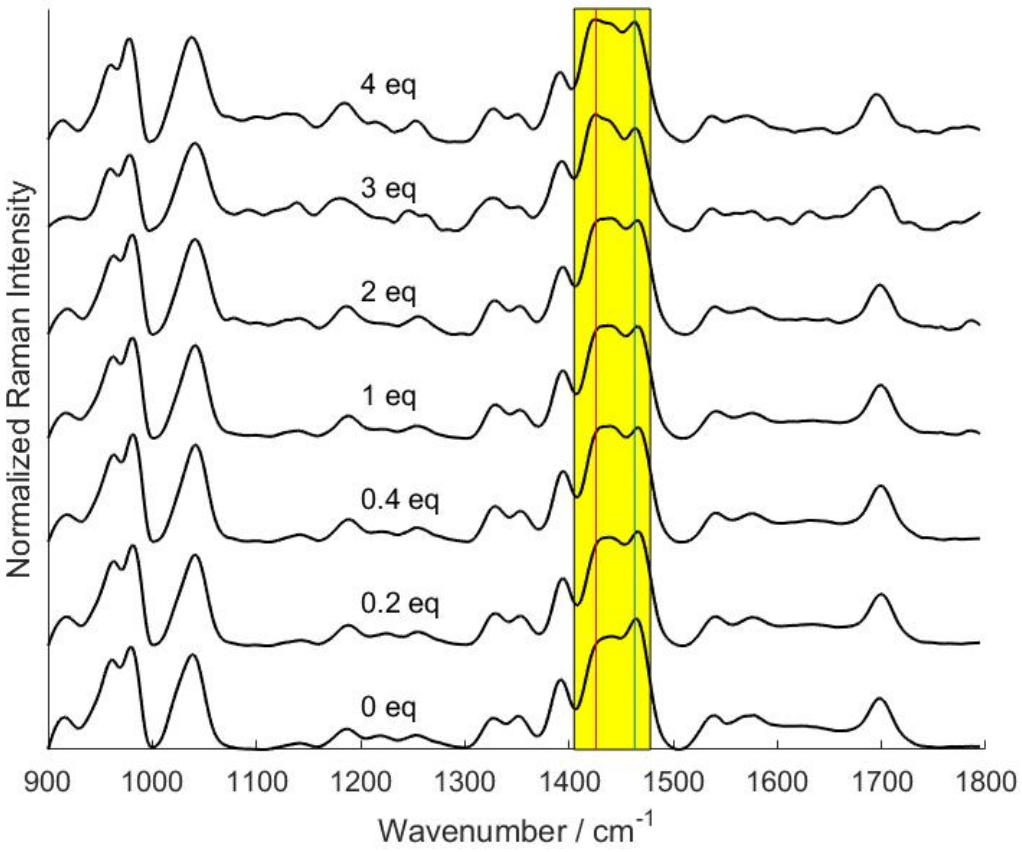


Figure 5.22: UVRR titration of the three-armed GCP-based supramolecular ligand (compound **1** kept constant at 100 μM) with increasing equivalents of the protein leucine zipper. The highlighted region includes the diagnostic Raman bands for protein recognition.

calculations on GCP model compounds showed the assignment of the two bands to a GCP bending mode mainly involving the guanidinium part and pyrrole-NH bending, respectively. In order to correct the absorption

and re-absorption effect upon protein addition, we used the peak at ca. 1413 cm^{-1} as an internal standard for re-scaling the Raman spectra. This band is assigned to the normal mode of pyrrole hydrogens in the 2- and 3-position which is not involved in binding of the peptide/protein and is therefore unaffected by the addition of leucine zipper.
Interestingly, notable spectral changes are observed in the marker region (intensity ratio at 1467 cm^{-1} and 1429 cm^{-1}) already upon addition of only 0.2 equivalents of the protein leucine zipper. These changes, which are the main difference between the current results and earlier UVRR binding studies on peptide recognition, can be attributed to the large number of carboxylate binding sites of leucine zipper (13 acidic side chains plus C-terminus). We assume that the initial complexation of the most reactive/sterically accessible carboxylates causes the acceleration of binding at the first initial point of titration.

5.4.3 Qualitative multivariate data analysis

Two methodologies of multivariate data analysis, NMF and MCR-ALS were employed to identify how the spectral changes of the intensity of Raman marker bands occur during different range of protein concentration. As mentioned in section 4.3 about the working procedure of the multivariate data analysis methods, the decomposition of the experimental data matrix into two matrices of concentration (**C**) and spectra (**S**) of the components usually starts by an initial estimation of either **S** or **C** which then is followed by optimizing the concentration and spectra of the components iteratively using the available information about the system. For a successful resolution of the experimental data, we first need to estimate the number of the components. Since there may be more than two structurally different species in the mixtures, because of multivalency of both ligand and protein, the exact interpretation of the experimental results is difficult, especially when their occurrence strongly depends on the experimental conditions.
In the supramolecular binding studies by UVRR spectroscopy, no matter how many binding sites are available on the protein, only the changes related to the ligand can be monitored under the resonance of the applied excitation wavelength. In the previous binding studies, it was assumed that each UVRR spectrum is a linear combination of two components, "free" and "complexed" ligand, and the concentration contribution of these two components was determined by NMF. This assumption was used because both ligand and binding partner (tetrapeptide) had only one binding site. For molecular recognition of a supramolecular ligand including three GCP binding motifs, we can assume that potentially three kinds of ligand complexes may exist in each mixture: the sum of ligand molecules bound

to protein through one arm ($\sum_i L_i P$), two arms ($\sum_i L_i P_2$), or it is also probable that all three arms participate in binding ($\sum_i L_i P_3$). Therefore, in spite of the tremendous number of stoichiometry which leads to lots of possible complex components, these three complexes and free ligand are involved in UVRR spectra. The subscript i refers to the number of ligands bound to one protein which can not be discriminated by UVRR and therefore will not be taken into account in our binding evaluation. Generally, in multivariate data analysis by NMF and MCR-ALS, we used two initial assumptions for the number of components. First, we considered the overlapped and indistinguishable contribution of all types of bound ligands in each UVRR spectrum. Therefore, like the previous binding studies, two species of "free" and "bound" ligand were assumed to be the contributors to the mixture spectra. By another assumption, we tried to discriminate the intermediate binding steps for the three armed ligand and evaluate their concentration profiles by MCR-ALS.

Analysis with the assumption of two components

By the first assumption, considering two components of the "free" and "complexed" ligand, we employed the spectra at the first and end point of the titration as initial guesses for the two species. Then, three calculations were carried out using multivariate data analysis. First, the calculation was performed on the smoothed, baseline-corrected and auto-scaled spectra. Secondly, the concentration profile and spectra of two components were derived from the smoothed and base-line corrected but non-scaled data. Finally, the second-derivative spectra were analyzed by multivariate data analysis. The latest approach can be done only by ALS because of the negative values in the second-derivative spectra. While the results from the first approach were considered for further quantitative data analysis, the outcome of the last two approaches was used as a complementary results for exploiting more information about the underlying principles of the binding. Both methods of NMF and ALS were employed for the first multivariate analysis on fully pre-treated spectra. In Figure 5.23 the UVRR spectra for two components calculated by NMF are shown. Figure 5.24 shows two related pairs of binding curves for the free and complexed ligand determined by NMF and ALS from the UVRR spectra in Figure 5.22. Obviously, both NMF and ALS approaches led to the same resolved concentration profiles with minor differences. All spectra were normalized to the mean value of neat ligand spectrum, this is why the values for the concentration ratio (y-axis) are between 0 and 1, indicating the fraction of neat ligand involved in binding event. No information about the saturation point can be extracted from the calculated results and it does

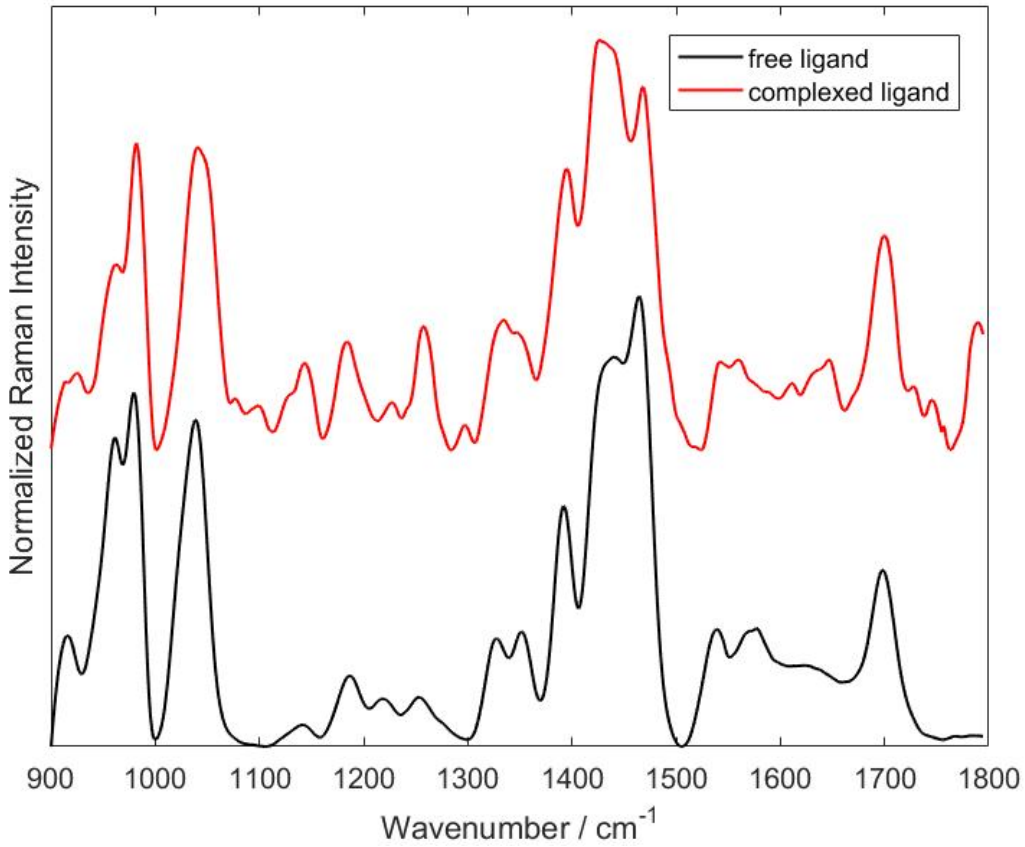


Figure 5.23: The spectra of "free" and "complexed" ligand calculated by NMF.

certainly not imply that at 400 μM of leucine zipper one has already achieved saturation - it is simply the end point of our titration. At leucine zipper concentrations larger than 400 μM, the quality of the UVRR spectra is significantly degraded due to absorption/reabsorption as well as autofluorescence from the protein. From experimental condition, only two constraints of non-negativity and closure (the sum of concentration ratios for the involved components in each spectrum/mixture is constant) were applied in both calculation methods. Hence, MCR-ALS and NMF were both capable to retrieve the relative contributions of the species with minimum prior information about the system. In comparison, both methods show the same trend for the concentrations of the two components. When the concentration of protein reaches nearly 200 μM, the concentration curves of free and complexed ligand meet each other and both get the half maximal ratio. This point can be defined as *inflection point* at which the growth rate of complexation reaches its maximum value. Consequently, for protein concentrations higher than 200 μM, the complexation ratio has been decreased, as shown by red and green curves. As a result, the trend of these curves can be divided into three phases: an early accelerating phase, a linear phase, and a final

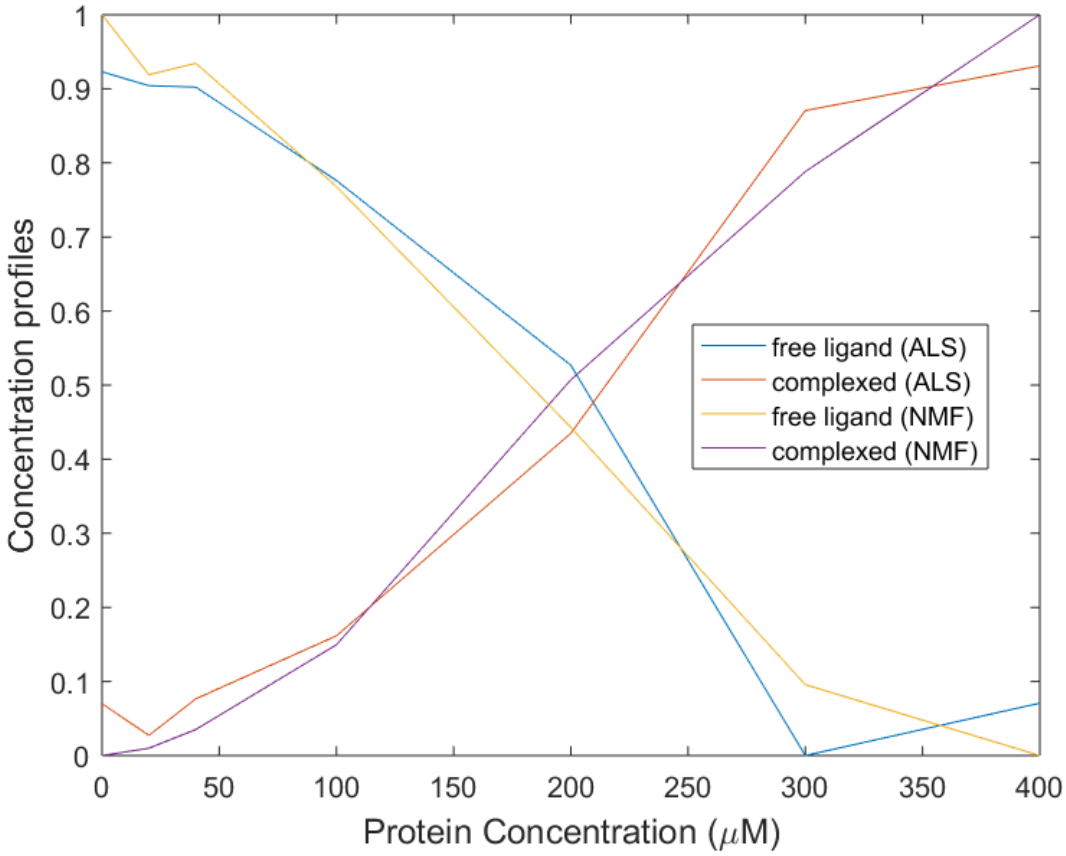


Figure 5.24: Binding curve: concentration ratio of complexed and free ligand as a function of protein concentration as determined by MCR-ALS and NMF, respectively.

declining stage. Therefore, the growth pattern typically follows a sigmoidal curve. A positively cooperative curve exhibits a sigmoidal shape [68]. Basically in a cooperative binding, it is assumed that addition of ligand to a macromolecule with multiple sites changes the affinity for all remaining unoccupied sites after the first binding. Since we have multiple sites on both ligand and protein, each of them may show a positive cooperativity. In other words, binding one GCP to one of binding sites on the protein should ease the binding of second arm and it also can happen for the protein. However, the cooperativity effect for the binding sites of the protein can not be monitored by UVRR spectroscopy because we used a ligand-based technique. Moreover, the three arms are completely identical and at this stage we could not differentiate their binding to monitor such an effect upon protein addition. Therefore, positive cooperativity could not cause the resulted sigmoidal shape. From another point of view, based on DFT calculations we can monitor two different kinds of possible binding in GCP motif as they are monitored by two different bands. Hence, we can consider two classes of binding: one for the ion-pairing monitored by the vibrational change of the guanidinium (at 1467 cm^{-1}) and the other

one the class of bound ligand in which the hydrogen bonding from the NH-pyrrole contribution (at 1429 cm^{-1}) makes the binding stronger. This situation can be assumed as binding to nonequivalent sites which can be observed in the spectra in which the intensity ratio of the two Raman marker bands in Figure 5.22, ca. 1467 cm^{-1} Raman intensity (green line) to the intensity of Raman band at ca. 1429 cm^{-1} (red line), is less than one. Also the third spectrum (ligand concentration = 80μM) in Figure 5.21 clearly shows the contribution of NH-pyrrole in binding. In comparison, in both cases of ligand and protein titration, there are spectra in which both Raman marker bands clearly change and support the idea of devoting our model to the category of binding to non-equivalent sites. The sites which are apparently identical (three identical GCP motifs), can display non-equivalency depending on the involved interactions, only ion-pairing or both ion-pairing and hydrogen-bonds. While the spectrum calculated for the complexed ligand by NMF in Figure 5.23 shows the contribution of both interactions in a complex molecule which is confirmed by ALS for the calculated spectra shown in Figure 5.25, the binding curves suggest the existence of two kinds of complexation either as a step-wise (multiple equilibria) or in an independent manner (binding to non-equivalent sites).

The mentioned assumed models (multiple equilibria or binding to non/equivalent sites) became even more evident when we analyzed the untreated spectra. Shown in Figure 5.26 (bottom) is the spectrum of bound ligand compared to the neat ligand spectrum, both of which are not self-absorption corrected. In another words, we applied both algorithms on the nonscaled data which are not normalized to the Raman band at 1413 cm^{-1}. The calculation results on these spectra for finding the bound ligand contribution over adding protein is displayed in Figure 5.26 (top). From the results calculated by NMF (dark blue line) two steps of complexation can be discriminated during the titration process between 0 and 400 μM. Obviously, the first step happens with high rate of complexation until addition of protein by two equivalents of ligand. After this point (protein concentration of 200 μM), a gradual increase of the bound ligand concentration is observed. The profile calculated by ALS (light blue curve) also represents a separation of two steps of binding at a protein concentration of 200 μM.

