

**Deanship of Graduate studies
Al-Quds University**

**Analysis of the Secondary Structure of the
Transmembrane Domain of SARS CoV E Protein using
FTIR Spectroscopy**

By

Qassem Mousa Abo-Remeleh

M.Sc.Thesis

Jerusalem-Palestine

1428/2007

**Analysis of the Secondary Structure of the
Transmembrane Domain of SARS CoV E Protein using
FTIR Spectroscopy**

By

Qassem Mousa Abo-Remeleh

B.Sc. Biology and Medical Technology

College of Science and Technology-Palestine

Supervisors

Dr. Mutaz Akkawi

Dr. Abd-Alkreem Alsharif

**Thesis submitted in partial fulfillment of requirements
for the degree of Master of science, Department of
Biochemistry & Molecular Biology
Al-Quds University**

1428/2007

Al-Quds University
Deanship of Graduate Studies
Department of Biochemistry & Molecular Biology

Thesis Approval

Analysis of the Secondary Structure of the Transmembrane Domain of SARS CoV E Protein using FTIR Spectroscopy

Prepared By: Qassem Mousa Abo-Remeleh

Registration No.: 20410107

Supervisors: Dr. Mutaz Akkawi

Dr. Abd-Alkreem Alsharif

Master thesis submitted and accepted:

The names and Signatures of examining Committee members are as follows:

- | | | |
|-----------------------|--------------------------|----------------|
| 1. Head of Committee: | Dr. Mutaz Akkawi | Signature..... |
| 2. Cosupervisor: | Dr. Abd-Alkreem Alsharif | Signature..... |
| 3. Internal Examiner: | Dr. Mohammed Abo-Alhaj | Signature..... |
| 4. External Examiner: | Dr. Sameir Alnajde | Signature..... |

Jerusalem-Palestine

1428/2007

Dedication

To my great teachers

To my family

To my lovely wife

To my friends

Declaration:

I certify that this thesis submitted for the degree of Master is the result of my own research, except where otherwise acknowledged, and that this thesis (or any part of the same) has not been submitted for a higher degree to any other university or institution.

Signed.....

Qassem Mousa Abo-Remeleh

Date:

Acknowledgments

I would like to express my grateful thanks and appreciation for my supervisor Dr. Mutaz Akkawi for his direction and patience, for his support for his continuous encouragement, for all what he did. I would at the same time express my grateful appreciation for Dr. Abd-Alkareem Alsharif for his support and fruitful discussion concerning the results and the suggestions. I have to appreciate the help and support of the dearest person to me Shereen which her role was very supportive and important to the success of this work for her continues encouragement. I would like to express my thanks for Al-Quds University for doing this work at the biochemical research lab directed by Dr. Mutaz Akkawi.

I would like to thank the group of prof. Shy Arkin of Hebrew University for allowing us to perform some of the experiments at their laboratories.

I would like to thank also Monther Abou Remeleh for helping me to get some scientific papers and supporting me.

My heartfelt thanks go to my family and to my friends and to my teachers. This work could not have been accomplished without their support.

Table of contents

List of Tables	IX
List of Figures	XI
List of Appendices	XV
Abbreviations and Units	XVII
Abstract	XIX
CHAPTER I	1
1.1. The SARS Coronavirus.....	2
1.1.1. General Overview.....	2
1.1.2. SARS Coronavirus Envelope (E) protein.....	3
1.2. Membrane Proteins	5
1.2.1. The Biology of Membrane Proteins	5
1.2.2. General Structure of Membrane Proteins	6
1.2.3. Structural Analysis of Membrane Proteins	7
1.3. Peptide Synthesis and Purification	8
1.3.1. Peptide Synthesis.....	9
1.3.1.1. Liquid-phase Synthesis.....	9
1.3.1.2. Solid-phase Synthesis	9
1.3.1.2.1. Fmoc SPPS.....	10
1.3.2. Peptide Purification.....	12
1.3.2.1. Reverse Phase Chromatography	12
1.4. Vibrational Spectroscopy.....	13
1.4.1. ATR-FTIR Spectroscopy.....	13
1.4.2. Site-specific Isotope Labeling.....	16
1.4.3. Site-specific Infrared Dichroism	17
1.5. Prediction of Transmembrane Domains	20
1.6. Global Search Molecular Dynamics Simulation	20
CHAPTER 2	22
2. Purpose of the Work	22
2.1. Main Purposes of this Study:.....	23
CHAPTER 3	24
3. Materials and Methods.....	25
3.1. Instrumentation	25

