



ScuDo

Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Energy Engineering (29th Cycle)

Terahertz Vibrations in Proteins: Experimental and Numerical Investigation

By

Andrea Bassani

Supervisors:

Prof. Alberto Carpinteri
Prof. Giuseppe Lacidogna

Doctoral Examination Committee:

Prof. M. Rizzi, Referee, Università del Piemonte Orientale “A. Avogadro”.
Prof. S. Yoshida, Referee, Southeastern Louisiana University.
Dr. F. Rossi, Università del Piemonte Orientale “A. Avogadro”.
Prof. U. Lucia, Politecnico di Torino
Prof. B. Montrucchio, Politecnico di Torino

Politecnico di Torino
2017

Declaration

I hereby declare that the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Andrea Bassani

2016

Dedicated to my mum, my dad and my beloved Simona

Acknowledgments

The research presented in this thesis was carried out at the Department of Structural, Geotechnical and building Engineering (DISEG), Politecnico di Torino, in cooperation with the Department of Physics, Università La Sapienza, Roma, and with technical support of Horiba Scientific.

At the end of these three years I really like to thank my supervisors Prof. Alberto Carpinteri and Prof. Giuseppe Lacidogna for their constant support during my PhD, for their knowledge and motivation.

Their experience and guidance really inspired me.

I would express my thankfulness also to my colleague Gianfranco for his constant help with Lusas, without his help everything would have been much more difficult. Special thanks to my parents: few rows cannot express all my love and my immense gratitude for all the sacrifices you have done for me and on my behalf during my studies.

This thesis is also for the rest of my family: Ilaria, Marco, Fra, Ale, Zio Silvano, Zia Mariagrazia, Zio Claudio e Zia Anna, not forgetting my grandparents Nonno Giovanni, Nonna Carla, Nonna Mina and Nonno Egidio... Thanks to everyone!

Special thanks to my girlfriend and future wife Simona from the deep of my heart: your constant smile and your genuine love have always helped me to face difficulties and to put everything in the right perspective.

Sincere gratitude to my office mate and friend Oscar: work is always lighter with a

bit of laughter!

Last but not least, I also thank my colleagues Federico, Fabio, Giuseppe and Manuela for the fun we had during these years.

Abstract

The principal goal of this Doctorate thesis is to study high frequency vibrations (in the range between Gigahertz and Terahertz) in nanoscopic biological structures such as proteins. In particular, the idea of this thesis is to found, by means of experimental sessions and numerical simulations, natural frequencies of entire proteins or of large portions of that.

The mechanical behaviour of proteins is receiving an increasing attention from the scientific community. Recently it has been suggested that mechanical vibrations play a crucial role in controlling structural configuration changes (folding) which govern proteins biological function. The mechanism behind protein folding is still not completely understood, and many efforts are being made to investigate this phenomenon. Complex Molecular Dynamics simulations and sophisticated experimental measurements are conducted to investigate protein dynamics and to perform protein structure predictions; however, these are two related, although quite distinct, approaches. Here we investigate the linearly free dynamics (frequencies and modes) of proteins by Modal Analysis. The input mechanical parameters are taken from the literature. We first give an estimate of the order of magnitude of the natural frequencies of protein crystals by considering both classical wave mechanics

VIII

and structural dynamics formulas. Afterwards, we perform modal analyses of some relevant chemical groups and of the full lysozyme and Na-K ATPase proteins. The numerical results are compared to experimental data, obtained from both *in-house* and literature Raman measurements.

Our present investigations are devoted to understand if stimulating protein samples with a laser that excites resonant mechanical vibrations (say, in the THz range) may induce variations in the vibrational spectra due to possible conformational changes of protein structure.

Contents

▪	Acknowledgements.....	V
▪	Summary.....	VII
1	Dynamic Behaviour of Proteins: an Overview.....	1
1.1	Introduction and Historical Review.....	1
1.2	Theoretical natural frequencies: Wave Propagation vs Structural Dynamics.....	7
2	Biology of Proteins.....	13
2.1	Introduction: Peptide Bond, Primary, Secondary, Tertiary and Quaternary Structure.....	13
2.2	Folding: a Complex and not Totally Understood Process.....	21
2.3	Lysozyme: Structure and Functions.....	26
2.4	Na-K ATPase (Sodium-Potassium Pump): Structure and Functions.....	28
3	Dynamics of Nano-Mechanical Systems.....	33
3.1	Dynamics of Discrete Systems.....	33
3.2	Dynamic Instability: Modal Analysis with Second Order Effects.....	43

3.2.1	Influence of the Load on the Natural Frequency.....	44
3.2.2	Discrete Mechanical Systems with One Degree of Freedom.....	45
3.2.3	Discrete Mechanical Systems with Two Degrees of Freedom.....	51
3.3	Elementary Models of Proteins.....	60
4	Raman Spectroscopy: Experimental Campaigns on Lysozyme and Na-K ATPase.....	65
4.1	Elements of Raman Spectroscopy.....	65
4.2	Classical Mechanics Interpretation.....	69
4.3	Quantum Mechanics Interpretation.....	74
4.4	Use of Raman Spectroscopy for Proteins Study.....	76
4.5	Experimental Campaign on Lysozyme.....	77
4.6	Experimental Campaign on Na-K ATPase.....	83
5	Numerical Simulations: Lattice Models of Lysozyme and Na-K ATPase.....	93
5.1	Mechanical Model: Covalent Bond, Harmonic Approximation, Numerical Parameters.....	94
5.2	Results of Numerical Simulations of Lysozyme: Amino Acids and Global Vibrations.....	103
5.3	Results of Numerical Simulations of Na-K ATPase.....	106
6	Conclusions.....	111
	References.....	115

Chapter 1

Dynamic behaviour of proteins: An overview

The mechanical behaviour of proteins is receiving an increasing attention from the scientific community. In this Chapter we shall provide some general concepts on the subjects of dynamic behaviour of proteins, including a brief historical review over some techniques used to study these phenomena. Afterwards, we will give the objectives and motivation of the study.

1.1 Introduction and historical review

Recently it has been suggested that mechanical vibrations play a crucial role in controlling structural configuration changes (folding) which govern proteins biological function. The mechanism behind protein folding is still not completely understood, and many efforts are being made to investigate this phenomenon. Complex Molecular Dynamics simulations and sophisticated experimental measurements are conducted to investigate protein dynamics and to perform protein structure predictions; however, these are two related, although quite distinct, approaches. Here we investigate the linearly free

dynamics (frequencies and modes) of crystallized powder, lyophilized powder and rehydrated powder of proteins by Modal Analysis. The input mechanical parameters are taken from the literature. We first give an estimate of the order of magnitude of the natural frequencies of proteins by considering both classical wave mechanics and structural dynamics formulations. Afterwards, we perform modal analyses of some relevant chemical groups and of the full lysozyme and Na-K ATPase proteins. The numerical results are compared to experimental data, obtained from both *in-house* and literature Raman measurements.

Understanding the mechanical behaviour of proteins has important implications in different applied sciences, from biology and medicine, to biochemistry, pharmacology, and bioengineering. The proteins three-dimensional structure controls crucial biological processes such as mitosis, mechanotransduction, immune response, cell metabolism, neural signal transmission, oxygen transportation, etc. (*Alberts et al. 2002, Anfinsen 1972, Tymoczko et al. 2002, Saunders et al. 2010*). In all these processes, protein dynamics plays a crucial role (*Alberts et al. 2010*). Moreover, several neurodegenerative diseases, as well as many allergies, are believed to be caused by incorrect folding of some proteins (*Selkoe 2003, Hammarström et al. 2003, Chiti et al. 2006, Johnson et al. 2005*). For these reasons, the mechanisms behind protein folding are objective of investigation since long time (*Levinthal 1968, Anfinsen 1973, Kim et al. 1990, Shortle 1996, Jackson 1998, Van den Berg et al. 1999,2000, Deechongkit 2004, Kubelka et al. 2004, Ellis 2006, Rose et al. 2006, Ojeda-May et al. 2010*).

Nowadays, both experimental and computational techniques are used to investigate protein dynamics. As for the former, several standard non-crystallographic techniques are adopted, including Protein nuclear magnetic

resonance spectroscopy (PNMRS), Dual polarisation interferometry (DPI), high time resolution measurements (including neutron scattering), Vibrational circular dichroism (VCD), Proteolysis, and Optical tweezers (*Huyghues-Despointes et al. 2001, Cross et al. 2008, Bu et al. 2001, Baumruk et al. 1993, Minde et al. 2012, Park et al. 2005, Mashaghi et al. 2014, Jagannathan et al. 2013, Jakobi et al. 2011, Jagannathan et al. 2012*). A related, although quite different, approach for studying protein vibration and structure is the use of Molecular Dynamics simulations. In this case, *de novo* or *ab initio* techniques are adopted to investigate protein folding (*Compiani et al. 2013, Bryngelson et al. 1995, Leopold et al. 1992, Robson et al. 2008, Rizzuti et al. 2013, Schaefer et al. 1998, Kmiecik et al. 2007, Adhikari et al. 2012, Lindorff-Larsen et al. 2011*). Linear normal mode analysis (i.e., Modal Analysis) is also used in modeling the dynamics of biomolecular complexes: main pros and cons of this technique are discussed in (*Ma et al. 2005*).

Single-molecule manipulation has allowed the forced unfolding of multidomain proteins (*Sotomayor et al. 2007, Oberhauser et al. 2001, Schlierf et al. 2004, Liphardt et al. 2001*). Staple *et al.* developed a statistical mechanics theory for the stretching and unfolding of multidomain biopolymers (*Staple et al. 2009*). In that paper the theory, valid essentially for any molecule that can be unfolded in the atomic force microscope (AFM) (*Binnig et al. 1986*), was applied to reproduce the force-extension curves of both titin and RNA hairpins as an example.

It has been suggested that mechanical vibrations that involve the whole protein play a crucial role in controlling structural configuration changes (folding). In particular, underdamped low-frequency collective vibrational modes in proteins have been proposed as being responsible for efficiently

directing biochemical reactions and biological energy transport (*Turton et al. 2014*). Therefore, the existence of delocalized vibrational modes and their involvement in biological function has become objective of investigation. Vibrational modes of proteins, both crystallized and in solution, have been found in the fields of GHz and THz. For example, in lysozyme a $3\text{--}4\text{ cm}^{-1}$ ($\sim 100\text{ GHz}$) highly delocalized hinge-bending mode that opens and closes the binding cleft was found by normal mode calculations (*Levitt et al. 1985*, *Brooks et al. 1985*). At the same time, some authors report that collective modes may occur also in the TeraHertz range. For example, in (*Turton et al. 2014*) one reads that “there have been suggestions that TeraHertz-frequency underdamped collective modes of the protein may direct the system along the correct path on a highly complex potential-energy surface (*Deak et al. 1998*), and may be responsible for the low-loss transport of energy through the protein (*Davydov 1973*), which may facilitate biological function through phonon-like modes (*Liu et al. 2008*)”.

Several attempts to measure the vibrational spectra of proteins in the TeraHertz range have been done with the aim of finding peaks that could then be assigned to biochemically relevant motions in the protein (*Turton et al. 2014*). Before 1990 there was a gap in the electromagnetic spectrum, $0.1\text{--}1.5\text{ THz}$ ($3\text{--}50\text{ cm}^{-1}$), referred to as the “TeraHertz Gap”, that was difficult to utilise due to the lack of a suitable light source. The invention of TeraHertz time-domain spectroscopy (THz-TDS) enabled measurements using photons between $0.1\text{--}1.5\text{ THz}$ ($3\text{--}50\text{ cm}^{-1}$) that were previously impractical. Nowadays spectroscopic analysis using frequencies between $0.1\text{--}15\text{ THz}$ ($3\text{--}500\text{ cm}^{-1}$) is being utilized by the biochemistry community and is giving scientifically interesting information (*Falconer et al. 2012*). Recently, far-infrared and THz-TDS experiments on lysozyme crystals have successfully

identified underdamped delocalized vibrational modes in the TeraHertz range (*Acbas et al. 2014, Niessen et al. 2015*). However, the biochemical relevance of such modes can not be proved since the crystal packing and hydration level modify the protein dynamics (*Turton et al. 2014*). In fact, the same technique has proven to be largely unsuccessful when used for studying proteins in solution due to the very large absorption by liquid water (*Turton et al. 2014, He et al. 2008*). In contrast, inelastic neutron scattering (*Lerbret et al. 2013, Roh et al. 2006*) and spontaneous Raman scattering (*Genzel et al. 1976, Urabe et al. 1998, Perticaroli et al. 2010, Hedoux et al. 2006*) have been efficiently used to study proteins in solution since they do not suffer from water absorption. However, these techniques become unreliable at low frequencies (<1 THz) due to the very strong Rayleigh peak from elastic scattering, and again it has not been possible to assign biochemical relevance to any peaks observed (*Turton et al. 2014*). To encompass this problems, *Turton et al.* adopted Femtosecond optical Kerr-effect (OKE) spectroscopy for measuring the depolarized Raman spectrum in the time domain; this technique in practice is superior at low frequencies, as it does not suffer from a large Rayleigh peak (*Hunt et al. 2007*). Although broad-range spectroscopic measurements have been attempted, the challenge is to observe specific modes. As a matter of fact, calculations show that functional conformational change in many biomolecular systems can be simulated using only the first few collective vibrational modes of the system (*Karplus et al. 2005, Bahar et al. 2005*). These correlated motions are the low frequency vibrational modes that extend throughout the macromolecule. The problem lies in the selection of the most relevant modes. However, as stated in (*Turton et al. 2014*), the role of TeraHertz-frequency vibrational modes in mediating biological functions in

general is not known at present. Therefore the question is still open, and new experimental and numerical investigations are needed.

A recently proposed conceptual interpretation of protein folding suggests that it could be regarded as a dynamic nano-buckling (with snap-through) (*Carpinteri 2014a, 2015*). According to this hypothesis, the instability causing the abrupt configuration change can be induced by electro-chemical forces, acting on an oscillating structure. As it is known in structural mechanics, the force required to produce instability in an oscillating mechanical system is lower than the critical force for the same system at rest; on the other hand, if an external harmonic excitation has a frequency that matches one of the natural frequencies of the system, then resonance occurs, and the vibration amplitude increases (diverges if the system is undamped): in this case the force required to produce instability tends to zero (see Chapter 3) (*Bažant et al. 1991, Xie 2006, Virgin 2007, Carpinteri 2013, 2014b*).

Aim of the present thesis is to establish a comparison between experimental measurements of long-range vibrations of two proteins and their numerical simulation by a numerical Lattice Model. We focused on the Chicken Egg White Lysozyme (CEWL), a well-known enzyme that, and on Na-K ATPase, also known as “sodium-potassium pump”. We performed broad-range Raman spectroscopy measurements on crystallized powder samples of Lysozyme and on dry and wet lyophilized powder ATPase, and compared the results to literature data. Therefore, we modelled the said proteins as a three-dimensional lattice using the Finite Element code Lusas; a classic Modal Analysis was run. A correlation between the resonant peaks found in the experiments and the corresponding vibrations modes given by the numerical simulation was looked for and some relevant cases are discussed in the paper. Furthermore, modal analyses were also performed, using the same Finite

Element code, on the structure of three well-known amino acids, i.e. phenylalanine, tyrosine, and tryptophan. The agreement between numerical and experimental results is discussed in the Chapter 5 also for these cases.

What reported in the present thesis represents the first stage of a long-term project, the main objective of which is to investigate the effects of mechanical resonance and instability phenomena in proteins under physiological conditions. Future steps will therefore regard the study of the dynamic response of proteins stimulated with frequencies near the natural resonance.

1.2 Theoretical natural frequencies: Wave propagation vs Structural dynamics

First step of our work, before experimentation and simulation, was to try to obtain natural frequencies of nanoscopic structures like proteins theoretically. To achieve this result we moved through two different paths: the theory of wave propagation and the laws governing structural dynamics.

Longitudinal expansion/contraction (tension/compression) waves may be originated in solids, in addition to transverse (shear) waves, as a result of different causes. These are generally called pressure waves, or phonons when their particle nature is emphasized, and travel at a speed which is characteristic of the medium. For most of the solids (and fluids) they present an order of magnitude of 10^3 m s^{-1} . Their wavelength λ can not exceed the maximum size of the body (or portion of the body) involved. Therefore, knowing the travelling speed v and the dimensional scale, the order of magnitude of frequency f of

the pressure wave can be evaluated by means of the well-known relationship

$$(Feynman et al. 1964) f = \frac{v}{\lambda}.$$

This equation is only a simplification to obtain the order of magnitude of frequencies.

In a more correct way we can say that nanostructures, as proteins can subtend “confined acoustical phonons”. Confined acoustical phonons are low-frequency standingwaves corresponding approximately to the acoustical phonons of the material at a wavelength $\lambda = 2Nd$ where d is the smallest dimension of the nanostructure and N is an integer. The lowest frequency acoustical mode of a globular protein of radius R has a wavelength $\lambda \approx 2R$, (\approx denotes the order of magnitude) which is much larger than the distances between the atoms because R varies typically between 1 and 10 nm. Therefore, the properties of a confined acoustical mode do not depend on the details of the interactions between the atoms and can be described by regarding the protein as an effective elastic medium.

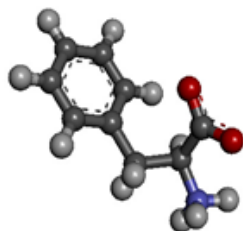
In elastic continuum theory, the lowest frequency of the confined longitudinal acoustical phonon of a sphere of radius R is given by

$$f = v / 2R \tag{1}$$

If the protein is represented by an elastic sphere of radius $R = 10$ nm and $v = 10^3$ m s⁻¹ we deduce $f \approx 500$ GHz.

Eq. (1) allows to obtain the order of magnitude of the eigenfrequencies of expansion/contraction vibration modes in solids at different dimensional scales. Thus, by assuming a constant pressure wave speed $v \approx 10^3$ m/s, we

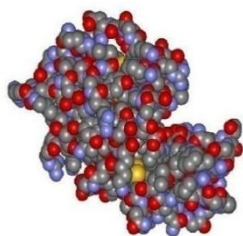
obtain for proteins vibration (Fig. 1.1): (i) $f \approx 10^{13}$ Hz for localized eigenvibrations of chemical groups or amino acids ($R \approx 10^{-10}$ m); (ii) $f \approx 10^{12}$ Hz for eigenvibrations of small proteins or extended portions of large proteins ($R \approx 10^{-9}$ m); (iii) $f \approx 10^{11}$ Hz for global eigenvibrations of large proteins ($R \approx 10^{-8}$ m).



Chemical groups / Amino Acids

$$R \approx 10^{-10} \text{ m}$$

$$f \approx 10^{13} \text{ Hz}$$



Small proteins/Large protein portions

$$R \approx 10^{-9} \text{ m}$$

$$f \approx 10^{12} \text{ Hz}$$



Large proteins

$$R \approx 10^{-8} \text{ m}$$

$$f \approx 10^{11} \text{ Hz}$$

Fig. 1.1 Proteins eigenvibration at different dimensional scales: correlation between the characteristic dimension, R , and frequency, f , by assuming a constant pressure wave speed $v \approx 10^3$ m/s.

On the other hand, the same results can be achieved by the theory of mechanical vibrations of elastic systems. Considering only expansion/contraction deformations, the order of magnitude of the vibration frequency can be evaluated as:

$$f \approx \frac{1}{2\pi} \sqrt{\frac{k}{m}}, \quad (2)$$

being k and m the apparent axial (tensile/compressive) stiffness and the mass of the system, respectively. Thus, applying Eq. (2) to proteins yields again ($k \approx 10^2 \text{ Nm}^{-1}$): (i) $f \approx 10^{13}$ Hz for chemical groups or amino acids ($m \approx 10^{-26}$ kg); (ii) $f \approx 10^{12}$ Hz for small proteins or extended portions of large proteins ($m \approx 10^{-24}$ kg); (iii) $f \approx 10^{11}$ Hz for large proteins ($m \approx 10^{-22}$ kg). For example, considering entire proteins, we have $f \cong 7.8 \times 10^{11}$ Hz for lysozyme ($m \cong 2.37 \times 10^{-23}$ kg; average diameter about 4–5 nm) or $f \cong 2.1 \times 10^{11}$ Hz for Na^+/K^+ -ATPase ($m \cong 1.78 \times 10^{-22}$ kg; average diameter about 4 nm, length about 16 nm). On the other hand, if we consider very small portions, like amino acids or chemical groups, we obtain frequencies of tens of TeraHertz (see Chapters 4-5). In order to make some more precise calculation, let us consider the simple case of a diatomic covalent bond.

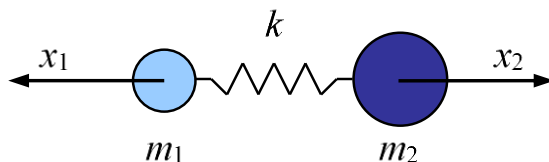


Fig. 1.2 Single C-C bond idealized as a mass-spring linear system

Considering the single C-C bond, the system appears as in Fig. 2.1, with $m_1 = m_2 = m \cong 1.99 \times 10^{-26}$ kg and $k = k_{C-C} \cong 180 \text{ N m}^{-1}$ (Ashby *et al.* 2014). Excluding rigid motions, vibration frequency (see also (Clough *et al.* 1975)) can be evaluated as follows:

$$f_{C-C} = \frac{1}{2\pi} \sqrt{\frac{2k_{C-C}}{m_C}} \cong 2.14 \times 10^{13} \text{ Hz} \quad (713 \text{ cm}^{-1}) \quad (3)$$

The previous value is in perfect agreement with the literature (see (Movasaghi *et al.* 2007)). At the same time, Eq. (3) can also be used to evaluate the bond stiffness when the mass and the vibration frequency of the diatomic molecule are known. For example, for the disulfide bridge (S-S bond) we have $m_S \cong 5.31 \times 10^{-26}$ kg and $f_{S-S} \cong 1.57 \times 10^{13}$ Hz (524 cm^{-1}) (see (Movasaghi *et al.* 2007)), and thus the spring stiffness should be:

$$k_{S-S} = 2\pi^2 m_S f_{S-S}^2 \cong 260 \text{ Nm}^{-1} \quad (4)$$

In conclusion, we remark that pressure waves may induce mechanical resonance in protein crystals (and in proteins in general) if their frequencies match the natural vibration frequencies of the atomic lattice.

Once we have understood that theoretical values of natural frequencies should range from 10^{11} Hz to 10^{13} Hz moving from big proteins to amino acids or chemical groups we started to plan experimental campaigns to measure these frequencies on different proteins.

Chapter 2

Biology of proteins

Proteins constitute most of a cell's dry mass; they not only represent the building blocks, but execute nearly all the cell's functions.

From many points of view proteins are the most complex and sophisticated molecules known; versatility of proteins is truly amazing.

In this chapter we shall introduce proteins structure from a biological point of view, moving from peptide bond and amino acids to 3D structures.

In the second part of the chapter we analyze the two proteins studied in this thesis: Lysozyme and Na-K ATPase (sodium-potassium pump).

2.1 Introduction: peptide bond, primary, secondary, tertiary and quaternary structure

Proteins are large biomolecules consisting at least of one chain of amino acids (20 amino acids are present in nature), each one linked to another by a covalent peptide bond (Figure 2.1); for this reason proteins are also called polypeptides.

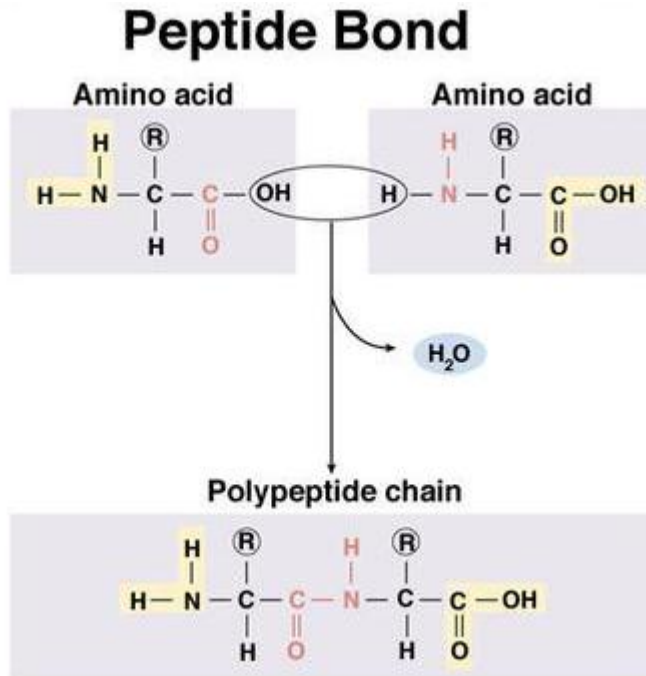


Figure 2.1: Peptide bond is synthesized when the carboxyl group of one amino acid molecule reacts with the amino group of the other amino acid molecule, causing the release of a molecule of water (H₂O)

Among the 20 amino acids present in nature 18 are composed by carbon, hydrogen, oxygen and nitrogen, while in methionine and cysteine also sulphur is present (Table 2.1).

