radiation

MDPI

Review

Terahertz Spectroscopic Analysis in Protein Dynamics: Current Status

Tiziana Mancini [1], Rosanna Mosetti [1], Augusto Marcelli [2,3], Massimo Petrarca [4,5], Stefano Lupi [1,2] and Annalisa D'Arco [1,2,4,*]

[1] Physics Department, University of Rome 'Sapienza', P.le Aldo Moro 5, 00185 Rome, Italy; tiziana.mancini@uniroma1.it (T.M.); mosetti.1766934@studenti.uniroma1.it (R.M.); stefano.lupi@roma1.infn.it (S.L.)
[2] National Institute for Nuclear Physics—Laboratori Nazionali di Frascati (INFN-LNF), Via E. Fermi 54, 00044 Frascati, Italy; augusto.marcelli@lnf.infn.it
[3] Rome International Centre for Materials Science Superstripes, Via dei Sabelli 119A, 00185 Rome, Italy
[4] Department of Basic and Applied Sciences for Engineering (SBAI), University of Rome 'Sapienza', Via Scarpa 16, 00161 Rome, Italy; massimo.petrarca@uniroma1.it
[5] National Institute for Nuclear Physics–Section of Rome 'Sapienza', P.le Aldo Moro 2, 00185 Rome, Italy
* Correspondence: annalisa.darco@roma1.infn.it

Simple Summary: Proteins are large biomolecules carrying out several different indispensable activities in all living organisms. As their chemical composition and structure lead to specific functionalities, possible variations alter conformational transition and cause dysfunctions and pathological diseases. An emerging technique for their detection and characterization is THz spectroscopy. In this review, we illustrate its potentialities and drawbacks in protein study. We report the most remarkable experimental works of the last decades, constituting decisive steps for the scientific and technological progress in THz spectroscopy for biomolecules. The importance of information provided by THz spectroscopy is outlined, conforming it as powerful and useful technique for topical open questions, too.

Abstract: Proteins play a key role in living organisms. The study of proteins and their dynamics provides information about their functionality, catalysis and potential alterations towards pathological diseases. Several techniques are used for studying protein dynamics, e.g., magnetic resonance, fluorescence imaging techniques, mid-infrared spectroscopy and biochemical assays. Spectroscopic analysis, based on the use of terahertz (THz) radiation with frequencies between 0.1 and 15 THz (3–500 cm^{-1}), was underestimated by the biochemical community. In recent years, however, the potential of THz spectroscopy in the analysis of both simple structures, such as polypeptide molecules, and complex structures, such as protein complexes, has been demonstrated. The THz absorption spectrum provides some information on proteins: for small molecules the THz spectrum is dominated by individual modes related to the presence of hydrogen bonds. For peptides, the spectral information concerns their secondary structure, while for complex proteins such as globular proteins and viral glycoproteins, spectra also provide information on collective modes. In this short review, we discuss the results obtained by THz spectroscopy in the protein dynamics investigations. In particular, we will illustrate advantages and applications of THz spectroscopy, pointing out the complementary information it may provide.

Keywords: terahertz; terahertz spectroscopy; amino acids; peptide; proteins; conformation; protein dynamics

Citation: Mancini, T.; Mosetti, R.; Marcelli, A.; Petrarca, M.; Lupi, S.; D'Arco, A. Terahertz Spectroscopic Analysis in Protein Dynamics: Current Status. *Radiation* **2022**, *2*, 100–123. https://doi.org/10.3390/radiation2010008

Academic Editor: Krishan Kumar

Received: 31 December 2021
Accepted: 1 February 2022
Published: 7 February 2022

1. Introduction

Amino acids/proteins participate in the formation of cells and tissues, and they are responsible for life activities in living organisms, such as enzymes, antibodies, signaling and transporting molecules, cell membrane, metabolic or catalytic functions, etc. [1–4].

Notably, they have a crucial role in various fields and applications, such as health, medicine and food safety. For example, rapid assessment of the contents and types of proteins in food can improve its quality, and the presence of specific proteins and their dynamics in blood samples can identify a pathological process or a disease at an early state. An important role of proteins is also played in viral pathogenesis, where membrane proteins are generally responsible for infectious processes, anchoring the virus to human receptors, as happens for the SARS-CoV-2 coronavirus [5,6].