3.2. Materials and Chemicals.....	26
3.3. Methods.....	28
3.3.1. ¹³ C= ¹⁸ O Labeling.....	28
3.3.2. Fmoc Protection.....	29
3.3.3. Protein Synthesis.....	31
3.3.4. Protein Purification.....	32
3.3.5. Lyophilization (Freeze Drying).....	35
3.3.6. Reconstitution of Peptide.....	35
3.3.7. FTIR Spectroscopy.....	35
3.4. Protein Prediction Programs.....	36
CHAPTER 4.....	38
4. Results and Discussion.....	38
4.1. Results.....	39
4.1.1. FTIR spectroscopy shows that the E protein of SARS CoV is α -helix.....	39
4.1.2. 26 residues of the E protein are embedded in the lipid bilayer.....	43
4.1.3. Data calculations and analysis.....	47
4.1.4. SARS CoV E protein has a unique transmembrane helical hairpin model with two kinks within the lipid bilayer.	52
4.1.5. Calculation of the dichroism (R).....	61
4.1.6. Protein Prediction Programs.....	61
4.1.6.1. The PSIPRED Protein Structure Prediction Server	62
4.1.6.2. Predict Protein Server.....	62
4.1.6.3. Membrane Protein Secondary Structure Prediction Server.....	63
4.2. Discussion.....	64
REFERENCES.....	73
APPENDICES.....	80

List of Tables

Table (3.1): The solvent system used for recrystallization of Fmoc protected and $^{13}\text{C}=^{18}\text{O}$ labeled amino acids.....	30
Table (3.2): The gradient of the pumped solvents used in the HPLC purification process.	33
Table (4.1) Determination of the Position of the peak's centers of the helix and the labeled site in H_2O saturated samples (A).....	46
Table (4.2): Determination of the Position of the peak's centers of the helix and the labeled site in H_2O saturated samples (B).....	47
Table (4.3): Determination of the Position of the peak's centers of the helix and the labeled site in D_2O saturated samples (A).....	49
Table (4.4): Determination of the Position of the peak's centers of the helix and the labeled site in D_2O saturated samples (B).....	50
Table (4.5): Calculation of the area under the peaks of the helix and the labeled site in H_2O saturated samples (A).....	54
Table (4.6): Calculation of the area under the peaks of the helix and the labeled site in D_2O saturated samples (A).....	55
Table (4.7): The dichroism ratio (R) for the labeled site and for the helix in both H_2O and D_2O saturated air samples (A).....	56

Table (4.8): Calculation of the area under the peaks of the helix and the labeled site
in H₂O saturated samples (B).....57

Table (4.9): Calculation of the area under the peaks of the helix and the labeled site
in D₂O saturated samples (B).....58

Table (4.10): The dichroism ratio (R) for the labeled site and for the helix in both
H₂O and D₂O saturated air samples (B).....59

List of Figures

Figure (1.1): Peptide synthesis protocol.....	11
Figure (1.2): FTIR spectrum shows Peaks that correspond to α -helical, β -strand and random coil peptide segment.....	15
Figure (1.3): The trapezoid ATR element and the incident polarized light.....	15
Figure (1.4): (a) A schematic diagram of a helix containing two C=O bonds.....	19
Figure (1.4): (b) The explicit diagram used in the mathematical derivations of a helix tilt.....	19
Figure (3.1): Schematic diagram shows the labeling system for the amino acids with $^{13}\text{C}=^{18}\text{O}$ labeling.....	29
Figure (3.2): Diagram of the SARS coronavirus E protein, outlining the synthesized transmembrane domain region from residue 7 to 38.....	31
Figure (3.3): Diagram of the synthesized transmembrane domain region from residue 7 to 38 of the SCoV E protein.....	31
Figure (3.4): Schematic diagram shows the gradient of the solvents used in the HPLC process.	33
Figure (3.5): HPLC chromatogram of the crude peptide using UV detector.....	34
Figure (3.6): HPLC chromatogram of the crude peptides using Fluorescence detector.....	34

Figure (4.1): ATR- FTIR spectrum of the SARS Coronavirus E protein transmembrane domain peptide reconstituted in lipid bilayer.	40
Figure (4.2): ATR- FTIR spectra of isotopically labeled site in SARS coronavirus E protein transmembrane domain peptide reconstituted in lipid bilayer at amino acid No. 21.....	41
Figure (4.3): ATR- FTIR spectra of isotopically labeled sites with $^{13}\text{C}=^{18}\text{O}$ in SARS Coronavirus E protein transmembrane domain peptide reconstituted in lipid bilayer focusing on the amide I region, showing all labels	42
Figure (4.4): ATR- FTIR spectra of isotopically labeled site in SARS coronavirus E protein transmembrane domain peptide reconstituted in lipid bilayer focusing on the amide I region and amide II region, showing the reduction in the amide II peak (H/D) exchange.....	44
Figure (4.5.a): Schematic diagram showing the Position of peak's centers of the helix and the labeled site with $^{13}\text{C}=^{18}\text{O}$ labels of the synthesized TMD of the SCoV E protein in ATR-FTIR spectrum in H_2O saturated air sample(A).....	48
Figure (4.5.b): Schematic diagram showing the Position of peak's centers of the helix and the labeled site with $^{13}\text{C}=^{18}\text{O}$ labels of the synthesized TMD of the SCoV E protein in ATR-FTIR spectrum in H_2O saturated air sample(B).....	48