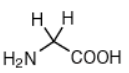
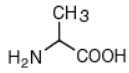
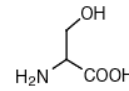
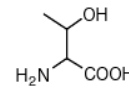
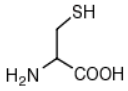
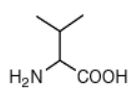
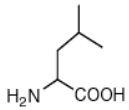
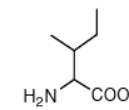
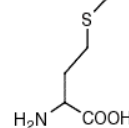
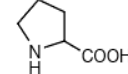
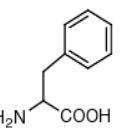
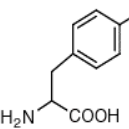
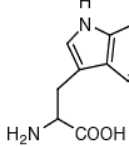
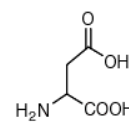
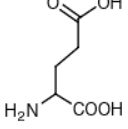
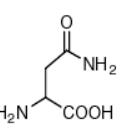
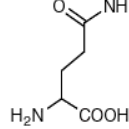
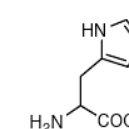
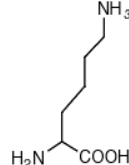
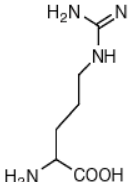
Small		Nucleophilic		
				
Glycine (Gly, G) MW: 57.05	Alanine (Ala, A) MW: 71.09	Serine (Ser, S) MW: 87.08, pK _a ~ 16	Threonine (Thr, T) MW: 101.11, pK _a ~ 16	Cysteine (Cys, C) MW: 103.15, pK _a = 8.35
Hydrophobic				
				
Valine (Val, V) MW: 99.14	Leucine (Leu, L) MW: 113.16	Isoleucine (Ile, I) MW: 113.16	Methionine (Met, M) MW: 131.19	Proline (Pro, P) MW: 97.12
Aromatic		Acidic		
				
Phenylalanine (Phe, F) MW: 147.18	Tyrosine (Tyr, Y) MW: 163.18	Tryptophan (Trp, W) MW: 186.21	Aspartic Acid (Asp, D) MW: 115.09, pK _a = 3.9	Glutamic Acid (Glu, E) MW: 129.12, pK _a = 4.07
Amide		Basic		
				
Asparagine (Asn, N) MW: 114.11	Glutamine (Gln, Q) MW: 128.14	Histidine (His, H) MW: 137.14, pK _a = 6.04	Lysine (Lys, K) MW: 128.17, pK _a = 10.79	Arginine (Arg, R) MW: 156.19, pK _a = 12.48

Table 2.1: 20 amino acids divided in 7 sub-families. For every amino acid are given chemical structure, name, and molecular weight and acid dissociation constant if present.

There are many thousands of protein types, but each one has its unique amino acids sequence. It's possible to identify a polypeptide backbone, made by the repeating

sequence of atoms along the core of the chain, and different side chains constituted by portions of amino acids not involved in peptide bonds but important to give specific characteristics to the chain (positive/negative charge, polar/non polar and so on) (Figure 2.2).

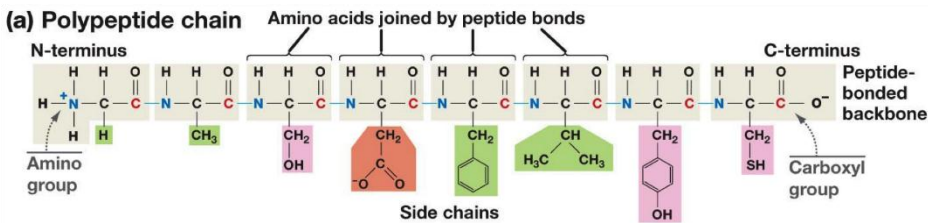


Figure 2.2: Protein consist of a polypeptide backbone with attached side chains of various types.

Atoms of the chains behave almost like hard spheres with a radius (van der Waals radius); for this reason possible bond angles are greatly limited and three dimensional arrangements of atoms are severely restricted.

With “folding” we describe conformational changes that lead a protein by its linear chain conformation to a three dimensional state without which it could not perform any of its functions

The folding of a protein chain, however, is restricted not also by steric reasons but also by many different non-covalent bonds, like Hydrogen bonds, electrostatic attractions and van der Waals attractions.

Non-covalent bonds are weaker than covalent (from 30 to 300 times) but acting together they contribute to the stability of each folded shape. (Figure 2.3)

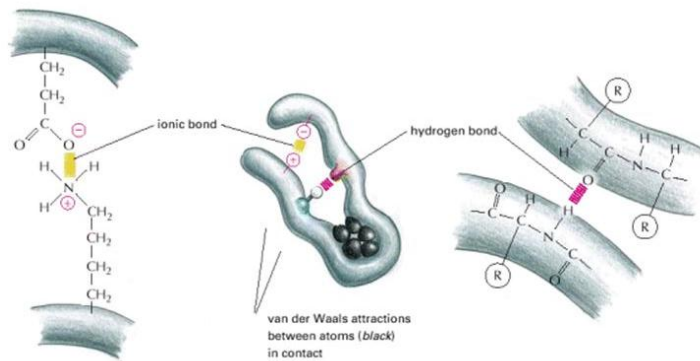


Figure 2.3: Three types of non covalent bonds presents in proteins

The final folded structure, also called conformation, is generally the ones that minimizes its free energy.

All the information needed by the protein to fold in its specific three dimensional shape are contained in the amino acid sequence. Each protein has its specific conformation but this shape can change a little when the protein interacts with other molecules such as other proteins, chemical elements, bacteria etc. This change in shape is absolutely fundamental to the function of the protein.

When we observe three dimensional structures of many different proteins it's clear that, despite the unique conformation of each type, two regular folding patterns are present. The first one, called α helix, is abundant in skin, hair, nails and horns, the second one, called β sheet, is abundant in the core of many proteins (Figure 2.4).

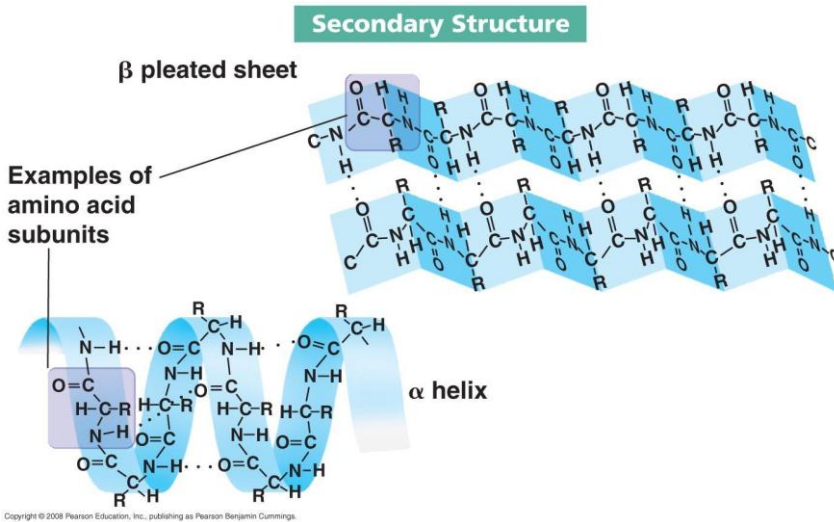


Figure 2.4: Two types of spatial organization in secondary structure of the backbone chain

Biologists often distinguish four steps of organization for proteins structures. Amino acid sequence is called primary structure, chains forming α helices and β sheets are called secondary structures, three dimensional shape is known as tertiary structure. If a particular protein is formed by more polypeptide chains, the entire structure is known as quaternary structure. (Figure 2.5)

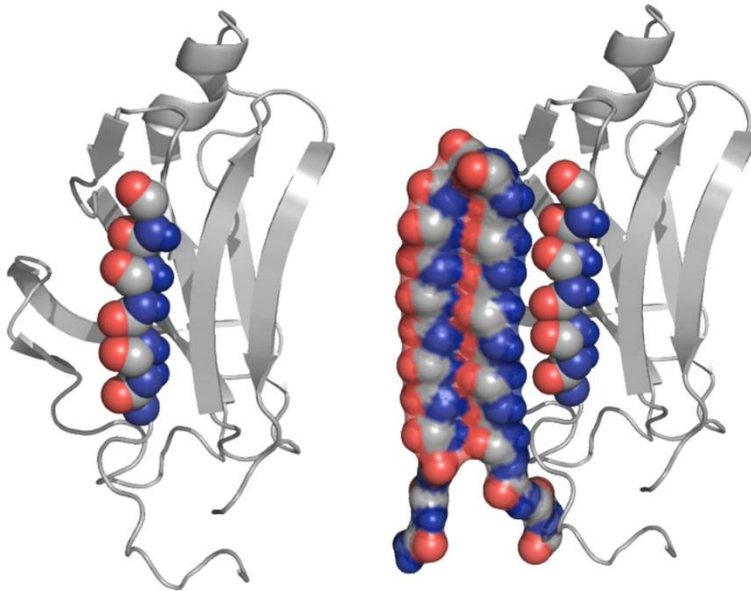


Figure 2.5: Quaternary structure depicted highlighting both atoms both secondary structures

Figure 2.6 summarizes all the four stages of the proteins structures.

Since every amino acid is chemically different and each one can occur in any position of a protein chain there are 20^n distinct possible polypeptide chains n amino acids long. Considering a typical protein length of about 300 amino acids, a cell could make more than 20^{300} different polypeptide chains. This number is enormous but only a small portion of that can be achieved since not all three dimensional conformation are possible, in fact proteins are so highly precisely built that the change of a few atoms, may be also one, in one amino acid can destroy the structure of the entire molecule.

Proteins play a large number of functions within the body including the catalysis of numerous metabolic reactions, DNA replication, response to stimuli and the transport of molecules.

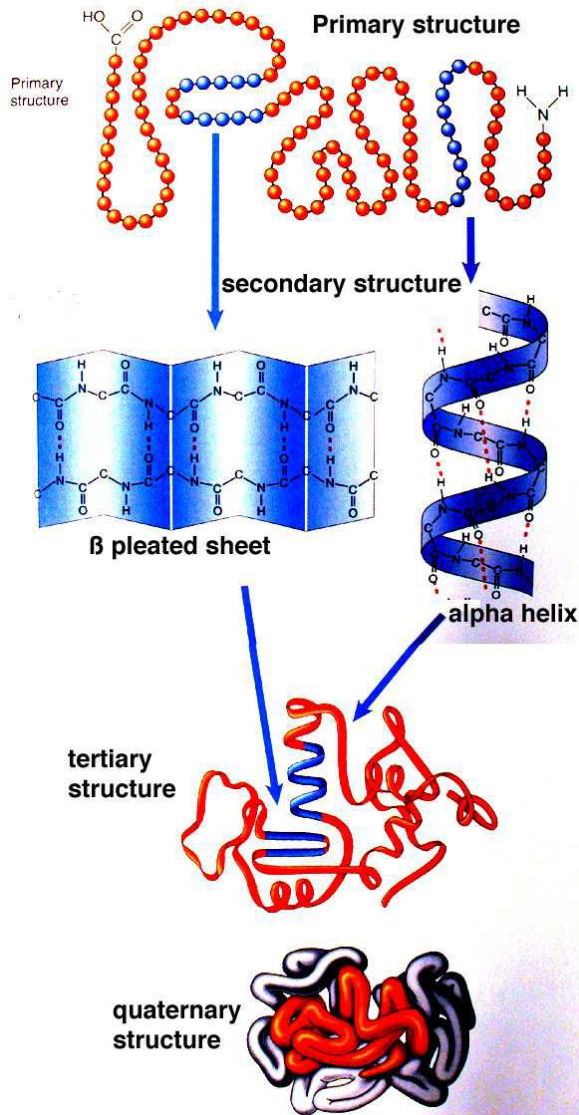


Figure 2.6: Sequence of the steps of the three dimensional shape of a protein moving from primary to quaternary structure

2.2 Folding: a complex and not totally understood process

How a protein folds to its functional state is a biological process that has occupied the mind of scientists for many years.

The “protein folding problem” consists of three strictly related pieces of a bigger puzzle: (1) What is the folding code? (2) What is the folding mechanism? (3) Can we predict the three dimensional structure of a protein knowing only its primary structure? Once described as an enormous challenge, protein folding has seen great progress in the last years. Nowadays, foldable proteins and nonbiological materials like polymers are designed quite easily and move toward successful applications. 3D structures of small proteins are now well predicted by computer algorithms. And now there is a reasonable answer for the question “how a protein can fold so quickly?”: a protein solves its global optimization problem by solving a series of smaller local optimization problems, growing and assembling the final structure from peptide fragments (*Dill, 2008*).

Moreover the protein folding process can be divided in two major problems. The first, introduced through the work of Anfinsen, refers to the folding/refolding of ribonuclease A (*Anfinsen et al. 1961, 1963*). In its studies Anfinsen showed that a polypeptide chain can spontaneously fold to its lowest free-energy without the help of any other biological system. This observation reminds to the idea expressed before that all the information necessary for protein folding is encoded in its own amino acid sequence (primary structure).

Two major conclusions followed from Anfinsen's work. First of all, it gave the start to the large research group of *in vitro* protein folding that has come to understand native structures by experiments inside test tubes rather than inside cells. Second, the Anfinsen idea implies a sort of division of works between biology and physic/chemistry: evolution can modify an amino acid sequence, but the folding equilibrium and kinetics of that sequence are then matters of physic and chemistry. Before the 1980s, the protein folding was seen as a sum of many different smaller interactions, such as van der Waals forces, hydrogen bonds, ion pairs, and water-mediated hydrophobic interactions. A key idea was a sort of cascade: the primary sequence encoded secondary structures, which then encoded tertiary structures (*Anfinsen et al. 1975*) and so on. However, thanks to statistical mechanical modeling and computer algorithms, a different point of view composed by three sub-ideas emerged in the end of 1980s: (a) there is a dominant component to the folding code, that it is the hydrophobic interaction, (b) the folding code is equally distributed both locally and nonlocally in the sequence, and (c) a protein's secondary structure is seen as much a consequence of the tertiary structure as a cause of it (*Dill 1990, 1999*).

Knowing that native proteins are only 5–10 kcal/mol more stable than their denatured states, it is obvious that no type of intermolecular force can be neglected in folding and structure prediction (*Yang et al. 2007*). Although it remains very hard to separate in a clean and rigorous way some types of interactions from others. Here are some examples. Folding process is not likely to be dominated by electrostatic interactions among charged side chains cause many proteins have relatively few charged residues; they are concentrated in high-dielectric regions on the protein surface. Protein stabilities and so folded structure, tend to be independent of pH (near neutral) and salt concentration, and charge mutations typically lead to small

effects on structure and stability. Hydrogen-bonding interactions are important, because essentially all possible hydrogen-bonding interactions are generally satisfied in native structures. Hydrogen bonds among backbone amide and carbonyl groups are important components of all secondary structures, and many studies of mutations in different solvents estimate their strengths to be around 1–4 kcal/mol (*Byrne et al. 1975, Fersht et al. 1985*) or even stronger (*Auton et al. 2007, Deechongkit et al. 2004*). Similarly, tight packing in proteins implies that also Van der Waals interactions are important (*Chen et al. 2001*).

However, the central question of the folding problem is: “there is a dominant factor that explains why any two proteins, for example, lysozyme and ribonuclease, have different native structures?”.

The answer of this question must be conserved in the side chains, not in the backbone, because one protein differs from another through the side chains. For example, there is evidence that hydrophobic interactions play an important role in protein folding (*Wolfenden 2007, Cordes et al. 1997, Hecht et al. 2004, Kamtekar et al. 1993, Kim et al. 1998, Wurth et al. 2006*).

Another great question about folding path is: what stabilizes secondary structures? Before any protein structure was known, Linus Pauling (*Pauling et al. 1951*) understood that proteins might have α -helices. Studies of lattice models (*Chan et al. 1990, Chikenji et al. 2006, Dill et al. 1995*) and tube models (*Banavar et al. 2007, 2002, Micheletti et al. 1999*) have highlighted that secondary structures in proteins are, first of all, stabilized by the chain compactness, an indirect consequence of the hydrophobic force to collapse. Helical and sheet are the only regular ways to pack a linear chain into a tight space.

Anyway our comprehension of the forces of folding remains quite incomplete.

The second major problem refers to the kinetics and dynamics of folding. Levinthal in 1968 studied for first this topic (*Levinthal 1966, 1969*). He has proved that folding process can't consist in a search of all the possible conformational configurations, because it would take an enormous amount of time, around 10⁵² sec for a protein with ~100 amino acid residues, like Lysozyme, (whereas is known that a protein folds in few milliseconds); this is called "Levinthal Paradox".

C. Levinthal first observed, in 1968, that proteins can search and fall very quickly to native states, often in microseconds. How do they do so quickly? It was suggested that if we understood the mechanism of protein folding, it could lead to computer algorithms to predict native structures from their amino acid sequences.

Other random questions like the following have driven the research: how can all the denatured molecules in a beaker find the same native structure, starting from different conformations? What states are not searched? Is folding a hierarchical process (*Baldwin et al. 1999, Kim et al. 1982*)? Are present folding nuclei (*Dyson et al. 2006, Matheson et al. 1978*)?

Several models have emerged: diffusion-collision model (*Karplus et al. 1979, 1994*), the nucleation-condensation mechanism (*Fersht 1997*), hierarchical diffusion-collision model (*McCammon et al. 1977, Myers et al. 2001*), hierarchic condensation (*Lesk et al. 1981*), foldons (*Callender et al. 1998, Krishna et al. 2006*), native-like topologies (*Debe et al. 1999, Makarov et al. 2002*).

These models are not mutually exclusive.

In spite of the numerous folding studies, both theoretical and experimental, this is still an open field and many questions remain.

New methods like threading (*Elofsson et al. 1996*) and homology modelling (*Moult et al. 1995, Sali 1995*) will probably allow for the prediction of the native state,

solving at the same time the first problem without giving a complete description of how a polypeptide acquires its 3D structure in a so short time.

When these methods are successful, it will allow to go beyond bioinformatics to (1) predict conformational changes, important for drug discovery; (2) understand protein mechanisms of action, folding processes, enzymatic catalysis, and other processes that require much more than the static native structure; (3) understand how proteins react to solvents, different pH, salts, denaturants, and so on.

2.3 Lysozyme: Structure and functions

In the present thesis two proteins were analyzed, lysozyme and Na-K ATP-ase, the so called sodium-potassium pump.

Lysozyme is a rather small protein, also known as N-acetylmuramide glycanhydrolase, consisting of 129 amino acids and having a molecular weight equal to 14.3 kDa. Its function is to hydrolyze some types of polysaccharides, including those that constitute the cell membrane of bacteria (*Fleming 1922*). Lysozyme is very abundant in nature and is found in a large number of biological secretions such as saliva, tears, human milk, etc. In humans this protein is encoded by the gene LYZ. It is also present within cytoplasmic granules of macrophages. An abundant source of lysozyme is also the egg white (*Yoshimura et al. 1988*). The antibacterial properties of the egg white –later attributed to the lysozyme contained within it– were discovered at first by Laschtschenko in the early 20-th Century (1907); however, the lysozyme was described for the first time in 1922 by Alexander Fleming, who coined the name which nowadays commonly identifies this protein. Its enzymatic function is held by catalyzing the hydrolysis of 1,4-beta-liNa-K ATPaseges between N-acetylmuramic acid and N-acetyl-D-glucosamine residues

in a peptidoglycan, and between N-acetyl-D-glucosamine residues in chitodextrins (*Peters et al. 1989*).

The three-dimensional structure of hen egg white lysozyme was described for the first time in 1965 by David Chilton Phillips, who obtained a 2-ångström (200 pm) resolution model via X-ray crystallography (*Blake et al.1965*). The shape of this enzyme is globular (compact), with an average diameter of 4-5 nm. It has been the first enzyme containing all the 20 amino acids the 3D structure of which was obtained. In 2007 Steve Kent obtained the first chemical synthesis of lysozyme at the university of Chicago. (Figure 2.7)

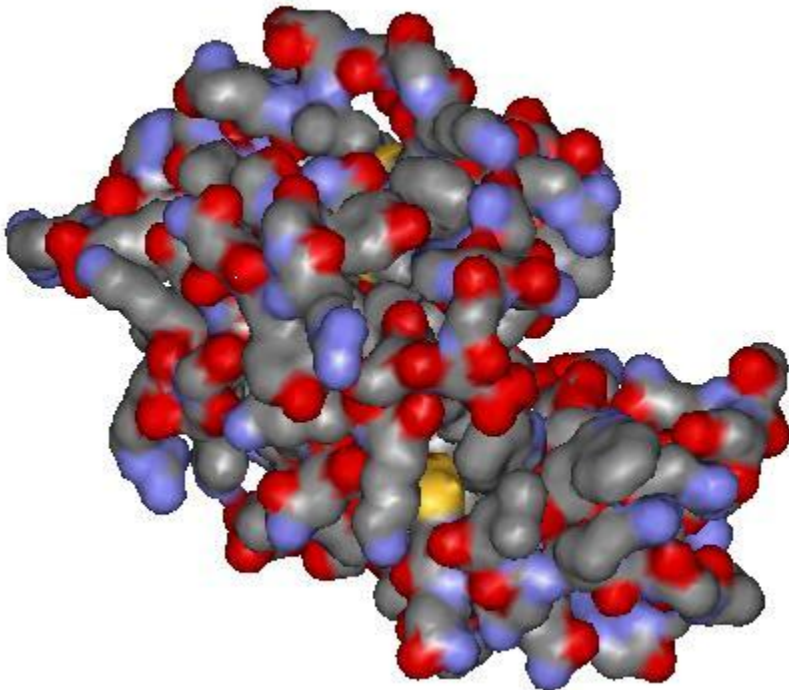


Figure 2.7: Three dimensional structure of lysozyme

2.4 Na-K ATPase (sodium-potassium pump): Structure and functions

Na⁺,K⁺-adenosine triphosphatase (NA-K ATPASE) is the most prominent member of the P-type adenosine triphosphatase (ATPase) family that includes sarcoplasmic reticulum Ca²⁺-ATPase and gastric H⁺,K⁺-ATPase, among others. This protein is also known as sodium potassium pump, is found in the plasma membrane of virtual all animal cells. NA-K ATPASE pumps three Na⁺ ions out and two K⁺ ions into the cell per molecule of ATP hydrolysed and thereby creates concentration gradients across the plasma membrane. It is primarily a Na⁺ ion pump, as K⁺ can be substituted equally well with other monovalent metal ions and even organic cations (*Ratheal et al., 2010*), whereas only Li⁺ and H⁺ can substitute partially for Na⁺ (*Blostein, 1985*). The gradients of Na⁺ and K⁺ ions created and maintained by NA-K ATPASE are used for many purposes, including the generation of action potentials along nerves, and as an energy source for secondary active transport.

NA-K ATPASE is expressed in all animal cells and is finely regulated. In addition to the catalytic unit (the α -subunit) of approximately 1,000 residues NA-K ATPASE contains a heavily glycosylated β -subunit of about 300 residues and a tissue-specific auxiliary regulatory subunit of approximately 70-180 residues. The latter are known as FXYD proteins and modulate Na-k ATPase function according to the specific needs of a given tissue. Furthermore, Na-k ATPase is now recognised as a key player in cell adhesion. Finally, wrong expression and activity of Na-k ATPase may be implicated in the development and progression of various types of cancer

(*Mijatovic et al. 2007*). Thus, Na-k ATPase is now recognized as an important therapeutic target (*Prassas et al. 2008*).

As seen from the brief discussion of Na-k ATPase physiological function, it is a complex system that is pivotal in the regulation of humans cell homeostasis.