Proteins are polypeptide chains formed by the progressive condensation reactions that create the peptide bond between the α-amino group and the α-carboxyl group of two amino acids [7,8]. The amino acid sequence assumes structural hierarchy, allowing the macromolecules to bend and generating typical three-dimensional protein structures: primary, secondary, tertiary and quaternary [7,8]. These structures are correlated to specific protein functionalities. Then, amino acids mutations and alterations of the primary sequence can produce anomalies and transitions to different conformational states linked to functionally relevant phenomena. Small differences in the primary structure imply completely different biological functions; for example, oxytocin and vasopressin are two oligopeptides with similar structure but different biological functionalities. Still, in patients with sickle cell anemia, a residue of glutamic acid in the chain is replaced by a valine residue due to a genetic error.

Therefore, the evaluation of protein kinds and quantities is an important index of protein function analysis, clinical diagnosis, recovery check, quality testing of biological products, etc. [9]. Several methods are used for studying proteins and their dynamics: techniques such as ultraviolet (UV) absorption, Lowry and Bradford assays, bicinchoninic acid and Kjeldahl methods, to mention a few [10–17], are used for the protein quantification; other significant semi-quantitative analyses to evaluate the protein kinds include chromatography methods, electrophoresis and immunoassays [18–25]. Among these, in particular, the protein immunoassay is conventionally used for studying gene expression at the protein level, detection of the activity of antibodies and early diagnosis of diseases. The high specificity and sensitivity are guaranteed by using the specific expression information of the homologous antibody of the protein that determines its presence and its amount. The antibody method involves several sophisticated, time-consuming, tedious and expensive processes. These criticalities emerged especially during the pandemic crisis of 2020, where the demand for immune analysis for early COVID-19 diagnosis has grown exponentially [26,27]. Thus, reliable and label-free protein analysis methods could support the conventional approaches improving detection efficiency in addition to reduction in the screening time and cost. For a three-dimensional visualization of protein conformational structures in cell systems, spectroscopy and microscopy approaches can be employed such as fluorescence spectroscopy and microscopy, atomic spectroscopy, confocal [28] and multi-photon microscopy [29], optical nonlinear imaging [30–33], super resolution methods, electron microscopy, etc. [34,35].

Spectroscopy and imaging based on THz radiation, ranging from tens of gigahertz (GHz) to several THz, are new methods which only in recent years emerged with their potential for protein detection and characterization (Figure 1). Biological vibrations and rotations, involving intra/inter molecular weak hydrogen bonds, van der Waals forces, conformational changes, non-bonded hydrophobic and hydrophilic interactions [4,36], are associated with vibrational and rotational energy that mostly lie in the THz domain. The optical parameters such as absorption coefficient and refractive index, characterizing the resonances, reflect the conformational changes and other large-scale deformations involving charge movement and relocation inside the biomolecule. Relative to proteins, the capability of THz to qualify and quantify them is recognized [37,38].

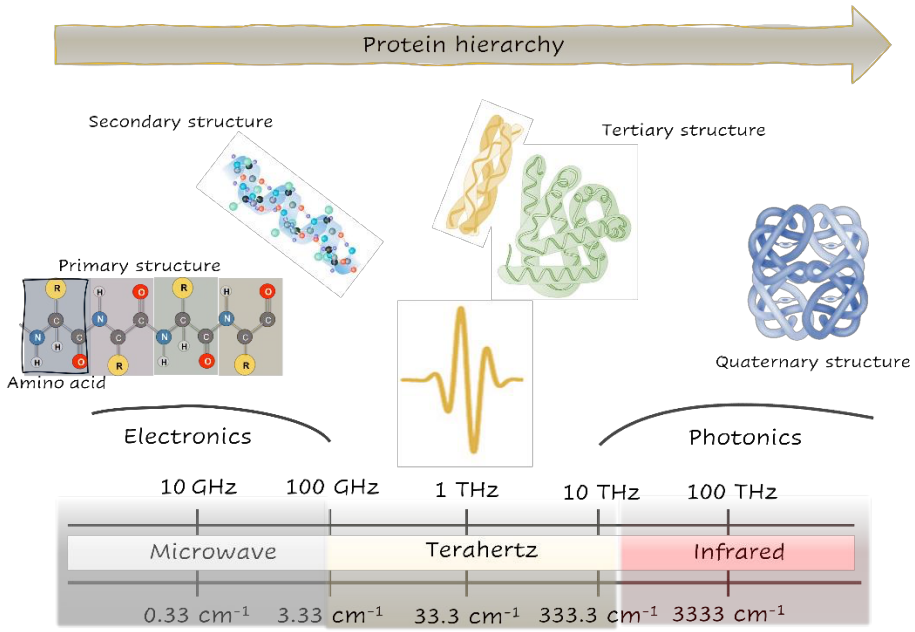


Figure 1. THz region and applications in the protein field of THz radiation. The THz domain typically lies between microwave and infrared regions, which refers to the frequency band spanning 0.1–15 THz (from 3 cm^{-1} to 500 cm^{-1}). The generation of THz radiation was successfully developed with both electronics and photonics, including the up-conversion of electronic radiofrequency sources and the down-conversion of optical sources. THz technology detects proteins, illustrated by their structure hierarchy.