Figure (4.6.a): Schematic diagram showing the Position of peaks centers of the helix and the labeled site with $^{13}\text{C}=^{18}\text{O}$ labels of the synthesized TMD of the SCoV E protein in ATR-FTIR spectrum in D_2O saturated air sample(A).....	51
Figure (4.6.b): Schematic diagram showing the Position of peaks centers of the helix and the labeled site with $^{13}\text{C}=^{18}\text{O}$ labels of the synthesized TMD of the SCoV E protein in ATR-FTIR spectrum in D_2O saturated air sample(B).....	51
Figure (4.7): Diagram of the SCoV E protein hydrophobic region outlining the two points with expected turn.....	53
Figure (4.8.a): Schematic diagram shows the dichroism ratio (R) for the labeled site and for the helix in both H_2O and D_2O saturated air samples(A).....	60
Figure (4.8.b): Schematic diagram shows the dichroism ratio (R) for the labeled site and for the helix in both H_2O and D_2O saturated air samples(B).....	60
Figure (4.9): The PSIPRED Protein Structure Prediction Server result.....	61
Figure (4.10): Schematic diagram for the PSIPRED Protein Structure Prediction Server results.	62
Figure (4.11): Predict Protein server results.....	62
Figure (4.12): Membrane Protein Secondary Structure Prediction Server results.....	63
Figure (4.13): Ribbon diagram of the structural model derived for the transmembrane domain of SCoV E protein proposed by (Arbely <i>et al</i>).....	65

Figure (4.14): Schematic diagram of the structural model derived for the transmembrane domain of SCoV E protein which contain two kinks in the position 26 and 31 Phe and Leu respectively.....	66
Figure (4.15): Schematic diagram of the structural model derived for the transmembrane domain of SCoV E protein which contain three expected kinks in the position No. 14, 26 and 31 Phe and Leu respectively.....	67
Figure (4.16): Negative staining electron micro-graphs of lipid DMPC Vesicles. (<i>Arbely et al</i>).....	68
Figure (4.17): Secondary structure probabilities for SARS-CoV E protein from Protein structure prediction.....	69
Figure (4.18): Simulated structure of a-helix	69

List of Appendices

Figure (1): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 13 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	82
Figure (2): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 14 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	83
Figure (3): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 17 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	84
Figure (4): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 18 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	85
Figure (5): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 19 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	86
Figure (6): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 20 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	87
Figure (7): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 21 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	88
Figure (8): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 22 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	89
Figure (9): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 23 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	90
Figure (10): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 24 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	91

Figure (11): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 25 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	92
Figure (12): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 26 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	93
Figure (13): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 27 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	94
Figure (14): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 28 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	95
Figure (15): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 29 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	96
Figure (16): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 31 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	97
Figure (17): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 33 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	98
Figure (18): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 34 labeled with ($^{13}\text{C}=\text{}^{18}\text{O}$).....	99

Abbreviations and Units

Abbreviations:

ATR-FTIR — Attenuated Total internal Reflection Fourier Transform

Infra Red.

ER — Endoplasmic Reticulum.

HPLC — High performance liquid chromatography

E protein — Envelope protein.

TFA — Trifluoroacetic acid

SPPS — Solid phase peptide synthesis

SSID — Site-specific infrared dichroism

Boc — Tert-butoxycarbonyl

M protein — Membrane protein.

S protein — Spike protein.

N protein — Nucleocapsid protein.

FTIR — Fourier Transform Infrared.

FMOC — 9-fluorenylmethoxycarbonyl.

FMOC-Cl — 9-fluorenylmethoxycarbonyl chloride.

HIV-1 — Human Immunodeficiency Virus type 1.

HCoV — Human Coronavirus.

IBV — Infectious Bronchitis Virus.

IR — Infrared.

MCoV — Murine Coronavirus.

MHV — Mouse Hepatitis Virus.

NMR — Nuclear Magnetic Resonance.

SARS — Severe Acute Respiratory Syndrome.

SCoV — SARS Coronavirus.

TGEV — Transmissible Gastroenteritis Virus.

TM — Transmembrane.

TMD — Transmembrane Domain.

GSMDS — Global searching molecular dynamics simulations.

E. coli — Escherichia Coli.

EmrE — E. coli multidrug resistance efflux transporter.

a. a— Amino acids

A— Alanine .

R— Arginine.

N— Asparagine.

D— Aspartic acid.

E— Glutamic acid.

G— Glycine.

I— Isoleucine.