The reaction cycle of P-type ATPases is explained by the Post-Albers scheme (*Albers 1967*) or E1/E2 theory. Here, pumping action is achieved by alternating the affinity and the accessibility of the transmembrane (TM) cation binding sites (*Glynn, 1993*). In the E1 species (e.g., $E1 \cdot 3Na^+$), the TM cation binding sites have high affinity for Na^+ and face the cytoplasm, whereas in E2, the binding sites have low affinity for Na^+ but high affinity for K^+ and face the extracellular side. The system is expected to have two gates, one on the cytoplasmic side and the other on the extracellular side (*Gadsby, 2009*). The two gates must operate in coordination, and the cycle must contain a step in which both gates are closed so that the TM binding sites become inaccessible from either side of the membrane. The peculiarity of P-type ATPases is autophosphorylation and –dephosphorylation of the catalytic aspartate during the reaction cycle. Only after three Na^+ ions are bound is the cytoplasmic gate closed and locked by phosphoryl transfer to the aspartate (i.e., transition into the E1P state). In the E1P state, bound Na^+ ions are occluded. During the transition to E2P, the extracellular gate opens, and three Na^+ ions are released sequentially. Then, the ions to be countertransported (normally 2 K^+) occupy the binding sites. The extracellular gate is closed by the hydrolysis of the aspartylphosphate and locked in $E2 \cdot 2K^+$. This is a stable state, and the transition into E1 is very slow without ATP. (Figure 2.8)

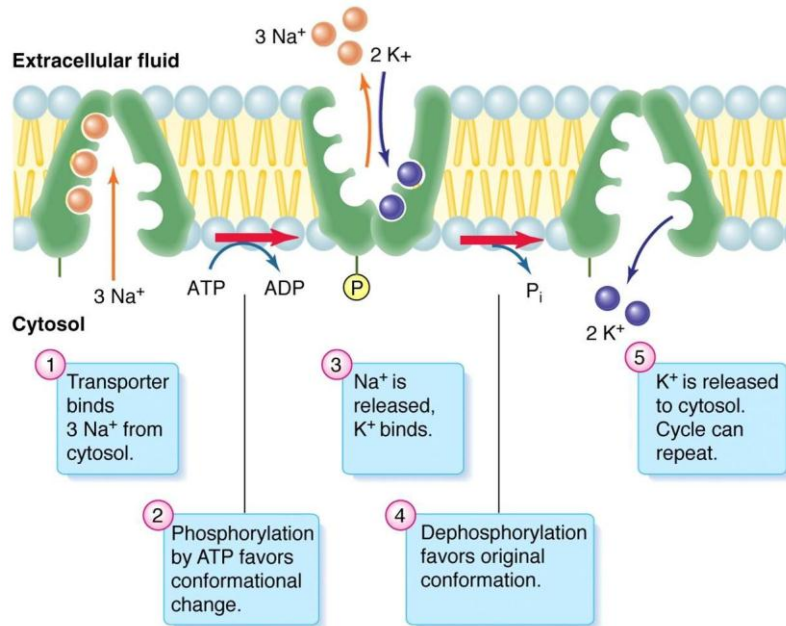


Figure 2.8: Na-K ATPase pump cycle

The three-dimensional structure of Na-K ATPase was described by Shinoda, Ogawa, Cornelius, who obtained a 2,4-ångström (240 pm) resolution model via X-ray crystallography (*Shinoda et al. 2009*). The shape of this enzyme is elongated, with two sort of horns that extend inside the cell and a globular part outside the cell. The entire protein has a main size of 15-16 nm, and counts 10131 atoms without Hydrogen. (Figure 2.9)

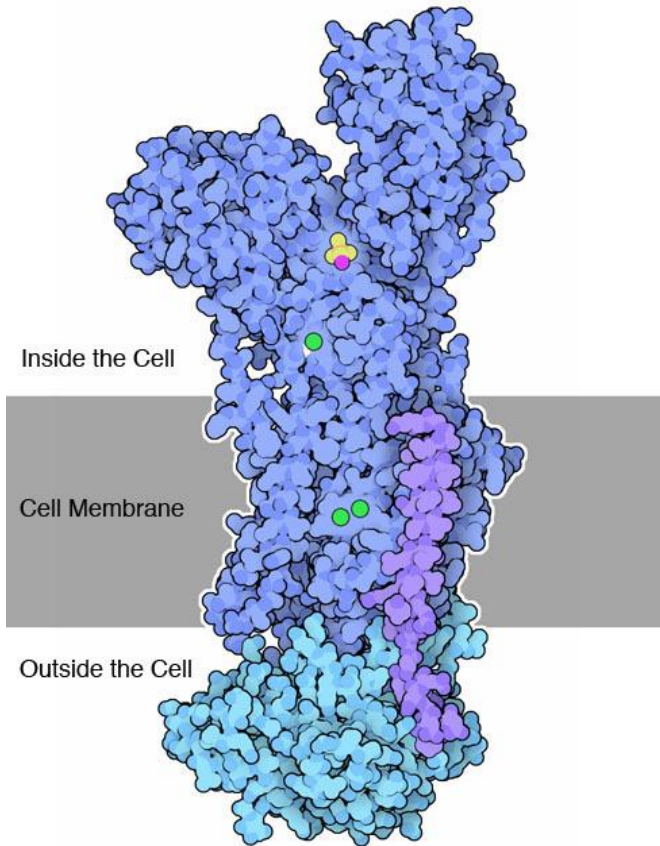


Figure 2.9: Na-K ATPase structure. In this picture are highlighted the three sectors into which is divided the protein: the first part lays inside the cell, the second one lays between the cell membrane, the last one is outside the cell.

Chapter 3

Dynamics of nano-mechanical systems

In this Chapter we shall provide some general concepts on Dynamics, discussing in the first part Dynamics of discrete system of one and more degrees of freedom, while in the second part we introduce the modal analysis with second order effects, in particular the influence of the load on the natural frequencies of discrete mechanical systems.

In the latter part of the Chapter, we apply these concepts to elementary models of proteins to obtain a general idea of their natural frequencies.

All concepts presented in this Chapter are borrowed and rearranged from Carpinteri, “Dinamica delle Strutture”, Pitagora Editrice Bologna.

5.1 Dynamics of discrete systems

The main goal of the dynamics of structures is to present methods for analyzing

stresses and deflections in any given structural system, when it is subjected to arbitrary dynamic loadings. This objective can be considered as an extension of standard methods of structural analysis, which are generally concerned with static loadings. These become thus functions of time, as well as the structural response. The dynamic loadings acting on a structure can be periodic or nonperiodic. The simplest periodic loading has the sinusoidal variation, which is also termed as harmonic. On the other hand, nonperiodic loadings may be either short-duration or impulsive loadings, as those generated by explosions, or long-duration loadings, as might result from an earthquake.

If a structure is subjected to a static load, the internal moments and shearing forces, as well as the deflection shape, depend only upon this load, by established principles of internal force equilibrium.

On the contrary, if the load is applied dynamically, the structural response depends also on the inertial forces, which oppose the accelerations producing them. If the motion is so slow to neglect both damping and inertial forces, the analysis can be considered to be static instant-by-instant, although loading and structural response are both time-dependent.

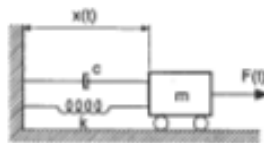


Fig 3.1: discrete system, one degree of freedom, composed by mass, spring and damper

The equation of motion of a single mass, subjected to an elastic force and a damping

force (Figure 3.1), can be expressed as:

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = 0 \quad (3.1)$$

where x is the linear elastic spring elongation, which depends on time t (the dot over the function represents the time derivative), m is the mass, c is the damping constant and k is the spring stiffness. Relationship (3.1) represents the well-known dynamic equation: *force = mass × acceleration*. In fact, both the active forces, $-kx$ and $-c\dot{x}$, result to be negative in case of positive elongations and velocities, respectively. Another interpretation which can be given to Equation (3.1) is by means of D'Alembert's principle, according to which each mass is in equilibrium in its frame of reference, once subjected to all the active and inertial forces. The latter oppose the acceleration and are equal to the product of the acceleration itself times the mass.

When the forces applied to the mass are not external, but only internal (elastic and damping forces) and inertial, the motion of the system is called free vibration. The solution of Equation (3.1) takes the following form:

$$x(t) = C e^{st} \quad (3.2)$$

Substituting Equation (3.2) into Equation (3.1) yields:

$$(ms^2 + cs + k)Ce^{st} = 0 \quad (3.3)$$

Dividing by mCe^{st} and introducing the notation

$$\omega^2 = \frac{k}{m} \quad (3.4)$$

Equation (3.3) becomes

$$s^2 + \frac{c}{m}s + \omega^2 = 0 \quad (3.5)$$

Here we have to face two case, in the first one, where we found $c = 0$ we will talk about undamped free vibrations.

In this case, the two solutions of Equation (3.5) are

$$s = \pm i\omega \quad (3.6)$$

where i is the imaginary unit. The system response is thus given by:

$$x(t) = C_1 e^{i\omega t} + C_2 e^{-i\omega t} \quad (3.7)$$

Since, according to Euler's formula, we have:

$$e^{\pm i\omega t} = \cos \omega t \pm i \sin \omega t \quad (3.8)$$

Equation (3.7) can be rewritten as

$$x(t) = A \sin \omega t + B \cos \omega t \quad (3.9)$$

where the two constants A and B can be expressed by means of the initial conditions.

In fact, since:

$$x(0) = B \quad (3.10a)$$

$$\dot{x}(0) = A\omega \quad (3.10b)$$

Equation (3.9) becomes (Figure 1.3):

$$x(t) = \frac{\dot{x}(0)}{\omega} \sin \omega t + x(0) \cos \omega t \quad (3.11)$$

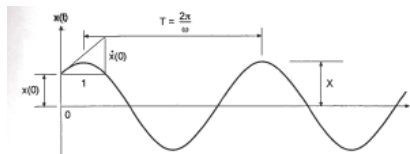


Fig 3.2: Harmonic motion and its parameters

This solution represents a simple harmonic motion and it is homogenous from a dimensional point of view, since the angular frequency (or angular velocity) ω has dimensions $[T]^{-1}$ and it is measured in radians per second. The ordinary frequency is measured in Hertz (cycles per second):

$$f = \frac{\omega}{2\pi} \quad (3.12)$$

whereas its reciprocal represents the period T :

$$T = \frac{1}{f} = \frac{2\pi}{\omega} \quad (3.13)$$

In addition to Equation (3.11), the motion can be described by the following expression:

$$x(t) = X \cos(\omega t - \varphi) \quad (3.14)$$

where the amplitude is given by

$$X = \sqrt{[x(0)]^2 + \left[\frac{\dot{x}(0)}{\omega}\right]^2} \quad (3.15)$$

and the phase angle by

$$\varphi = \arctan \frac{\dot{x}(0)}{\omega x(0)} \quad (5.16)$$

In the second case, where we found $c > 0$, we will talk about damped free vibrations. In this thesis, due to the nature of the chemical bonds and to the nature of the numerical simulation we performed, this type of motion is less important, so we will give only basic concepts.

In this case, the two solutions of Equation (3.5) are

$$s_{1,2} = -\frac{c}{2m} \pm \sqrt{\left(\frac{c}{2m}\right)^2 - \omega^2} \quad (3.17)$$

Three types of motion are represented by this expression, according to the quantity under the square-root, whether it is positive, negative or equal to zero. It is convenient to discuss first the case when the radical term vanishes.

$$1^{\text{st}} \text{ case: } \omega = \frac{c}{2m} \text{ (critically-damped system)}$$

The critical value of the damping coefficient is:

$$c_c = 2m\omega = 2m\sqrt{\frac{k}{m}} = 2\sqrt{km} \quad (3.18)$$

Note that this response does not include oscillation about the equilibrium position, but only an exponential decay towards this position (Figure 3.3).

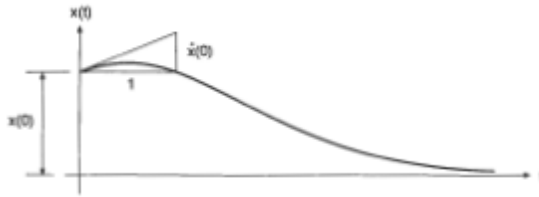


Fig 3.3: dynamic response in critically damped condition

It can be stated that the critically-damped condition represents the smallest amount of damping for which no oscillation occurs in the free-motion response.

2nd case: $c < 2m\omega$ (*undercritically-damped system*)

Since the quantity under the radical sign in Equation (3.17) is negative, it is convenient to describe damping in terms of a damping ratio ζ , which is the ratio of the given damping coefficient c to the critical value c_c :

$$\zeta = \frac{c}{c_c} = \frac{c}{2m\omega} \quad (3.22)$$

Introducing Equation (3.22) into Equation (3.17) leads to:

$$s_{1,2} = -\zeta\omega \pm \sqrt{(\zeta\omega)^2 - \omega^2} \quad (3.23)$$

with $0 < \zeta < 1$.

Equation (3.23) can be expressed as

$$s_{1,2} = -\zeta\omega \pm i\omega_D \quad (3.24)$$

where

$$\omega_D = \omega\sqrt{1-\zeta^2} \quad (3.25)$$

is the damped frequency. Its value is close to that related to the undamped frequency ω in practical cases, where generally $\zeta < 1/4$.

The dynamic response of an undercritically-damped system is obtained by substituting the solutions (2.24) into Equation (3.2):

$$x(t) = C_1 e^{-\zeta\omega t + i\omega_D t} + C_2 e^{-\zeta\omega t - i\omega_D t} = e^{-\zeta\omega t} \left(C_1 e^{i\omega_D t} + C_2 e^{-i\omega_D t} \right) \quad (3.26)$$

The term within brackets represents a simple harmonic motion.

A plot of the response of an undercritically-damped system subjected to an initial displacement $x(0)$ and starting with zero velocity is shown in Figure 3.4.

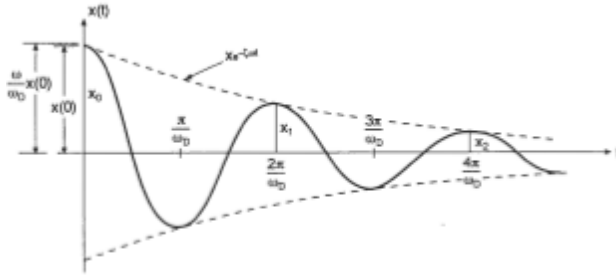


Fig 3.4: dynamic response in under-critically damped condition

The mass oscillates about the neutral position with an exponentially decreasing amplitude.

3rd case: $c > 2m\omega$ (*overcritically-damped system*)

In this case, $\zeta = c / c_c > 1$ and the response is similar to the motion of a critically-damped system. However, the asymptotic return to the neutral position is slower depending upon the amount of damping. Note that it is very unusual, under normal conditions, to have overcritically-damped structural systems.

Until now we have treated discrete systems with a single degree of freedom but in general, and in the protein structures that we analyze, structures must be described by discretized models with several degrees of freedom, and not by a single degree of freedom model. Indeed, the structures are continuous systems and would present an infinite number of degrees of freedom.

Consider a vibrating system formed by n masses m_i and by n springs in series, with stiffness k_i , $i=1, 2, \dots, n$ (Figure 3.5).

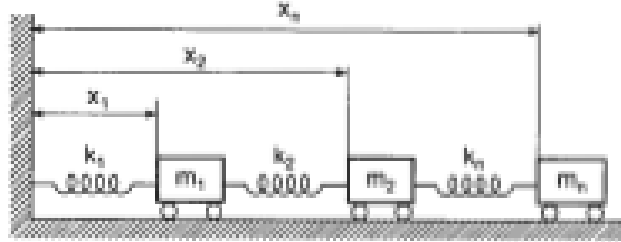


Fig 3.5: Mechanical system with n -degrees of freedom

There are n equations of motion, one for each vibrating mass:

$$\begin{aligned}
 m_1 \ddot{x}_1 + k_1 x_1 - k_2 (x_2 - x_1) &= 0 \\
 m_2 \ddot{x}_2 + k_2 (x_2 - x_1) - k_3 (x_3 - x_2) &= 0 \\
 m_3 \ddot{x}_3 + k_3 (x_3 - x_2) - k_4 (x_4 - x_3) &= 0 \\
 \vdots & \\
 m_n \ddot{x}_n + k_n (x_n - x_{n-1}) &= 0
 \end{aligned} \tag{3.27}$$

In Equations (3.27) viscous damping forces are neglected, as well as the presence of forcing actions. Equations (3.27) can be put in the compact form:

$$[M]\{\ddot{x}\} + [K]\{x\} = \{0\} \tag{3.28}$$

where $[M]$ and $[K]$ are the mass and stiffness matrices, respectively:

$$[M] = \begin{bmatrix} m_1 & 0 & 0 & \dots & 0 \\ 0 & m_2 & 0 & \dots & 0 \\ 0 & 0 & m_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & m_n \end{bmatrix} \tag{3.29a}$$

$$[K] = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & 0 & \dots & 0 \\ -k_2 & k_2 + k_3 & -k_3 & 0 & \dots & 0 \\ 0 & -k_3 & k_3 + k_4 & -k_4 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -k_n & k_n \end{bmatrix} \quad (3.29b)$$

Note that the mass matrix is diagonal, while the stiffness matrix results to be tri-diagonal.

Let us search for a solution of Equation (3.28) in the form:

$$\{x\} = \{X\} \sin(\omega t - \varphi) \quad (3.30)$$

By introducing Equation (3.30) into Equation (3.28), one obtains:

$$-\omega^2 [M] \{X\} \sin(\omega t - \varphi) + [K] \{X\} \sin(\omega t - \varphi) = \{0\} \quad (3.31)$$

that is

$$([K] - \omega^2 [M]) \{X\} = \{0\} \quad (3.32)$$

Equation (3.22) represents an eigenvalue problem, since the linear algebraic equation system is homogenous and the trivial solution lacks a physical meaning.

The determinant of the term in round brackets must be then equal to zero:

$$\text{Det}([K] - \omega^2 [M]) = 0 \quad (3.33)$$

The n -order polynomial equation in the unknown ω^2 which arises from condition (3.33) represents the characteristic equation of the elastic system. To each eigenvalue $\omega_i^2, i = 1, 2, \dots, n,$ corresponds an eigenvector $\{X_i\}$, but for a multiplicative constant.

It can thus be deduced that a vibrating system with n degrees of freedom has n eigen-

frequencies, as well as n mode shapes. Each vibrating mode shape corresponds to a different eigen-frequency. The lowest frequency is named as fundamental frequency. The related mode shape is called fundamental mode.

Equation (3.33) will remain the fundamental condition for the analysis, although matrices $[M]$ and $[K]$ will be no longer diagonal or tri-diagonal, but will be more complex. This complexity reflects the real and effective connection between vibrating masses, which are generally not in series as in Figure 3.5.

In particular we will see how protein structure will be composed by masses and springs in 3D space, without any repeating order.

3.2 Dynamic instability: Modal analysis with second-order effects

In previous section, we have seen how to calculate the natural frequencies of a linear elastic structure according to modal analysis. On the other hand, from Structural Mechanics we know how to obtain the buckling loads of a structure according to a linearized approach to the elastic stability analysis (*Bažant et al. 2003*). The smallest one among these loads (or load multipliers) corresponds to the critical load, associated to the condition of neutral equilibrium that separates the region of stability (load smaller than the critical load) from that of instability (load greater than the critical load).

From a mathematical point of view, both modal analysis and buckling analysis

are seen as eigenvalue problems: they both lead to two formally identical equations written in symbolic form, each one having, of course, different physical meanings. In this Chapter we will focus our analysis on structures subjected to loads that may cause static instability. Thus, taking into account the effect of geometric nonlinearity in the equations of motion through the geometric stiffness matrix, the problem can be reduced to a generalized eigenproblem where both the load multiplier and the natural frequency of the system are unknown. According to this approach, all the configurations intermediate between those of pure buckling and pure free vibrations can be investigated (Carpinteri and Paggi, 2013).

3.2.1 Influence of the Load on the Natural Frequency

In engineering applications there are a lot of situations where structures that undergo flexural vibrations are also loaded by a static axial load. Very common examples are structural members as columns, struts, and towers.

In the case of a slender elastic structure, the applied static load, even if it does not lead to instability, has an influence on the natural vibration frequencies of the structure. In other words, resonance frequencies result modified, in the sense that they are no longer those of the unloaded structure: in general they depend also on the kind and magnitude of the applied load. Consequence of this is, for example, that an external harmonic excitation to produce resonance should have a frequency that matches one of the natural frequencies of the loaded structure, and not of the unloaded one.

This fact requires a very careful analysis of the natural vibration frequencies, which must take into account the second order effects induced by the applied loads, in line

with the modal analysis with second-order effects.

In general, as we will better see later on, there are more complicated cases in which a load increases certain frequencies while others are reduced.

The described influence of the load on the natural frequencies appears in general, not only in the case of slender structures, but in this last case it has a larger influence, while in the other cases it is normally negligible because no important second order effects take place.

As we will see from the end of this Chapter and later on in Chapter 5, proteins will be represented as structures made only by masses and beams with their own stiffness, so calculation of natural frequencies and influence of external load on them is of enormous importance.

3.2.2 Discrete Mechanical Systems with One Degree of Freedom

Let us consider the mechanical system shown in Figure 3.6, consisting of two rigid rods connected by an elastic hinge of rotational rigidity k and constrained at one end by a pinned support and at the other by a roller support. A mass m is placed in correspondence of the intermediate elastic hinge and the system is loaded by a horizontal axial force N . Considering the absolute rotation φ of the two arms as the generalized coordinate, the total potential energy, W , and the kinetic energy, T , of the whole system are:

$$\begin{aligned}
 W(\varphi) &= \frac{1}{2}k(2\varphi)^2 - 2Nl(1 - \cos\varphi), \\
 T(\dot{\varphi}) &= \frac{1}{2}m\left[\frac{d}{dt}(l\sin\varphi)\right]^2 + \frac{1}{2}m\left[\frac{d}{dt}(l - l\cos\varphi)\right]^2 = \frac{1}{2}ml^2\dot{\varphi}^2.
 \end{aligned}
 \tag{3.34}$$

The equation of motion can be determined by writing Lagrange's equation:

$$\frac{\partial}{\partial t} \left(\frac{\partial T}{\partial \dot{\varphi}} \right) - \frac{\partial T}{\partial \varphi} = - \frac{\partial W}{\partial \varphi}. \quad (3.35)$$

In the present case, this yields:

$$ml^2 \ddot{\varphi} = -4k\varphi + 2Nl \sin \varphi, \quad (3.36)$$

which can be suitably linearized in correspondence of $\varphi = 0$:

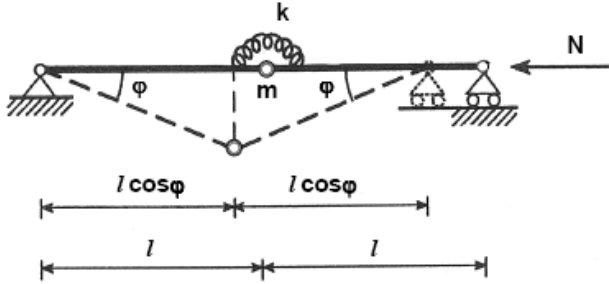


Figure 3.6: Scheme of the first one-degree of freedom system analyzed.

$$ml^2 \ddot{\varphi} = -4k\varphi + 2Nl\varphi. \quad (3.37)$$

Looking for the solution to Equation (3.37) in the general form $\varphi = \varphi_0 e^{i\omega t}$, where ω denotes the natural angular frequency of the system, we obtain the following equation which provides the conditions of equilibrium of the system:

$$\left(4k - 2Nl - \omega^2 ml^2 \right) \varphi_0 = 0. \quad (3.38)$$

A nontrivial solution to Equation (3.38) exists if and only if the term in brackets is

equal to zero. This critical condition corresponding to the bifurcation of the equilibrium establishes a one-to-one relationship between the applied axial force, N , and the angular frequency, ω :

$$N = \frac{2k}{l} - \frac{ml}{2}\omega^2. \quad (3.39)$$

Moreover, Equation (3.39) admits two important limit conditions for (respectively) $N=0$ and $m=0$. In the former case, Equation (3.39) gives the natural angular frequency of the system according to pure modal analysis, i.e. $\omega_1 = \sqrt{4k/ml^2}$. In the latter, the pure critical Eulerian load is obtained, i.e. $N_1 = 2k/l$.

Dividing Equation (3.39) by N_1 , we obtain the following relationship between N and ω in a nondimensional form:

$$\left(\frac{\omega}{\omega_1}\right)^2 + \left(\frac{N}{N_1}\right) = 1. \quad (3.40)$$

As a second example, let us consider the mechanical system shown in Figure 3.7, consisting of two rigid rods on three supports, of which the intermediate one is assumed to be elastically compliant with rigidity k .

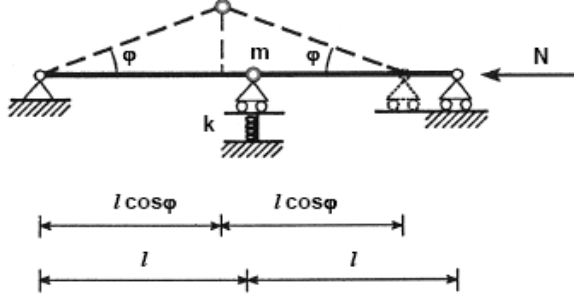


Figure 3.7: Scheme of the second one-degree of freedom system analyzed.