Here, we review results obtained by THz spectroscopy and imaging in the study of amino acids and protein dynamics. In the first sections of this review, we discuss the experimental techniques commonly used for protein detection, their limitations and some critical issues. Then, we illustrate the advantages and applications of THz radiation providing complementary information to existing techniques. Finally, we discuss perspectives and other potential THz contributions in the field of biomolecules.

2. Consolidate Technologies for Amino Acid/Protein Detection

There are two well-established methodological approaches to protein characterization: total protein nitrogen measurements and chemical and structural identification of a specific protein, schematically summarized in Figure 2. The determination of total protein nitrogen is a proximity tool for the early and rapid quantification of the protein content. The Kjeldahl [10] and Dumas methods are widely used for nutrition labeling and quality control [15–17]. Other techniques are used and preferred for their accuracy, low cost and great applicability in many fields where the identification of proteins is required, such as the bicinchoninic acid, biuret and Lowry methods as well as UV absorption. However, many of these are time-consuming and can only measure the total organic nitrogen present in the sample.

The most common immunoassays are enzyme-linked immunosorbent assays, lateral flow immunoassay and Western blotting, but other analytical technologies [16,18–25] have been adopted in biomedical and clinical studies. Some analytical methods for monitoring and detection are based on the ability to propagate/amplify biomolecules in cell cultures or on the detection of a specific antibody. However, these laboratory diagnostic techniques require time to detect the specific immune response and its quantification. Especially, when used for virus recognition or metabolic evaluation the task is challenging, and some immunoassays may give false positive results. Other techniques, such as microscopy, fluorescence and spectroscopy, are preferred in describing the protein dynamics. Here we briefly recall the most used techniques: X-ray crystallography, nuclear magnetic resolution

(NMR), Cryo-electron microscopy (Cryo-EM), fluorescence, circular dichroism (CD) and vibrational spectroscopies, including infrared (IR) and Raman ones.

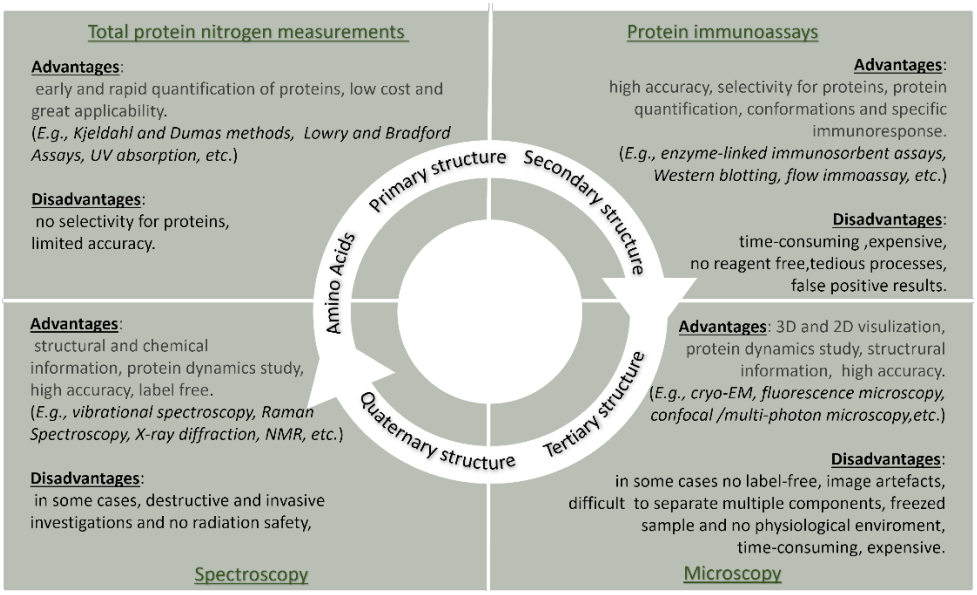


Figure 2. Schematic overview about the advantages and disadvantages of main techniques for protein characterization.