It is also possible to use the second-derivative spectra for ALS, but the constraint of non-negativity should be applied only on concentration profile rather than both concentration and spectra. The advantage of using second-derivative spectra is that they do not need the baseline correction. In this regard, the second derivative spectra of the smoothed experimental data were calculated in MATLAB and then were further analyzed by ALS. The results are displayed in Figure 5.27 showing the concentration profile (top) and the second-derivative spectra (bottom) of two components. The

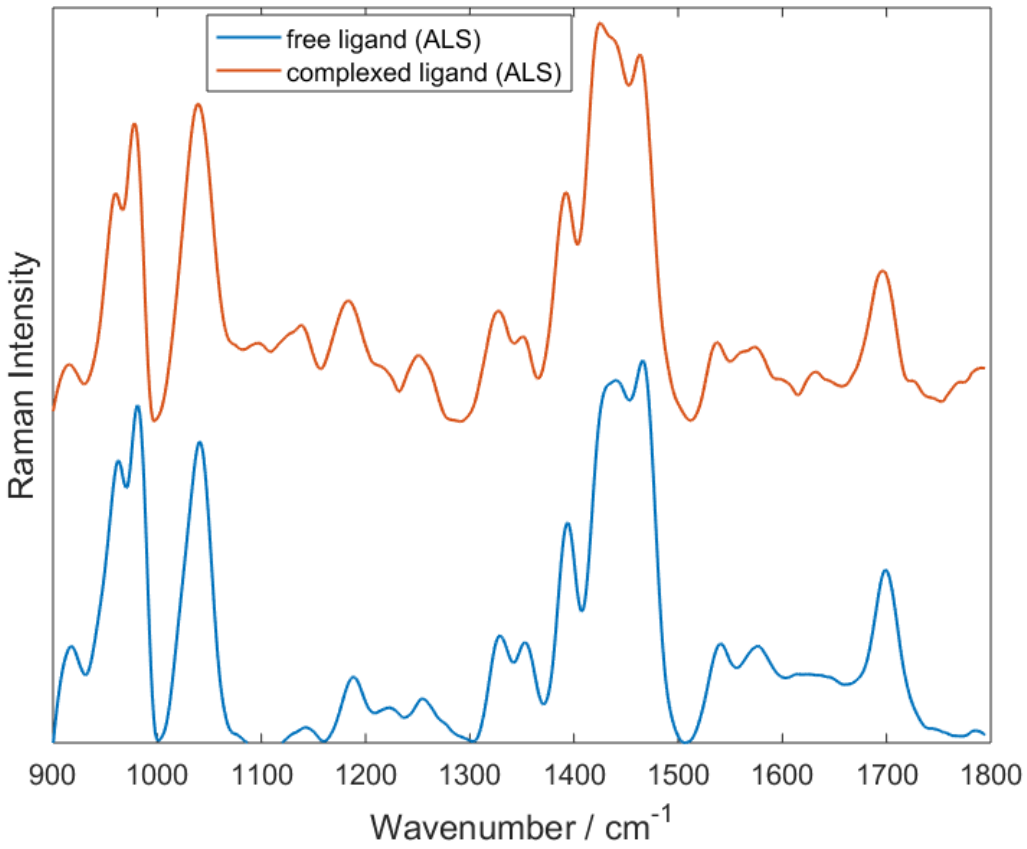


Figure 5.25: UVRR spectra of the free and complexed ligand as determined by MCR-ALS.

concentration profile of complexed ligand shows a notable increase already upon addition of only 20 μM of the protein leucine zipper. It confirms the anticipation of the initial complexation of the "hottest", i.e., most reactive/sterically accessible carboxylates with ligand. The concentration ratio of complexed ligand stays at a nearly constant level between 20 and 200 μM of protein concentration and then decreases from this point. From this result, the existence of another species, the concentration profile of which should increase from the protein addition at 200 μM, is strongly suggested. The second-derivative spectra can not be used by NMF since they have negative elements. This is the main advantage of ALS which is flexible for applying different constraints during the calculation and makes it more applicable for analyzing the variety of the spectra.
In addition to an apparent cooperativity that is more clear in the concentration profiles of the bound ligand in Figure 5.24, all the results from three calculations (displayed by concentration profile of complexed ligand in Figures 5.24, 5.26 and 5.27) are on this agreement that the protein concentration of 200 μM is a transition point between two different curvations of complexed ligand concentration profile. Moreover, the latest, analysis of the second derivative data, suggests that there should be more

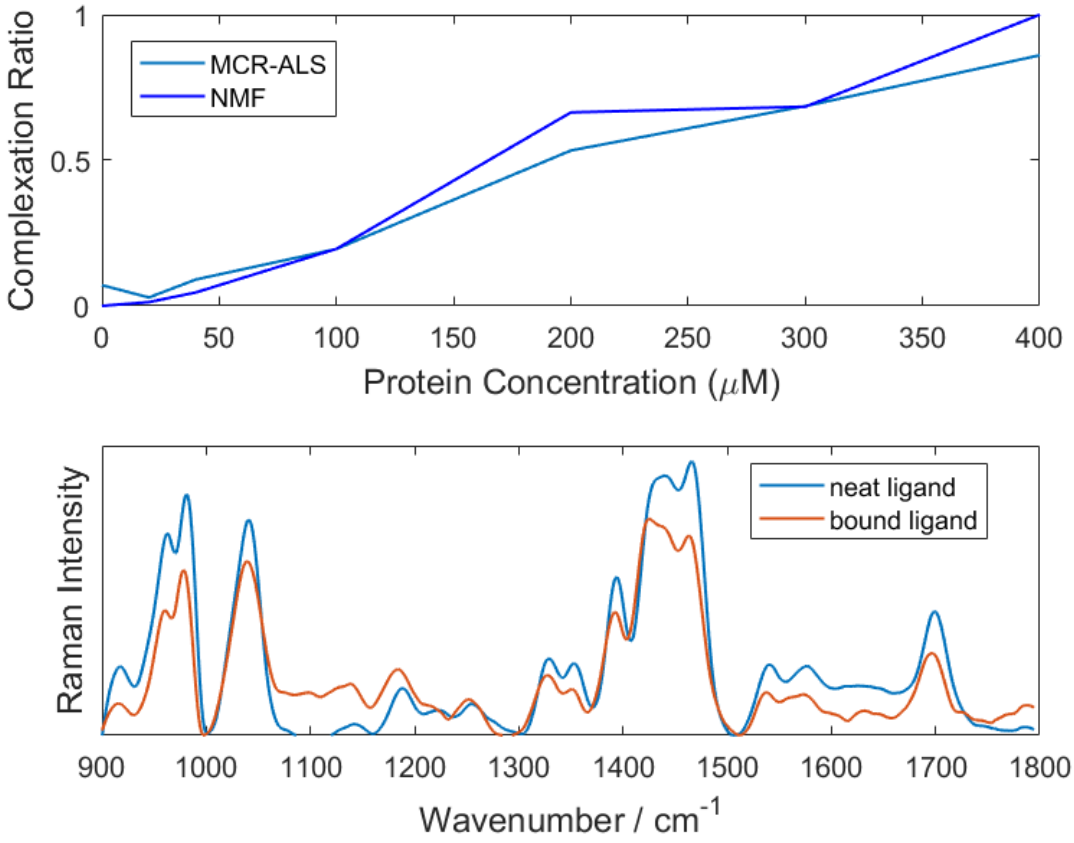


Figure 5.26: The analysis of the nonscaled UVRR data, (top) concentration profiles of complexed ligand calculated by NMF (dark blue) and ALS (light blue), (b) the UVRR spectra of neat and cmplexed ligand before self-absorption correction.

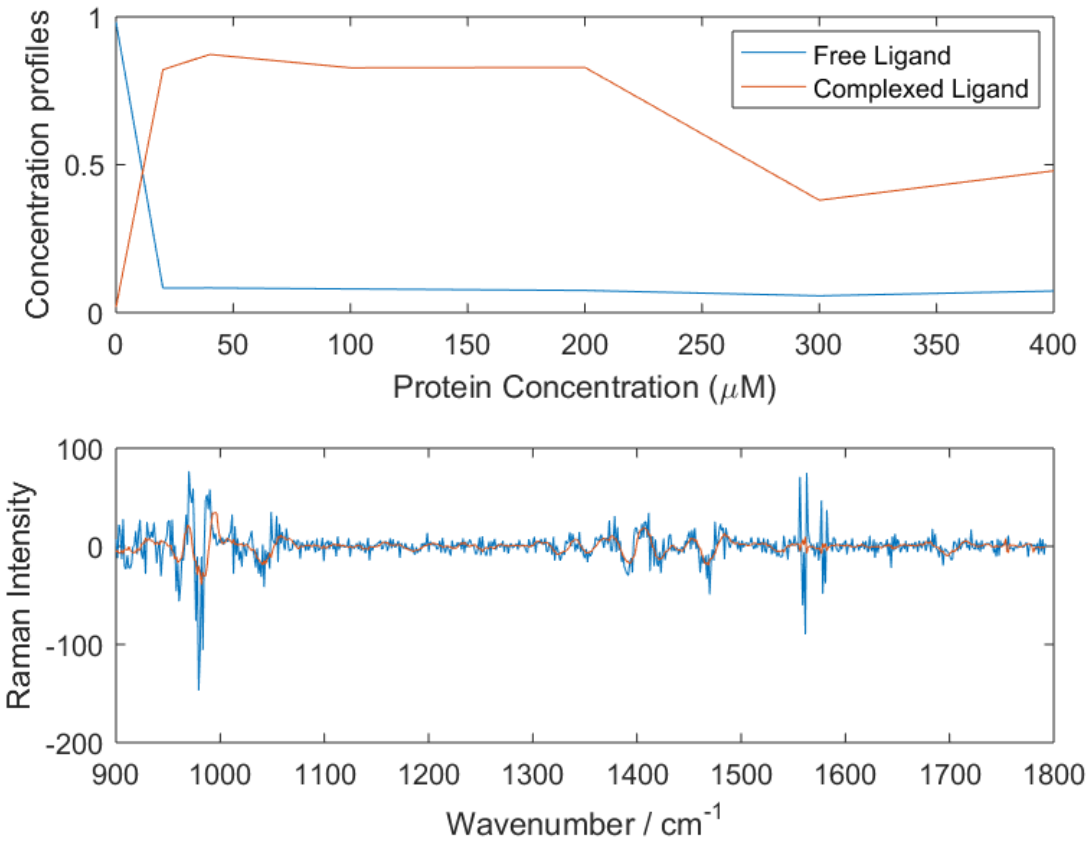


Figure 5.27: The analysis of the second-derivative spectra calculated by ALS, (top) concentration profiles of two components, (b) the second-derivative spectra of the neat and complexed ligand.

than one type of bound ligand.

Analysis with the assumption of more than two components

Assuming two components of "free" and "complexed" ligand - as an initial estimation of the number of species for further data analysis - was based on this fact that the spectra of all complexed molecules are overlapped and they can not be distinguished by UVRR. In fact, it was a general view of the binding process without looking at the intermediate formation of the complexes. Based on the obtained concentration profiles, especially from Figure 5.27, there is at least one hidden intermediate state for the process of binding. Moreover, when no exact prior information is available about the system, it is better to perform the analysis with different initial estimations, e.g., for the number of components. Therefore, we analyzed the data with initial estimation of more than two components in order to exploit the information of the complexation from the results of multivariate analysis.
For this purpose, evolving factor analysis (EFA) was used as a chemometric tool to monitor chemical processes. Having the spectra of mixtures in the rows of the data matrix, EFA sequentially performs the principal component analysis (PCA) on gradually adding the spectra of mixtures. This process is performed from top to bottom of data set (forward EFA) and from bottom to top (backward EFA) to investigate the emergence and the decay of the process contributions, e.g., the contributions of different components upon addition of protein. In this way EFA provides estimations of concentration regions for existence of each component and their evolution during the titration. These information is gradually known by recording a new response vector at each stage of the process. Figure 5.28 displays the way to plot the information provided by EFA. The interpretation of the information is possible by visualizing these plots. The theory of the procedure is explained briefly in section 2.3. The row number on the x-axis is the number of PCA calculation and it starts with 2 simply because at least two spectra are needed to initially run the calculation. A singular value decomposition (SVD) of the submatrix containing only the first two spectra of the original data matrix is made (blue line). Then, the third spectrum of the data matrix is added to the initial submatrix and the SVD is made for this new submatrix (red curve). This process is repeated until SVD is made over the whole submatrices. The whole graph (Figure 5.28, top) provides information about the appearance of the different components along the experiments. Each line represents a similar group of singular values stating the essence of one component on the related calculation region. For example, the line starting by three spectra at zero (third submatrix including the first four spectra) shows the combination

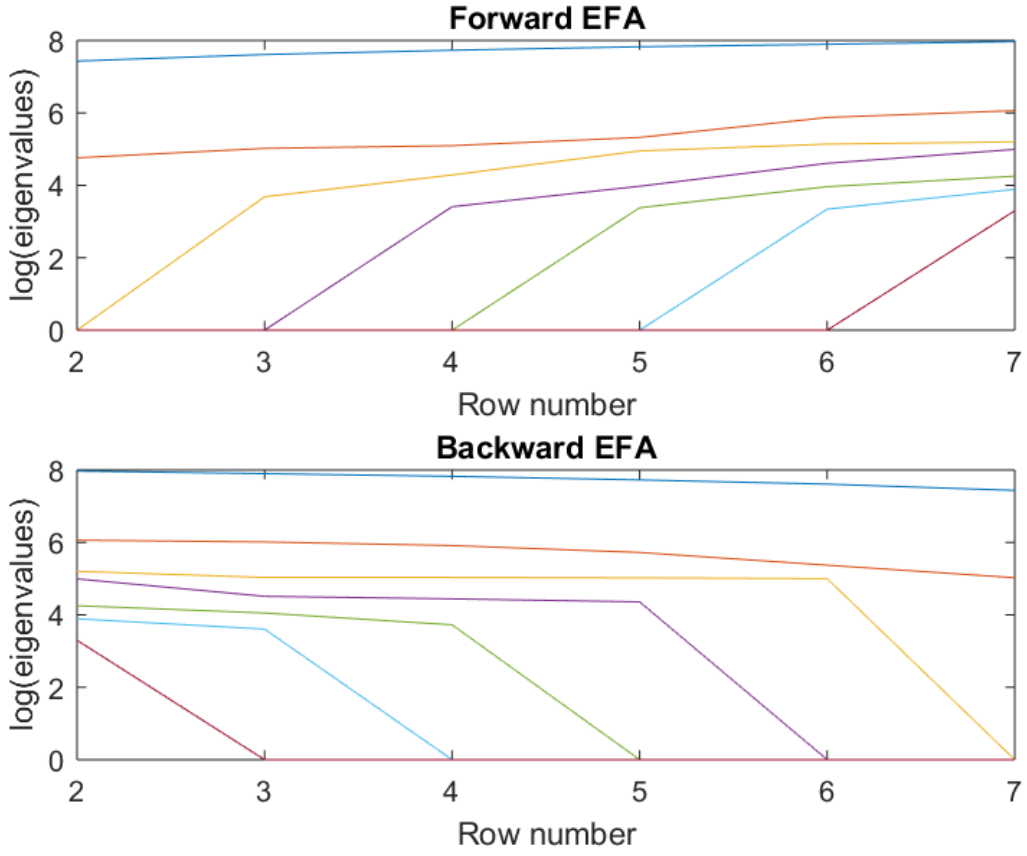


Figure 5.28: Grafical information about the complex formation derived from Evolving Factor Analysis (EFA): (top) forward EFA plot, (bottom) backward EFA plot. The row numbers on the x-axis are the numbers of spectra added in every step of calculation.

of three lines suggesting the essence of three components after addition of 1 mM of protein to the ligand. The same process is repeated for the backward EFA starting from the last two spectra providing the information about the disappearance of the different components shown at the bottom of Figure 5.28.

By combination of forward and backward plots, EFA estimates the concentration profiles for the preselected number of components. The forward and backward EFA suggest the pressence of three components. However, we calculated the concentration profiles for three and four species because it has been proven that the simultaneous presence of several closed reaction systems in an experiments leads unavoidably to the underestimation of the number of compounds [55]. The results are displayed in Figure 5.29 after normalization. Since the calculation of PCA starts with the first two spectra, the first point in which EFA calculates the concentration contribution of each component is the first protein titration (200 μM). Hence, at point zero, we set the number 1 for the concentration profile of neat ligand and zero for other components since we know the

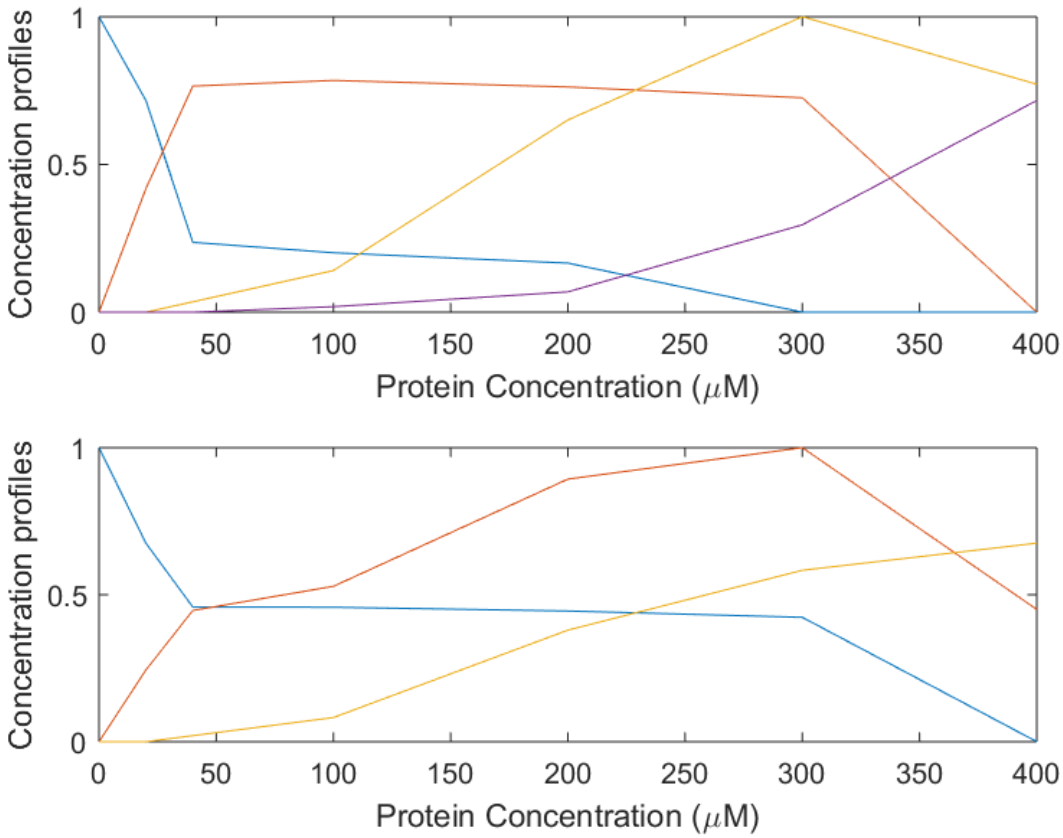


Figure 5.29: Estimation of concentration profiles by EFA for (top) four involved components and (bottom) three components.

complexed ligand does not exist before protein addition. Four components of free ligand, bound ligand through one, two and three arms are assumed to exist during the binding process for a three armed GCP based ligand. One possible formulation for writing a series of equilibria is as follows,

$$L + P \rightleftharpoons LP \quad ; \quad K_1 = \frac{[LP]}{[L][P]} \tag{5.8}$$

$$LP + p \rightleftharpoons LP_2 \quad ; \quad K_2 = \frac{[LP_2]}{[LP][P]} \tag{5.9}$$

$$LP_2 + p \rightleftharpoons LP_3 \quad ; \quad K_3 = \frac{[LP_3]}{[LP_2][P]} \tag{5.10}$$

where K_i are the association constants for each equilibrium. However, it should be noted that the manner in which the equilibrium constants are written says nothing certainly about the mechanism of the reactions [68]. This principle describes why the concentration profiles should be calculated with different assumptions for the number of components. For example, it happens sometimes that two reaction systems are so closed that two components would be simultaneously present in the same range of titration. We observe this situation in Figure 5.29 (top) for the concentration profiles of two components (plotted in violet and yellow) which both are available in the same range of concentration. A simple calculation showed that the mean value of the concentration profiles of these two components are nearly the same as the resulted profile of third component in calculation for three components, shown in yellow at the bottom of Figure 5.29. Therefore, by considering three components instead of four components for further calculation, no information will be lost about the whole system. The normalized concentration profiles derived from EFA were used as an initial estimation of "C-type" matrix for starting the calculation by MCR-ALS. The final optimized concentration profile and spectra are displayed in Figure 5.30. As can be seen in the figures, the concentration profile of partially bound ligand (red) has an intersection with the concentration profile of neat ligand (blue) at around 17 μM. This value was also extracted from Figure 5.27 and can be considered as the first dissociation constant in the equilibrium equations series. The decrease of partially bound ligand concentration is followed by the increase of the concentration of another component which is assumed as fully bound ligand. The concentration profile of partially bound ligand suddenly drops down when the concentration of protein exceeds 200 μM and it is in equilibrium with third component (assumed as the fully complexed ligand) at around 220 μM. A nearly similar amount can also be derived from Figure 5.29.