As in the previous case, a mass m is placed in correspondence of the intermediate hinge and the system is loaded by a horizontal axial force N . Considering the absolute rotation φ of the two arms as the generalized coordinate, the total potential energy, W , and the kinetic energy, T , of the whole system are:

$$\begin{aligned} W(\varphi) &= \frac{1}{2}k(l \sin \varphi)^2 - 2Nl(1 - \cos \varphi), \\ T(\dot{\varphi}) &= \frac{1}{2}ml^2\dot{\varphi}^2. \end{aligned} \quad (3.41)$$

Following the procedure discussed above, we determine the equation of motion by employing Lagrange's equation (3.35):

$$ml^2\ddot{\varphi} = -l \sin \varphi (kl \cos \varphi - 2N), \quad (3.42)$$

which can be suitably linearized in correspondence of $\varphi = 0$:

$$ml^2\ddot{\varphi} = -l\varphi(kl - 2N). \quad (3.43)$$

Looking for the solution to Equation (3.43) in the general form $\varphi = \varphi_0 e^{i\omega t}$, where

ω denotes the natural angular frequency of the system, we obtain the following condition of equilibrium of the system:

$$(kl^2 - 2Nl - \omega^2 ml^2)\varphi_0 = 0. \quad (3.44)$$

As in the previous example, by setting the term in brackets equal to zero, we obtain a one-to-one relationship between the applied axial force, N , and the angular frequency, ω :

$$N = \frac{kl}{2} - \frac{ml}{2}\omega^2. \quad (3.45)$$

This equation admits two important limit conditions for (respectively) $N = 0$ and $m = 0$. In the former case, Equation (3.45) gives the natural angular frequency of the system according to pure modal analysis, i.e. $\omega_1 = \sqrt{k/m}$. In the latter, the pure critical Eulerian load for buckling instability is obtained, i.e. $N_1 = kl/2$. Dividing Equation (3.45) by N_1 , we obtain the same relationship between the nondimensional terms N/N_1 and $(\omega/\omega_1)^2$ as in the previous example (see Equation (3.40)).

A graphical representation of the condition (3.40) in Figure 3.8 shows that the resonance frequency is a decreasing function of the compressive axial load.

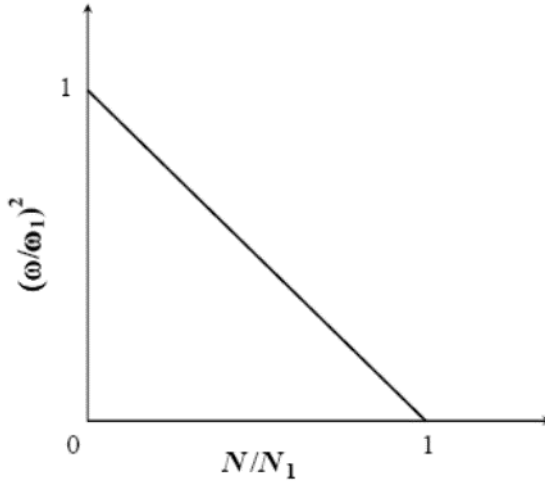


Figure 3.8: Nondimensional frequency squared vs. nondimensional axial force for the single degree of freedom systems analyzed.

This demonstrates, for the analyzed mechanical systems with a single degree of freedom, that resonance can take place for $\omega < \omega_1$, provided that the system is loaded by an axial force N given by Equation (3.40).

In addition, for the static case, the issue of stability or instability of the mechanical system in the correspondence of the bifurcation point, can be discussed by evaluating the higher order derivatives of the total potential energy W (Carpinteri, 1997).

3.2.3 Discrete Mechanical Systems with Two Degrees of Freedom

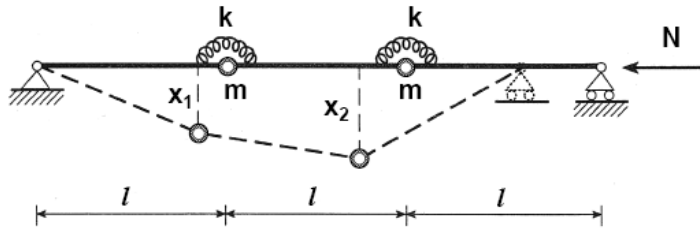


Figure 3.9: Scheme of the first two-degrees of freedom system analyzed.

Let us now consider the mechanical system with two degrees of freedom shown in Figure 3.9, consisting of three rigid rods connected by two elastic hinges of rotational rigidity k , and constrained at one end by a pinned support and at the other by a roller support. A mass m is placed in correspondence of the intermediate elastic hinges and the system is loaded by a horizontal axial force N . Assuming the vertical displacements x_1 and x_2 of the elastic hinges as the generalized coordinates, the total potential energy, W , and the kinetic energy, T , of the whole system are given by:

$$\begin{aligned}
W(x_1, x_2) &= \frac{1}{2}k \left[\left(\arcsin \frac{x_1}{l} - \arcsin \frac{x_2 - x_1}{l} \right)^2 \right. \\
&\quad \left. + \left(\arcsin \frac{x_2}{l} + \arcsin \frac{x_2 - x_1}{l} \right)^2 \right] \\
&\quad - Nl \left[3 - \cos \left(\arcsin \frac{x_1}{l} \right) - \cos \left(\arcsin \frac{x_2}{l} \right) \right. \\
&\quad \left. - \cos \left(\arcsin \frac{x_2 - x_1}{l} \right) \right], \\
T(\dot{x}_1, \dot{x}_2) &= \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_1^2 x_1^2 + \frac{1}{2}m\dot{x}_2^2 \\
&\quad + \frac{1}{2}m \left(\frac{2x_1\dot{x}_1}{l} + \frac{x_2\dot{x}_2}{l} - \frac{x_2\dot{x}_1}{l} - \frac{x_1\dot{x}_2}{l} \right)^2.
\end{aligned} \tag{3.46}$$

Performing a Taylor series expansion of Equation (3.46) about the origin, and assuming $x_1/l < 1/10$ and $x_2/l < 1/10$, we obtain:

$$\begin{aligned}
W(x_1, x_2) &\cong \frac{k}{2l^2} (5x_1^2 + 5x_2^2 - 8x_1x_2) - \frac{N}{l} (x_1^2 + x_2^2 - x_1x_2), \\
T(\dot{x}_1, \dot{x}_2) &\cong \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_2^2.
\end{aligned} \tag{3.47}$$

The equations of motion are identified by considering Lagrange's equations:

$$\frac{\partial}{\partial t} \left(\frac{\partial T}{\partial \dot{x}_i} \right) - \frac{\partial T}{\partial x_i} = - \frac{\partial W}{\partial x_i}, \quad i=1,2. \tag{3.48}$$

In matrix form, they are:

$$\begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{Bmatrix} + \begin{bmatrix} \frac{5k}{l^2} & -\frac{4k}{l^2} \\ -\frac{4k}{l^2} & \frac{5k}{l^2} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} - N \begin{bmatrix} \frac{2}{l} & -\frac{1}{l} \\ -\frac{1}{l} & \frac{2}{l} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}. \quad (3.49)$$

Looking for the solution to Equation (3.49) in the general form $\{q\} = \{q_0\} e^{i\omega t}$, where ω denotes the natural angular frequency of the system, we obtain the following equation, written in symbolic form:

$$\left(-\omega^2 [M] + [K] - N [K_g]\right) \{q_0\} = \{0\}, \quad (3.50)$$

where $[M]$, $[K]$ and $[K_g]$ denote (respectively) the mass matrix, the elastic stiffness matrix and the geometric stiffness matrix of the mechanical system. Their expressions can be simply obtained by comparing Equation (3.50) with Equation (3.49).

A nontrivial solution to Equation (3.50) exists if and only if the determinant of the resultant coefficient matrix of the vector $\{q_0\}$ vanishes. This yields the following generalized eigenvalue problem:

$$\det\left([K] - N [K_g] - \omega^2 [M]\right) = 0, \quad (3.51)$$

where N and ω^2 represent the eigenvalues. For this example, Equation (3.51) provides the following relationships between the eigenvalues ω^2 and N :

$$\omega^2 = \frac{k}{ml^2} - \frac{N}{ml}, \quad (3.52a)$$

$$\omega^2 = \frac{9k}{ml^2} - \frac{3N}{ml}. \quad (3.52b)$$

As limit cases, if $m = 0$, then we obtain the Eulerian buckling loads:

$$N_1 = \frac{k}{l}, \quad (3.53a)$$

$$N_2 = \frac{3k}{l}, \quad (3.53b)$$

whereas, if $N = 0$, we obtain the natural frequencies of the system:

$$\omega_1 = \sqrt{\frac{k}{ml^2}}, \quad (3.54a)$$

$$\omega_2 = \sqrt{\frac{9k}{ml^2}}. \quad (3.54b)$$

As far as the eigenvectors are concerned, the system (3.50) yields the eigenvectors corresponding (respectively) to the eigenfrequencies (3.52a) and (3.52b) as functions of N :

$$x_1 = \frac{4k/l - N}{6k/l - 3N} x_2, \quad (3.55a)$$

$$x_1 = \frac{4k/l - N}{14k/l - 5N} x_2. \quad (3.55b)$$

Dividing Equations (3.52a) and (3.52b) by ω_1^2 , we derive the following nondimensional relationships between the eigenvalues:

$$\left(\frac{\omega}{\omega_1}\right)^2 = 1 - \left(\frac{N}{N_1}\right), \quad (3.56a)$$

$$\left(\frac{\omega}{\omega_1}\right)^2 = \left(\frac{\omega_2}{\omega_1}\right)^2 - \frac{N_2}{N_1} \left(\frac{N}{N_1}\right). \quad (3.56b)$$

In analogy with the results for the single degree of freedom systems, a graphical representation of Equations (3.56a) and (3.56b) is provided in Figure 3.10.

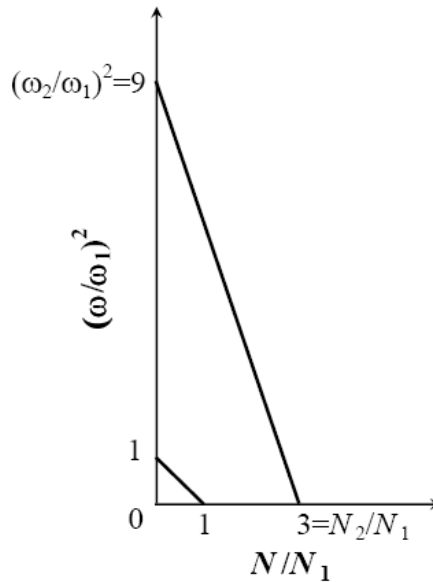


Figure 3.10: Nondimensional frequencies squared vs. nondimensional axial forces for the two-degrees of freedom system in Figure 3.9.

We notice that both the eigenfrequencies are decreasing functions of the

compressive axial load. Entering the diagram with a value of the nondimensional compressive axial force in the range $0 < N/N_1 < 1$, the coordinates of the points of the two curves provide the two modified resonance frequencies of the mechanical system. Axial forces larger than N_1 in the range $1 < N/N_1 < N_2/N_1$ can only be experienced if an additional constraint is introduced into the system in order to prevent the vertical displacement of the midpoint, allowing at the same time the rotation and the horizontal displacement. Moreover, we observe that the applied compressive load influences all the eigenfrequencies, and not just the first one. In particular, in the present example, the influence of the axial load is greater on the highest frequency than on the lower one.

As a second example of a system with two degrees of freedom, let us examine that of Figure 3.11, which consists of three rigid rods on four supports, of which the central ones are assumed to be elastically compliant with rigidity k . A mass m is placed in correspondence of the intermediate hinges and the system is loaded by a horizontal axial force N .

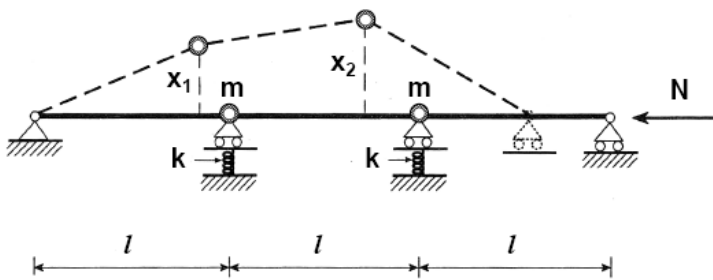


Figure 3.11: Scheme of the second two-degrees of freedom system analyzed.

Assuming the vertical displacements x_1 and x_2 of the elastic hinges as the

generalized coordinates, the total potential energy, W , and the kinetic energy, T , of the whole system are given by ($x_1/l < 1/10$ and $x_2/l < 1/10$):

$$\begin{aligned}
 W(x_1, x_2) &= \frac{1}{2}k(x_1^2 + x_2^2) - Nl \left[3 - \cos\left(\arcsin \frac{x_1}{l}\right) \right. \\
 &\quad \left. - \cos\left(\arcsin \frac{x_2}{l}\right) - \cos\left(\arcsin \frac{x_2 - x_1}{l}\right) \right] \\
 &\cong \frac{1}{2}k(x_1^2 + x_2^2) - \frac{N}{l}(x_1^2 + x_2^2 - x_1 x_2), \\
 T(\dot{x}_1, \dot{x}_2) &\cong \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_2^2.
 \end{aligned} \tag{3.57}$$

In this case, Lagrange's equations (3.48) yield the following matrix form:

$$\begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{Bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{Bmatrix} + \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} - N \begin{bmatrix} \frac{2}{l} & -\frac{1}{l} \\ -\frac{1}{l} & \frac{2}{l} \end{bmatrix} \begin{Bmatrix} x_1 \\ x_2 \end{Bmatrix} = \begin{Bmatrix} 0 \\ 0 \end{Bmatrix}. \tag{3.58}$$

Looking for the solution to Equation (3.58) in the general form $\{q\} = \{q_0\}e^{i\omega t}$, where ω denotes the natural angular frequency of the system, we obtain the following equation, written in symbolic form:

$$\left(-\omega^2 [M] + [K] - N [K_g] \right) \{q_0\} = \{0\}, \tag{3.59}$$

where $[M]$, $[K]$ and $[K_g]$ denote (respectively) the mass matrix, the elastic stiffness matrix and the geometric stiffness matrix of the mechanical system. As it can be readily seen, the geometric stiffness matrix for this problem is the same as that of the previous example.

A nontrivial solution to Equation (3.59) exists if and only if the determinant of the resultant coefficient matrix of the vector $\{q_0\}$ is equal to zero. This yields the following generalized eigenvalue problem:

$$\det\left([K] - N[K_g] - \omega^2[M]\right) = 0, \quad (3.60)$$

where N and ω^2 are the eigenvalues of the system. For this example, Equation (3.60) provides the following relationships between the eigenvalues:

$$\omega^2 = \frac{k}{m} - 3\frac{N}{ml}, \quad (3.61a)$$

$$\omega^2 = \frac{k}{m} - \frac{N}{ml}. \quad (3.62b)$$

As limit cases, if $m = 0$, we obtain the Eulerian buckling loads:

$$N_1 = \frac{1}{3}kl, \quad (3.63a)$$

$$N_2 = kl, \quad (3.63b)$$

whereas, if $N = 0$, then we obtain the natural frequencies of the system:

$$\omega_1 = \omega_2 = \sqrt{\frac{k}{m}}. \quad (3.64)$$

As far as the eigenvectors are concerned, the system (3.59) yields the eigenvectors corresponding (respectively) to the eigenfrequencies (3.61a) and (3.61b), as functions of the axial force, N :

$$x_1 = \frac{N/l}{5N/l - 2k} x_2, \quad (3.65a)$$

$$x_1 = \frac{N/l}{3N/l - 2k} x_2. \quad (3.65b)$$

Dividing Equations (3.61a) and (3.61b) by ω_1^2 , we derive the following nondimensional relationships between the eigenvalues:

$$\left(\frac{\omega}{\omega_1}\right)^2 = 1 - \left(\frac{N}{N_1}\right), \quad (3.66a)$$

$$\left(\frac{\omega}{\omega_1}\right)^2 = 1 - \frac{N_1}{N_2} \left(\frac{N}{N_1}\right). \quad (3.66b)$$

A graphical representation of Equations (3.66a) and (3.66b) is provided in Figure 3.12.

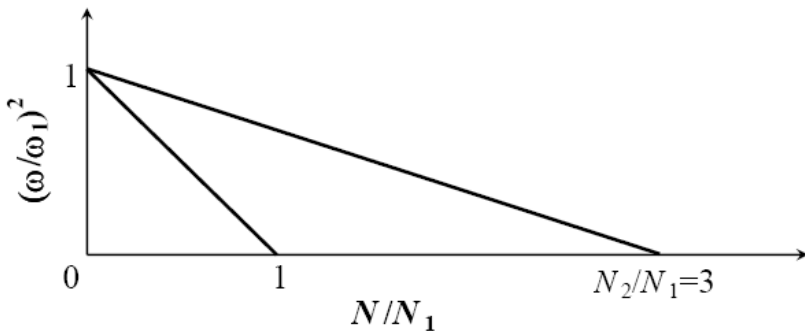


Figure 3.12: Nondimensional frequencies squared vs. nondimensional axial forces for the two-degrees of freedom system in Figure 3.11.

Also in this case, both the frequencies are decreasing functions of the compressive axial load. However, in the present example, the influence of the axial load is greater on the lower frequency of the system than on the higher one.

3.3 Elementary models of proteins

The first step of our research has been to model proteins and interatomic bonds in a mechanical form, the easiest possible, to try to obtain, with biological parameters derived by literature, an idea of the order of magnitude of their natural frequencies. Because the central topic of our research is Na-K ATPase, we used this protein to start.

Biological parameters of this protein were presented in previous Chapter, here we report raw data.

Na/K-ATPase properties:

- large protein
- molecular weight = 110 kDa (1.78×10^{-22} kg)
- average diameter ≈ 4 nm; length ≈ 16 nm
- 1028 amino acids (10131 atoms)

As evident from Figure 2.9 the shape of this protein is a sort of Y, but it's clear that one dimension is bigger than the others, so, for primary evaluations, we modelled entire protein with an elastic beam.

In the first model, figure 3.13, the beam is mass-less and the entire mass of the protein is concentrated at the end.

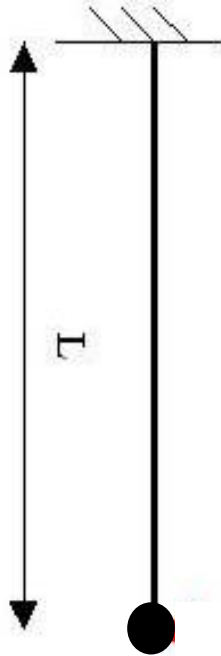


Fig 3.13: Longitudinal complete protein vibrations: concentrated mass where protein is shown schematically with an elastic beam with no mass, the entire mass is concentrated at one end.

For this mechanical scheme, that is the same one of Figure 3.1, longitudinal natural vibrations can be easily calculated by means of the well-known equation (see also Eq 3.4)

$$f = \frac{\omega}{2\pi} = \frac{\sqrt{\frac{k}{m}}}{2\pi} \quad (3.67)$$

where mass $m = 1.78 \times 10^{-22}$ kg, and the stiffness, from literature, $k = 200$ N/m (Lagace 2007).

For parameters mentioned above, natural frequency of the simplified scheme in figure 3.13, is equal to $f = 0.11 \times 10^{12}$ Hz = 0.11THz = 110 GHz.

This result, obtained with large simplifications of the real structure, is however in agreement with theory proposed in Chapter 1, section 1.2, where we established that the order of magnitude of natural frequencies for big proteins, like Na-K ATPase is, would be 10^{11} Hz.

The same protein can easily be schematized like an elastic beam with the mass of the protein distributed over the entire length, like in figure 3.14.

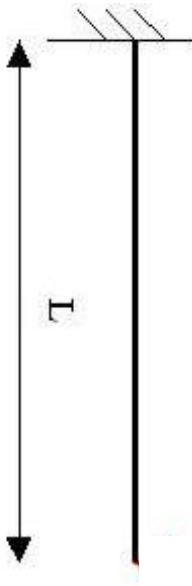


Fig 3.14: Longitudinal complete protein vibrations: protein is shown schematically with an elastic beam with the entire mass distributed over the length

In this case, to calculate the natural frequency of the structure must be used the formula

$$f = \frac{\omega}{2\pi} = \frac{1}{4L} \sqrt{\frac{EA}{\mu}} \quad (3.68)$$

where $L = 15.6 \times 10^{-9}$ m, cross sectional area $A = (\pi D^2 / 4) = 5.03 \times 10^{-17}$ m², linear density $\mu = M / L = 0.11 \times 10^{-13}$ kg/m, $E = (k \times L) / A = 3,10 \times 10^9$ N/m².

With these values, natural frequency of the simplified scheme in figure 3.14, is equal to $f = 0.27 \times 10^{12}$ Hz = 0.27THz = 270 GHz.

Also this value is of the same order of magnitude of the one obtained with concentrated mass and proposed in section 1.2.

We schematized not only entire protein, but we also modeled intermolecular bond using scheme in figure 3.15.

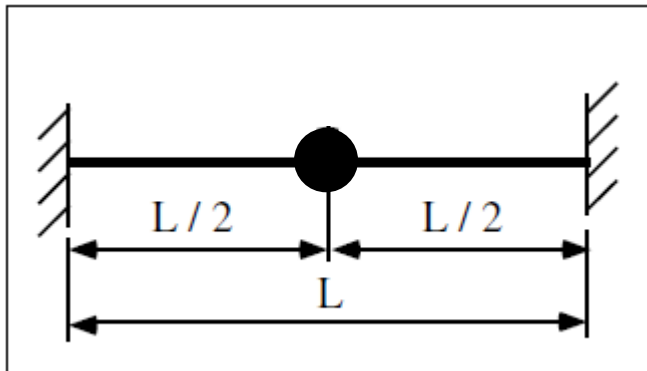


Fig 3.15: Natural vibrations at interatomic bond level.

For this type of model, natural frequency is obtained using equation 3.67, but in this case, parameters are different, because we have to consider not the entire protein but the average mass of a single atom and the average length of a bond:

$m = 1.76 \times 10^{-26}$ kg, $L = 0.2$ nm.

Natural frequency obtained is $f = 16.9 \times 10^{12}$ Hz = 16.9 THz, confirming hypothesis of section 1.2 for vibrations at bond level.

Chapter 4

Raman spectroscopy: Experimental campaigns on Lysozyme and Na-K ATPase

This chapter describes the experimental investigations carried out with the Raman technique on Lysozyme and Na-K ATPase.

In the first part of the chapter, after a brief historical overview, we should describe the basics of the technique giving also two different interpretations, the classical one and the quantum mechanics one.

In the second part of the chapter we will describe in detail the experimental sessions and analyse the results.

4.1 Elements of Raman Spectroscopy

Raman Spectroscopy provides information about molecular vibrations. This non-destructive technique consists in irradiating a sample with a light source (usually a laser) and collecting the scattered light (Figure 1) (*Raman 1928, Raman et al. 1928*). Most of the latter will be at the same frequency of the incident light (the so-called Rayleigh peak); a small percentage (10^{-5} %) will be shifted in frequency from the laser due to the light-molecule interaction. Plotting the intensity of this scattered light versus the frequency yields a Raman spectrum. Generally, the Rayleigh peak is located at 0 cm^{-1} , whereas other peaks will arrange themselves on the spectrum at frequencies that refer to the energy levels of different specific functional groups that compose the sample.

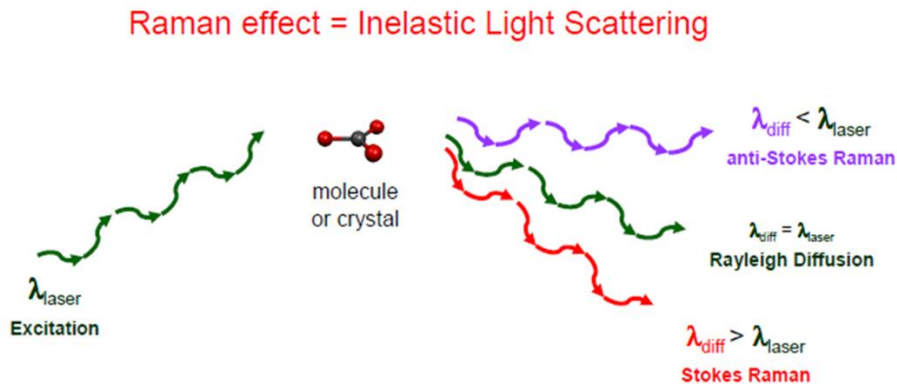


Figure 1: Scheme of Raman spectroscopy technique

Raman spectroscopy can be used both for identification and quantification purposes, as the spectra obtained are highly specific, real fingerprints for the elements. The identification of samples is usually performed by comparing

the spectra with numerous databases now available. Advantages of the Raman technique include requiring a limited sample preparation, and allowing the analysis of both solid samples and aqueous solutions (*Frezzotti et al.2012*).