X-ray crystallography, NMR and Cryo-EM mainly concern investigations of the three-dimensional structure of molecules. X-ray crystallography exploits the diffraction lattice for collecting an electron density map of the crystal. The need for a solid crystalline sample and the ionizing properties of X-rays represent limitations in the study of proteins in the physiological environment [39–42].

Most studies in organic chemistry involve the probing of hydrogen atoms with NMR [43–45]. The process of relaxation of nuclear magnetic spins from higher to lower energy levels after a certain time-span is the principle of NMR to achieve structural atomic information.

Cryo-EM provides a 3D structure of cells and macromolecules visualized with 5–8 nm spatial resolution. Although Cryo-EM allows high spatial resolutions, suitable for studying the cytoskeleton and the complex surface structures of viruses, the need of a frozen sample, although useful since it fixes the protein in a well-defined configuration, is not ideal to the investigate protein dynamics.

Fluorescence spectroscopy is an alternative method that works most accurately at very low concentrations of fluorophores, giving structural information about a single molecule or protein complex interactions. It exploits mainly the intrinsic and extrinsic protein fluorescence. It may contribute to reconstruct structural effects induced by pH and solvent composition, to measure lifetimes and temporal kinetics. Few proteins possess intrinsic fluorophores, such as tryptophan and tyrosine, but most of them are non-fluorescent. In many of these cases, an external fluorophore can be introduced by chemical coupling or non-covalent binding. This labelling technique allows the development of new fluorescent dyes with different excitation and emission properties. This has promoted the emergence of new techniques for the study of protein dynamics: quenching, fluorescence resonance energy transfer (FRET) [46,47], bioluminescence resonance energy transfer (BRET) [48], fluorescence recovery after photobleaching (FRAP) [49], fluorescence polarization and cross-correlation, and on all microscopy.

However, all spectroscopic/microscopic techniques have clear limitations. Indeed, requiring chemical labels, the biological functionalities can interfere with the protein dynamics; additionally, photo-bleaching can introduce artefacts and limit the measurement repeatability in the microscopic approach. Therefore, it may be desirable to implement

real-time, three-dimensional imaging with high spatial resolution, high sensitivity and high chemical selectivity for unlabeled living cells.

Circular dichroism (CD) involves the measurements of small differences in the UV–vis absorption values due to the chirality of biomolecules. The left- and right-handed polarized components of the incident light are absorbed differently by the optically active molecules, which yields a difference in the absorption parameters [50–52]. CD spectroscopy is largely adopted to verify the natural secondary structure of proteins and their changes induced by chemical–physical conditions.

Vibrational spectroscopy, including Fourier Transform Infrared (FTIR) and Raman spectroscopy, provides selective, non-destructive identification of the molecules, through the interrogation of molecular roto-vibrational and vibrational modes [53–55]. Both FTIR and Raman spectroscopy give similar information about a molecule: FTIR is associated with the change of dipole momentum and Raman with a change of the polarizability, respectively, as a consequence of the matter–light interaction. Their use is typically in chemical and biochemical research of small-molecule compounds and in the identification of synthesized compounds due to their intrinsic chemical selectivity. The nine characteristic vibrational bands (named amide A, B, I, II, III, . . . , VII, where amide I and II, between 1600 and 1700 cm^{-1} as well as 1500 and 1600 cm^{-1} respectively, are the major contributors in the IR spectrum) exhibited by the peptide bonds in the infrared region ensure the recognition of the protein backbone conformation and the assessment of the secondary structure of peptides and proteins [53]. Novel instruments combine the FTIR technique and the features of an evanescent field, probing the sample with the Attenuated Total Reflection (ATR) technique. ATR is frequently used for small amounts of solid and liquid samples, avoiding difficult preparation. Instead, time-resolved FTIR enables the observation of protein reaction at the sub-millisecond timescale, such as the study of the light-driven proton pump bacteriorhodopsin [53]. Nevertheless, linear Raman spectroscopy and microscopy are limited to weak signals and very long times for signal collection. Thus, only recently with the accessibility of ultrafast lasers, coherent Raman Scattering (CRS) techniques are going to assert themselves thanks to the sensitivity at the same molecular vibrations probed in spontaneous Raman spectroscopy and to the nonlinear dependence on the incoming light fields [30–33].

3. THz Technology for Protein Spectroscopy

In the last two decades, many efforts have been made to improve THz technology, e.g., THz sources and detectors [56–69], discovering new materials with strong THz response [70–73] and new customizable, flexible and compact devices for spectroscopy [74].