Additionally, we analyzed the second-derivative spectra by ALS with the assumption of three components. The results are displayed in Figure 5.31.

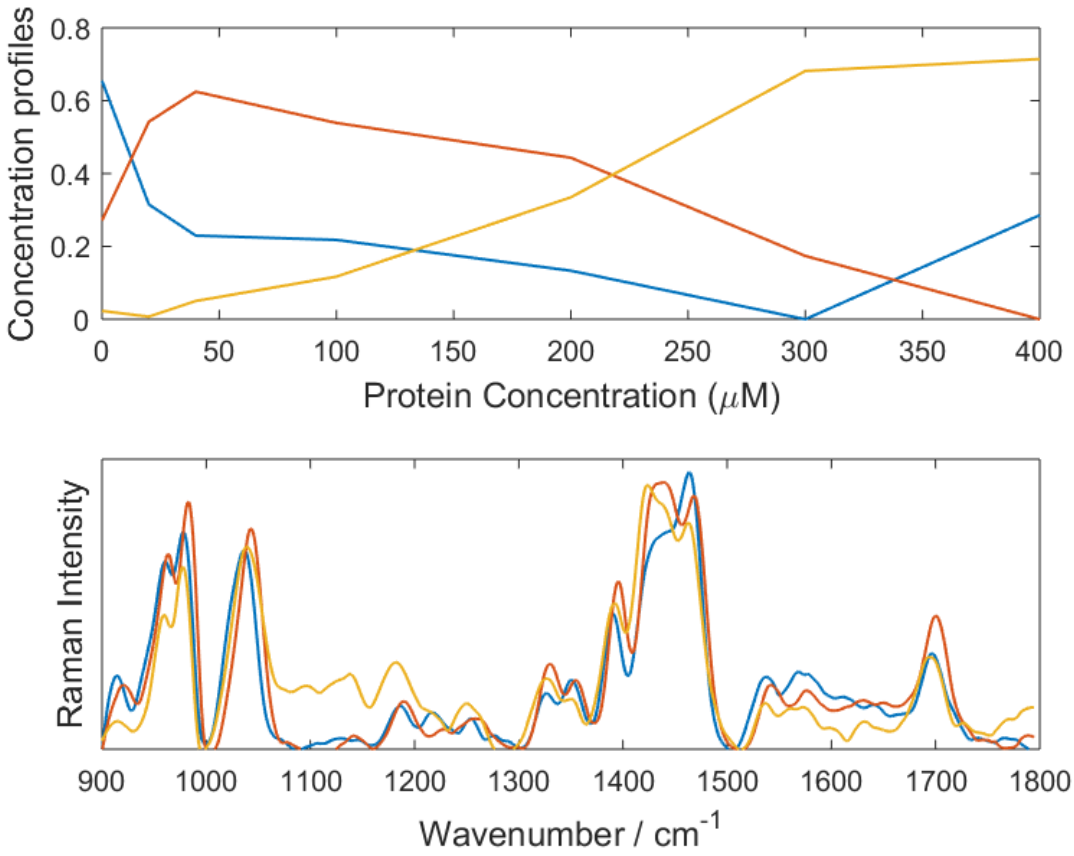


Figure 5.30: MCR-ALS results with the initial assumption of three components,(top) concentration profiles, (bottom) the calculated UVRR spectra of the components.

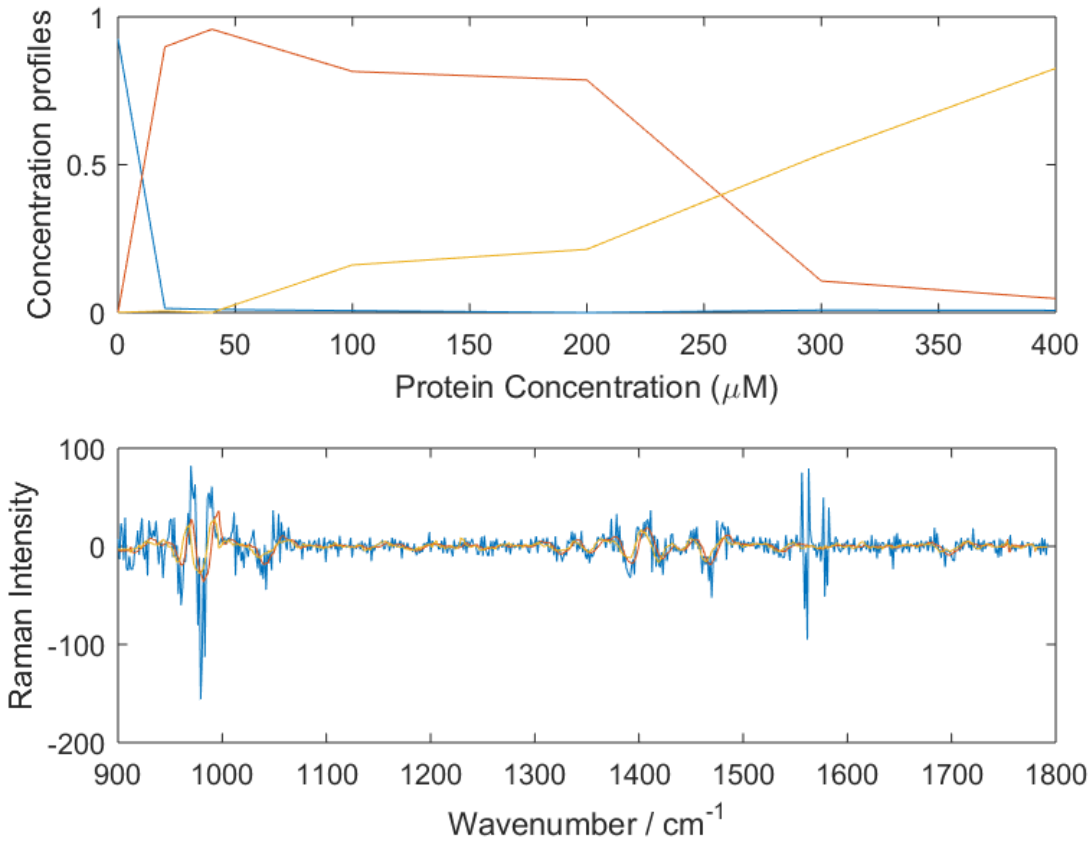


Figure 5.31: MCR-ALS results performed on second-derivative spectra with the initial assumption of three components, (top) concentration profiles, (bottom) the calculated UVRR spectra of the components.

The concentration profiles of three components (top) suggests two numbers of K_{d1} =15 μM and K_{d2} =250 μM as the dissociation constants of two reactions of forming two different forms of complexed ligand. The spectra of pure components for three species resolved by ALS are displayed in Figure 5.32. We specified them as the spectra for neat, partially and fully bound ligand. However, it should be mentioned again that the spectrum that we considered as that of the fully bound ligand may be a sum of different species overlapped in the related region of protein concentration. Even the concentration profile in the intermediate state (red) can not determine the type of partially bound ligand. The difference between two spectra of bound ligand (plotted in red and yellow) is mostly in the Raman band at 1419 cm^{-1} which was attributed to the binding mode of NH-pyrrole (based on the previous DFT calculation). The complex formation starts at the initial steps of protein addition which resulted in a very low K_d value (between 15 and 20 μM). However, the related calculated spectrum for this initial complex (red spectrum in Figure 5.32) indicates little contribution from NH-pyrrole while the obvious change occurs at 1427 cm^{-1} attributed to the guanidinium vibration. The NH-pyrrole

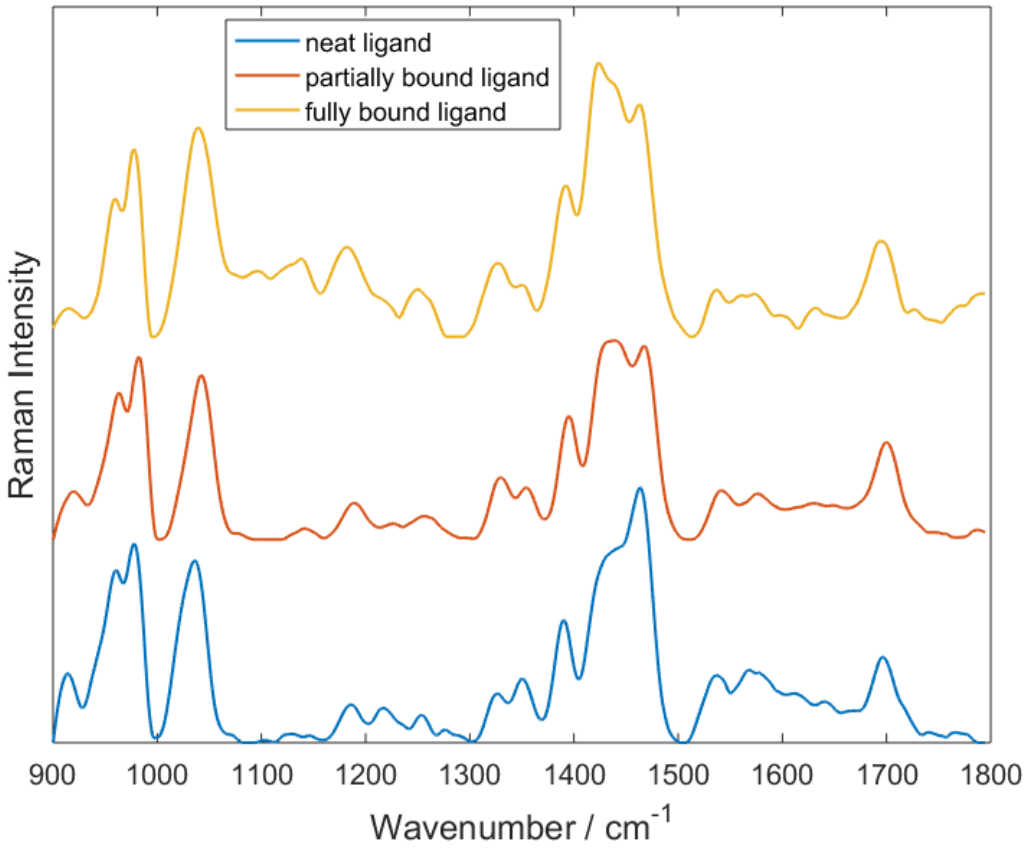


Figure 5.32: Three pure-components spectra calculated by ALS.

contribute in binding only at high concentrations of protein. The support for this claim is the spectrum specified as fully bound ligand (yellow in

Figure 5.32) showing the simultaneous contribution of both guanidinium and NH-pyrolle. Either these two types of interactions happen in a step-wise manner or independently, binding to nonequivalent sites can describe the sigmoidal behavior of the concentration profile of the bound ligand in Figure 5.24.
The various concentration profiles from different calculations and a non definite binding process make it difficult to define a certain model for our supramolecular system. Therefore, for quantitative analysis of the experimental data, we considered two general models. First, we look at the general behaviour of ligand during the complexation process without noticing the intermediate states, i.e., analysis with the assumption of two components. For this purpose, the results from the resolved data into two components (Figure 5.24) were used to be modeled. On the other hand, we use the dissociation constants defined by multivariate data analysis of three components in Figure 5.29 (bottom), Figure 5.30 and Figure 5.31 to plot the theoretical concentration profiles based on the multi-steps binding equations. At the end, the results of two models are compared.

5.4.4 Quantitative multivariate data analysis

One of the fundamental issues in supramolecular chemistry has always been the quantitative analysis of the intermolecular interactions of interest [89]. In this regard, we attempted to quantitatively analyze the calculated concentration profiles by a deeper look at the multiple equilibria equations. Assuming that each ligand molecule can bind to protein through its three arms, the general binding equation known as *Adair equation* [68] can be written as follow,

$$v = \frac{\sum_{i=1}^{i=3} iK_i[p]^i}{\sum_{i=0}^{i=3} K_i[p]^i} \tag{5.11}$$

where $K_i = [LP_i]/[L][P]^i$ are the equilibrium constants, and their determination with precision requires exact concentration of ligand, protein and complexed molecules. Basically, there should be also a subscript for ligand in Equation 5.11 since there are also multiple sites on the protein. Nevertheless, because only the vibration of ligand molecules are monitored by UVRR, from the spectroscopic point of view there is no difference between L_jP_i and jLP_i. Accordingly, we can claim that writing *Adair equation* in this form describe all of possible bound ligand which can be monitored by UVRR ligand-based technique. However, it is often more useful to see if the data can be fitted by some more restrictive equation that expresses a simple model and requires fewer adjustable parameters. Therefore, we need to compare the concentration profile of the complexed ligand with the general binding models of multiple equilibria for multisite

macromolecules to see which model can be matched with the determined data calculated by MCR-ALS and NMF.
The graph for the bound ligand concentration, from a two components binding studies by MCR-ALS and NMF (the concentration profiles of complexed ligand in Figure 5.24), can be described into three phases: an early accelerating phase, a linear phase, and a saturation. The latest were not achieved completely because of the limited titration due to the fluorescence. Therefore, the growth pattern typically follows a sigmoidal curve. While a positively cooperative binding exhibits a sigmoidal shape [68], obviously it is not the case for the cooperativity of the protein. Basically in a cooperative binding, it is assumed that addition of the ligand to a macromolecule with multiple sites changes the affinity for all remaining unoccupied sites after occurrence of the first binding. In a ligand-based technique, in which the concentration of complexed macromolecule increases upon titration of protein to a constant concentration of ligand, the effect of cooperativity for the binding sites of the protein can not be monitored, even though it may exist. From another point of view, we have also multiple sites on the ligand which may show a positive cooperativity because of binding with protein. It means binding one GCP to one of the binding sites on the protein should ease the binding of second arm. However, the three arms are completely identical and at this stage we could not differentiate their binding to monitor such an effect. On the other hand, the formulation for cooperative binding derived from the *Adiar equation* is so complex and needs adjusting some parameters which could not be defined by experiment. Hence, we decided to use a general form of sigmoidal function in order to model the growth curve of general concentration profiles.
Sigmoidal functions, whose graphs are "S-shaped" curves, appear in a variety of mathematical forms, like hyperbolic tangent, the "logistic" sigmoid and the "algebraic" sigmoid. Many biological dynamic processes, such as certain enzyme kinetic and population growth processes, develop almost step-wise which are a special class of sigmoid functions [90].
The general form of sigmoid function we applied is the "logistic" function which has been used for describing the growth behavior of many multilayer systems usually based on the parameter of time [91]. Here, the concentration of protein during the titration is selected as the variable of our system and logistic function

$$v = \frac{1}{(1 + e^{-k([p]-[p]_m)})} \tag{5.12}$$

where v is the ratio of complexed ligand which is expressed as a function of protein concentration $[P]$ and $[P]_m$ is the protein concentration required to achieve the half maximal saturation. It is defined as the inflection point

at which the growth rate reach its maximum value. The function was fitted to the data points in a "least square sense" by using *fminsearch* function in MATLAB. The result of fitting is shown in Figure 5.33, where the mean value of MCR-ALS and NMF-derived data points (the mean value of bound ligand, green and violet, curves in Figure 5.24, top) are depicted as a function of protein concentration. Since we used a general sigmoid