In 1928, when Sir Chandrasekhra Venkata Raman discovered the phenomenon that bears his name, only crude instrumentation was available. Sir Raman used sunlight as the source and a telescope as the collector; the detector was his eyes. That such a feeble phenomenon as the Raman scattering was detected was indeed remarkable.

Gradually, improvements in the various components of Raman instrumentation took place. Early research was concentrated on the development of better excitation sources. Various lamps of elements were developed (e.g.,helium, bismuth, lead, zinc) (*Kerschbaum 1914, Veskatesachar et al. 1930, Hibben 1939*). These proved to be unsatisfactory because of low light intensities. Mercury sources were also developed. An early mercury lamp which had been used for other purposes in 1914 by Kerschbaum (*Kerschbaum 1914*) was developed. In the 1930s mercury lamps suitable for Raman use were designed. Hibben developed a mercury burner in 1939, and Spedding and Stamm (*Spedding et al. 1942*) experimented with a cooled version in 1942. Further progress was made by Rank and McCartney (*Rank et al. 1948*), who studied mercury burners and their backgrounds. Hilger Co. developed a commercial mercury excitation source system for the Raman instrument, which consisted of four lamps surrounding the Raman tube. Welsh et al. (*Welsh et al. 1952*) introduced a mercury source in 1952, which became known as the Toronto Arc. The lamp consisted of a four-turn helix of Pyrex tubing and was an improvement over the Hilger lamp. Improvements in lamps were made by Ham and Walsh (*Ham et al. 1958*),

who described the use of microwave-powered helium, mercury, sodium, rubidium and potassium lamps. Stammreich (*Stammreich 1950a, 1950b, 1956a, 1956b*) also examined the practicality of using helium, argon, rubidium and cesium lamps for colored materials. In 1962 laser sources were developed for use with Raman spectroscopy (*Gilson et al. 1970*). Eventually, the Ar⁺ (351.1-514.5nm) and the Kr⁺ (337.4-676.4 nm) lasers became available, and more recently the Nd-YAG laser (1,064 nm) has been used for Raman spectroscopy.

Progress occurred in the detection systems for Raman measurements.

Whereas original measurements were made using photographic plates with the cumbersome development of photographic plates, photoelectric Raman instrumentation was developed after World War II.

The first photoelectric Raman instrument was reported in 1942 by Rank and Wiegand (*Rank et al. 1942*), who used a cooled cascade type RCA IP21 detector. The Heigl instrument appeared in 1950 and used a cooled RCA C-7073B photomultiplier. In 1953 Stamm and Salzman (*Stamm et al. 1953*) reported the development of photoelectric Raman instrumentation using a cooled RCA IP21 photomultiplier tube. The Hilger E612 instrument (*Ferraro 1967*) was also produced at this time, which could be used as a photographic or photoelectric instrument. In the photoelectric mode a photomultiplier was used as the detector. This was followed by the introduction of the Cary Model 81 Raman spectrometer (*Ferraro 1992*). The source used was the 3 kW helical Hg arc of the Toronto type. The instrument employed a twin-grating, twin-slit double monochromator.

Developments in the optical train of Raman instrumentation took place in the early 1960s. It was discovered that a double monochromator removed stray

light more efficiently than a single monochromator. Later, a triple monochromator was introduced, which was even more efficient in removing stray light.

Holographic gratings appeared in 1968, which added to the efficiency of the collection of Raman scattering in commercial Raman instruments.

These developments in Raman instrumentation brought commercial Raman instruments to the present state of the art of Raman measurements.

Now, Raman spectra can also be obtained by Fourier transform (FT) spectroscopy.

FT-Raman instruments are being sold by all Fourier transform infrared (FT-IR) instrument makers, either as interfaced units to the FT-IR spectrometer or as dedicated FT-Raman instruments.

4.2 Classical Mechanics Interpretation

Light, in the classic wave theory, is nothing but an electromagnetic radiation having an oscillating electric field capable of interacting with a molecule through its polarizability, i.e., the ability of electrons to interact with the electric field.

Figure 4.1 illustrates a wave of polarized electromagnetic radiation traveling in the z-direction. It consists of the electric component (x-direction) and magnetic component (y-direction), which are perpendicular to each other.

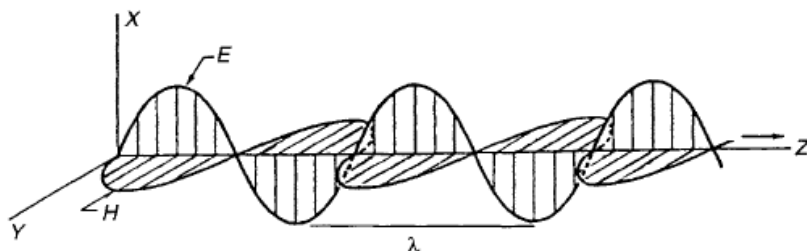


Figure 4.1: Plane polarized electromagnetic radiation

The electric field strength (E) at a given time (t) is expressed by

$$E = E_0 \cos 2\pi\nu t, \quad (4-1)$$

where E_0 is the amplitude and ν is the frequency of radiation as defined later. The distance between two points of the same phase in successive waves is called the "wavelength," λ , which is measured in units such as \AA (angstrom), nm (nanometer), and cm (centimeter). The relationships between these units are:

$$1 \text{ \AA} = 10^{-8} \text{ cm} = 10^{-1} \text{ nm} \quad (4-2)$$

Thus, for example, $4,000 \text{ \AA} = 400 \text{ nm}$.

The frequency, ν , is the number of waves in the distance light travels in one second. Thus,

$$\nu = \frac{c}{\lambda} \quad (4-3)$$

where c is the velocity of light ($3 \times 10^{10} \text{ cm/s}$). If λ is in the unit of centimeters, its dimension is $(\text{cm/s})/(\text{cm}) = 1/\text{s}$. This "reciprocal second" unit is also called the "hertz" (Hz).

The third parameter, which is most common to vibrational spectroscopy, is

the "wavenumber," $\tilde{\nu}$, defined by

$$\tilde{\nu} = \frac{\nu}{C} \quad (4-4)$$

The difference between ν and $\tilde{\nu}$ is obvious. It has the dimension of (1/s)/(cm/s) = 1/cm.

By combining (4-3) and (4-4) we have

$$\tilde{\nu} = \frac{\nu}{C} = \frac{1}{\lambda} \text{ (cm}^{-1}\text{)} \quad (4-5)$$

By combining (4-3) and (4-4), we obtain

$$\nu = \frac{c}{\lambda} = C\tilde{\nu} \quad (4-6)$$

As shown earlier, the wavenumber ($\tilde{\nu}$) and frequency (ν) are different parameters, yet these two terms are often used interchangeably. Thus, an expression such as "frequency shift of 50cm⁻¹" is used conventionally by Raman spectroscopists and we will follow this convention through this thesis.

To explain Raman effect in a classical way, let us consider a simple diatomic molecule idealized as a mass-spring linear system, wherein m_1 and m_2 represent the two masses of the molecule, x_1 and x_2 their displacements (in opposite directions), and k the elastic constant of the linear spring simulating the bonding force (Figure 4.2) (*Ferraro et al. 2003*).

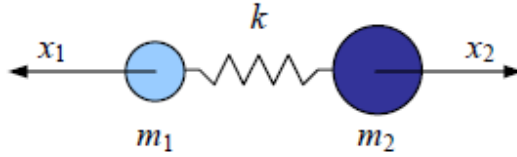


Figure 4.2: A diatomic molecule idealized as a mass-spring linear system

Reducing the system to an equivalent macromolecule and combining Hooke's law and Newton's second law of motion (or, alternatively, D'Alembert's principle), yields the equation of motion:

$$\frac{m_1 m_2}{m_1 + m_2} \left(\frac{d^2 x_1}{dt^2} + \frac{d^2 x_2}{dt^2} \right) + k(x_1 + x_2) = 0, \quad (4.7)$$

where t is time. Replacing the reduced mass $(m_1 m_2 / m_1 + m_2)$ with μ and the total (relative) displacement $(x_1 + x_2)$ with x , Eq. (4.7) can be rewritten as follows:

$$\mu \frac{d^2 x}{dt^2} + kx = 0. \quad (4.8)$$

Eq. (4.8) is the well-known equation of the harmonic oscillator, with general solution given by (for $\dot{x}(0) = 0$, but without loss of generality):

$$x = x_0 \cos(2\pi f_m t), \quad (4.9)$$

where x_0 is the vibration amplitude and f_m is the molecule vibration frequency (Hz), given by:

$$f_m = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}}, \quad (4.10)$$

where $\sqrt{k/\mu} = \omega_m$ is the angular frequency of vibration of the molecule (rad s^{-1}).

It is therefore clear that each molecule has its own characteristic (or natural) vibration frequency. In general, multi-atomic molecules have different frequencies, generated by the combination of atomic masses and characteristics of the interatomic bonds. These molecular vibrations are measurable through Raman spectroscopy since the polarizability of a molecule, α , is a function of the shift (displacement) x .

Returning to our elementary example, when light hits the diatomic molecule of Fig. 2.2, a dipole moment P arises. It is a function of polarizability and electric field, and is defined as follows:

$$P = \alpha E_0 \cos(2\pi f_0 t), \quad (4.11)$$

where E_0 and f_0 are respectively the intensity and the frequency of the electric field. The polarizability can be approximated, for small displacements, by the following linearized expression:

$$\alpha \approx \alpha_0 + x \left(\frac{\partial \alpha}{\partial x} \right)_{x=0}. \quad (4.12)$$

Substituting Eq.s (4.12) and (4.9) into Eq. (4.11), yields:

$$P = \alpha_0 E_0 \cos(2\pi f_0 t) + x_0 \cos(2\pi f_m t) E_0 \cos(2\pi f_0 t) \left(\frac{\partial \alpha}{\partial t} \right)_{x=0}. \quad (4.13)$$

Eq. (4.13) shows that there are two effects that arise from the interaction of a molecule with a source of light: the first component is the so-called Rayleigh radiation, which is the main component and has the same frequency of the incident light (f_0) – and therefore it is not interesting for the analysis –, while the second part is the real Raman component, expressed as a shift in frequency from the Rayleigh radiation and including the molecule eigenfrequency (f_m). Positive frequency shifts are called Stokes contributions, negative shifts are called Anti-Stokes contributions.

4.3 Quantum Mechanics Interpretation

Raman effect can also be explained by the following quantum mechanics interpretation (*Ferraro et al. 2003*). In this case, it is assumed that the effect originates from an inelastic scattering of a photon by a molecular bond. Following the Jablonski diagram (Figure 4.3), we can see that the effect appears like an excitation of the molecule to a virtual energy state by a photon (h is the Plank's constant). When this happens, there are three possible developments. In the first case, the molecule emits a photon at exactly the same energy of the incident photon, therefore returning to the original state; this is an elastic process and is called Rayleigh scattering. In the second case, the molecule relaxes by emitting a photon with an energy lower than the incident one, so the molecule persists in an excited energy state different from

the fundamental; in this case we speak of Stokes scattering. The third and last case occurs when the molecule in origin is not in the fundamental state but already in an excited state; the incident photon leads the molecule to a higher virtual energy state so the consequent relaxation to the ground state results in a photon with energy greater than that of the incident photon. This case is called Anti-Stokes scattering. As a consequence of the fact that many molecules at room temperature are in the ground state, Stokes shifting is much more frequent than Anti-Stokes.

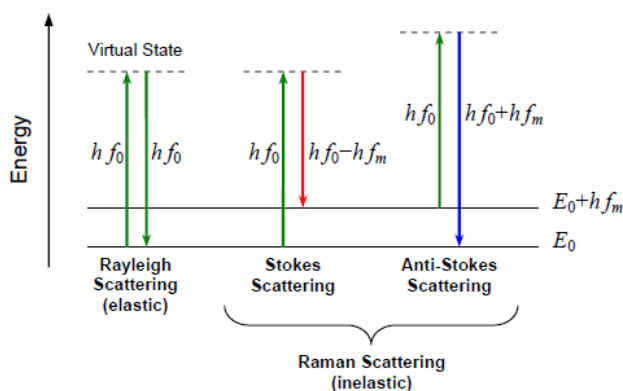


Figure 4.3: Jablonski diagram: explanation of Raman effect

It can be shown that the power of the scattered light, P_s , is directly proportional to the intensity of the incident light I_0 , and inversely proportional to the wavelength λ of the incident light raised to the fourth power:

$$P_s \propto \frac{I_0}{\lambda^4}. \quad (4.14)$$

Consequently, it would be desirable to work with short wavelength light sources and high-power, but unfortunately this is not always possible.

4.4 Use of Raman Spectroscopy for Proteins Study

The use of Raman spectroscopy for proteins study can be dated back to the 70's of the last century. The first studies were designed to investigate the structure of proteins (*Spiro et al. 1977*), assign the bands to the respective chemical groups, and so on. In the 90's the use of Raman spectroscopy was a bit put aside because of limits imposed by technological standards of the time (*Carey 1998*). On the other hand, towards the end of the 90's, the development of the CCD (Charge-Coupled Device) detector, better performing filters and electronics in general, have greatly increased the signal-to-noise ratio obtainable in the Raman spectra (*Carey 1999*). The improvement in sensitivity has also paved the way for the so-called micro-Raman, and the extraction of ligand-binding details from crystalline samples.

Proteins have over 20,000 degrees of freedom, and this results in more than 20,000 vibration modes. In facts, a protein on average is constituted by 500 amino acids (*Netzer et al. 1997*). Each amino acid has an average of 15 atoms; this brings the total number of atoms to about 7500. Since every atom in space has three degrees of freedom, this brings us to the more than 20,000 modes mentioned above for the entire protein. Usually a normal mode involves numerous internal coordinate variations. In other cases, these vibrations

involve only very few atoms, such as stretching of the C=O group, for example. As a consequence, the spectral richness is so high as not to allow the identification of each single mode of vibration due to overlapping of some modes (*Deng et al. 1999*). Therefore, a selection of the most relevant peaks in relation to the objective of the investigation is fundamental in the study of proteins through Raman spectroscopy. For example, information on the secondary structure of the protein can be deduced from the region called Amide I (~50 THz), dominated by the contribution of the backbone C=O. On the other hand, it is also possible to focus on specific areas to follow the changes of groups involved in specific protein reactions (*Callender et al. 1994, Robert et al. 1996*).

4.5 Experimental Campaign on Lysozyme

Raman spectroscopy measurements were performed on hen egg white lysozyme crystallized powder. Lysozyme samples (Worthington Biochemical Corporation, Product Code: LYSF, Lot Number: 34K15116B; 100% protein, two times crystallized) were received and tested in the form of a dialyzed, lyophilized powder, containing $\geq 8,000$ units per mg dry weight. According to the unit definition reported in the certificate of analysis, one unit is equal to a decrease in turbidity of 0.001 A_{450 nm} per minute at 25°C, pH 7.0 using a 0.3 mg/ml suspension of *Micrococcus lysodeikticus* cells (Product Code: ML) as substrate. were purchased from Worthington Biochemical Corporation and stored in a refrigerator at 2-8°C. Samples of lysozyme appear as in Fig. 4.4, where a ten-euro-cents coin is put close to it to give an idea of the dimension. Both the experiments, on

Lysozyme and on Na-K ATPase, were performed in the High pressure spectroscopy (HPS) laboratory at the Physics Department of Sapienza University in Rome, with the technical support of the company Horiba Scientific.

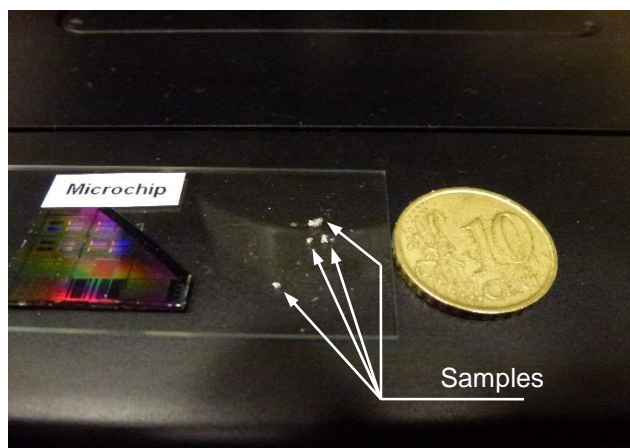


Figure 4.4: Samples of lysozyme crystallized powder

The experimental apparatus consists in a Raman spectrometer LabRAM HR evolution, a last-generation device produced by Horiba Scientific (Fig. 4.5). It mounts a confocal microscope equipped with three objectives (20x, 50x, 100x) (Fig. 4.5), coupled to a 633 nm wavelength He-Ne laser (Fig. 4.6a). The spectrometer disposes of two diffraction gratings, having resolutions equal to 600 lines / mm ($< 3 \text{ cm}^{-1}$) and to 1800 lines / mm ($< 1 \text{ cm}^{-1}$). In order to analyze low frequencies ($< 10\text{-}15 \text{ THz}$), ultra low frequency (ULF) filters of the latest generation were adopted (Fig. 4.6b). The ULF module allows

Raman spectroscopic information in the sub-100 cm^{-1} region, with measurements down to 5 cm^{-1} routinely available.

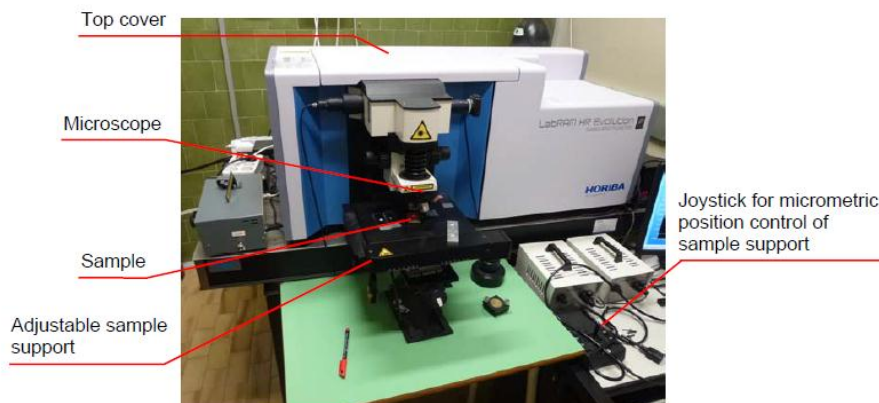


Fig. 4.5 Raman spectrometer HR evolution

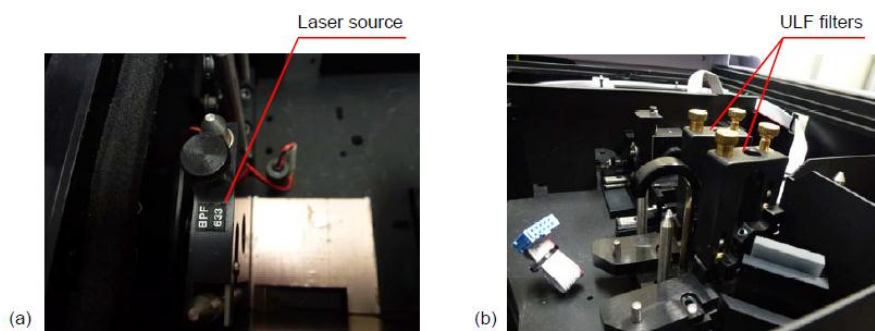


Fig. 4.6a/b Spectrometer components (located under the top cover of Fig. 4.5): (a) red laser source; (b) ULF filters module

This technology claims the highest performances of the single spectrometer: measurements are obtained in just a few seconds or minutes, without any limitation in the higher wavenumber region; Stokes and Anti-Stokes spectral

features can be simultaneously measured, providing additional information (Rapaport *et al.* 2010).

A preliminary calibration of the spectrometer was done prior to the measurements by means of a *in-house* code developed on purpose, and adopting the emission spectrum of a Neon lamp as a reference.

Fig. 4.7 shows the Raman spectrum obtained for lysozyme crystallized powder samples. Data were processed with LabSpec software, supplied with the spectrometer used. Acquisition was made using the 600 lines / mm diffraction grating, 100x objective and 200 μm confocal hole. Four different spectral ranges were analyzed, i.e. grating centered at 650, 1400, 2100 and 2900 cm^{-1} , and then results were collected in a unique spectrum. Each result is the mean of 10 measures with 60 s acquisition time each. Low-degree (≤ 3) polynomial functions were subtracted to each original spectrum in order to eliminate the background (polynomial baseline correction).

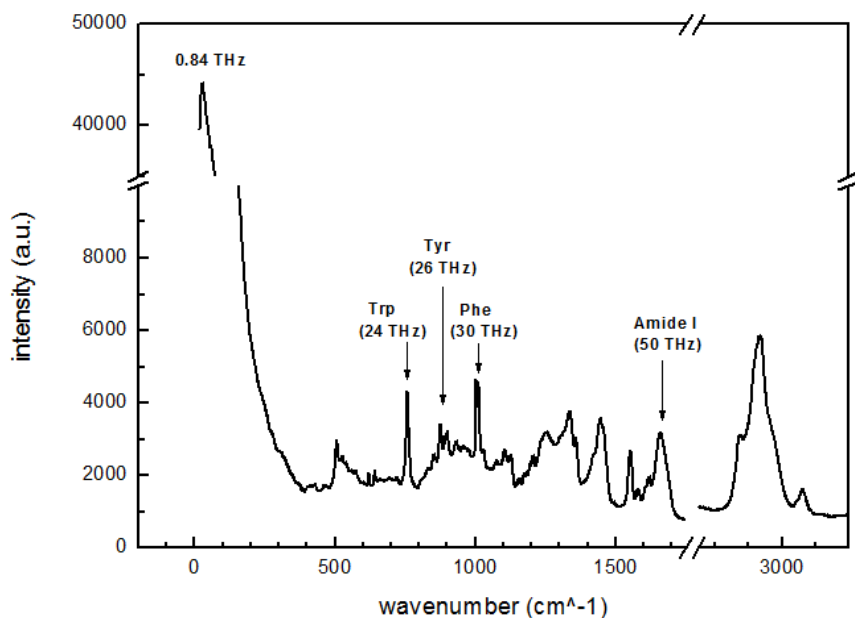


Fig. 4.7 Raman spectrum of crystallized lysozyme powder obtained using ULF filters

The results are in agreement with those of the literature, see for example (*Movasaghi et al. 2007*). Peaks related to three amino acids (tryptophan, tyrosine and phenylalanine) as well as the peak of amide I are highlighted in Fig. 4.7.

Amide I is characteristic of all proteins and is of relevance in biology since it refers to the vibration of the protein backbone. The above-mentioned three amino acids were selected as reference to validate the numerical model (see Chapter 5).

Thanks to the use of ULF filters it has been possible to investigate also low-frequency vibrations. A large peak in correspondence to 0.84 THz is clearly visible in Fig. 4.7 (top-left). The same peak is better highlighted in Figure 4.8.

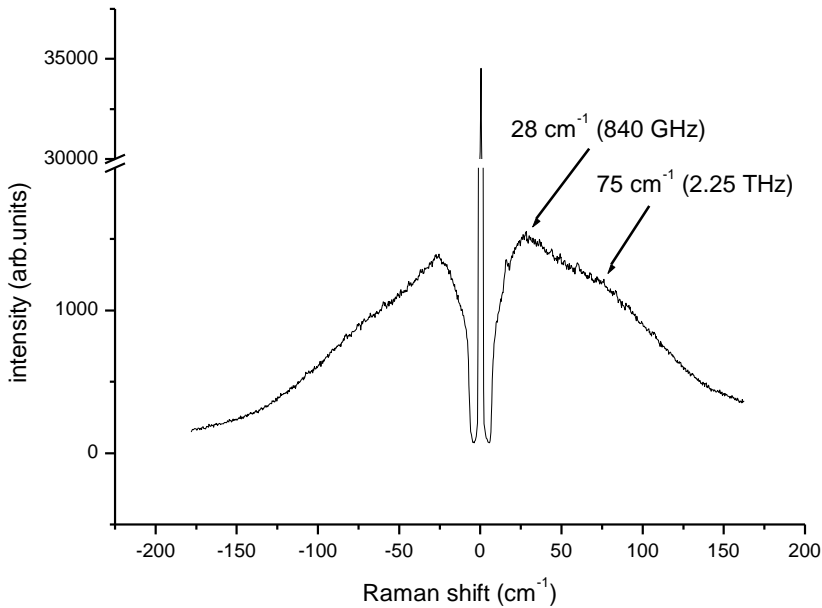


Fig. 4.8 Crystallized lysozyme powder Raman spectrum centered at 0 cm⁻¹

Fig. 4.8 shows the Raman spectrum of lysozyme around the origin.