These systems have promoted the diffusion of THz radiation for spectroscopy applications. Thus, it was applied with success in various scientific fields e.g., gas sensing [75–80], chemical and pharmaceutical analysis [81,82], condensed matter [83–85], identification of crystalline polymorphs, microelectronics and security [86–89], agri-food industry [90] and cultural heritage [91], including the emerging support in the field of biomedicine and bio-imaging [92–99], etc. Thus, THz spectroscopy has become a valuable tool for rapid and non-invasive detection thanks to many advantages that make it particularly appealing for probing the intermolecular structure and dynamics of biomolecules [95,100,101]. In fact, THz radiation lies between microwave and infrared, being characterized by low photon energy (4 meV @ 1 THz), and it provides energy levels related to rotational and vibrational molecular modes and intermolecular vibrations, such as hydrogen bonds [102,103]. For example, external lattice vibrations typically dominate the low-frequency region, and these modes are influenced by the crystalline arrangement of molecules. This occurs for different polymorphs where the unique THz spectral features can be used as identifying fingerprints [104]. THz waves are sensitive to the conformation and structure of proteins [105,106], and they can be used for providing useful information in combination with quantum-mechanical theory [101,107]. Moreover, the low photon energy is too low to heat materials or to induce atoms/molecules ionization. For this reason, it is a non-ionizing

radiation, suitable and attractive for noninvasive biological applications and biomedical imaging [97]. At variance, other biochemical/molecular techniques, such as ultraviolet or X-rays using higher-energy photons (from few eV to ~keV), may damage any biological sample [81,101,108,109].

Furthermore, THz radiation strongly interacts with polar molecules [74,110,111], such as water. Especially in THz imaging, the intense absorption, due to water molecules [95], represents a limiting condition for the THz penetrability inside fresh tissues (from tens to hundreds of microns) [94], reducing the diagnostic capability in vivo only to superficial layers. On the contrary, nonpolar materials (such as paper, cloths and plastic) are usually transparent in the THz range [89,112,113].

Many layouts and materials can be used for THz signal collection in THz spectroscopy [114–116], exhibiting high performance in terms of the signal-to-noise ratio with coherent detection mode [116]. Because THz spectroscopy is insensitive to the thermal background, it shows a high signal-to-noise Ratio (SNR), and it does not require the use of cooled detectors [116,117]. Concerning the coherent detection mode, the THz electric field is directly assessed. Therefore, both amplitude and phase of the THz pulse electric field can be simultaneously measured, and the optical parameters, including sample absorption coefficient and refractive index, can be extracted without using Kramers–Kronig relations [101,116].

4. Current Status of THz Spectroscopy for Research on Amino Acids and Short-Chain Peptides

Amino acids are the backbones of proteins as well as fundamental elements to assemble cells and tissues. They are involved in all cellular metabolic and repairing processes and are responsible for various pathological genesis. The structural differences between amino acids depend on the side chain R and its polarity. Most THz biomolecule investigations have been focused on amino acids in the low-frequency spectral interval [118–131], see Table 1. All 20 standard α-amino acids under the same conditions revealed a correlation between molecular structures and spectral peaks between 0.2 and 3.0 THz (6.67–100 cm^{-1}) [122]. The amino acids were studied in the solid phase, mixing them with a polyethylene powder in various proportions. For example, high proportions of amino acids to polyethylene (1:1 and 2:1) [123,124], as well as low proportions (1:15 and 1:9) [125,126], are used. From these, it emerged that the proportion of amino acids to THz transparent material in the production of the pellet influences the spectrum, especially in the low-frequency region [123,127]. Taday et al. [127] characterized the temperature-dependent L-glutamic acid spectrum and compared it with the theoretical model proposed by Chiba et al. [132], but poor agreement was achieved. Actually, theoretical simulations are not able to model the complexity of hydrogen bonding and crystalline structure occurring at low frequency in biological molecular systems.

Table 1. Main THz absorption frequencies observed in the literature [118–134].