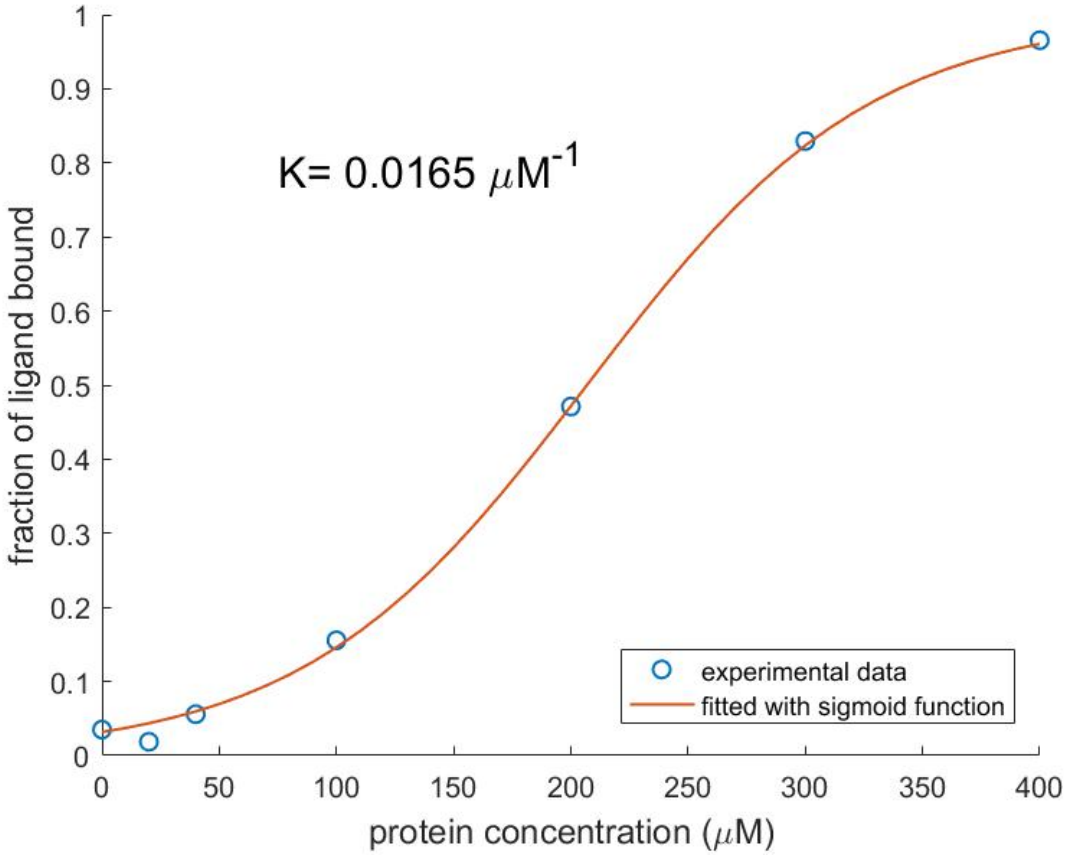


Figure 5.33: Binding curve determined from the UVRR titration in Figure 5.22. The data points are the normalized fraction of the complexed GCP-based ligand, obtained from the mean value of NMF and MCR-ALS results.

function, the binding constant measured by this formula (K_a = 0.0165 μM^{-1}) is not an exact binding constant but instead is a average of the multiple K_a values from multiple binding events (different complexation form). When we use the value of MCR-ALS or NMF separately instead of mean valued data points, very similar K_a values were obtained (0.0165 and 0.0172 μM, respectively).
In order to model the concentration profiles of three components, we applied multi-steps equilibrium equations.

$$[L] = c_L \frac{K_{d1}K_{d2}}{[P]^2 + K_{d1}[P] + K_{d1}K_{d2}} \tag{5.13}$$

$$[LP] = c_L \frac{K_{d1}[P]}{[P]^2 + K_{d1}[P] + K_{d1}K_{d2}} \tag{5.14}$$

$$[LP_2] = c_L \frac{[P]^2}{[P]^2 + K_{d1}[P] + K_{d1}K_{d2}} \tag{5.15}$$

In these equations, K_{d1} and K_{d2} are the dissociation constants and c_L is the total concentration of ligand. Again, it is necessary to mention, writing the equilibrium equations in this manner says nothing about how the complexed ligand molecules are formed. It is just a known illustration of multiple equilibria which we used for modeling the concentration profiles of three components.
Figure 5.34 shows the graph of these three equations. Each subplot displays the theoretical concentration contribution of three components with different couple of dissociation constants. The three couple values of

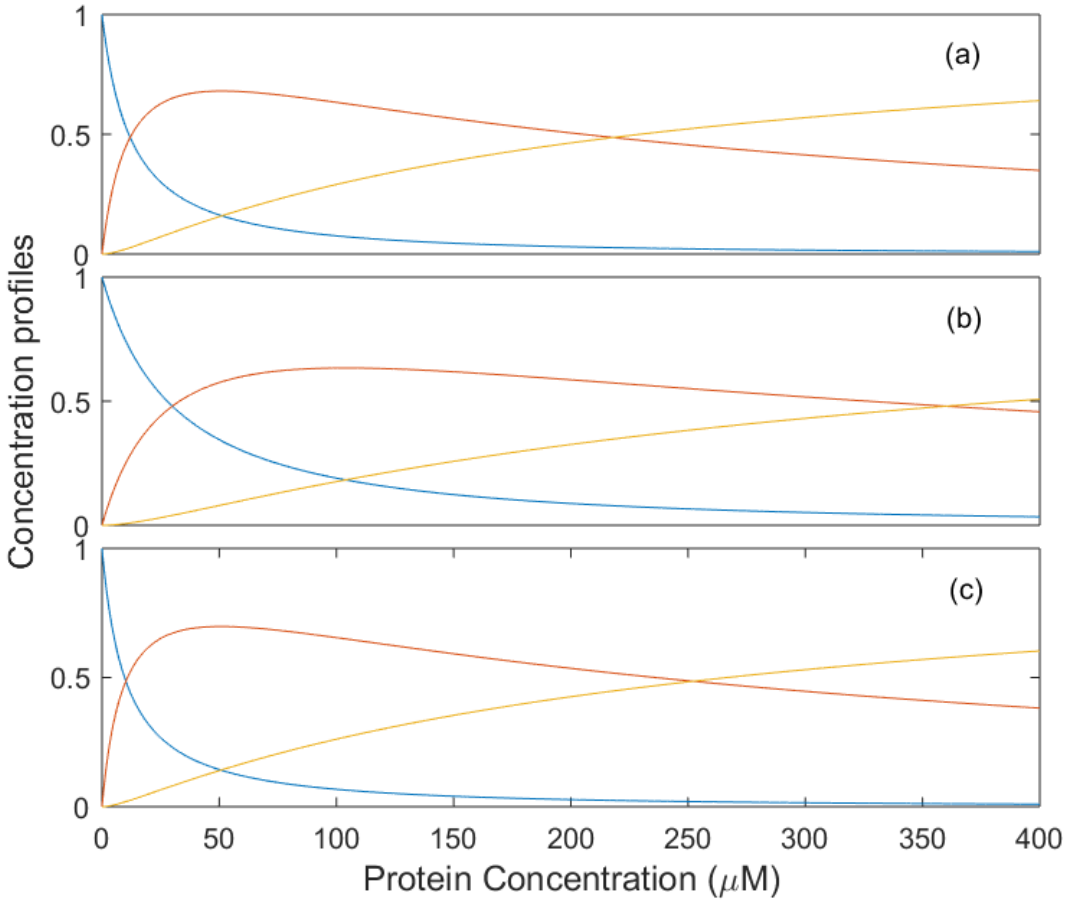


Figure 5.34: The modeling of three components concentration profiles using multi-steps equilibrium equations with (a) K_{d1}=12.5 and K_{d2}=217 μM derived from MCR-ALS, (b) K_{d1}=40 and K_{d2}=360 μM derived from EFA, (c) K_{d1}=10 and K_{d2}=257 μM derived from second-derivative MCR-ALS.

K_{d1} and K_{d2} were derived from multivariate data analysis which resolved the experimental data into three components by MCR-ALS, EFA and

second-derivative MCR-ALS shown in (a), (b) and (c), respectively. All three figures show the concentration contributions of two kinds of bound ligand at 400 μM. Among the three sets of concentration profiles, the concentration profiles shown in (b) are more compatible with its original calculated data in Figure 5.29. Moreover, its K_d values are in agreement with the association constant calculated by sigmoid function. The mean value of two association constants $K_1=\frac{1}{40}=0.025\ \mu M^{-1}$ and $K_2=\frac{1}{360}=0.0027\ \mu M^{-1}$ is equal to the K_a calculated in Figure 5.33.
The difference between two association constants confirms the binding to nonequivalent sites on the ligand molecule. Though all three arms are identical, instead of two different binding sites, we can consider two classes of binding: one for the ion-pairing monitored by the vibrational change of the guanidinium and the other one for the bound ligand in which the hydrogen bonding from the NH-pyrrole also contributes to binding. This assumption makes much sense based on the spectra in Figure 5.32, in which the intensity ratio of the two Raman marker bands (ca. 1467 cm^{-1} Raman intensity ascribed to the vibrational mode of guanidinium to the intensity of Rman band at ca. 1429 cm^{-1} attributed to NH-pyrrole vibration) is the main difference between the spectra of two bound ligand molecules.

5.4.5 Molecular docking simulation

To provide insights into the possible binding modes of the tri-armed GCP-based ligand, a series of docking simulations on different regions of leucine zipper was performed by a collaboration partner in computational biochemistry (CRC1093, subproject A8, Dr. Pandian Sokkar). Two regions were identified containing patches of acidic residues for the docking calculations against the positively charged multivalent ligand (AutoDock Vina15 [92] and MGLTools16 [93] were used). A flexible docking approach was employed, in which dihedral angles of the side chains of the acidic and basic residues are allowed to rotate. The multivalent ligand was fully flexible. Five docking calculations were performed on each region. Although tentative, the docking studies clearly suggest that the tri-armed ligand binds strongly to leucine zipper. A potential binding site was found to be in the middle segment of the leucine zipper dimer (Figure 5.35, left). There, the supramolecular ligand interacts with the sidechains of GLU20, GLU27, GLU30 and GLU31 of chain 1, GLU31 and GLU29 of chain 2 (Figure 5.35, bottom right). Interestingly, the carboxylate group of GLU31 is encapsulated by one GCP as well as by the ammonium N-H (top right).
Another important binding motif was found at the N-terminal segment of the Leucine Zipper dimer. There, the guanidinium groups of the tri-armed GCP-based receptor interact with the acidic residues GLU5 of chain 1, GLU2 and ASP6 of chain 2, while the amide oxygen atom of the ligand

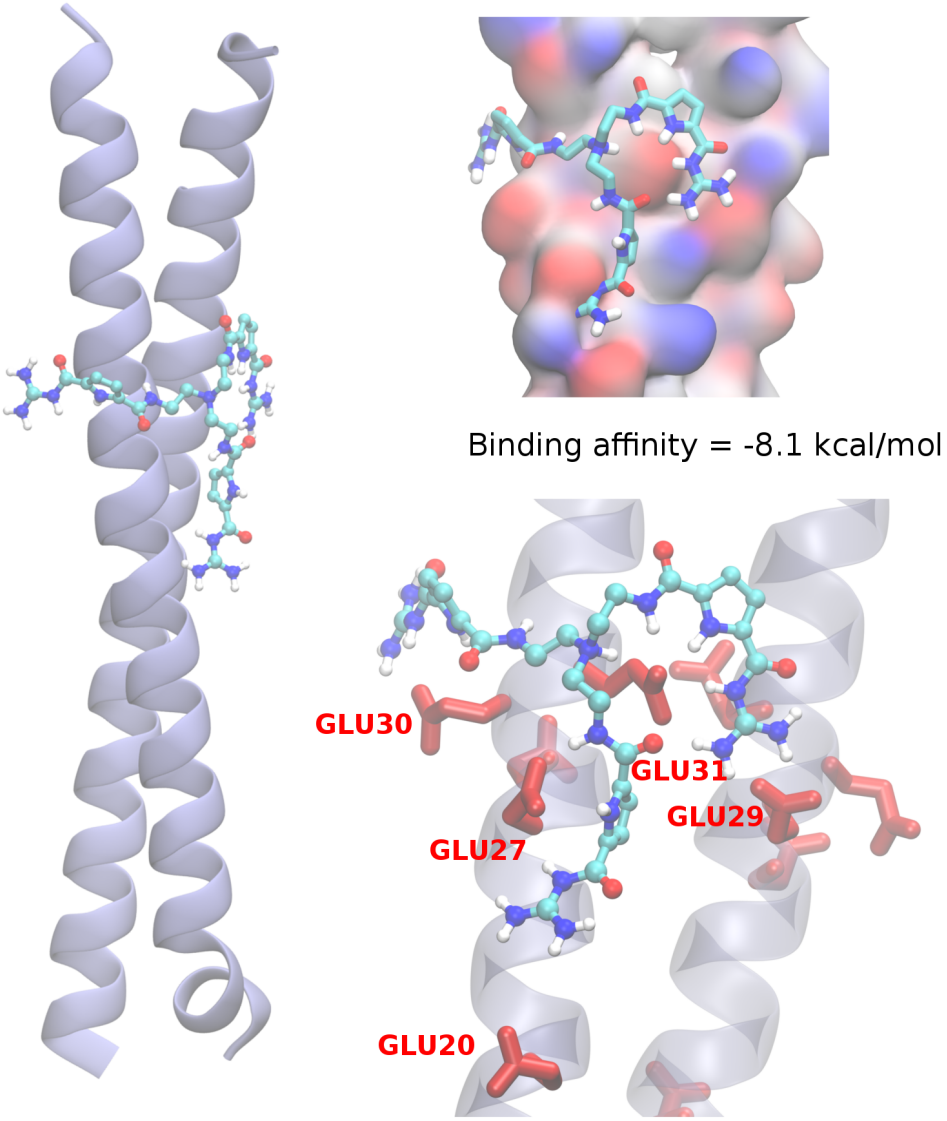


Figure 5.35: Tentative binding site of the tri-armed GCP-based supramolecular ligand with the leucine zipper dimer. Overview of the entire protein–ligand complex (left). Corresponding molecular surface representation with protein residues colored by element type (oxygen atoms: red, nitrogen: blue)(top right). Glutamate sidechains involved in protein–ligand interaction are shown (bottom right).

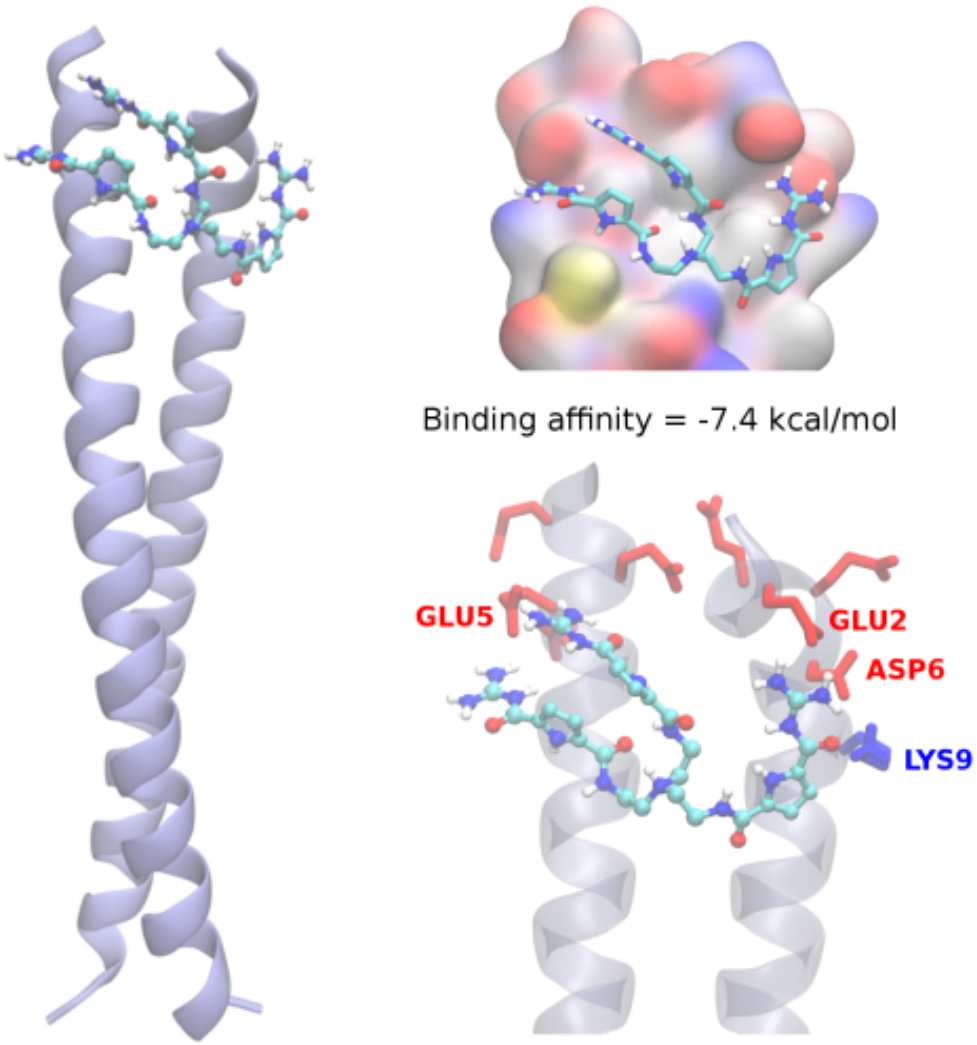


Figure 5.36: Predicted binding mode of the tri-armed GCP-based supramolecular ligand with the leucine zipper dimer near N-terminus. Overview of the entire protein-ligand complex (Left). Corresponding molecular surface representation, same color code as in Figure 5.35 (top right). Individual acidic and basic side chains involved in protein-ligand interactions are shown (bottom right).

interacts with LYS9 of chain 2 (Figure 5.36). A rigid docking with a search space covering the entire leucine zipper dimer showed that the earlier binding mode was favorable.
By comparing the docking simulation with the multivariate data analysis results, binding of protein to non-equivalent sites of ligand would make more sense. This assumption can be seen more clearly in Figure 5.35 in which the contribution of hydrogen bonds in capsulating GLU31 makes difference between the binding of three arms.

5.5 UVRR spectra of molecular twezeers

In addition to supramolecular ligand including GCP motif, we used molecular tweezers as a different class of host molecules to prove the concept of using UVRR technique in supramolecular chemistry. The long-term goal was extending the ligand-based approach used in this technique for monitoring the molecular recognition of other kinds of supramolecular ligand. Although molecular tweezers has been developed in its structure from its early design (phosphate molecular tweezers) by using different anions and linkers for the purpose of site-specifity in targeting proteins (reviewed in section 2.1 and section 3.3), even the recent biological applications were reported by using the phosphate tweezers (CLR01) [94–96]. Moreover, it is a good example for showing the basic principles of "process specificity" of the molecular tweezers in peptide recognition. Therefore, we used phosphate tweezers to investigate its interaction with a tri-peptide including one lysine and two alanine (KKA) shown in Figure 5.37, left and right, respectively.
With our UVRR experimental setup, applying 266 nm excitation wavelength, we failed to record a Raman spectrum of molecular tweezers since it was completely obscured by fluorescence. Therefore, we used a back-scattering microscopic setup (in the laboratory of physical chemistry at Jena university, Prof. Jürgen Popp) with an excitation wavelength of 244 nm to record the first UVRR spectrum of molecular tweezers. The spectra of neat molecular tweezers, at 520 μM and its mixture with 9 equivalents of KKA are displayed in Figure 5.38.