It is the result of 10 mediated 60 s-acquisitions, each one obtained with the 1800 lines / mm grating, centered at 0 cm⁻¹. The large peak at 28 cm⁻¹ (0.84 THz) and a broad shoulder at 75 cm⁻¹ (2.25 THz) are indicated in the same diagram. Also these peaks at lower frequencies are in agreement with the literature (*Genzel et al. 1976*).

These peaks are not assigned in literature to any chemical group, as they correspond to vibrations involving the entire protein or large portions of it.

Such delocalized or global vibrations were investigated numerically by modeling the entire lysozyme (see Chapter 5).

4.6 Experimental Campaign on Na-K ATPase

Raman spectroscopy measurements were performed on Na-K ATPase lyophilized powder and solution. Na-K ATPase powder were purchased from Sigma-Aldrich Corporation and stored in a refrigerator at 2-8°C.

Samples of Na-K ATPase appear as in Fig. 4.9, where samples both in powder (a) and in solution (b) are presented.

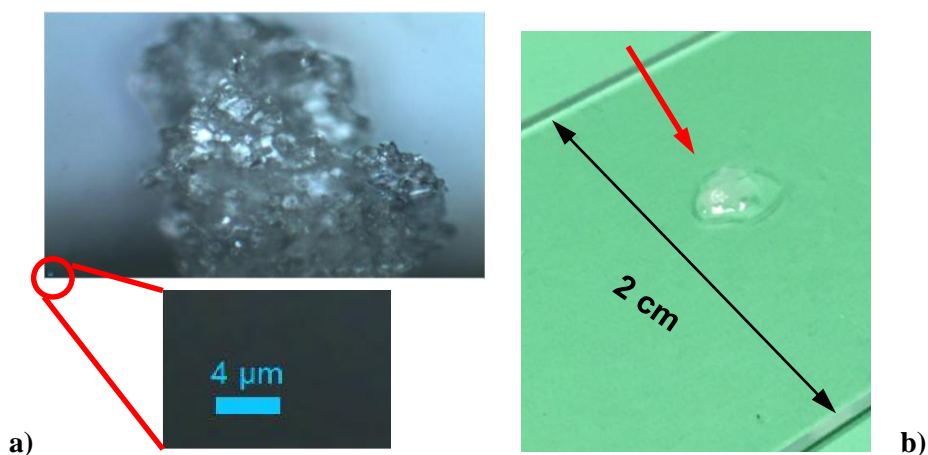


Fig. 4.9: a) Na-K ATPase lyophilized powder b) Na-K ATPase solution,

Fig. 4.10 shows the Raman spectrum obtained for Na-K ATPase lyophilized powder sample.

Also for these samples, like for lysozyme, data were processed with LabSpec software, supplied with the used spectrometer. Acquisition was made using the 600 lines / mm diffraction grating, 100x objective and 200 μm confocal hole. Five different spectral ranges were analyzed, i.e. grating centered at a 650, 1400, 2100, 2900 and 3400 cm^{-1} , and then results were collected in a unique spectrum. Each result is the mean of 10 measures with 60 s acquisition time each. Low-degree (≤ 3) polynomial functions were subtracted to each original spectrum in order to eliminate the background (polynomial baseline correction).

The results are in agreement with those of the literature, see for example (*Movasaghi et al. 2007*).

Na-K ATPase is a less studied protein (at the spectroscopic investigation level) with respect to lysozyme. Therefore, in addition to the total spectrum obtained by the analysis, in table 4.1 it is shown the assignment of the characteristic peaks as a fingerprint of the protein (*Movasaghi et al. 2007*).

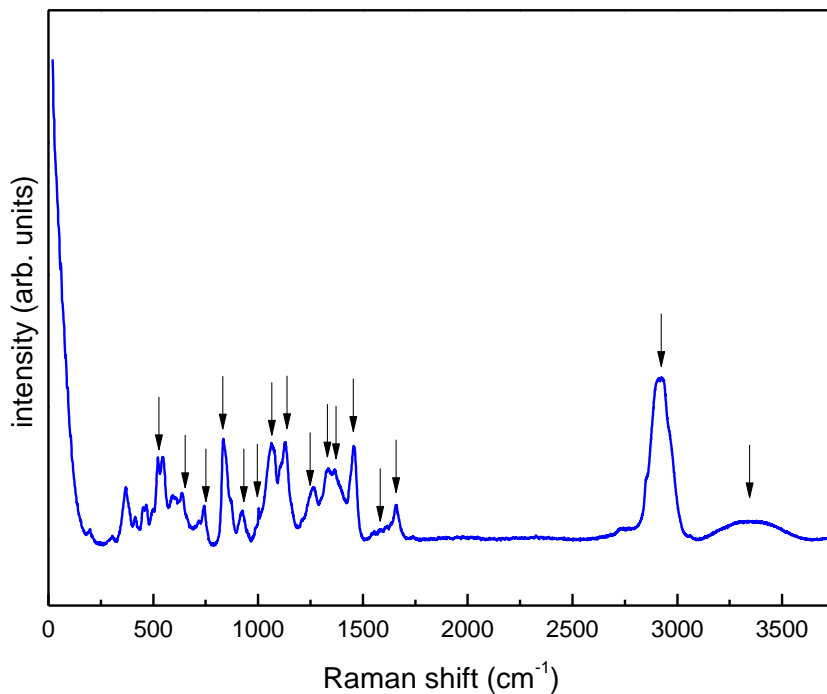


Figure 4.10: Complete raman spectrum of Na-K ATPase powder

Frequency (cm ⁻¹)	Assignment
524	S-S stretching
640	C-C twisting in Tyr
740	C-S stretching
831	Tyr
930	C-C stretching
1004	Phe
1067	Pro
1125	Trp

1260	Amide III
1335	CH ₃ -CH ₂ wagging
1365	Trp
1454	C-H bending
1585	Tyr, Phe
1620-1670	Amide I
2900-3000	CH ₂ -CH ₃ stretching
3200-3600	H ₂ O stretching

Table 4.1: correspondence between Na-K ATPase peaks on the spectrum and chemical groups or amino acids.

In the spectrum a high frequency band (3200-3600) cm⁻¹ is evidenced, it can be identified as a spectroscopic signal compatible with water. Even if the analyzed sample under lyophilized conditions, to confirm that this band is really due to the hydration water around the protein, called water shell, we hydrated protein, by depositing a drop of MilliQ water on the sample powder. The spectrum acquired on the protein treated in this way is shown in black in Figure 4.11.

The comparison shows that the sample of ATPase powder contains a part of water of hydration.

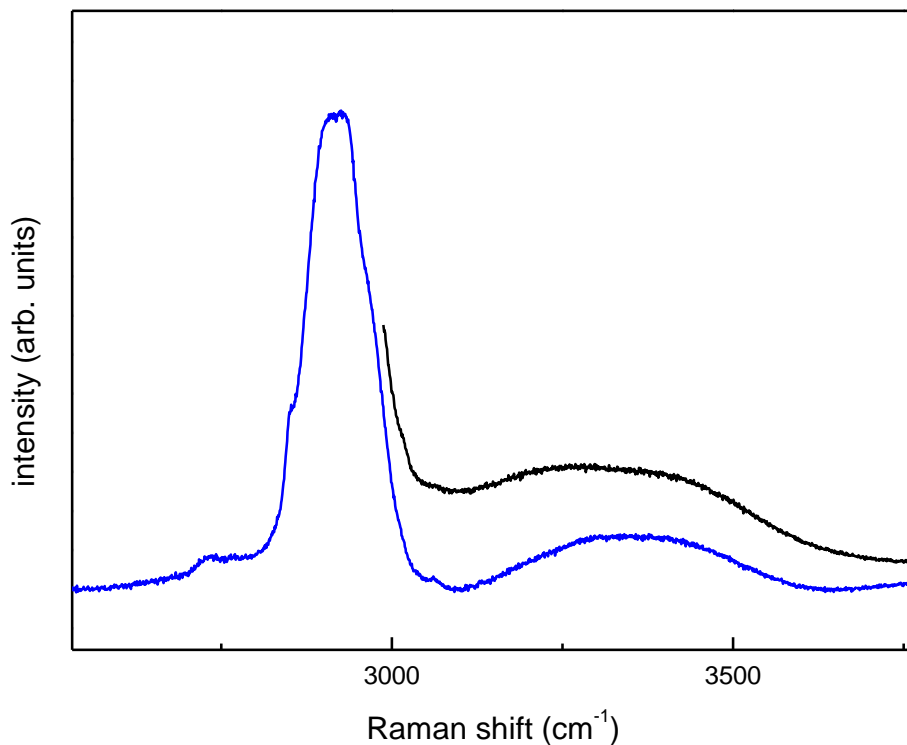


Figure 4.11: comparison of water peaks between lyophilized powder (blue) and hydrated sample (black)

Thanks to the use of ULF filters it has been possible also to investigate low-frequency vibrations.

Fig. 4.12 and Fig 4.13 show Stokes and Antistokes ATPase spectrum obtained with grating with 600 lines / mm (spectral resolution of <3 cm⁻¹) and with 1800 lines / mm (spectral resolution of <1 cm⁻¹) respectively.

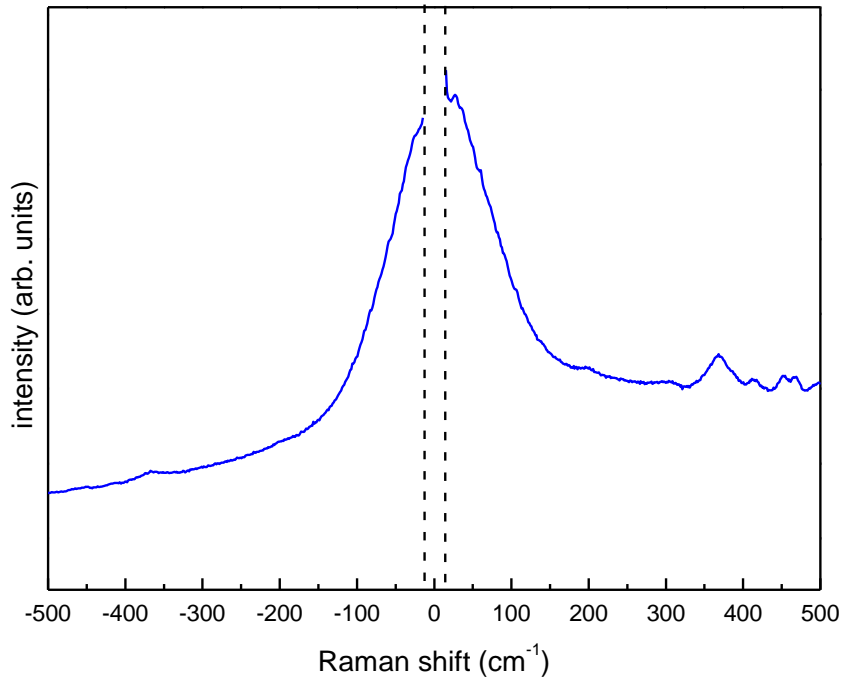


Fig. 4.12 Lyophilized Na-K ATPase Raman spectrum centered at 0 cm⁻¹. (grating 600 lines/mm)

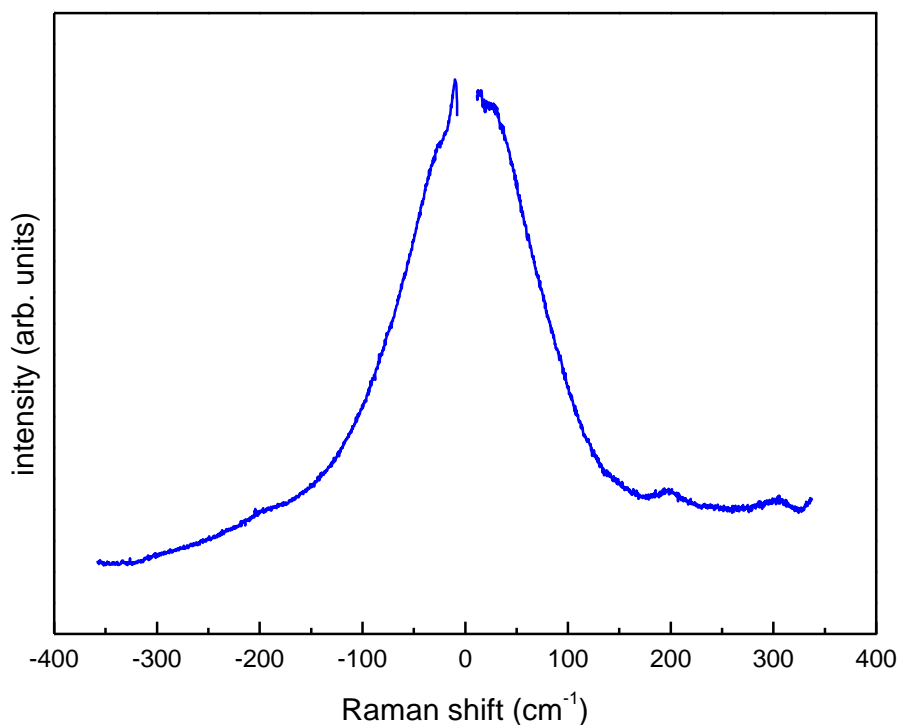


Fig. 4.13 Lyophilized Na-K ATPase Raman spectrum centered at 0 cm⁻¹. (grating 1800 lines/mm)

In both figures the central part of spectra around 0 cm⁻¹ is not present. As a matter of fact, in this portion the Rayleigh peak, due to the laser source, cannot be filtered by ULF, therefore this part is totally uninteresting.

By analyzing the spectra it is important to consider both Stokes and Antistokes regions; in fact the presence on both semi-axis of the same peak is a measure of the goodness of the test.

Peaks are clearly visible in correspondence of 27, 190, 300 cm^{-1} (0.81, 5 and 9 THz)

These peaks are not assigned in literature to any chemical group, as they should correspond to vibrations involving the entire protein or large portions of it.

Such delocalized or global vibrations were investigated numerically by modeling the entire Na-K ATPase (see Chapter 5).

Measurements on the sample in aqueous solution were repeated. Through successive measures we chose an amount of solute that would allow the collection of a Raman spectrum of good quality using the grating of 600 lines/mm.

Also for this acquisition the complete spectrum (in red in Figure 4.13) was collected merging 5 spectral ranges (i.e. grating centered at 750, 1400, 2100, 2850, 3400 cm^{-1}). The spectrum ATPase in aqueous solution is shown in Figure 4.13 in comparison with the one previously obtained from the powder. By the figure it can be seen as water acts like a sort of dumper, not shifting frequencies but lowering peaks. Obviously, adding water to the powder, peak around 3300 cm^{-1} , which indicates vibrations of water molecules, arises.

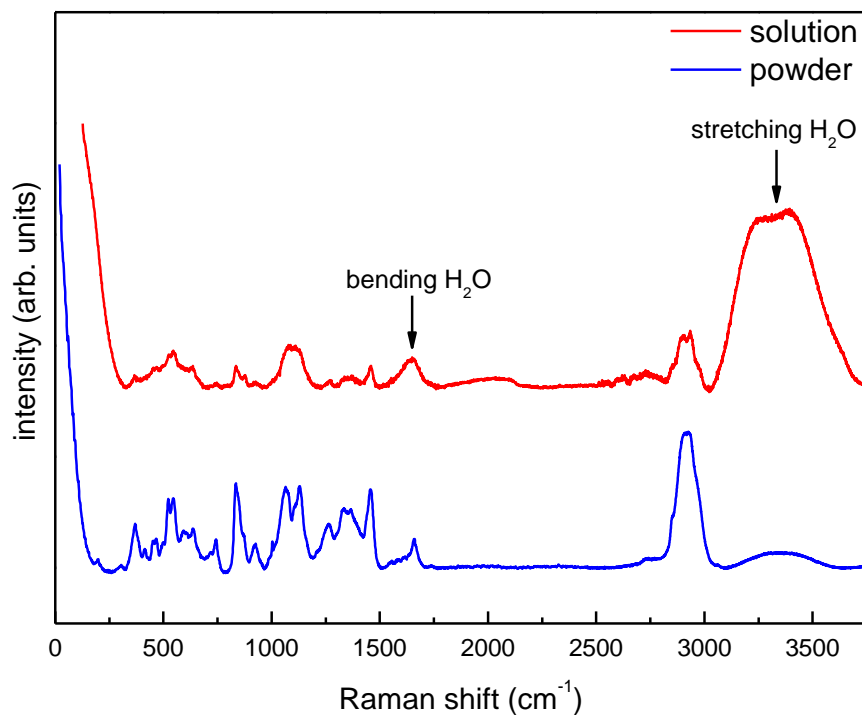


Figure 4.13: comparison of complete spectra of Na-K ATPase in solution (red line) and in powder (blue line).

With the same grating (600 lines / mm) the spectrum shown in Figure 4.14 was collected around origin. In this case, peaks clearly visible in dry samples are not present; this fact is probably caused by absorbance of water. Anyhow a low shoulder around 190 cm⁻¹, like the one seen on dry powder, is visible.

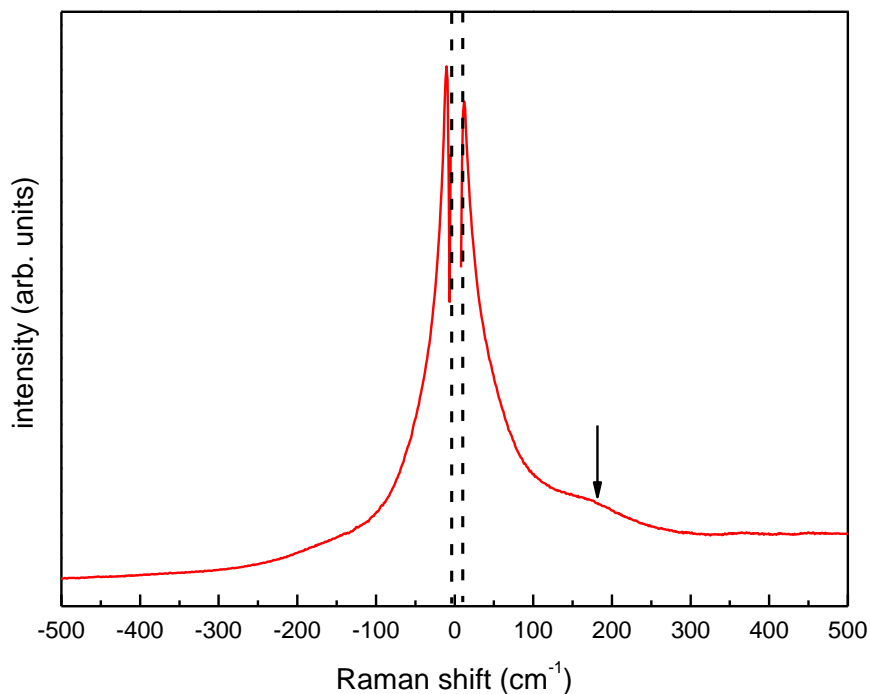


Fig. 4.14: hydrated Na-K ATPase Raman spectrum centered at 0 cm^{-1} . (grating 600 lines/mm).

In this Chapter the results of the experimental campaigns on Lysozyme and Na-K ATPase, have been presented. In addition to providing the complete Raman spectra of both protein, the attention has been focalized on low frequency regions of these spectra. In these regions, in fact, lie peaks related to vibrations involving the global protein or large parts of it: these vibrations will be deeply examined in the following chapter by means of numerical simulations.

Chapter 5

Numerical simulations: Lattice models of Lysozyme and Na-K ATPase

In this chapter the results of numerical simulations on Lysozyme and Na-K ATPase are shown.

In the first part of the chapter the software used for simulations is presented and all the parameters necessary to run the program are discussed.

In the second part of the normal modes of the two proteins are illustrated. Furthermore, the mode shapes corresponding to the vibration frequencies of three amino acids present in both proteins (Tryptophan, Tyrosine and Phenylalanine) and detected by the Raman spectra are highlighted.

5.1 Mechanical Model: Covalent bond, harmonic approximation, numerical parameters

Molecular forces are difficult to be modelled. However, for small displacements, attractive/repulsive forces between atoms can generally be considered as linearly elastic (*Feynman et al. 1964*).

Bonds between the same type of atom are covalent bonds, and bonds between atoms having electronegativities which differ by a little are also predominantly covalent. For covalent bonds, bond energies and bond lengths depend on many factors; however, there is a general trend in that the shorter the bond length, the higher the bond energy. The potential energy associated to a covalent bond is made of two parts, and can be empirically expressed as (*Pauling 1960*):

$$U(r) = -\frac{A}{r^m} + \frac{B}{r^n} \quad (\text{with } m < n, m > 2), \quad (5.1)$$

where A and B are positive constants (depending on the nature of interacting atoms), and r is the distance between atoms. The first and the second terms of the right-hand side of Eq. (5.1) represent the attractive and the repulsive energy, respectively. The former prevails at large interatomic distances (large r), while the latter dominates at short distances (small r); note that the right-

hand side of Eq. (5.1) diverges for $r \rightarrow 0$. For intermediate distances, a balance between the two terms is expected, with the consequent presence of a minimum of the potential energy (Fig. 5.1a). The equilibrium position (distance), r_0 , is therefore defined by setting the first derivative of function $U(r)$ equal to zero (minimum of potential energy):

$$\frac{dU}{dr} = 0 \Leftrightarrow r = r_0. \quad (5.2)$$

The interatomic force, $F = -dU/dr$, is attractive for $r > r_0$ ($dU/dr > 0$), repulsive for $r < r_0$ ($dU/dr < 0$), and null for $r = r_0$ (Fig. 5.1a). Other empirical expressions of the potential $U(r)$ can be adopted. For instance, the Morse potential has been specifically developed to describe the covalent bond between a couple of atoms (*Morse 1929*). However, unlike that of Eq. (5.1), the Morse potential does not diverge for $r \rightarrow 0$ and consequently it must be used for interatomic distances not much smaller than the equilibrium distance. The study of the small-amplitude atomic oscillations around the equilibrium position requires considering the second-order Taylor series expansion of the potential (harmonic approximation):

$$U(r) \cong U(r_0) + \left(\frac{dU}{dr} \right)_{r=r_0} (r - r_0) + \frac{1}{2} \left(\frac{d^2U}{dr^2} \right)_{r=r_0} (r - r_0)^2. \quad (5.3)$$

$$U(r) \cong U(r_0) + \left(\frac{dU}{dr} \right)_{r=r_0} (r-r_0) + \frac{1}{2} \left(\frac{d^2U}{dr^2} \right)_{r=r_0} (r-r_0)^2.$$

$$U(r) \cong U(r_0) + \left(\frac{dU}{dr} \right)_{r=r_0} (r-r_0) + \frac{1}{2} \left(\frac{d^2U}{dr^2} \right)_{r=r_0} (r-r_0)^2. \text{ Recalling Eq.}$$

(5.2) and defining the force constant (stiffness) k as:

$$k := - \left(\frac{dF}{dr} \right)_{r=r_0} = \left(\frac{d^2U}{dr^2} \right)_{r=r_0}, \quad (5.4)$$

the harmonic approximation of the potential becomes (Fig. 10b):

$$U(r) \cong U(r_0) + \frac{1}{2} k (r-r_0)^2. \quad (5.5)$$

Therefore, near r_0 we obtain the linear expression for the interatomic force (Fig. 10b):

$$F = -k(r-r_0) = -kx, \quad (5.6)$$

where we have set $x = (r-r_0)$. On these bases,

$$\mu \frac{d^2x}{dt^2} + kx = 0. \quad (5.7)$$

results fully justified.