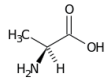
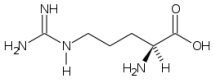
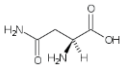
Amino Acids	THz Absorption Frequencies (THz)
 Alanine (Ala)	2.21, 2.56, 2.72, 2.91, 3.14, 3.37, 4.17
 Arginine (Arg)	0.99, 1.45, 2.02, 2.62, 3.51, 3.77, 4.40
 Asparagine (Asn)	1.64, 1.85, 2.06, 2.26, 2.62, 3.11, 3.57, 3.92, 4.90

Table 1. *Cont.*

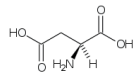
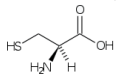
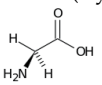
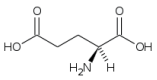
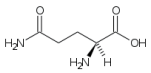
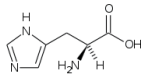
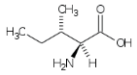
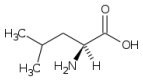
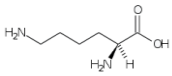
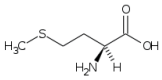
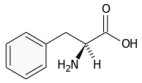
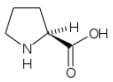
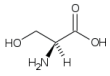
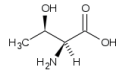
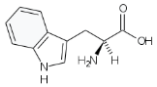
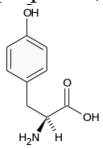
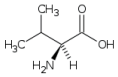
Amino Acids	THz Absorption Frequencies (THz)
 Aspartic acid (Asp)	1.35, 1.71, 2.58, 3.01, 3.26, 3.98, 4.41, 5.36
 Cysteine (Cys)	1.40, 1.70, 2.33, 2.78, 2.94, 3.23, 3.64, 4.44, 4.78, 5.87
 Glycine (Gly)	1.83, 2.30, 2.51, 2.70, 4.07
 Glutamic acid (Glu)	1.21, 2.03, 2.23, 2.48, 2.64, 2.80, 3.26, 3.58, 4.0, 4.50
 Glutamine (Gln)	1.70, 2.28, 2.50, 3.37, 4.08, 4.92
 Histidine (His)	0.88, 1.72, 2.08, 2.44, 2.80, 3.00, 3.39, 3.96, 4.33, 5.30
 Isoleucine (Iso)	0.30, 0.85, 1.08, 1.42, 1.72, 2.41, 2.74, 3.54, 4.27, 5.26
 Leucine (Leu)	0.66, 0.84, 1.46, 1.64, 2.14, 2.56, 2.74, 2.88, 2.96, 3.68, 5.11
 Lysine (Lys)	0.90, 1.26, 1.79, 2.25, 2.64
 Methionine (Met)	1.01, 1.79, 2.70, 2.94, 3.77
 Phenylalanine (Phe)	1.25, 2.02, 2.52, 2.76, 4.16
 Proline (Pro)	1.69, 2.00, 2.64, 3.12, 3.62, 4.05, 4.69
 Serine (Ser)	1.97, 2.41, 2.71, 3.12, 3.98, 4.34

Table 1. *Cont.*

Amino Acids	THz Absorption Frequencies (THz)
 Threonine (Thr)	1.11, 1.42, 2.12, 2.61, 3.06, 3.33, 3.75, 4.44, 4.98, 5.30
 Tryptophan (Trp)	0.91, 1.19, 1.44, 1.85, 2.26, 2.57, 3.22, 3.69, 4.02, 4.85
 Tyrosine (Tyr)	0.95, 1.92, 2.06, 2.65, 2.82, 3.31, 3.48, 3.96, 4.32, 4.75, 5.13, 6.22
 Valine (Val)	1.11, 1.68, 2.12, 2.22, 2.52, 2.64, 2.84, 3.53, 4.39

Yi et al. [133] employed THz-TDS and IR spectroscopy to study all 20 amino acids, accurately collecting their spectra over a wide range from 0.2 to 6 THz (6.67 to 200 cm^{-1}). Mixtures in different proportions of amino acids with polyethylene powder, transparent to THz radiation, were made. Their absorbance spectra were characterized to evaluate the optimal proportion for each amino acid. Among the fifteen aliphatic amino acids, L-glycine (L-gly), L-alanine (L-ala), L-valine (L-val), L-isoleucine (L-iso) and L-leucine (L-leu) have a similar structure, the only structural difference being the side chain. Nevertheless, they exhibit differences in the THz absorption spectra: L-val has prominent absorption at 1.11 THz and 4.39 THz; for L-iso, the main absorption peaks were shifted to lower frequencies at 0.85 and 4.27 THz; and finally for L-leu, at 0.66 and 3.68 THz. The authors correlated the differences due to the longer side chains that shift the absorption peaks to lower frequencies [133]. Moreover, the aromatic L-Tyrosine (L-tyr), L-phenylalanine (L-phe) and L-tryptophan (L-try) present a phenyl group on the side chain, and although they have similar structures, the THz spectra are different. A broad absorption band at 3.22 THz is observed in the L-try spectrum, and three absorption bands at 2.02, 2.76 and 4.16 THz appear in the L-phe absorption spectrum. L-tyr includes two symmetric peaks around them, with splitting attributable to the –OH group of L-tyr. For L-ala and L-threonine (L-thr) the main absorption peaks are located at 2.91 THz and 3.33 THz, respectively. The accurate study by Yi et al. is in a good agreement with previous works [118–125,127–131] ranging between 0.1 and 3 THz. Thus, they established an absorption spectral database of protein constituents in the solid phase and demonstrated that THz spectroscopy is able to distinguish different spectral features of amino acids with a very similar structure.