Observed by DFT calculation and band assignments, the vibrational mode of the tweezers's cavity is enhanced by this exciting line. However, the idea of applying ligand-based UVRR technique for studying the binding between ligand and partner molecules is based on enhancing the vibrational mode of the ligand chromophoric group which is mainly contributed in binding. In molecular tweezers, while the net results of the contribution of both cavity and linker decides the stability of the complex, the vibration of the cavity is enhanced by UV excitation wavelength. In other words, an exciting line in resonance with the phosphate group is required to monitor

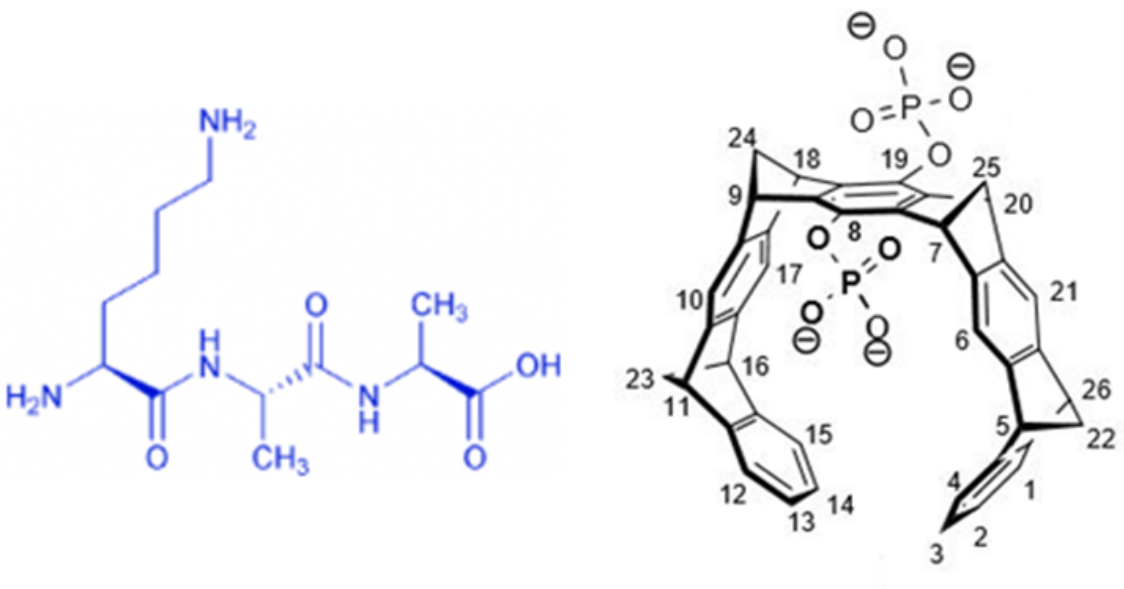


Figure 5.37: The chemical structure of (laft) tri-peptide used as a binding partner, (left) molecular tweezers.

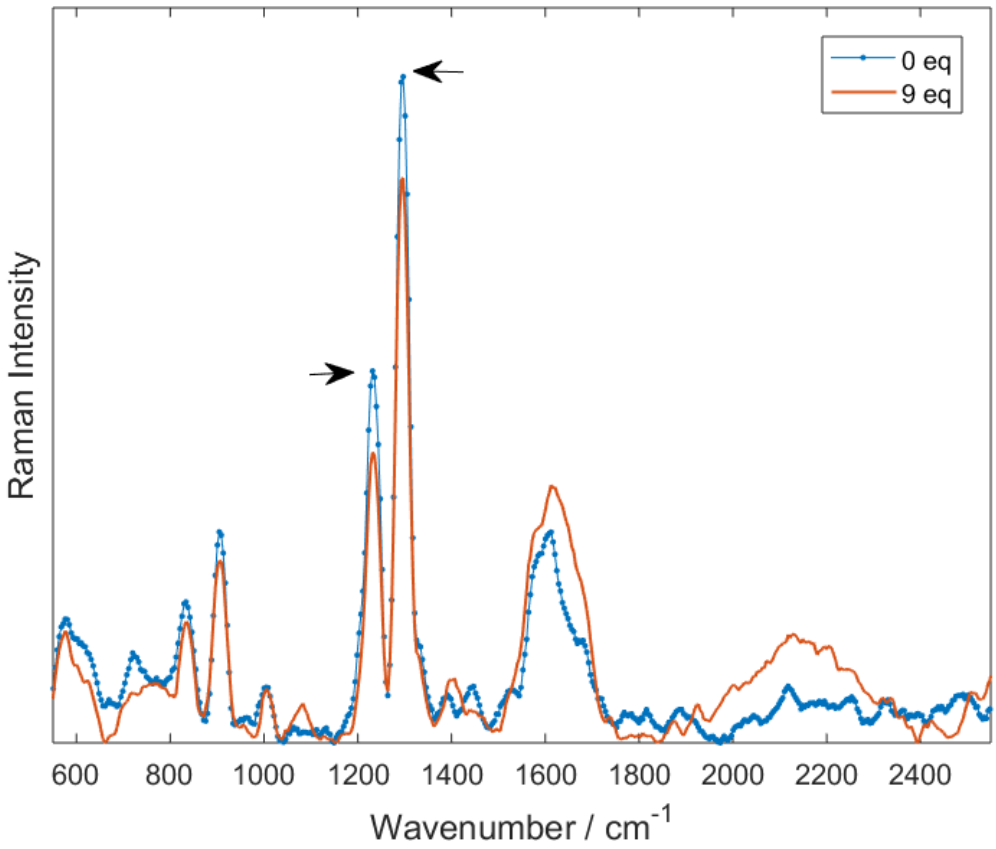


Figure 5.38: UVRR spectrum of molecular tweezers excited by 244 nm

its lysine- and arginine-selectivity. Nonetheless, the cavity itself has also a notable effect in binding. The inclusion of a guest into the tweezers's cavity is a result of hydrophobic interaction between CH groups of the guest

and π-electron rich tweezers's cavity. This cation-π interaction between the cavity and lysine or arginine is possible to be evaluated by UVRR spectroscopy. In particular, from band assignment in DFT calculation, the two intense Raman bands, between 1200 and 1400 cm^{-1} in Figure 5.38, are attributed to the benzene ring connected to the linkers. These bands, which are originated from the stretching mode along the axes of linkers (connecting axes between C8 and C19 in Figure 5.37), are more influenced by adding the binding partner. The intensity of these Raman marker bands decreases upon addition of KAA only after adding high equivalent of partner molecule. An analysis of the spectra by MCR-ALS, shown in Figure 5.39, displays a concentration change of complexed tweezers after addition of partner molecule (KAA) by 6 equivalents of the molecular tweezers's concentration. Because of the instability of the laser power in that time, which can be observed in the intensity fluctuation in some titration points at the bottom of Figure 5.39, the accurate binding studies could not be accomplished.

In general, there is a promise for applying UVRR technique to monitor

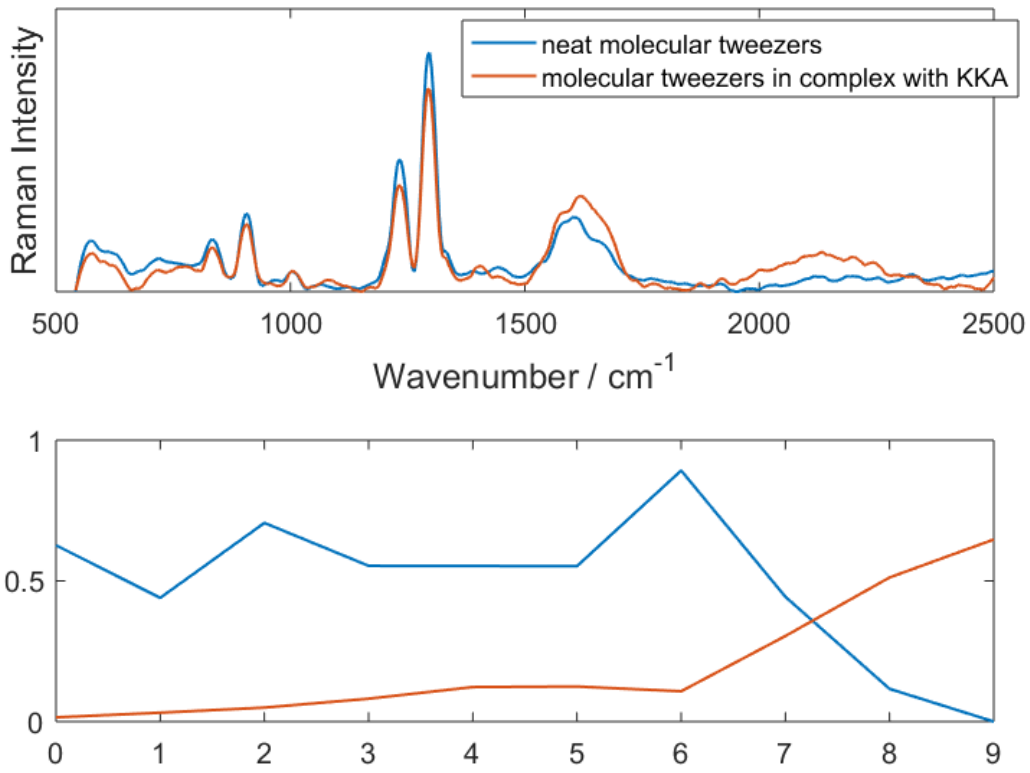


Figure 5.39: (top) The spectra and (bottom) concentration profiles of neat and complexed tweezers, resolved by MCR-ALS.

and evaluate the binding between molecular tweezers and peptide through the vibrational changes of the tweezers's cavity which plays a crucial role

in forming the complex. However, performing more experiments by using additional excitation wavelengths in resonance with the linkers, it should be possible to evaluate the binding from different point of interactions, i.e, the contribution of ion pairing and hydrophobic interactions guided by the linkers and cavity, respectively. These experiments are highly recommended for the future binding studies of the molecular tweezers.

Chapter 6

Comparison and outlook

When talking about developing a technique for the application of a special field of study, besides the basic principles of the technique, the system under study becomes significantly important. Moreover, the main development in this project compared to the previous binding studies by UVRR spectroscopy was extending the approach from peptide to protein recognition by supramolecular multivalent instead of monovalent ligand. Therefore, in this chapter, we will first look at the supramolecular system under investigation with the emphasize on the structural differences between leucine zipper and dermcidin as well as two multivalent supramolecular ligands. This comparison would be helpful for our final outlook in the view of the fact that binding studies between leucine zipper and compound **2** was more successful than binding studies between dermcidin and compound **1**. In another comparing discussion, the key differences between two multivariate data analysis methods will be briefly discussed since it was a significant point to use an additional multivariate data analysis method for analyzing the multi-steps complexation process when there was no prior information about how two molecules bind was available. The conclusions from these comparisons and discussions can broaden our outlook on this project.

Protein leucine zipper versus dermcidin recognition

Electrostatic interactions provide the main strength of the complexation of oxo anions. Therefore, the higher the number of carboxylates (negatively charged side chains) on the protein surface, the stronger complexation with the GCP-based supramolecular ligand. The number of negatively charged residues in dermcidin is 9 compared to 13 glutamic and aspartic acid available in the structure of the protein leucine zipper. Therefore, though both structures have negative net charge, the net charge of leucine zipper is bigger than that of dermcidin (-3 compared to -2). In addition, the clustering of these negatively charged residues is also

important for increasing the probability of the binding. As it was mentioned in section 4.2, the protein leucine zipper possesses an interesting coiled coil structure which brings the residues on two adjacent monomers closer to each other, making an ideal dimer structure for multi-armed supramolecular ligand to bind to anionic group on different adjacent monomers. Since understanding the leucine zipper dimer structure permits us to figure out its interaction with ligand in the residue level, here we look at the interhelical interactions with a focus on the positions of the charged residues.

Shown in Figure 6.1 is a helical wheel representation of leucine zipper.

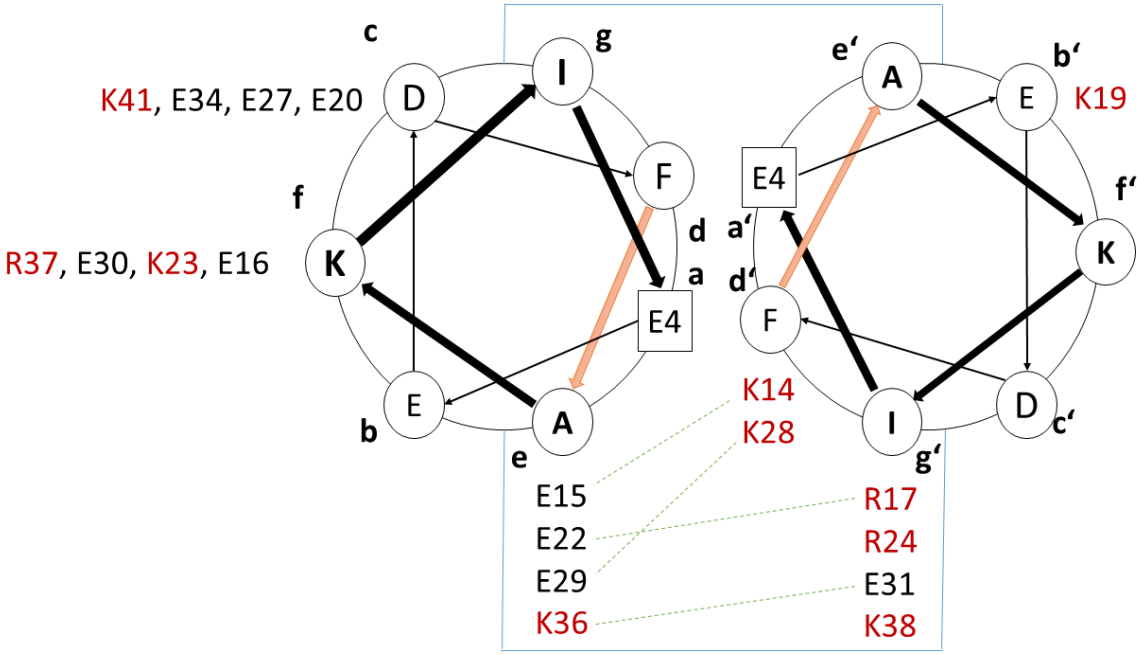


Figure 6.1: Cross-sectional wheel diagram of coiled coil leucine zipper I_α.

The propagation of the polypeptide chain is directed into the page from N to C terminus with parallel and homo chains. The complementary hole of "$a - a' - d - d'$" makes the core packing geometry. The central box containing the hydrophobic residues at "a", "d", "e" and "g" positions on one chain and those at "a'", "d'", "e'" and "g'" positions on another chain signify the complete dimerization interface. Most of the a and d positions are occupied by leucine and isoleucine, forming the hydrophobic core. However, there are four non-Leu and non-Ile at the "d" position which also provide interhelical interactions. The side chains of F7 and F7$'$ make a π stacking interaction, tethering the N-termini together. K14 and K28 from one monomer form an interhelical salt bridge with E15$'$ and E29$'$ from the other monomer, respectively. Finally, C42 from one monomer forms a disulfide bond with C42$'$ from the other monomer, fastening the C-terminus of leucine zipper dimer [79]. The amino acids in the region of interface are buried because of the "knob to holes" pattern. However,

the residues that flank this predominantly hydrophobic face (i.e., e and g positions) can provide additional interactions. These interactions likely contribute in dimerization process through the formation of interhelical ion pairs between residues in the g_i and e'_{i+5} positions. For example, the electrostatic interactions between R17 and E22 and also between $E31$ and K36 are supposed to be stabilizing interactions as shown in Figure 6.2. Moreover, Glu15 and Glu29 have already participated in making interchain salt bridges.