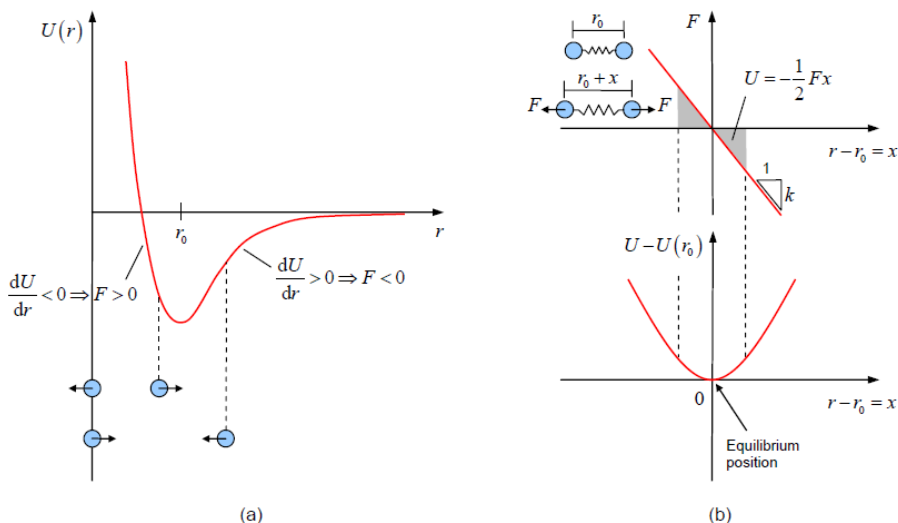


Fig. 5.1 Model of interatomic bond: potential energy versus interatomic distance (a), and harmonic approximation near the equilibrium distance (two atoms joined by a spring) (b)

Numerical lattice models were built for some amino acids, full lysozyme and full Na-K ATPase. LUSAS finite element code was adopted (*Lusas 2015*). Ball-and-stick models were obtained, with masses concentrated at the nodes (atoms) and interatomic bonds replaced by linear-elastic, massless beams. Only covalent bonds were considered, neglecting all weaker bonds. The beam ends were internally clamped at each node in order to prevent rotational labilities (bending and torsion internally rigid constraints). Only the elastic axial behaviour of the beams was taken into account: this was achieved by providing each beam with bending and torsion rigidities much greater than axial rigidity, so to inhibit both bending and torsion deformations, in addition

to shear deformation. In general, the axial stiffness k_i of the i -th beam is expressed as [99]:

$$k_i = \frac{EA_i}{l_i}, \quad (5.8)$$

being E (FL⁻²) the Young's modulus (equal for all beams), A_i (L²) and l_i (L) the cross-sectional area and the beam length, respectively. Each value was set so to reproduce the desired bond stiffness. Bending and (primary) torsion beam stiffness depend both on length l_i and on bending, EI_{ij} , and torsion, GJ_i , rigidities, respectively; I_{ij} (L⁴) denotes the second moment of area (moment of inertia) of the cross-section with respect to the principal axis j ($j = 1, 2$), G (FL⁻²) is the shear modulus, and J_i (L⁴) denotes the St. Venant torsion constant (*Carpinteri 1997*). Starting from beam lengths l_i and local axial, bending, and torsion rigidities, the local (element) stiffness matrix \mathbf{K}_e can be computed (*Bathe 1982, Zienkiewicz et al. 2005, Cheng 2001*). On the other hand, knowing the point masses m_k ($k = 1, 2, \dots, n$) allows to obtain also the local mass matrix \mathbf{M}_e (*Bathe 1982, Zienkiewicz et al. 2005, Cheng 2001*). By rotating, expanding, and assembling the local matrices of stiffness, \mathbf{K}_e , and mass, \mathbf{M}_e , the corresponding global matrices \mathbf{K} and \mathbf{M} are obtained. In our case, once given the required input data, all computations were automatically done by the adopted software.

Therefore, the undamped free dynamics of a structural assembly with n degrees of freedom is governed by the following homogeneous linear system (generalized eigenproblem) (Clough *et al.* 1975, Cheng 2001)

$$\begin{pmatrix} \mathbf{K} - \omega^2 \mathbf{M} \\ n \times n & n \times n \end{pmatrix} \begin{matrix} \mathbf{x} \\ n \times 1 \end{matrix} = \begin{matrix} \mathbf{0} \\ n \times 1 \end{matrix}, \quad (5.9)$$

the non-trivial solution of which gives the eigenvalues ω^2 (square of the natural angular frequencies of the system) by solving the algebraic equation of order n (characteristic equation) resulting from the following condition:

$$\det(\mathbf{K} - \omega^2 \mathbf{M}) = 0. \quad (5.10)$$

Once the eigenvalues have been determined, Eq. (5.9) allows to obtain the eigenvectors \mathbf{x} (vibration modes). A system with n degrees of freedom will therefore have n natural frequencies as well as n vibration modes. The solution of the eigenvalue-eigenvector problem was obtained by the adopted finite element code. All models were left unconstrained in the 3D space; thus, six rigid motions had to be excluded from each analysis.

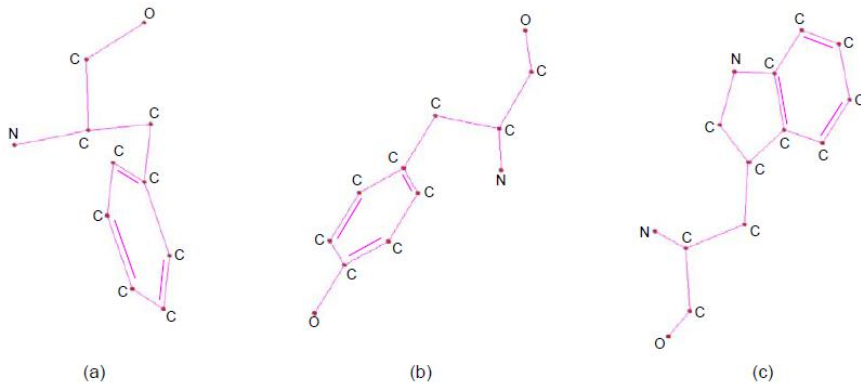


Fig. 11 Finite element lattice model of amino acids: phenylalanine (a), tyrosine (b), and tryptophan (c)

Atom masses and bond stiffnesses are listed in Tab. 5.1. Interatomic distances' order of magnitude is 10^{-10} m. Since we are dealing with powers of ten which are not appropriate for computer calculations, a numerical scaling was introduced. We remark that this scaling must not be intended as a physical scaling, being it just an artifice introduced for ease of computation. The adopted numerical scaling is listed in Tab. 5.2. According to that scaling, the conversion between the real (subscript r) and the numerical (subscript n) frequencies (in Hz) results to be the following:

$$f_r = f_n \times 10^{13}. \quad (5.11)$$

The input parameters introduced in the analyses are listed in Tab. 5.3. Note that the values of Young's modulus and cross-sectional area in Tab. 5.3 have

no physical basis: they were chosen such that, through Eq. (5.8), the required axial stiffness is obtained.

3D 3-node thin beam elements were used, with 3 subdivisions per element; however, the number of subdivisions has no influence on our analysis since, as pointed out before, we are only interested in the axial behaviour (bending/torsion deformations are disregarded; see the related inertia constants in Tab. 5.3).

Atom	Atom mass (kg)	Bond	Bond stiffness (N m ⁻¹)
Carbon (C)	1.994475×10 ⁻²⁶	C–C	180
		C=C	320
Nitrogen (N)	2.325918×10 ⁻²⁶	C–N	160
Oxygen (O)	2.656698×10 ⁻²⁶	C–O	190

Tab. 5.1 Atom masses and bond stiffnesses

Physical quantity	Scaling
Mass, M (kg)	numerical = real ×10 ²⁶
Distance, L (m)	numerical = real ×10 ¹⁰
Stiffness, F L ⁻¹ (N m ⁻¹)	numerical = real ×10 ⁰

Tab. 5.2 Scaling between real and numerical quantities

Mass	Elastic properties of beams	Geometric properties of beams cross-section
Point (atom) mass, kg $m_C = 1.994475$ $m_N = 2.325918$ $m_O = 2.656698$	Young modulus, $N m^{-2}$ $E = 3 \times 10^5$ Poisson ratio* , -0.3	Area, m^2 $A_{C-C} = 0.9240 \times 10^{-3}$ $A_{C-C} = 1.4293 \times 10^{-3}$ $A_{C-N} = 0.7840 \times 10^{-3}$ $A_{C-O} = 0.9057 \times 10^{-3}$
Beam mass, $kg m^{-3}$ $\rho = 0$	* irrelevant for present analysis	Principal second moments of area, m^4 $I_1 = I_2 = 1 \times 10^6$ Torsion constant, m^4 $J = 1 \times 10^6$

Tab. 5.3 Input data for numerical analysis

5.2 Results of numerical simulations of Lysozyme: Amino acids and global vibrations

Figure 5.2 shows the mode shapes corresponding to the vibration frequencies indicated in the Raman spectrum of Figure 4.7 (see Trp, Tyr, and Phe peaks). These modes mainly involve deformations (breathing) of the rings. The numerical frequencies in Figure 5.2 are very close to the corresponding experimental values.

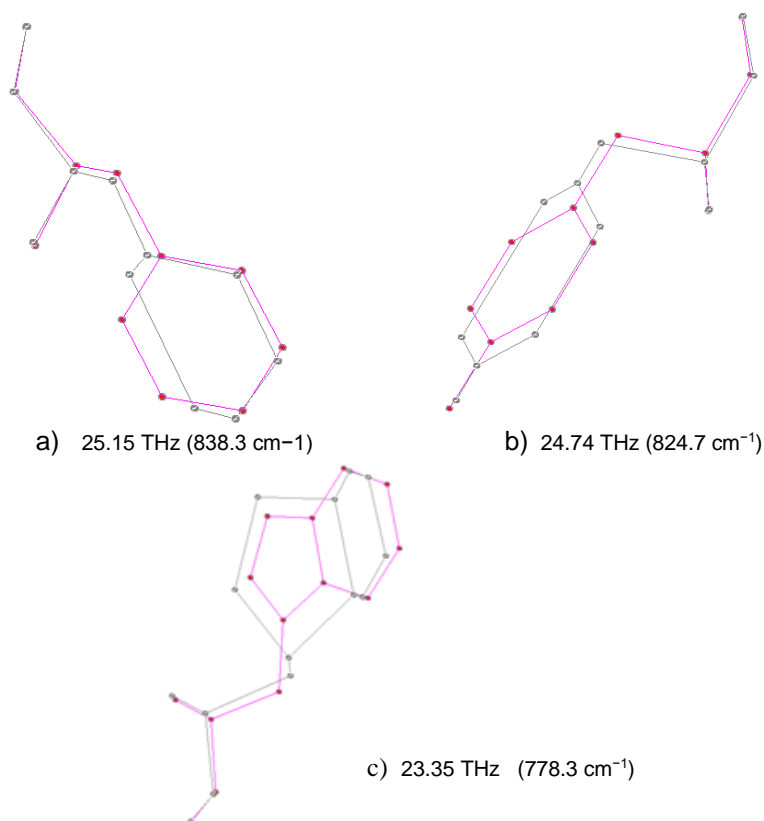


Fig. 5.2 Numerical mode shapes and related frequencies corresponding to ‘Phe’, ‘Tyr’, and ‘Trp’ peaks in Figure 4.7: phenylalanine (a), tyrosine (b), and tryptophan (c)

Fig. 5.3 shows the numerical model of lysozyme implemented in LUSAS. To import the entire geometry of the protein on our software, a connectivity matrix was created, in MATLAB, based on atom coordinates taken from the RCSB Protein Data Bank by identification code 4YM8 for the type of Lysozyme we modelled (<http://www.rcsb.org/pdb>).

The result is a tree-like three-dimensional structure. 1020 connections were created among the 1000 point masses (atoms) constituting the enzyme. The average interatomic distance is equal to 0.141 nm. A .dxf file was created by MATLAB and therefore imported in LUSAS to construct the structural geometry.

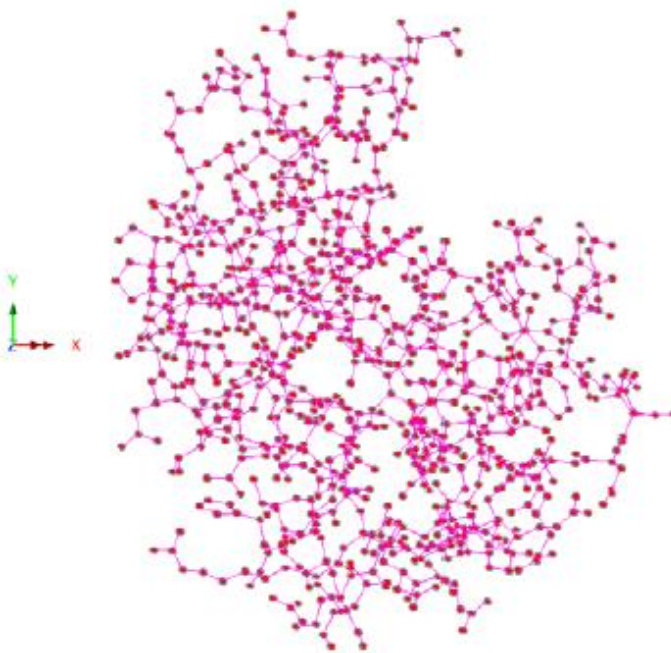
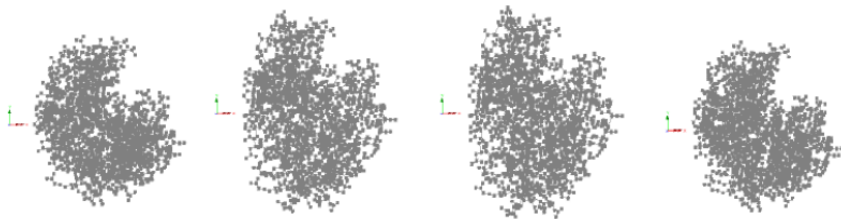


Fig. 5.3 Lysozyme finite element lattice model

A constant value of the point mass $m_a = 2.37$ kg was used for all atoms (actual average atom mass equal to 2.37×10^{-26} kg); the beam mass density ρ was set equal to zero. The elastic properties and the bending and torsion constants are the same as in Tab. 5.3. The cross-sectional area A was set equal to 9.4×10^{-4} m², constant for all beams. The default value of 4 subdivisions per beam element was used. The adopted numerical scaling is the same as before (see Tab. 5.2), so Eq. (5.11) still holds.

Using the area $A = 9.4 \times 10^{-4}$ m², Young's modulus $E = 3 \times 10^5$ N m⁻², and the average distance $l = 1.41$ m, Eq. (5.8) yields an average axial (bond) stiffness k equal to 200 N m⁻¹.

Fig. 5.4 shows the mode shape found by the numerical model for a frequency close to the large peak on the left of the Raman spectrum shown in Fig. 4.7. The experimental and numerical frequencies are respectively equal to 0.84 THz (28.0 cm⁻¹) and 0.88 THz (29.3 cm⁻¹). This mode is clearly a global one, that involves at least half protein. The frequency spectrum is broad: other global modes were found, as well as many localized modes. In general, the higher frequencies are related to localized vibration modes.

**Fig. 5.4** Global vibration mode of lysozyme at 0.88 THz (29.3 cm⁻¹).

5.3 Results of numerical simulations of Na-K ATPase

Fig. 5.5 shows the lattice model of Na-K ATPase implemented in LUSAS. Using the same strategy explained for lysozyme using coordinates of RCSB Protein Data Bank, code 2ZXE, a connectivity matrix was created in MATLAB,

The result is a Y-like three-dimensional structure. Among the 10131 point masses (atoms) constituting the protein, 10351 connections were created. The average interatomic distance is equal to 0.149 nm. A .dxf file was created by MATLAB and therefore imported in LUSAS to construct the structural geometry.

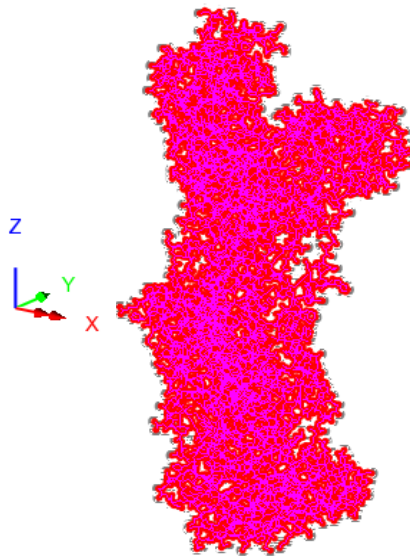


Fig. 5.5 Na-K ATPase finite element lattice model

For Na-K ATPase model a constant value of the point mass $m_a = 1.74$ kg was used for all atoms (the actual average atom mass is equal to 1.74×10^{-26} kg); the beam mass density ρ was set equal to zero. The elastic properties and the bending and torsion constants are the same as in Tab. 4.3. The cross-sectional area A was set equal to 9.4×10^{-4} m², constant for all beams. The default value of 4 subdivisions per beam element was used. The adopted numerical scaling is the same as before (see Tab. 5.2), so Eq. (5.11) still holds. Using the area $A = 9.4 \times 10^{-4}$ m², Young's modulus $E = 3 \times 10^5$ N m⁻², and the average distance $l = 1.41$ m, Eq. (5.8) yields an average axial (bond) stiffness k equal to 200 N m⁻¹.

As it is evident from the spectra of lysozyme and of ATPase, peaks related to the chemical groups are the same (for details see section 5.2), consequently is more interesting to numerically simulate vibration modes at low frequency, those that involve large portions of protein or the entire protein.

As explained in chapter 4, in the spectral region near to the origin, below 500 cm⁻¹, peaks are located around 27, 190, 300 cm⁻¹ (0.81, 5 and 9 THz). Therefore, by using a lattice model, these frequencies and vibrational modes are investigated.

Fig. 5.6 shows the mode shape found by the numerical model for a frequency close to the peak at 27 cm⁻¹. The experimental and numerical frequencies are respectively equal to 0.81 THz (27.0 cm⁻¹) and 0.82 THz (27.3 cm⁻¹). This mode is clearly a global one, that involves all the top-left part of the protein that extends inside the cell.

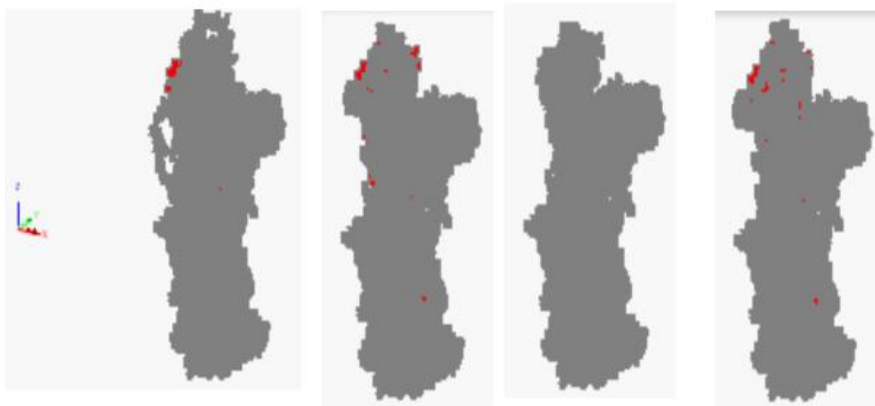


Fig. 5.6: top left vibration mode at 0.82 THz

Fig. 5.7 shows the mode shape found by the numerical model for a frequency close to the peak at 190 cm^{-1} . The experimental and numerical frequencies are respectively equal to 5.7 THz (190 cm^{-1}) and 5.73 THz (191 cm^{-1}).

Also in this case, vibration involves all the top-right part of the protein that, as the previous one, extends inside the cell.

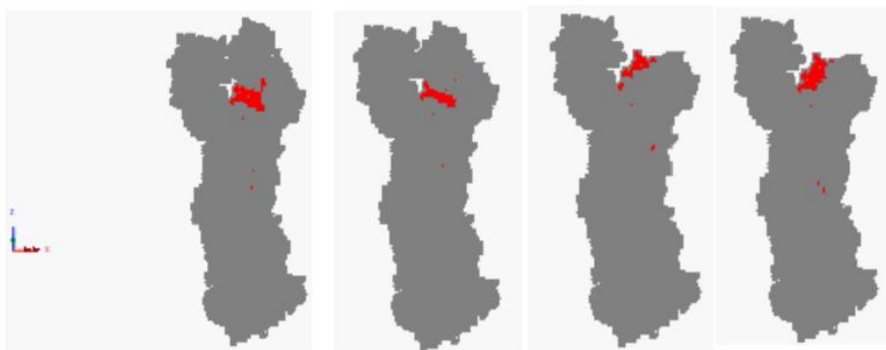


Fig. 5.7: top right vibration mode at 5.73 THz

Fig. 5.8 shows the mode shape found by the numerical model for a frequency close to the peak at 300 cm^{-1} . The experimental and numerical frequencies are respectively equal to 9 THz (300 cm^{-1}) and 9.06 THz (302 cm^{-1}).

In this case, even the mode involves a large part of the protein, this one belongs to the bottom of the protein, the part that extends outside the cell.

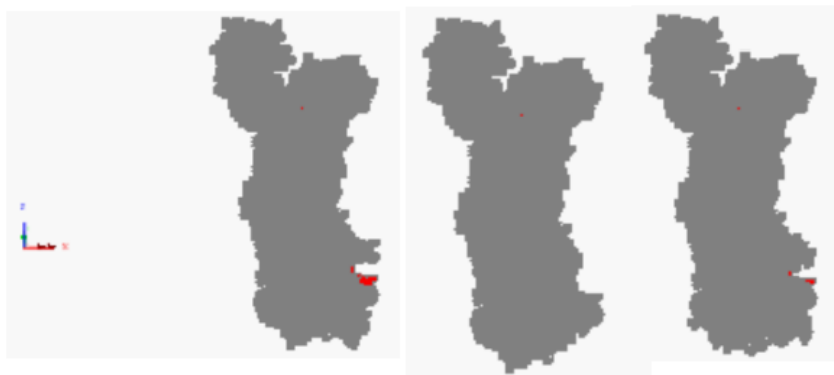


Fig. 5.8: bottom vibration mode at 5.73 THz

As appears clearly from numerical simulations, all these low frequency modes, refer to large portions of protein, while modes at higher frequencies, around tens of TeraHertz, are characteristic of small part of protein, such as chemical groups, amino acids, water molecules and so on.

This conclusion is in perfect agreement with the hypothesis of Chapter 1 and 3 (see sections 1.2 and 3.4).

Chapter 6

Conclusions

Important protein biological functions are driven by structural configuration changes, i.e. folding. This is a complex phenomenon not completely understood at present. Recent studies have argued that proteins mechanical vibrations can play a significant role: we cannot exclude that large conformational changes may occur as the result of nano-instabilities induced by resonant mechanical oscillations.

In this thesis, the free dynamics of crystallized powder, protein powder and hydrated powder of two types of protein (Lysozyme and Na-K ATPase) is analyzed experimentally and numerically.

Raman spectroscopy measurements were conducted and the results are compared to the corresponding ones from the literature, where literature was present. In particular, the use of ULF filters allowed to investigate also the lower frequencies, under 400 cm^{-1} , (i.e., those corresponding to global vibrations).

Moreover, the linear normal modes of well-known amino acids contained in

proteins (phenylalanine, tyrosine, tryptophan) as well as those of the whole proteins, both Lysozyme and Na-K ATPase, were investigated by Modal Analysis.

The approach adopted here belongs to classical mechanics. More precisely, we referred to biological materials and, starting from data of mechanical properties taken from the literature (e.g., mass, bond stiffness/energy), we evaluated the natural frequencies and mode shapes corresponding to small-amplitude vibrations around the equilibrium configuration.

The obtained results show a very good correspondence to the experimental evidences. The resonant frequencies found range from hundreds of GHz (entire protein or large portions of it) to tens of THz (chemical groups/amino acids).

In our numerical simulations we found values for resonant frequencies lower than the ones visible on Raman spectra, despite the presence of Ultra Low Frequency filters: this fact refers to the reason that, even with ULF filters we can get near the origin of the spectrum, frequencies lower than tens of Gigahertz are anyway covered by the Rayleigh peak. It could be very interesting to investigate experimentally these vibrations with others equipment, i.e. near field microscopy or femtosecond optical Kerr-effect (OKE) spectroscopy.

Numerical analyses like those conducted here can be performed with standard Finite Element programs and require a computational cost lower than molecular dynamics simulations (*de novo* or *ab initio*).

The results, although less general than those given by molecular dynamics analyses, can easily give precious information about the vibrational behaviour of proteins around a specific geometric configuration.

As a matter of fact, it is possible to investigate vibration mode shapes (previously unknown) corresponding to specific frequencies identified by Raman measurements

or other techniques. This can be useful to characterize new proteins as well as unknown behaviours of common proteins.

This thesis does not have the will to be exhaustive about a vast and still relatively unknown topic as the vibrations at the nanoscale in biological materials and their possible implications in cell metabolism. This work is the first step in a larger project that tries to bring together the knowledge of engineers and biologists.

As we have seen in Chapter 3, when a structure vibrates near its natural frequencies, instability can occur easily, and instability can cause a change of configuration; transferring this concept to proteins structure, we shall think that the right stimuli at the right frequencies could cause changes in the 3D structure...consequences of this fact are all to be discovered.

References

- Acbas, G.; Niessen, K. A.; Snell, E. H.; Markelz, A. G. (2014). Optical measurements of long-range protein vibrations. *Nat Commun*, 5, 3076. doi:10.1038/ncomms4076.
- Adhikari, A.N.; Freed, K.F.; Sosnick, T.R. (2012). De novo prediction of protein folding pathways and structure using the principle of sequential stabilization. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (43), 17442–7.
- Albers RW. (1967). Biochemical aspects of active transport. *Annu Rev Biochem.*; 36:727-56.
- Alberts, B.; Bray, D.; Hopkin, K.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. (2010). Protein Structure and Function. *Essential cell biology*, 3rd ed.; Garland Science: New York, NY, 2010; pp. 120–170.
- Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walters, P. (2002). The Shape and Structure of Proteins. In *Molecular Biology of the Cell*, 4th ed.; Garland Science: New York and London,
- Anfinsen, C.B. (1972). The formation and stabilization of protein structure. *The Biochemical Journal*, 128 (4), 737–49.