Another interesting aspect regards enantiomers, chiral molecules which are non-superimposable mirror images of each other, isotopologues and polymorphism [134]. In this context, THz spectroscopy is sensitive to crystalline structures. In pharmaceutical and pharmacological science, the use of amino acids and small peptide chains is required; the knowledge of their crystallinity is a topic issue from studies of changes in medicinal aging and to the detection of illegal substances. In particular, drugs bind to the human body thanks to receptors which are protein complexes [135]. In addition, the THz spectra of DL-leu and the two polymorphs of DL-val have been measured, and computer calculations have been implemented to reliably distinguish the nuances of THz spectra similar to solid-state systems [136,137]. Yamaguchi et al. [92] demonstrated THz sensitivity in enantiomers recognition, showing how the THz absorption spectra of L-(D-) and DL-ala are quite different; L- and D-ala show two absorption bands located at 74.4 and 85.7 cm^{-1} [92],

and racemic compound DL-ala exhibits one absorption band at 41.8 cm^{-1}. Therefore, THz spectroscopy is considered an off-line tool for assessing crystallinity, also in co-lyophilized amino acid/gelatin mixtures (L-ala, serine and valine) [128]. The physical/chemical conditions are critical for amino acids, and in particular, amino acid medicine is strongly susceptible to the storage and transportation states. Herein, the induced changes affect several properties of amino acids, e.g., stability, solubility, mechanical and physic-chemical properties, and alter the protein dynamics. Several experimental as well as theoretical studies carried out addressing those issues [138–143]. Aqueous solutions of amino acids have been extensively studied through different experimental techniques [144–149] and molecular dynamics simulations [150]. On the other side, there have been a few dielectric relaxation (DR) studies, covering MHz-GHz frequency domain, to probe the behavior of hydrated amino acids [151–153].

One of the amino acids extremely susceptible to change is Lysine (Lys), which has a key role in human metabolism and various physiological functions, such as promotion of human growth and enhancement of immunity. With the addition of water molecules, the surrounding environment of Lys is changed, affecting its collective molecular vibrations. Exploring the molecular collective modes at low frequency, Bian et al. [154] found that Lys and Lys·H_2O have distinct spectral features in the range between 0.3 and 2.5 THz. The interpretation of the results was entrusted by theoretical calculations based on Density Function Theory (DFT) [155,156]; the water molecule is recognized to locate the carboxyl group and the amino group of hydrated Lys molecule, while its vibrational modes are produced by a dihedral torsion or bond angle bend of molecular chains. An equivalent description was also achieved for differences shown in THz spectral features of L-phe and for L-phe·H_2O [157], associating them to different crystalline structures and to the presence of hydrogen bond interaction in the case of L-phe·H_2O.

The hydration process of amino acids is the basis for a further understanding of the behavior of proteins in the aqueous environment. Samanta et al. [158] studied five amino acids dissolved in phosphate buffers. The optical absorption parameters dependent on the frequency, coefficient and refractive index were distinguished from the contribution of the solute, solvent and water of hydration. The departure of the absorption coefficient from the ideal behavior, $\Delta\alpha$, is associated with the section of the hydrated water molecules around the amino acid molecules, which differ from the absorption of water in mass up to 5–7 Å from a single solute molecule. Interaction with water molecules is mainly due to -COO- and $-NH^{3+}$. They found a negative value of $\Delta\alpha$ for all amino acids, with the exception of Gly, meaning this is a structure breaker. Arg, Ser, Trp, Lys and Asp are instead structure makers, leading to the formation of an ordered water molecule network around single amino acids. Furthermore, the anomalous behavior of Gly was described and explained. Fitting real and imaginary parts of the solution dielectric constant with a multiple Debye model, three time scales were obtained (8.7 ps, 80 and 200 fs). They were associated with cooperative rearrangement of the hydrogen bonded network for 8.7 ps and to competitive quick jump of under-coordinated water and small angular rotation preceding a large angle jump for 200 fs. Finally, the contribution at 80 fs was due to hydrogen-bond bending and the related transverse acoustic phonons, which propagate in a direction normal to the hydrogen bonds in between two neighboring water molecules. The first two are found to depend on both hydrophobic and hydrophilic residues of the considered amino acid, and their trends were analyzed for all molecules.