This detailed analysis of the leucine zipper structure can be applied for

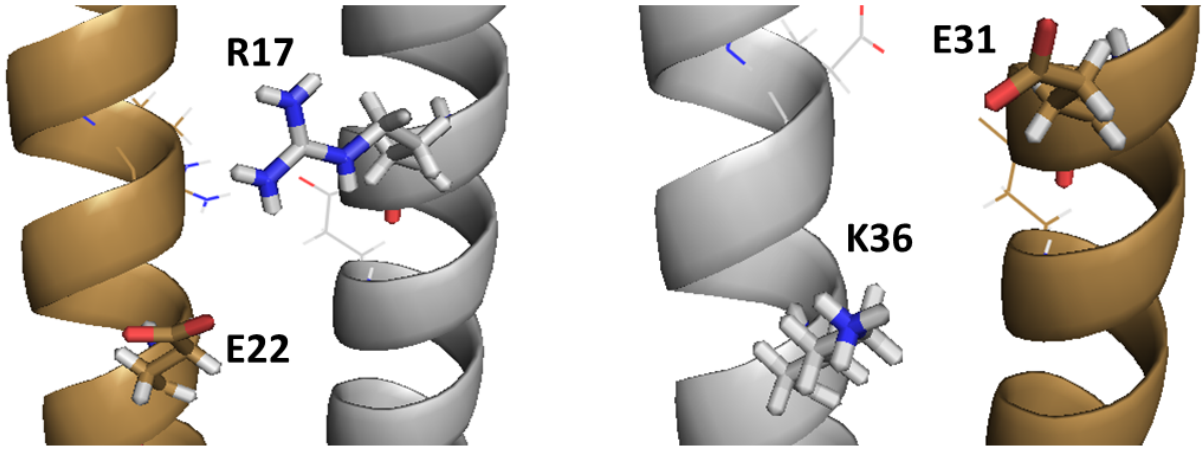


Figure 6.2: Stabilization of leucine zipper I_α dimer interactions by interhelical ion pairs between residues in the g_i and e'_{i+5} positions

justification of possible binding modes calculated by molecular docking. From 13 negatively charged residues which are counted as the candidates for binding with GCP motif via their carboxylate side chains, one (GLU4) is buried in the region of the interface. The glutamic acids at "e" and "g" positions, though take part in the dimerization, still have the possibility to be recognized by GCP motif. It is even assumed that the overall hydrophobic characteristic of this central box makes it an ideal background for hydrogen bond interactions. So, if the glutamic acids placed in this region are flanked in a suitable angle and flexible enough in their conformation, they can provide a right complementary geometry for making hydrogen bonds with GCP motif, since directionality is an important factor in this kind of interactions. One example of this hypothesis is displayed in Figure 5.35 where the molecular docking shows the encapsulation of GLU31 by a combination of hydrogen bonds with GCP and central ammonium. We see in the wheel diagram that Glu31 is located in the central box at position "g". Compared to Glu15 which form salt bridge with Lys14 (see Figure 6.3, right), it seems that Glu31 has more flexibility because it is separated from its ion pair partner (Lys36) by a longer distance. On the other hand, Glu22 has the same flexibility condition as Glu31 regarding its interaction with Lys17 (see Figure 6.2,

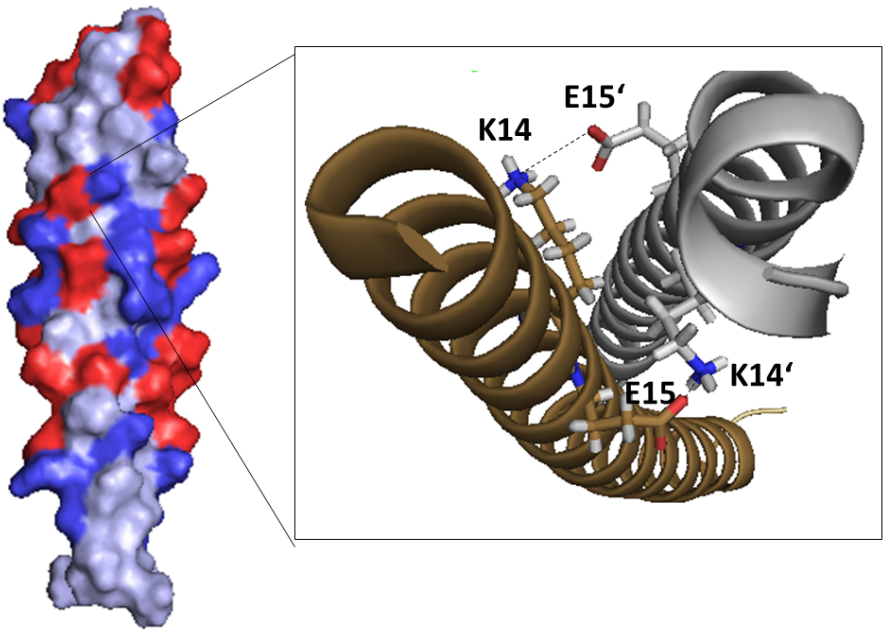


Figure 6.3: Electrostatic surface of Leucine Zipper (left), basic and acidic residues forming interchain salt bridges (right).

left). However, perhaps the directionality of Glu31 is more appropriate than that of Glu22, so that it is preferred for making hydrogen bonds with ligand (according to molecular docking). Because it was shown in earlier studies that the carboxylate groups which allow maximum interactions are those which can make a complete co-planarity with GCP binding sites. Nonetheless, the three arms work together to make the binding happen. So, how the protein binds with ligand is not decided only by the position and directionality of one single amino acid. For example, the electrostatic interaction between other glutamic and aspartic acids which are mainly distributed at positions "c" and "f" in the wheel diagram, with the guanidinium of one arm may also has an influence on the preference of Glu31 for making hydrogen bonds. It is enough to see that Glu31 follows Glu30 which can bind to another GCP making the binding stronger, while the adjacent amino acid to Glu22 is Lys23 which can be the source of repulsion force toward the ligand.

On the other hand, for having an efficient binding, especially for the electrostatic interactions, the distributed charge in the structure of the receptor play a major role in targeting the protein. The representation of leucine zipper by its molecular surface, shown at the left of Figure 6.3, displays the distribution of both electronegative (red) and electropositive (blue) potentials across the surface. Although the electrostatic protein surface features clusters of negatively charged amino acids nearly elsewhere, their localization and the distance between their side chains with positively charged residues are the important factors for recognition by GCP motif

from supramolecular ligand, especially because of the importance of the hydrogen bonds in the recognition process by GCP motif. Being in neighborhood with hydrophobic residues permit the charged residues to make hydrogen bonds with GCP motif with less opposite influence of water molecules. These points can justify the docking results, in which only the residues edged from the hydrophobic interchain surface in leucine zipper and partially surrounded by hydrophobic residues were identified as potential binding sites to be packed by one GCP included in three-armed ligand via hydrogen bonds (Figure 5.35). Similarly, both negative and positive side chains are distributed over all helices in dermcidin (see Figure 3.3). However, the main difference in the location of charged residues in the structure of two proteins comes from their vicinity with hydrophobic amino acids. As shown in the dermcidin hydrophobic surface in Figure 6.4, the charged residues are clustered on one side of the helix and the hydrophobic residues are clustered on the other side, thus the structure of dermcidin in amphipathic [97]. In leucine zipper, the charged residues are distributed among the hydrophobic residues in the interchain area (Figure 6.5).
In general terms, the vicinity of two helices in the context of coiled coil structure, on one hand, and the well defined sequential position of the charged residues among the hydrophobic residues, on another hand, which itself is also a result of leucine zipper coiled coil structure, makes it a relatively ideal partner molecule for multiarmed GCP supramolecular ligand. These properties are missed in the structural geometry of dermcidin. However, depending on the positions of target residues on leucine zipper and the potential electrostatic repulsion or attraction, their binding with ligand can destabilize or stabilize the dimer, an important issue which should be taken into account because leucine zipper domain meditates homodimerization and cellular targeting of kinase PKG and disrupting the dimerization of the PKG I_α leucin zipper domain abrogates targeting of the kinase.

Supramolecular ligand compound 1 versus compound 2

The main features of two supramolecular ligands, compound **1** (Figure 4.2) and compound **2** (Figure 4.3), were represented in section 4.1. Here, based on the results from their binding studies with proteins, we look up the structural differences of two molecules which we think might cause these results. At first glance, the major differences are the number of GCP motif and the length of two compound's linker. Although both have three positive charges, one of them in compound **1** is only guanidinium without being attached by pyrrole and carbonyl. The long linker of compound **1** makes it more flexible than compound **2**. However, this flexibility

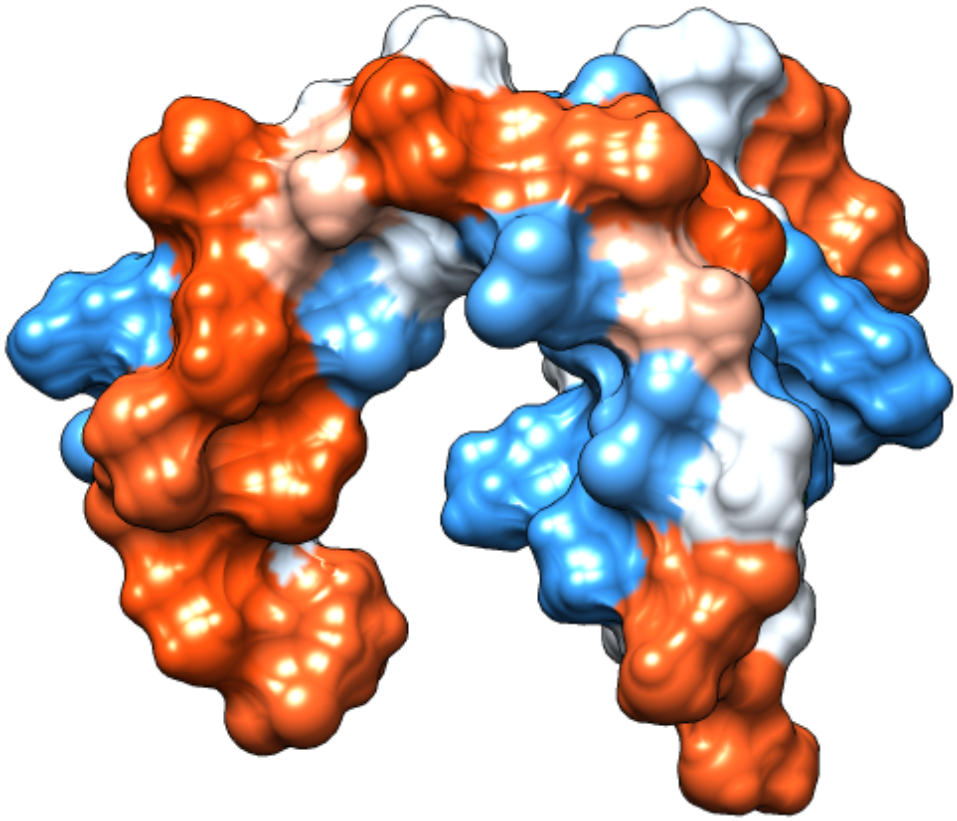


Figure 6.4: Electrostatic surface of the protein dermcidin with negatively charged residues in red, positively charged residues in blue and neutral in white

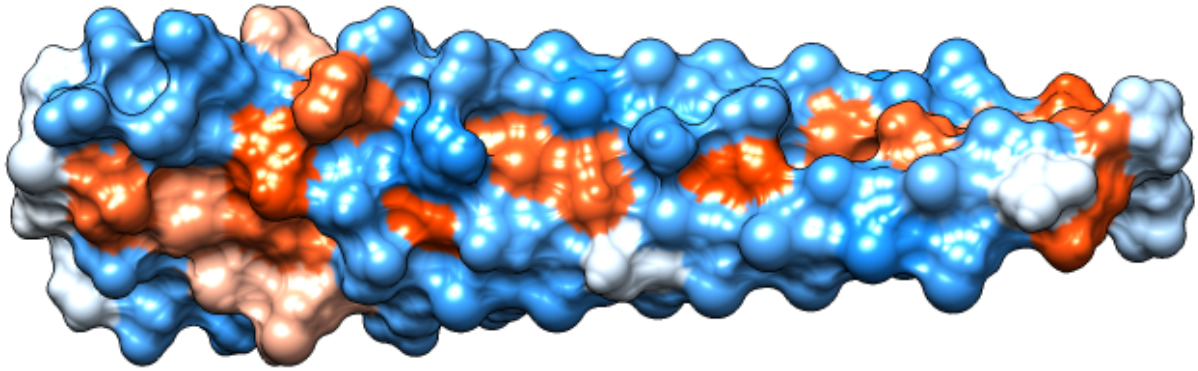


Figure 6.5: Electrostatic surface of the protein leucine zipper with negatively charged residues in red, positively charged residues in blue and neutral in white

seems to act as a negative point for binding with dermcidin, in which the negatively charged residues are distributed among the four close helices. In comparison, the attachment of three GCPs by a small but flexible linker in compound **2** has made it a suitable binder for leucine zipper dimer. Apparently, the two homo adjacent helices provide an appropriate geometry for multi-armed supramolecular ligand. Especially, the central ammonium N-H can help one GCP motif to encapsulate the carboxylate groups, as we see in Figure 5.35. This efficiency of the scaffold for helping the GCP arms by involving in hydrogen bonding is a result of small linker. A longer linker would not allow this ammonium to take part in hydrogen bond in the context of encapsulation. Another major feature which can have an influence on molecular recognition by compound **1**, either positive or negative, is the carboxylate group available in the structure of this multivalent supramolecular ligand. This group imposes a positive effect if it can access the positively charged residues on the target protein. However, it is also possible that this group makes a self-aggregation which then causes a negative effect on binding. The later effect was observed when we prepared the ligand solution with high concentration.
In conclusion, for understanding the detail of the interactions between protein and supramolecular ligand, it is helpful to zoom in their structural characteristics with an emphasize on the position of the binding sites, as we did during the last two sections. However, it should be always reminded that each factor can not define the outcome of the binding process by itself. Each conformation and geometrical characteristic in one molecule in complement with those of its partner molecule should work together and, at the same time, compete with environmental condition in the favor of a strong binding. An unambiguous analysis of the binding can only be made by crystal structure of the complex molecule.

Multivariate data analysis NMF versus MCR-ALS

In this thesis, two multivariate data analysis methods, NMF and MCR-ALS, were applied to analyze experimental spectra. In particular, we presented the application of MCR-ALS for exploring the information about the probable intermediate states in a multi-equilibria binding. Because there was no prior information about the manner by which two molecules bind, we analyzed the data from different perspectives, e.g., analyzing with different initial estimation of the number of components or analyzing the second derivative data. For this purpose, MCR-ALS was a suitable candidate because it is broadly used in multivariate data analysis of multi-equilibra case studies for different experimental data. While the basic principles of all multivariate analysis techniques are nearly the same, MCR-ALS is known for its flexibility in applying different constraints on

the system. As an example of utilizing this flexibility, we can mention the multivariate analysis of spectra from UVRR titration of supramolecular ligand by protein dermcidin (Figure 5.15), where there was a need for defining a new closure constraint, since the spectrum of dermcidin had to be also considered in the analysis. Instead of rewriting or changing the algorithm for NMF to add a new constraint for limiting the concentration of protein, there is a possibility in MCR-ALS for user-defined constraint during the process of calculation. This fundamental potential also helped us to define the number of species during the calculation in order to resolve the information, e.g., spectra and concentration profiles, of each component.
In multi-component data analysis, due to the overlap of individual components and a presence of a baseline, the intensities have contributions from more than one component. As a result, the calculated concentrations of each individual component have contributions from other components. In fact, the overlap of spectra results in an under-resolved concentration and over-resolved spectra. This is the main reason for eliminating the baseline before starting the calculation. However, because second derivative spectra usually have sharper and better-resolved peaks, compared to the conventional spectra, the overlap of the spectra is reduced [57]. Therefore, we calculated the second derivative spectra to be analyzed by MCR-ALS. NMF is not able to analyze the second derivative spectra because of the negative values.
Generally, using the same constraints, both methods derived nearly similar concentration profiles for two components from UVRR spectra (Figure 5.24). The additional analysis by considering the assumption of three components was performed to calculate the concentration profile of different potentially bound ligands. The outcome of these calculations suggested the model of binding to non-equivalent binding sites in supramolecular multi-armed ligand which was also justified by molecular docking. Although, the three-armed ligand has identical arms including GCP motif, the two potential sorts of binding including hydrogen-bonds and electrostatic interactions can make these arms non-equivalent. This model can be the source of sigmoidal shape in the binding curve.

Outlook

Overall, we demonstrated the concept of UVRR binding studies for small protein recognition. While earlier studies focused on small tetrapeptides (4 amino acids), we now extended the size of the molecular system by more than one order of magnitude. The challenge of facing with fluorescence was overcome by choosing the protein with minimum number of aromatic amino acids in its sequence. The sensitivity of UVRR in

probing the small structural change of GCP motif allowed us to monitor the binding between a supramolecular multivalent ligand and biologically highly relevant proteins, leucine zipper with 47 and dermcidin with 49 amino acids. Even the preliminary studies for optimizing the experimental condition, e.g., pH and concentration, was performed mainly by UVRR spectroscopy. For qualitative analysis of the spectral change, we presented a new series of multivariate data analysis to extract the spectra and concentration profiles of the involved components with different initial assumptions as well as the evolution of the binding process. Applying different analytical techniques with various initial estimations enabled us to overcome the limited information priorly available about the system and propose a binding model based on qualitative analysis of the data and molecular docking. Moreover, while the successful binding studies were performed between a supramolecular multivalent GCP-based ligand and protein leucine zipper, we made some efforts for probing the binding event between molecular tweezers and small peptide.
With this proof of concept, the future of this project is promising for binding studies between more complex molecular systems including GCP motif. However, it will need to overcome the disturbing autofluorescence generated upon 266 nm cw laser excitation and also to address the vibrational spectroscopic differentiation between the different binding arms of multivalent supramolecular ligands. On the other hand, an initial monitoring of spectral change of the molecular tweezers by UVRR provides the basic information for starting a new project for binding studies of molecular recognition by this supramolecular ligand.

Chapter 7

Summary and conclusion

Ultraviolet Resonance Raman (UVRR) spectroscopy was employed for label-free monitoring of molecular recognition of protein by supramolecular ligands. The multivalent supramolecular ligands armed with guanidinium carbonyl pyrrole (GCP), known as an efficient binder for carboxylate group by a combination of electrostatic and hydrogen bonds, were developed and synthesized by Schmuck and coworkers. The binding properties of this motif in peptide recognition was examined in previous UVRR studies providing a valuable information including the spectral change of GCP motif due to binding with partner molecule (tetrapeptide) backed up with DFT calculation and band assignment. This information stimulated the extension of the technique from monitoring the peptide recognition to protein recognition. The basic principle remained the same: the selective and sensitive monitoring of GCP by enhancing its vibrations through its electronic resonance with UV excitation wavelength is used for recording its spectral change upon addition of the partner molecule. The selective determination of GCP motif by UVRR spectroscopy is of more interest in protein recognition because the whole system contains more amino acid groups. In this regard, the recognition of two proteins, dermcidin and leucine zipper was investigated by UVRR spectroscopy by two multivalent supramolecular ligands (compound **1** and compound **2** shown in Figure 4.2 and Figure 4.3).

Finding the optimum condition for the main UVRR titration experiments consisted of two major titrations: pH-dependent UVRR titration and concentration-dependent UVRR experiments. The first was necessary because the strong electrostatic interaction is a result of protonated GCP. Although we knew the suitable pK_a value for protonated GCP motif from previous studies (6.5), the multivalent supramolecular ligand supposed to have more than two forms of protonated and deprotonated GCP due to the presence of three GCP motifs which results in having partially protonated ligand. Interestingly, nearly same value ($pK_a = 6.42$) was determined for fully protonated ligand as shown in Figure 5.12. In

addition, another value ($pK_a = 7.5$) was found in the equilibrium between partially protonated and fully deprotonated states. The advantage of this experiment was finding the range of pH for various protonation states of multivalent supramolecular ligand which can be especially helpful for the future binding studies. While two states for partially protonated ligand were anticipated, only one concentration profile was suggested for the intermediate state (three states in general). The failure of resolving four spectra could be due to our experimental condition, e.g., chosen pH interval (0.5). If this assumption is not the case, then we can conclude that the three armed supramolecular ligand only exist in three protic states (instead of four) which is an interesting case and it would have a significant influence on the interpretation of the results of the binding studies. For having more reliable results, a pH titration with small pH interval is suggested for the future binding studies of multivalent supramolecular ligand.

The second preliminary experiment, concentration dependent UVRR titration, was performed with Li-40 and the results were extended for compound **1** and compound **2** (shown in Figure 5.8). With acquiring the nonlinear function of Raman intensity versus concentration, we found the optimum range of sample concentration within which the attenuation of the scattered light is minimal. Moreover, the quantitative knowledge of ligand self-absorption is important since the relative intensities of Raman bands can be changed because of self-absorption while the change due to binding between two molecules should be considered for binding studies. In binding studies with constant concentration of ligand, the latter effect is neglected and the main purpose of concentration-dependent experiments is finding the optimum concentration. However, for the binding studies with more complex systems for which a protein-based titration experiment with bigger range of concentration variation is inevitable, the self-absorption correction is necessary. Therefore, applying the concentration-dependent nonlinear function of Raman intensity in the final multivariate data analysis is suggested for such a system.