-
- Anfinsen, C.B. (1973). Principles that govern the folding of protein chains. *Science* 181 (4096), 223–30.
- Ashby, M.; Shercliff, H.; Cebon, D. (2014). *Materials: engineering, science, processing and design*, 3rd ed., Butterworth-Heinemann,.
- Bahar, I.; Rader, A. J. (2005). Coarse-grained normal mode analysis in structural biology. *Current Opinion in Structural Biology*, 15, pp. 586–592.
- Bathe, K. J. (1982). *Finite Element Procedures in Engineering Analysis*. Prentice-Hall Inc., Englewood Cliff, N.J.
- Baumruk, V.; Keiderling, T.A. (1993). Vibrational Circular Dichroism of Proteins in H₂O Solution. *Journal of the American Chemical Society*, 115, 6939–6942.
- Bažant, Z. P.; Cedolin, L. *Stability of Structures: Elastic, Inelastic, Fracture, and Damage Theories*, Oxford University Press, 1991.
- Binnig, G.; Quate, C.F.; Gerber, Ch. (1986). Atomic force microscope. *Physical Review Letters*, 56, 930–3.
- Blake, C.C.; Koenig, D.F.; Mair, G.A.; North, A.C.; Phillips, D.C.; Sarma, V.R. (1965). Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. *Nature* 206 (4986): 757–61.
- Blostein R. (1985). Proton-activated rubidium transport catalyzed by the sodium pump. *J Biol Chem. Jan 25;260(2):829-33*.
- Brooks, B.; Karplus, M. (1985). Normal modes for specific motions of macromolecules: application to the hinge-bending mode of lysozyme. *Proc. Natl Acad. Sci. USA*, 82, 4995–4999.
- Bryngelson, J.D.; Onuchic, J.N.; Socci, N.D.; Wolynes, P.G. (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins*, 21 (3), 167–95.
- Bu, Z.; Cook, J.; Callaway, D.J. (2001). Dynamic regimes and correlated structural dynamics in native and denatured alpha-lactalbumin. *Journal of Molecular*

- Biology*, 312 (4), 865–73.
- Callender RH, Dyer RB, Gilmanishin R, Woodruff WH. (1998). Fast events in protein folding: the time evolution of primary processes. *Annu Rev Phys Chem.* 49:173-202.
- Callender, R.; Deng, H. (1994). Nonresonance Raman difference spectroscopy: a general probe of protein structure, ligand binding, enzymatic catalysis, and the structures of other biomacromolecules. *Annu. Rev. Biophys. biomol. Struct.* 23, 215–245.
- Carey, P. R. (1999). Raman spectroscopy, the sleeping giant in structural biology, awakes. *J. biol. Chem.* 274, 26625–26628.
- Carey, P. R. Raman spectroscopy in enzymology: the first 25 years. (1998). *J. Raman Spectrosc.* 29, 7–14.
- Carpinteri A (May 30, 2014). Distinguished Lecture in Solid Mechanics presented at the California Institute of Technology.
- Carpinteri, A. (1997). *Structural Mechanics: A Unified Approach*. Chapman & Hall, London.
- Carpinteri, A.; Lacidogna, G.; Manuello, A. (2015). *Acoustic, Electromagnetic, Neutron Emissions from Fracture and Earthquakes*, Springer.
- Carpinteri, A.; Malvano, R.; Manuello, A.; Piana, G. (2014). Fundamental frequency evolution in slender beams subjected to imposed axial displacements. *Journal of Sound and Vibration*, 333, 2390–2403.
- Carpinteri, A.; Paggi, M. (2013). A theoretical approach to the interaction between buckling and resonance instabilities. *Journal of Engineering Mathematics*, 78 (1), 19–35.
- Cheng, F. Y. (2001). *Matrix Analysis of Structural Dynamics. Applications and Earthquake Engineering*. Marcel Dekker, Inc., New York.

-
- Chiti, F.; Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75, 333–66.
- Clough, R. W.; Penzien, J. (1975). *Dynamics of Structures*, McGraw-Hill, New York
- Clough, R. W.; Penzien, J. (1975). *Dynamics of Structures*, McGraw-Hill, New York.
- Compiani, M.; Capriotti, E. (2013). Computational and theoretical methods for protein folding. *Biochemistry*, 52 (48), 8601–24.
- Cross, G.H.; Freeman, N.J.; Swann, M.J. (2008). Dual Polarization Interferometry: A Real-Time Optical Technique for Measuring (Bio)molecular Orientation, Structure and Function at the Solid/Liquid Interface. In *Handbook of Biosensors and Biochips*. John Wiley & Sons, Hoboken, New Jersey; Part Four. Transducer Technologies for Biosensors.
- Davydov, A. S. (1973). The theory of contraction of proteins under their excitation. *J. Theor. Biol.* 38, 559–569.
- Deak, J.; Chin, H.; Lewis, C.; Miller, R. (1998). Ultrafast phase grating studies of heme proteins: observation of the low-frequency modes directing functionally important protein motions. *J. Phys. Chem. B*, 102, 6621–6634.
- Debe DA, Carlson MJ, Goddard WA. (1999).The topomer-sampling model of protein folding. *3rd Proc Natl Acad Sci U S A*.Mar 16; 96(6):2596-601.
- Deechongkit, S.; Nguyen, H.; Powers, E.T.; Dawson, P.E.; Gruebele, M.; Kelly, J.W. (2004). Context-dependent contributions of backbone hydrogen bonding to beta-sheet folding energetics. *Nature*, 430 (6995), 101–5.
- Deng, H.; Callender, R. (1999). Raman spectroscopic studies of the structures, energetics, and bond distortions of substrates bound to enzymes. *Methods Enzymol.* 308, 176–201.

- Dyson HJ, Wright PE, Scheraga HA. 2006. The role of hydrophobic interactions in initiation and propagation of protein folding. *Proc Natl Acad Sci U S A*. Aug 29; 103(35):13057-61.
- Ellis, R.J. (2006). Molecular chaperones: assisting assembly in addition to folding. *Trends in Biochemical Sciences*, 31 (7), 395–401.
- Elofsson, A. et al. (1996). A study of combined structure/sequence profiles. *Fold Des*, 1(6): p. 451-61.
- Falconer, R. J.; Markelz, A. G. (2012). Terahertz Spectroscopic Analysis of Peptides and Proteins. *J Infrared Milli Terahz Waves*, 33, 973–988.
- Ferraro J. R. (1967). Raman Spectroscopy, Theory and Practice. H. A. Szymanski, ed. pp. 44-81. Plenum Press, New York,
- Ferraro J. R., R. Jarnutowski, and D. C. Lankin. (1992). *Spectroscopy* 7, 30.
- Ferraro, J. R.; Nakamoto, K.; Brown, C. W. (2003). *Introductory Raman Spectroscopy*, 3rd ed.; Elsevier.
- Fersht AR.(1997). Nucleation mechanisms in protein folding. *Curr Opin Struct Biol*. Feb; 7(1):3-9.
- Feynman, R. P.; Leighton, R. B.; Sands, M. (1964). The Feynman Lectures on Physics, Addison–Wesley, USA.
- Feynman, R. P.; Leighton, R. B.; Sands, M. (1964). *The Feynman Lectures on Physics*, Addison–Wesley, USA.
- Fleming, A. (1922). On a remarkable bacteriolytic element found in tissues and secretions. *Proceedings of the Royal Society B* 93 (653), 306–317.
- Frezzotti, M. L.; Tecce, F.; Casagli, A. (2012). Raman spectroscopy for fluid inclusion analysis. *Journal of Geochemical Exploration* 112, 1–20.

- Gadsby DC. (2009). Ion channels versus ion pumps: the principal difference, in principle. *Nat Rev Mol Cell Biol.* 2009 May;10(5):344-52. doi: 10.1038/nrm2668. Epub Apr 2.
- Genzel, L. *et al.* (1976). Low-frequency Raman spectra of lysozyme. *Biopolymers*, 15, 219–225.
- Genzel, L.; Keilmann, F.; Martin, T. P.; Winterling, G.; Yacoby, Y.; Fröhlich, H.; Makinen, M. W. (1976). Low-Frequency Raman Spectra of lysozyme. *Biopolymers* 15 (1): 219–225.
- Gilson T. R. and P. J. Hendra. (1970). *Laser Raman Spectroscopy. pp. 1-266. Wiley-Interscience, London.*
- Glynn. I. M. (1993). Annual review prize lecture. 'All hands to the sodium pump'. *J Physiol.Mar;* 462: 1–30.
- Gräter F, Grubmüller H. (2007). Fluctuations of primary ubiquitin folding intermediates in a force clamp. *J Struct Biol. Mar;* 157(3):557-69.
- Ham N. S. and A. Walsh. (1958). *Spectrochim. Acta* 12, 88.
- Hammarström, P.; Wiseman, R.L.; Powers, E.T.; Kelly, J.W. (2003). Prevention of transthyretin amyloid disease by changing protein misfolding energetics. *Science*, 299 (5607), 713–6.
- He, Y.; Ku, P. I.; Knab, J. R.; Chen, J. Y.; Markelz, A. G. (2008). Protein dynamical transition does not require protein structure. *Phys. Rev. Lett.* 101, 178103.
- Hedoux, A. *et al.* (2006). Evidence of a two-stage thermal denaturation process in lysozyme: A Raman scattering and differential scanning calorimetry investigation. *J. Chem. Phys.* 124, 014703.
- Hibben J. H.. (1939). *The Raman Effect and Its Chemical Application. Reinhold PubUshing Corp., New York.*

- Hoang L, Bedard S, Krishna MM, Lin Y, Englander SW. (2002). Cytochrome c folding pathway: kinetic native-state hydrogen exchange. *Proc Natl Acad Sci U S A. Sep 17, 99(19):12173-8*
- Hunt, N. T.; Jaye, A. A.; Meech, S. R. (2007). Ultrafast dynamics in complex fluids observed through the ultrafast optically-heterodyne-detected optical-Kerr-effect (OHD-OKE). *Phys. Chem. Chem. Phys.*, 9, 2167–2180.
- Huyghues-Despointes, B.M.P.; Pace, C.N.; Englander, S.W.; Scholtz, J.M. (2001). Measuring the Conformational Stability of a Protein by Hydrogen Exchange. *Methods in Molecular Biology*. Murphy, K.P.; Humana Press, Totowa, New Jersey, pp. 69–92.
- Jackson, S.E. (1998). How do small single-domain proteins fold? *Folding & Design*, 3 (4), R81–91.
- Jagannathan, B.; Elms, P.J.; Bustamante, C.; Marqusee, S. (2012). Direct observation of a force-induced switch in the anisotropic mechanical unfolding pathway of a protein. *Proceedings of the National Academy of Sciences of the United States of America*, 109 (44), 17820–5.
- Jagannathan, B.; Marqusee, S. (2013). Protein folding and unfolding under force. *Biopolymers*, 99 (11), 860–9.
- Jakobi, A.J.; Mashaghi, A.; Tans, S.J.; Huizinga, E.G. (2011). Calcium modulates force sensing by the von Willebrand factor A2 domain. *Nature Communications*, 2:385.
- Jin, H., Sciammarella, C., Yoshida, S., Lamberti, L. (2015). Advancement of Optical Methods in Experimental Mechanics. *Proceedings of the 2014 Annual Conference on Experimental and Applied Mechanics*.
- Johnson, S.M.; Wiseman, R.L.; Sekijima, Y.; Green, N.S.; Adamski-Werner, S.L.; Kelly, J.W. (2005). Native state kinetic stabilization as a strategy to ameliorate

- protein misfolding diseases: a focus on the transthyretin amyloidoses. *Accounts of Chemical Research*, 38 (12), 911–21.
- Karplus M, Weaver DL. (1979). Diffusion-collision model for protein folding. *Biopolymers*.18:1421–37.
- Karplus M, Weaver DL. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci. Apr*; 3(4):650-68.
- Karplus, M.; Kuriyan, J. (2005). Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, pp. 6679-6685.
- Kerschbaum F. P., Z. (1914). *Instrumentenk* 34, 43
- Kim, P.S.; Baldwin, R.L. (1990). Intermediates in the folding reactions of small proteins. *Annual Review of Biochemistry*, 59 (1), 631–60.
- Kmiecik, S.; Kolinski, A. (2007). Characterization of protein-folding pathways by reduced-space modelling. *Proceedings of the National Academy of Sciences of the United States of America*, 104 (30), 12330–5.
- Krishna MM, Maity H, Rumbley JN, Lin Y, Englander SW. (2006). Order of steps in the cytochrome C folding pathway: evidence for a sequential stabilization mechanism. *J Mol Biol. Jun 23*; 359(5):1410-9
- Kubelka, J.; Hofrichter, J.; Eaton, W.A. (2004). The protein folding ‘speed limit’. *Current Opinion in Structural Biology*, 14 (1), 76–88.
- Leopold, P.E.; Montal, M.; Onuchic, J.N. (1992). Protein folding funnels: a kinetic approach to the sequence-structure relationship. *Proceedings of the National Academy of Sciences of the United States of America*, 89 (18), 8721–5.
- Lerbret, A.; Hedoux, A.; Annighöfer, B.; Bellissent-Funel, M. C. (2013). Influence of pressure on the low-frequency vibrational modes of lysozyme and water: a complementary inelastic neutron scattering and molecular dynamics simulation study. *Proteins*, 81, 326–340.

- Lesk AM, Rose GD. (1981). Folding units in globular proteins. *Proc Natl Acad Sci U S A. Jul;78(7):4304-8.*
- Levinthal, C. (1968). Are there pathways for protein folding? *Journal de Chimie Physique et de Physico-Chimie Biologique*, 65, 44–45.
- Levitt, M.; Sander, C.; Stern, P.S. (1985). Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* 181, 423–447.
- Lindorff-Larsen, K.; Piana, S.; Dror, R.O.; Shaw D.E. (2011). How fast-folding proteins fold. *Science*, 334 (6055), 517–20.
- Liphardt, J.; Onoa, B.; Smith, S.B.; Tinoco, I.Jr.; Bustamante, C. (2001). Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292, 733–7.
- Liu, D. *et al.* (2008). Studies of phononlike low-energy excitations of protein molecules by inelastic X-ray scattering. *Phys. Rev. Lett.* 101, 135501.
- LUSAS (2015). *User Reference Manual (version 15.1)*, FEA Ltd., Kingston upon Thames, Surrey, UK.
- Ma, J. (2005). Usefulness and Limitations of Normal Mode Analysis in Modeling Dynamics of Biomolecular Complexes. *Structure*, 13, 373–380.
- Makarov DE, Keller CA, Plaxco KW, Metiu H. (2002). How the folding rate constant of simple, single-domain proteins depends on the number of native contacts. *Proc Natl Acad Sci U S A. Mar 19; 99(6):3535-9.*
- Mashaghi, A.; Kramer, G.; Lamb, D.C.; Mayer, M.P.; Tans, S.J. (2014). Chaperone action at the single-molecule level. *Chemical Reviews*, 114 (1), 660–76.
- Matheson RR, Jr, Scheraga HA. (1978). A method for predicting nucleation sites for protein folding based on hydrophobic contacts. *Macromolecules.* 11:819–29.

-
- Mattevi A, Rizzi M, Bolognesi M, (1996). New structures of allosteric proteins revealing remarkable conformational changes. *Current opinion in structural biology*. 6: 824-829.
- McCammon JA, Gelin BR, Karplus M. (1977). Dynamics of folded proteins. *Nature*. Jun 16; 267(5612):585-90.
- Mijatovic Tatjana, Eric Van Quaquebeke, Bruno Delest, Olivier Debeir, Francis Darro, Robert Kiss. (2007). Cardiotonic steroids on the road to anti-cancer therapy. *Biochimica et Biophysica Acta 1776*; 32–57
- Minde, D.P.; Maurice, M.M.; Rüdiger, S.G. (2012). Determining biophysical protein stability in lysates by a fast proteolysis assay, FASTpp. *PLoS One*, 7 (10), e46147.
- Morse, P. M. (1929). Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Phys. Rev.* 34, 57–64.
- Moult, J. et al. (1995). A large-scale experiment to assess protein structure prediction methods. *Proteins*, 1995. 23(3): p. ii-v.
- Movasaghi, Z.; Rehman, S.; Rehman, I. U. (2007). Raman spectroscopy of biological tissues. *Applied Spectroscopy Reviews*, 42 (5), 493–541.
- Myers JK, Oas TG. (2001). Preorganized secondary structure as an important determinant of fast protein folding. *Nat Struct Biol*. Jun; 8(6):552-8.
- Netzer, W. J.; Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388, 343–349.
- Niessen, K. A.; Xu, M.; Markelz, A. G. (2015). Terahertz optical measurements of correlated motions with possible allosteric function. *Biophys Rev*, 7, 201–216. doi:10.1007/s12551-015-0168-4
- Oberhauser, A.F.; Hansma, P.K.; Carrion-Vazquez, M.; Fernandez, J.M. (2001). Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc Natl Acad. Sci.*, 98, 468–72.

- Ojeda-May, P.; Garcia, M.E. (2010). Electric field-driven disruption of a native beta-sheet protein conformation and generation of a helix-structure. *Biophysical Journal*, 99 (2), 595–9.
- Park, C.; Marqusee, S. (2005). Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nature Methods*, 2 (3), 207–12.
- Pauling, L. 1960. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press, Ithaca, New York.
- Perticaroli, S. *et al.* (2010). Broadband depolarized light scattering study of diluted protein aqueous solutions. *J. Phys. Chem. B*, 114, 8262–8269.
- Peters, C.W.; Kruse, U.; Pollwein, R.; Grzeschik, K.H.; Sippel, A.E. (1989). The human lysozyme gene. Sequence organization and chromosomal localization. *Eur. J. Biochem.* 182 (3), 507–16.
- Prassas II, Diamandis EP. (2008). Novel therapeutic applications of cardiac glycosides. *Nat Rev Drug Discov. Nov*; 7(11):926-35. doi: 10.1038/nrd2682. Epub Oct 24.
- Protein Data Bank. Available online: <http://www.rcsb.org/pdb>.
- Raman, C. V. (1928). A new radiation. *Indian Journal of Physics* 2, 387–398.
- Raman, C. V.; Krishnan, K. S. (1928). The optical analog of the Compton effect. *Nature* 121, 711.
- Rank D. H. and J. S. McCartney. (1948). *Opt. Soc. Am.* 38, 279
- Rank D. H. and R. V. Wiegand. (1942). *Opt. Soc. Am.* 32, 190.
- Rapaport, A.; Roussel, B.; Reich, H.J.; Adar, F.; Glebov, A.; Mokhun, O.; Smirnov, V.; Glebov, L. (2010). Very Low Frequency Stokes and Anti-Stokes Raman Spectra Accessible with a Single Multichannel Spectrograph and Volume Bragg Grating Optical Filters. *ICORS, AIP Conf. Proc.* 2010, 1267: 808–809.

- Ratheal IM , Virgin GK , Yu H , Roux B , Gatto C , Artigas P . (2010). Selectivity of externally facing ion-binding sites in the Na/K pump to alkali metals and organic cations. *Proc Natl Acad Sci U S A*. 26 ottobre; 107 (43): 18718-23.
- Rizzuti, B.; Daggett, V. (2013). Using simulations to provide the framework for experimental protein folding studies. *Archives of Biochemistry and Biophysics*, 531 (1-2), 128–35.
- Robert, B. (1996). Resonance Raman studies in photosynthesis – chlorophyll and carotenoid molecules. In *Biophysical Techniques in Photosynthesis.*; Amesz, J.; Hoff, A. J.,Eds.; Kluwer Academic Publishers, Dordrecht, pp. 161–176.
- Robson, B.; Vaithilingam, A. (2008). Protein folding revisited. *Progress in Molecular Biology and Translational Science*, 84, 161–202.
- Roh, J. H. *et al.* (2006). Influence of hydration on the dynamics of lysozyme. *Biophys. J.*, 91, 2573–2588.
- Rose, G.D.; Fleming, P.J.; Banavar, J.R.; Maritan, A. A backbone-based theory of protein folding. *Proceedings of the National Academy of Sciences of the United States of America* **2006**, 103 (45), 16623–33.
- Rupley, J. A. (1967). The Binding and Cleavage by Lysozyme of N-acetylglucosamine Oligosaccharides. *Proceedings of the Royal Society B* 167 (1009), 416–428.
- Sali, A. (1995). Modeling mutations and homologous proteins. *Curr Opin Biotechnol*, 6(4): p. 437-51.
- Saunders, R.; Deane, C.M. (2010). Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Research (Oxford University Press)*, 38 (19), 6719–28.
- Schaefer, M.; Bartels, C.; Karplus, M. (1998). Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. *Journal of Molecular Biology*, 284 (3), 835–48.

- Schlierf, M.; Li, H.; Fernandez, J.M. (2004). The unfolding kinetics of ubiquitin captured with single-molecule forceclamp techniques. *Proc. Natl Acad. Sci.*, 101, 7299–304.
- Selkoe, D.J. (2003). Folding proteins in fatal ways. *Nature*, 426 (6968), 900–4.
- Shinoda T, Ogawa H, Cornelius F, Toyoshima C. (2009). Crystal structure of the sodium-potassium pump at 2.4 Å resolution. *Nature. May 21;459(7245):446-50. doi: 10.1038/nature07939.*
- Shortle, D. (1996). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB Journal*, 10 (1), 27–34.
- Sotomayor, M.; Schulten, K. (2007). Single-molecule experiments in vitro and in silico. *Science*, 316, 1144–8.
- Spedding F. H. and R. F. Stamm. (1942). *Chem. Phys.* 10, 176
- Spiro, T. G.; Gaber, B. P. (1977). Laser Raman scattering as a probe of protein structure. *Annu. Rev. Biochem.* 46, 553–572.
- Stamm R. F. and C. F. Salzman. (1953). *Opt. Soc. Am.* 43, 126.
- Stammreich H., R. Forneris, and K. Sone. (1955). *Chem. Phys.* 23, 1972
- Stammreich H., R. Forneris, and Y. Tavares. (1956). *Chem. Phys.* 25, 580, 1277 and 1278
- Stammreich H.. (1950). *Experientia* 6, 224.
- Stammreich H.. (1950). *Phys. Rev.* 78, 79.
- Stammreich H.. (1956). *Spectrochim. Acta* 8, 41.
- Staple, D.B.; Payne, S.H.; Reddin, A.L.C.; Kreuzer, H.J. (2009). Stretching and unfolding of multidomain biopolymers: a statistical mechanics theory of titin. *Physical Biology*, 6, 025005. doi:10.1088/1478-3975/6/2/025005 (11pp).
- Turton, D.A.; Senn, H.M.; Harwood, T.; Laphorn, A.J.; Ellis, E.M.; Wynne, K. (2014). Terahertz underdamped vibrational motion governs protein-ligand

binding in solution. *Nature Communications*, 5, 3999.
doi:10.1038/ncomms4999

Tymoczko, J.L.; Stryer, L. (2002). Protein Structure and Function. *Biochemistry*.
W. H. Freeman, San Francisco.

Urabe, H.; Sugawara, Y.; Ataka, M.; Rupprecht, A. (1998). Low-frequency Raman
spectra of lysozyme crystals and oriented DNA films: dynamics of crystal
water. *Biophys. J.* 74, 1533–1540.

Van den Berg, B.; Ellis, R.J.; Dobson, C.M. (1999). Effects of macromolecular
crowding on protein folding and aggregation. *The EMBO Journal*, 18 (24),
6927–33.

Van den Berg, B.; Wain, R.; Dobson, C.M.; Ellis, R.J. (2000). Macromolecular
crowding perturbs protein refolding kinetics: implications for folding inside
the cell. *The EMBO Journal*, 19 (15), 3870–5.

Veskatesachar B. and L. Sibaiya. (1930). *Indian J. Phys.* 5, 747

Virgin, N. L. (2007). *Vibration of Axially Loaded Structures*. Cambridge University
Press.

Welsh H. L., M. F. Crawford, T. R. Thomas, and G. R. Love. (1952). *Can. J. Phys.*
30, 577.

Xie, W.C. (2006) *Dynamic Stability of Structures*. Cambridge University Press.

Yoshimura, K.; Toibana, A.; Nakahama. K. (1988). Human lysozyme: sequencing
of a cDNA, and expression and secretion by *Saccharomyces cerevisiae*.
Biochem. Biophys. Res. Commun. 150 (2), 794–801.

Zienkiewicz, O. C.; Taylor, R. L. (2005). *The Finite Element Method for Solid and
Structural Mechanics* 6th ed.; Elsevier, Amsterdam.