Low-frequency vibrational modes are strongly influenced by size and long-range order of the molecule. As consequence, small biological molecules, such as amino acids, tend to have distinct, relatively isolated features compared to polymers, such as short chain of peptides, with a larger number of modes. Thus, the THz spectral characteristics of the peptide reflect their amino acid composition, permutation sequence, intermolecular hydrogen bond and crystal structure. For example, the simplest molecule with a peptide bond, crystalline N-methyl acetamide, shows bands at 3.6 THz (120 cm^{-1}) and 6.0 THz (200 cm^{-1}) [159]. Yamamoto et al. used THz-TDS to study the absorption coefficient and

refractive index of simple amino acids structures, glycine and L-alanine, compared with their corresponding polypeptides in the 0.21–1.65 THz (7–55 cm^{-1}) domain [160]. Poly-glycine was observed to have peaks at 1.365 THz (45.5 cm^{-1}) induced by the interaction of the chain different from the poly-L-alanine. The differences indicate that poly-glycine has a longer-range, ordered structure of solid-state crystals than poly-L-alanine does.

Kutteruf et al. [107] also demonstrated the great potential of THz spectroscopy for solid-phase short-chain peptide sequences, in the spectral region 1–15 THz. Pure solid samples of low molecular weight protein fragments pressed into polyethylene pellets were studied at 77 K and 298 K. Highly structured THz absorption spectra were typically observed with sharp spectral features determined by molecular symmetries and structure. In the same work, authors proved how quantum mechanical calculations, applied for the isolated diglycine species, can match experimental results if the description includes solid-state hydrogen-bonding and intermolecular interactions.

Likewise, Neu et al. [161] studied six polypeptide chains with very similar primary and secondary structures in the low-frequency region up to 2.5 THz. The result modeled with DFT calculations was used to calculate the dynamic motions of several peptides and to visualize the corresponding displacements of the strongest resonances.

Notably, the difference is clear between amino acids and polypeptides existing in the THz frequency dynamics range. The physical quantities show a different dependence by the length of the peptide chain, indicating that they reflect different interactions.

5. Current Status of THz Spectroscopy for Protein Research

As mentioned above, it is clear that molecular conformation affects the biological activity and functionality of proteins [162], so it is strategic to characterize the native protein conformation. With the advance in THz technologies, research on proteins and their conformation in the THz spectral region has been extensively carried out [163–166]. All findings demonstrated that the collective vibration mode related to the protein molecular conformation is optically active in the THz range [167,168]. Markelz and co-authors collected the THz spectrum of lyophilized powder bovine serum albumin (BSA) of pressed pellets in the range between 0.06 and 2.00 THz (2–66.6 cm^{-1}). Paciaroni et al. focused the attention on vibrational collective dynamics of a dry perdeuterated maltose-binding protein in the THz domain [169]. Sun et al. used THz spectroscopy to study the binding of hemagglutinin protein and broadly neutralizing monoclonal antibodies in the liquid environment [170].

THz spectroscopy is, among others, a convenient method for probing conformational changes in photoactive proteins [171–173]. Castro–Camus and Johnston [171] investigated conformational changes in photoactive yellow protein (PYP) suspended in a physiological buffer solution. The sample was uniformly illuminated by high-intensity blue (450 nm) light emitting diodes to trigger a conformational change from the ground state of the photoactive yellow protein (called pG) to the photo-intermediate (pB). Changes in the protein structure were probed by THz radiation, which is far from triggering any modifications in the protein itself. Absorption differences were observed, such as an increase in the absorption of the pB configuration compared to that in the pG state in the range 0.25–2 THz (8–66.6 cm^{-1}).

An interesting study was conducted by Han et al. using both THz spectroscopy and imaging [174]. These attempted to identify proteins and quantify their content. The frequency-dependent refractive index of six proteins (rn21, rn22, rn28, n42, n43 and n53 with concentration of 2.0 μg/μL and quantity of 8 μg) were initially evaluated with THz-TDS in the spectral region 0.1–3.5 THz (3.33–117 cm^{-1}). In Figure 3, the refractive index of these six proteins is reported in the sub-frequency band.