Two multivariate data analysis methods, NMF and MCR-ALS, were applied for analyzing both the pH dependent and the binding UVRR studies, with the main advantages of using the whole available spectral information and having no *a priori* information of the system. In particular, two methods were used and compared for analysis of the main ligand-based binding studies of compound **2** and protein leucine zipper for finding the concentration profile and pure spectra of two components, free and complexed ligand. Analyzing the data with the assumption of multi-step equilibria was performed by MCR-ALS. We applied this assumption in order to show the right binding model for the observed sigmoid function of the concentration profile of bound ligand resulted from

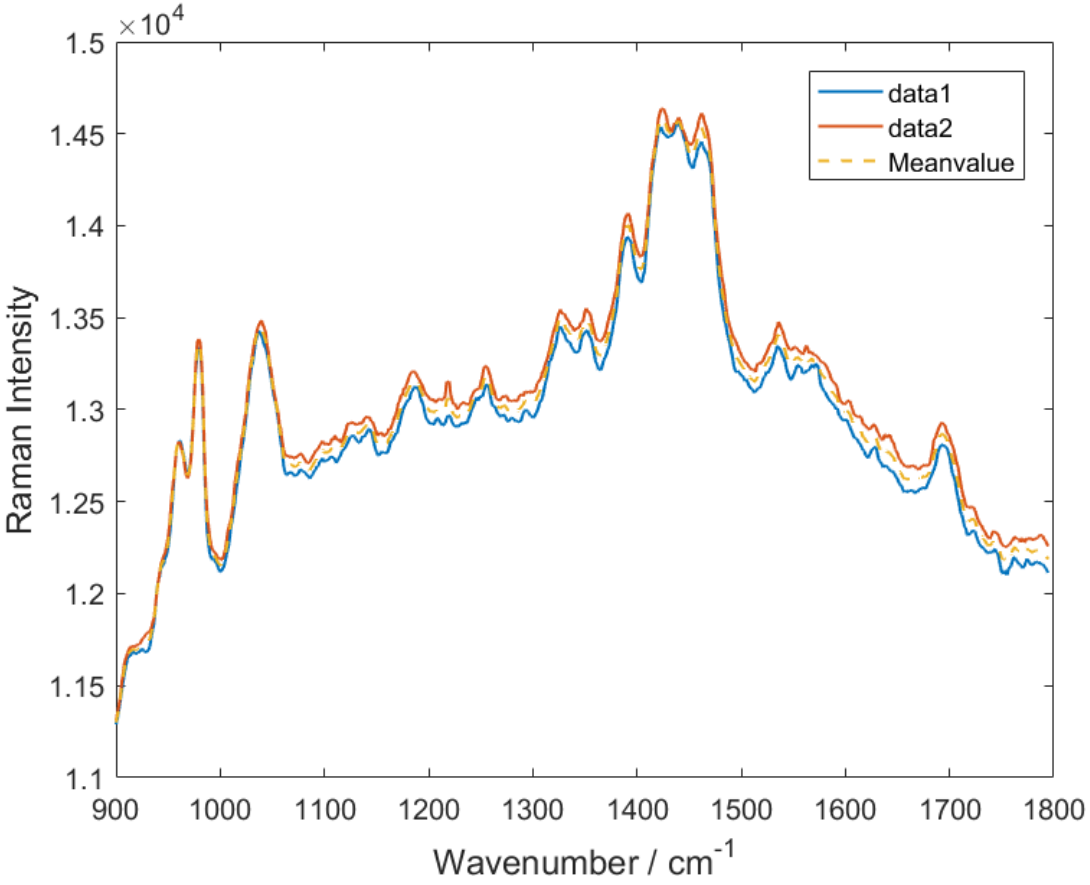
two component analysis. Based on the two different calculated binding constants from multi-steps binding analysis and with considering the molecular docking results, binding to non-equivalent binding site caused the sigmoidal shape of bound ligand concentration profile.
In addition to GCP-based supramolecular ligands and with the purpose of extending the UVRR binding studies for a different class of supramolecular ligands, we performed some experiments to evaluate the binding between molecular tweezers and a tripeptide containing lysine. Because of the interference of the fluorescence resulted from the benzene rings of the cavity in the Raman spectra with 266 nm excitation, we used 244 nm excitation for this experiment. The elementary results showing the intensity change of the Raman bands of the central benzene connected to the linker was promising for the future binding studies of molecular tweezers. However, since the lysine/arginine recognition by molecular tweezers is guided by different parts of the molecule (cavity and linkers) and their vibration are enhanced by different excitation wavelengths, the binding studies of molecular tweezers are more challenging than that of GCP-based ligand and separate UVRR experiments performed by various excitation wavelengths are suggested.
In this thesis, UVRR spectroscopic binding studies for label-free molecular recognition of proteins dermcidin and leucine zipper by supramolecular ligand were successfully demonstrated. In addition to applying the sensitivity of resonance Raman spectroscopy for probing the large systems, multivalency was investigated by UVRR. The multivalent supramolecular ligand with more than one GCP motif was selectively enhanced by UV wavelength to determine their binding effect. In the case of three identical arms containing the GCP motif (compound **2**), the good choice of multivariate analysis method with appropriate initial assumptions allowed us to extract the structural information of the complexation according to molecular docking. By using MCR-ALS backed up with the information from the chemical process acquired by EFA, the pure spectra and concentration profiles were calculated. Moreover, the qualitative analysis of the calculated concentration profiles led to the specification of an average binding constant using defined sigmoid function and two stepwise binding constants by multi-step equilibria.

Appendix A

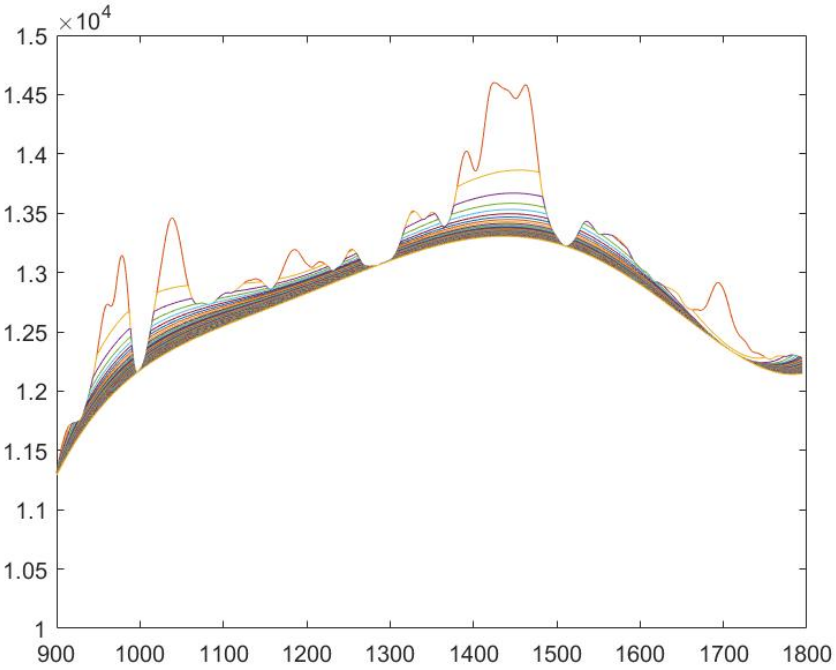
Data pre-treatment

A.1 Binding study - compound 2 : LZ, (1:4 eq)

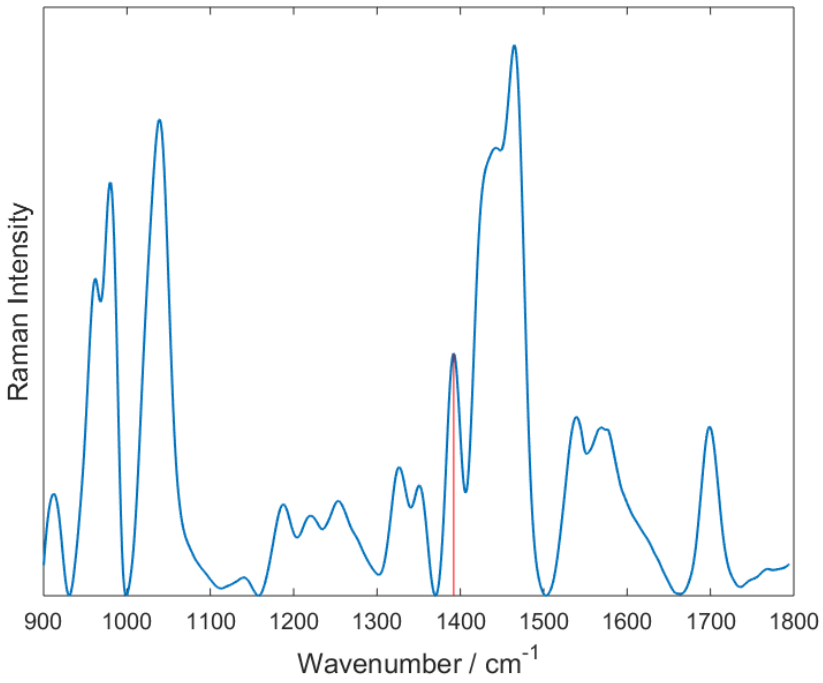
A.1.1 Raw spectrum



A.1.2 Smoothed and baseline-corrected spectrum



A.1.3 The internal standard Raman band in the spectrum of neat ligand



A.2 Matlab code - Modpoly: modified polynomial for baseline correction

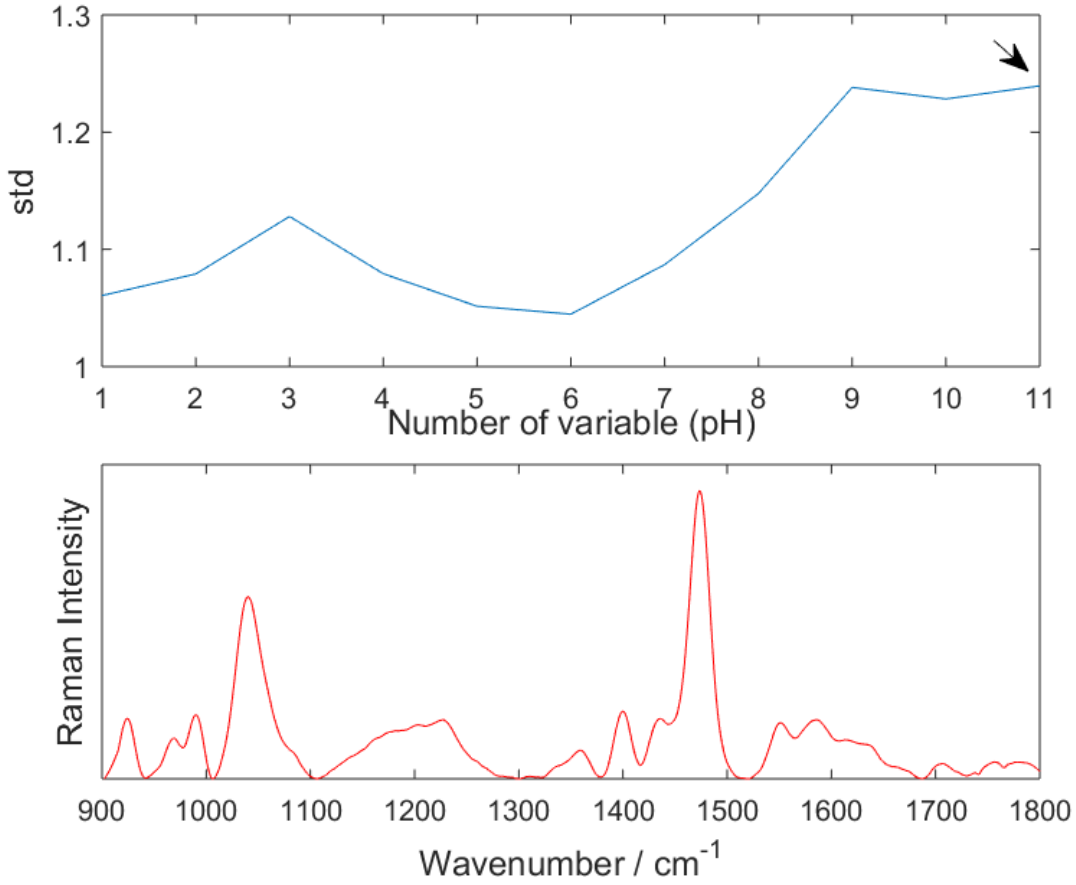
```
function [ Modpoly ] = Modpoly(xdata,ydata,n)
%written by Banafshe Zakeri
% xdata and ydata are matrices with i*1 dimension
ydata=[zeros([length(xdata),1]),ydata];
j=2;
count=0;
while norm(ydata(:,j)-ydata(:,j-1))>0.005
    O=zeros([length(xdata),1]);
    [c,S,mu]=polyfit(xdata,ydata(:,j),n);
    Pfit=polyval(c,(xdata-mu(1))/mu(2));
for i=1:length(xdata)
    if ydata(i,j)<Pfit(i)
      O(i)=ydata(i,j);
   else O(i)=Pfit(i);

    end
end
ydata=[ydata,O];
j=j+1;
count=count+1
end

    plot(xdata,ydata)
   Modpoly=ydata(:,j)
end
```

Appendix B

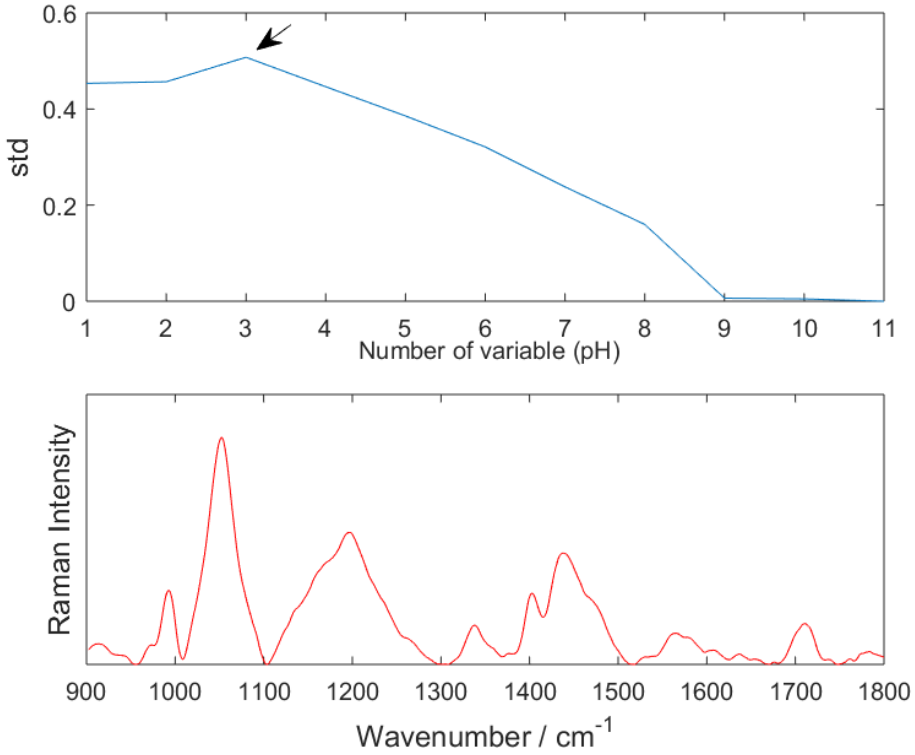
Multivariate data analysis

B.1 Pure variable method for pH-dependent experiments

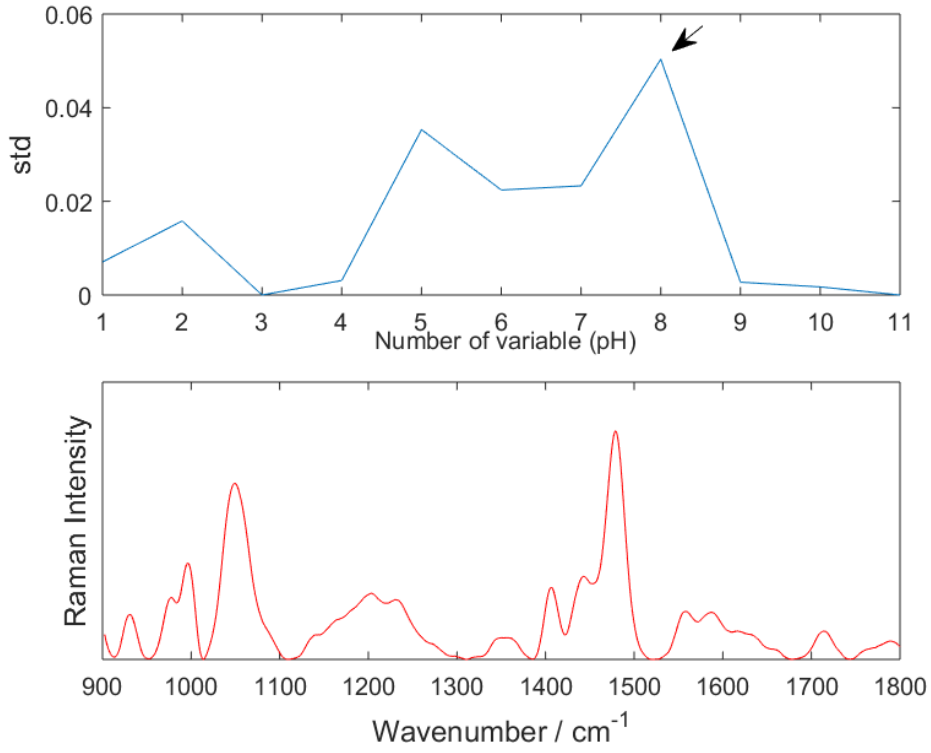
B.1.1 First pure spectrum, pH=11



B.1.2 Second pure spectrum, pH=3



B.1.3 Third pure spectrum, pH=7



Bibliography

[1] Johnson, D. S. and Boger, D. L. *Comprehensive Supramolecular Chemistry*, volume 4. (1996).

[2] Steed, J. W. and Atwood, J. L. *Supramolecular Chemistry.* (2009).

[3] Kubik, S. *Chem. Soc. Rev.* **39**, 3648 (2010).

[4] Gale, P. A., Howe, E. N., and Wu, X. *Chem.* **1**, 351–422 (2016).

[5] Schmuck, C. *Chem. Eur. J.* **6**, 709–718 (2000).

[6] Schmuck, C. *Coord. Chem. Rev.* **250**, 3053–3067 (2006).

[7] Schmuck, C. *Synlett* , 1798–1815 (2011).

[8] Klärner, F. G., Kahlert, B., Nellesen, A., Zienau, J., Ochsenfeld, C., and Schrader, T. *J. Am. Chem. Soc.* **128**, 4831–4841 (2006).

[9] Fokkens, M., Schrader, T., and Klärner, F. G. *J. Amer. Chem. Soc.* **127**, 14415–14421 (2005).

[10] Klärner, F. G. and Schrader, T. *Acc. Chem. Res.* **46**, 967–978 (2013).

[11] Schmuck, C. and Wich, P. *Angew. Chem. Int. Ed.* **45**, 4277–4281 (2006).

[12] Küstner, B., Schmuck, C., Wich, P., Jehn, C., and SK. *Phys. Chem. Chem. Phys.* **9**, 4598–4603 (2007).

[13] Niebling, S., Srivastava, S. K., Herrmann, C., Wich, P. R., Schmuck, C., and Schlücker, S. *Chem. Commun.* **46**, 2133 (2010).

[14] Niebling, S., Kuchelmeister, H. Y., Schmuck, C., and Schlücker, S. *Chem. Commun.* **47**, 568–570 (2011).

[15] Niebling, S., Kuchelmeister, H. Y., Schmuck, C., and Schlücker, S. *Chem. Sci* **3**, 3371 (2012).

[16] Niebling, S. *University of Osnabrück, Diss* (2013).

[17] Smith, B. *Synthetic Receptors for Biomolecules.* (2015).

[18] Moreira, I. S., Fernandes, P. A., and Ramos, M. J. *Proteins: Struct., Funct., Bioinf.* **68**, 803–812 (2007).

[19] Moiani, D., Cavallotti, C., Famulari, A., and Schmuck, C. *Chem. Eur. J.* **14**, 5207–5219 (2008).

[20] Schmuck, C. and Hei, M. *ChemBioChem* **4**, 1232–1238 (2003).

[21] Schmuck, C. and Heil, M. *Org. Biomol. Chem.* **1**, 633–666 (2003).

[22] Schmuck, C., Rupprecht, D., and Wienand, W. *Chem. Eur. J.* **12**, 9186–9195 (2006).

[23] Schmuck, C. and Geiger, L. *Chem. Commun.* , 1698–1699 (2004).

[24] Schmuck, C. and Geiger, L. *J. Am. Chem. Soc.* **127**, 10486–10487 (2005).

[25] Mondal, M. and Hirsch, A. K. H. *Chem. Soc. Rev.* **44**, 2455–2488 (2015).

[26] Schrader, T., Bitan, G., and Klärner, F.-G. *Chem. Commun.* **52**, 11318–11334 (2016).

[27] Attar, A., Chan, W. T. C., Klärner, F. G., Schrader, T., and Bitan, G. *BMC Pharmacol. Toxicol* **15**, 1–14 (2014).

[28] Dutt, S., Wilch, C., Gersthagen, T., Talbiersky, P., Bravo-Rodriguez, K., Hanni, M., Sánchez-García, E., Ochsenfeld, C., Klärner, F. G., and Schrader, T. *J. Org. Chem.* **78**, 6721–6734 (2013).

[29] Dutt, S., Wilch, C., Gersthagen, T., Wölper, C., Sowislok, A. A., Klärner, F. G., and Schrader, T. *Eur. J. Org. Chem.* , 7705–7714 (2013).

[30] Bier, D., Rose, R., Bravo-Rodriguez, K., Bartel, M., Ramirez-Anguita, J. M., Dutt, S., Wilch, C., Klärner, F. G., Sanchez-Garcia, E., Schrader, T., and Ottmann, C. *Nat. Chem* **5**, 234–239 (2013).

[31] Trusch, F., Kowski, K., Bravo-Rodriguez, K., Beuck, C., Sowislok, A., Wettig, B., Matena, A., Sanchez-Garcia, E., Meyer, H., Schrader, T., and Bayer, P. *Chem. Commun.* **52**, 14141–14144 (2016).

[32] Talbiersky, P., Bastkowski, F., Klärner, F.-G., and Schrader, T. *J. Am Chem. Soc.* **130**, 9824–9828 (2008).

[33] Das, R. S. and Agrawal, Y. K. *Vib. Spectrosc* **57**, 163–176 (2011).

[34] Carey, P. R. *Biochemical applications of Raman and resonance Raman spectroscopies*, volume 112. (1984).

[35] Papadopoulos, P. *Colloid. Polym. Sci.* **286**, 487–487 (2008).

[36] Ferraro, J. R., Nakamoto, K., and Brown, C. W. *Introductory Raman Spectroscopy.* (2003).

[37] Maes, G. *Handbook of Raman Spectroscopy. From the Research Laboratory to the Process Line*, volume 59. (2003).

[38] Vrabie, V., Gobinet, C., Piot, O., Tfayli, A., Bernard, P., Huez, R., and Manfait, M. *Biomed Signal Process Control* **2**, 40–50 (2007).

[39] Esteban, M., Ariño, C., Díaz-Cruz, J. M., Díaz-Cruz, M. S., and Tauler, R. *TrAC, Trends Anal. Chem.* **19**, 49–61 (2000).

[40] Windig, W., Gallagher, N. B., Shaver, J. M., and Wise, B. M. *Chemom. Intell. Lab. Syst.* **77**, 85–96 (2005).

[41] Jaumot, J., Eritja, R., and Gargallo, R. *Anal. Bioanal. Chem.* **399**, 1983–1997 (2011).

[42] Schachtner, R. *University of Regensburg, Diss* (2010).

[43] Tauler, R. *J. Chemom.* **15**, 627–646 (2001).

[44] Lee, D. D. and Seung, H. S. *Adv Neural Inf Process Syst* , 556–562 (2001).

[45] Shashilov, V. A., Ermolenkov, V. V., and Lednev, I. K. *Inorg. Chem.* **45**, 3606–3612 (2006).

[46] Diewok, J., De Juan, A., Tauler, R., and Lendl, B. *Appl. Spectrosc.* **56**, 40–50 (2002).

[47] Diewok, J., De Juan, A., Maeder, M., Tauler, R., and Lendl, B. *Anal. Chem.* **75**, 641–647 (2003).

[48] Abdollahi, H., Maeder, M., and Tauler, R. *Anal. Chem.* **81**, 2115–2122 (2009).

[49] Varmuza, K. and Peter, F. *Introduction to multivariate statistical analysis in chemometrics.* (2008).

[50] Savitzky, A. and Golay, M. J. *Anal. Chem.* **36**, 1627–1639 (1964).

[51] Cormack, I. G., Mazilu, M., Dholakia, K., and Herrington, C. S. *Appl. Phys. Lett.* **91** (2007).

[52] Lieber, C. A. and Mahadevan-Jansen, A. *Appl. Spectrosc.* **57**, 1363–1367 (2003).

[53] Zhao, J., Lui, H., Mclean, D. I., and Zeng, H. *Appl. Spectrosc.* **61**, 1225–1232 (2007).

[54] Hyvärinen, A., Karhunen, J., and Oja, E. *Independent Component Analysis*, volume 21. (2001).

[55] De Juan, A. and Tauler, R. *Anal. Chim. Acta* **500**, 195–210 (2003).

[56] De Juan, A., Navea, S., Diewok, J., and Tauler, R. *Chemom. Intell. Lab. Syst.* **70**, 11–21 (2004).

[57] Shashilov, V. A., Xu, M., Ermolenkov, V. V., and Lednev, I. K. *J. Quant. Spectrosc. Radiat. Transfer* **102**, 46–61 (2006).

[58] Windig, W. *Chemom. Intell. Lab. Syst.* **36**, 3–16 (1997).

[59] Wich, P. R. and Schmuck, C. *Angew. Chem. Int. Ed.* **49**, 4113–4116 (2010).

[60] Bukowska, J. and Piotrowski, P. *Optical Spectroscopy and Computational Methods in Biology and Medicine*. (2014).

[61] Huang, C.-Y., Balakrishnan, G., and Spiro, T. G. *J. Raman Spectrosc.* **37**, 277–282 (2006).

[62] Wen, Z. Q. *J. Pharm. Sci.* **96**, 2861–2878 (2007).

[63] Oladepo, S. A., Xiong, K., Hong, Z., Asher, S. A., Handen, J., and Lednev, I. K. *Chem. Rev.* **112**, 2604–2628 (2012).

[64] Cristina Stanca-Kaposta, E., Gamblin, D. P., Screen, J., Liu, B., Snoek, L. C., Davis, B. G., and Simons, J. P. *Phys. Chem. Chem. Phys.* **9**, 4444 (2007).

[65] Jiang, Q.-q., Bartsch, L., Sicking, W., Wich, P. R., Heider, D., Hoffmann, D., and Schmuck, C. *Org. Biomol. Chem.* **11**, 1631 (2013).

[66] Ehlers, M., Grad, J. N., Mittal, S., Bier, D., Mertel, M., Ohl, L., Bartel, M., Briels, J., Heimann, M., Ottmann, C., Sanchez-Garcia, E., Hoffmann, D., and Schmuck, C. *ChemBioChem* , 591–595 (2017).

[67] Grad, J. N., Gigante, A., Wilms, C., Dybowski, J. N., Ohl, L., Ottmann, C., Schmuck, C., and Hoffmann, D. *J. Chem. Inf. Model.* **58**, 315–327 (2018).

[68] Holde, K. E. V., Johnson, W. C., and Ho, P. S. *Principles of Physical Biochemistry.* (2006).

[69] Rava, R. P. and Spiro, T. G. *J. Phys. Chem.* **89**, 1856–1861 (1985).

[70] Mosier-Boss, P. A., Lieberman, S. H., and Newbery, R. *Appl. Spectrosc.* **49**, 630–638 (1995).

[71] De Luca, A. C., Mazilu, M., Riches, A., Herrington, C. S., and Dholakia, K. *Anal. Chem.* **82**, 738–745 (2010).

[72] Schmuck, C. and Wich, P. R. *Topics Curr. Chem.* **277**, 3–30 (2007).

[73] Jiang, Q.-Q., Sicking, W., Ehlers, M., and Schmuck, C. *Chem. Sci.* **6**, 1792–1800 (2015).

[74] Schmuck, C. and Bickert, V. *J. Org. Chem.* **72**, 6832–6839 (2007).

[75] Sinha, S., Lopes, D. H., Du, Z., Pang, E. S., Shanmugam, A., Lomakin, A., Talbiersky, P., Tennstaedt, A., McDaniel, K., Bakshi, R., Kuo, P. Y., Ehrmann, M., Benedek, G. B., Loo, J. A., Klärner, F. G., Schrader, T., Wang, C., and Bitan, G. *J. Am. Chem. Soc.* **133**, 16958–16969 (2011).

[76] Herzog, G., Shmueli, M. D., Levy, L., Engel, L., Gazit, E., Klärner, F. G., Schrader, T., Bitan, G., and Segal, D. *Biochemistry* **54**, 3729–3738 (2015).

[77] Zheng, X., Liu, D., Klärner, F. G., Schrader, T., Bitan, G., and Bowers, M. T. *J. Phys. Chem. B* **119**, 4831–4841 (2015).

[78] Vöpel, T., Bravo-Rodriguez, K., Mittal, S., Vachharajani, S., Gnutt, D., Sharma, A., Steinhof, A., Fatoba, O., Ellrichmann, G., Nshanian, M., Heid, C., Loo, J. A., Klärner, F. G., Schrader, T., Bitan, G., Wanker, E. E., Ebbinghaus, S., and Sanchez-Garcia, E. *J. Amer. Chem. Soc.* **139**, 5640–5643 (2017).

[79] Qin, L., Reger, A. S., Guo, E., Yang, M. P., Zwart, P., Casteel, D. E., and Kim, C. *Biochemistry* **54**, 4419–4422 (2015).

[80] Harbury, P., Zhang, T., Kim, P., and Alber, T. *Science* **262**, 1401–1407 (1993).

[81] Reger, A. S., Yang, M. P., Koide-Yoshida, S., Guo, E., Mehta, S., Yuasa, K., Liu, A., Casteel, D. E., and Kim, C. *J. Biol. Chem.* **289**, 25393–25403 (2014).

[82] Surks, H. K., Mochizuki, N., Kasai, Y., Georgescu, S. P., Tang, K. M., Ito, M., Lincoln, T. M., and Mendelsohn, M. E. *Science* **286**, 1583–1587 (1999).

[83] Neumann, W. *Fundamentals of dispersive optical spectroscopy systems.* SPIE-The international Society for Optical Engineering, (2014).

[84] Liu, C. and Berg, R. W. *Appl. Spectrosc. Rev* **48**(5), 425–437 (2013).

[85] Strekas, T. C., Adams, D. H., Packer, A., and Spiro, T. G. *Appl. Spectrosc.* **28**, 324–327 (1974).

[86] Srivastava, S. K., Niebling, S., Küstner, B., R. Wich, P., Schmuck, C., and Schlücker, S. *Phys. Chem. Chem. Phys.* **10**, 6770 (2008).

[87] Hirose, K. *Journal of Inclusion Phenomena and Macrocyclic Chemistry* **39**, 193–209 (2001).

[88] Schalley, C. A. *Analytical Methods in Supramolecular Chemistry* , 1–16 (2012).

[89] Thordarson, P. *Chem. Soc. Rev.* **40**, 1305–1323 (2011).

[90] Martinez, A. S., González, R. S., and Terçariol, C. A. S. *Physica A* **387**, 5679–5687 (2008).

[91] Tsoularis, A. and Wallace, J. *Math Biosci* **179**(1), 21–55 (2002).

[92] Trott, O. and Olson, A. *J. Comput. Chem.* **31**, 455–461 (2010).

[93] Morris, G. M., Ruth, H., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. *J. Comput. Chem.* **30**, 2785–2791 (2009).

[94] Trusch, F., Kowski, K., Bravo-Rodriguez, K., Beuck, C., Sowislok, A., Wettig, B., Matena, A., Sanchez-Garcia, E., Meyer, H., Schrader, T., and Bayer, P. *Chem. Commun.* **52**, 14141–14144 (2016).

[95] Xu, N., Bitan, G., Schrader, T., Klärner, F. G., Osinska, H., and Robbins, J. *Am. Heart J.* **6**, 1–13 (2017).

[96] Bier, D., Mittal, S., Bravo-Rodriguez, K., Sowislok, A., Guillory, X., Briels, J., Heid, C., Bartel, M., Wettig, B., Brunsveld, L., Sanchez-Garcia, E., Schrader, T., and Ottmann, C. *J. Amer. Chem. Soc.* **139**, 16256–16263 (2017).

[97] Jung, H. H., Yang, S. T., Sim, J. Y., Lee, S., Lee, J. Y., Kim, H. H., Shin, S. Y., and Kim, J. I. *BMB Rep* **43**, 362–368 (2010).

List of Figures

Publication

B. Zakeri, S. Niebling, A. G. Martinez, P. Sokkar, E. Sanchez-Garcia, C. Schmuck and S. Schlücker, *Molecular recognition of carboxylates in the protein leucine zipper by a multivalent supramolecular ligand: residue-specific, sensitive and label-free probing by UV resonance Raman spectroscopy*, Phys. Chem. Chem. Phys.**20**, 1817–1820 (2018).

Presentation

B. Zakeri, A. G. Martinez, C. Schmuck and S. Schlücker, *An important application of UVRR spectroscopy in supramolecular chemistry*, International OSA Network of Students (IONS) and the Conference on Optics, Atoms and Laser Applications (KOALA), Australia, Brisbane, October 2017.

Poster presentation

1. International symposium, supramulecular chemistry on proteins, Essen, September 2015.
2. 16th European Conference on the Spectroscopy of Biological Molecules, Bochum, September 2015.
3. Joint Graduate Student Symposium, Protein-Ligand Interactions, Hannover, September 2016.
4. International symposium, supramulecular chemistry on proteins, Essen, September 2017.

Acknowledgment

My sincere thanks, first of all, is dedicated to the German Research Foundation (Collaborative Research Center, CRC 1093, "Supramolecular Chemistry on Proteins") not only for the financial support but also for allowing me to be a part of this project. During this collaboration, I was honored to work among experienced people in chemistry and biology in a well-designed program at CRC1093-graduate school and I am really thankful to all of them.

To this thesis, there are several people who have contributions either by fruitful collaborations or their help and support. Above all, our main collaboration with the group of organic chemistry (Prof. Dr. Carsten Schmuck, project A1) is deeply appreciated. I give my deepest thanks to Prof. Schmuck not only for this collaboration but also for his constructive suggestions as my second supervisor during our report meetings. For the main results of this collaboration, which is included in the major part of this thesis acquired by using their supramolecular ligands, I would like to thank Dr. Alba Gigante Martinez for synthesizing and purification of the samples. From their group, I also give my thanks to Dr. Poulami Jana for her kind help whenever I had a question. Our another collaborative organic chemistry group was the group of Prof. Dr. Thomas Schrader (project A3). I give my thanks for his constructive suggestions during running the project and also his student Andrea Sawislok for preparing the sample. In addition, I appreciate our short but fruitful collaboration with the group of computational chemistry (Prof. Dr. Elsa Sanchez Garcia, project A8) and especially Dr. Pandian Sokkar who performed the molecular docking. Last but not least, I give my sincere thanks to Prof. Dr. Sebastian. Schlücker for supervising this project (A9) and also all people in our group especially Dr. Axel Hoffmann and Bernd Walkenfort for their technical support in the laser lab. Moreover, I would like to thank Dr. Stephan Niebling, the previous PhD student in the group of Prof. Schlücker, for providing the valuable information and sharing his experience with me from his previous research work.
Finally, I would like to thank my family for their endless and unconditional love and support.

Statement

I hereby declare on oath that I wrote this dissertation independently

"label-free and site-specific detection of protein recognition by supramolecular ligands in solutions using Ultraviolet resonance Raman spectroscopy"

I also did not use other sources and aids except the ones which are specified in this dissertation. I only submit this dissertation in this doctoral procedure at the university of Duisburg-Essen and I declare that this work, neither in the same nor in another form, has not been already submitted to another examination procedure.

Essen, 26. June, 2018