L— Leucine.

K— Lysine.

P— Proline.

S— Serine.

T— Threonine.

F— Phenylalanine.

Y— Tyrosine.

V— Valine.

C— Cysteine.

M— Methionine.

PSIPRED — Protein Structure Prediction Server.

R — Dichroism.

A \parallel — The absorption of parallel polarized light.

A \perp — The absorption of perpendicular polarized light.

Units:

cm⁻¹ — A reciprocal centimeter (wavenumber) used for the energy of vibrational levels.

mm — millimeter is a unit of length in the metric system, equal to one thousand of a meter.

nm — A nanometer is a unit of length in the metric system, equal to 10⁻⁹ of the meter.

kDa — kilo Dalton is a unit of mass equal to 1000 Daltons.

Abstract

One of the major obstacles facing the field of structural biology in the postgenomic era is the inherent difficulty of solving the structure of membrane proteins under native conditions. Membrane proteins share a common property; part of their structure is embedded in the lipid bilayer. This feature makes them attractive drug targets, which requires a detailed knowledge of the secondary structure of their transmembrane domain. Both crystallography and NMR still encounter difficulties in handling membrane proteins, so there is an urgent need for new biophysical methods and new insights in the biophysics of membrane proteins to solve the secondary structure of such proteins.

The outbreak of the severe acute respiratory syndrome (SARS) virus, July 2003, has presented a formidable challenge for the scientific community. As part of that effort, we decided to study the high resolution backbone structure of E transmembrane proteins of the SARS coronavirus, by Attenuated Total Internal Reflection (ATR) FTIR of eighteen of isotopically labeled sites with ($^{13}\text{C}=^{18}\text{O}$) of the synthesized sequence for the SARS coronavirus E protein transmembrane domain. ATR-FTIR spectroscopy is a well-established method for generating precise structural information on isotopically labeled membrane proteins embedded in a lipid bilayer. We used the new biophysical method site specific infrared dichroism (SSID), to investigate the structure and orientation of transmembrane α -helical bundles.

We postulate in this work that the E protein of SARS CoV is α -helix, and it has 26 residues embedded in the lipid bilayer, and the SARS CoV E protein is not a regular helix, but it adopts a unique transmembrane helical hairpin model, and the E protein has two possible kinks at residue No. 26 and 31 Phe and Leu respectively within the lipid bilayer, which is reported for the first time in this thesis. And it also has a possible kink in residue No 15 too. All the results were confirmed experimentally.

الملخص بالعربية

من العوائق التي تواجه البيولوجيا البنائية (structural biology) إيجاد المبنى الخاص للبروتينات الموجودة في الأغشية والذي يحاكي وجودها في بيئتها الطبيعية ، ألا وهي طبقتي الدهون المكونة للأغشية في الخلايا. البروتينات الموجودة في الأغشية تكون مغروسة بين طبقتي الدهون، الكارهة للماء، وهذه الميزة جذبت العديد من علماء تصنيع الادوية، ولكن لتصميم هذه الادوية هناك حاجة ماسه إلى معرفه معلومات عن المبنى الثانوي لهذا الجزء من البروتين الموجود داخل طبقتي الدهون، من ناحية المبنى و وجهته داخل الغشاء. الطريقتان المعروفتان حتى الآن لحل المبنى الثانوي للبروتينات الموجودة في الاغشية هما طريقه الرنين المغناطيسي النووي (NMR) وطريقه التصوير السيني البلوري (crystallography) ، ولكن حتى هذه اللحظة هذه الطرق تعاني من صعوبات جمة في التعامل مع البروتينات الموجودة في الأغشية و حل مبناها، ولذلك هناك حاجة ماسه إلى استخدام طريقه بيوفيزيائية جديده لدراسة البروتينات، و خاصة المبنى الثانوي للبروتينات الموجودة بالأغشية.

إن انتشار مرض السارس (SARS) في يوليو عام 2003 قد كان له تأثير كبير على المجتمعات و الشعوب، فقد أصبح هناك تحد كبير بالنسبة للعلماء لحل هذه المشكلة، و كمساهمة منا قررنا القيام بدراسة مفصلة ودقيقه لأحد البروتينات الغشائية المهمة لفيروس السارس، المسمى البروتين المعطفي (Envelop (E) protein) ، وقد استخدمنا طريقه جديده تسمى (ATR-FTIR) (الانعكاس الداخلي الكلي الموهن باستخدام تحويلات فوريير للأشعة تحت الحمراء)، و في هذه الطريقه نستخدم دلائل معدلة النظائر ($^{13}\text{C}=^{18}\text{O}$) للأحماض الأمينية داخل البيبتيد المراد دراسته والذي نقوم ببنائه .

هذه الطريقه تعطي معلومات كثيرة ودقيقه عن المبنى الثنائي للبروتينات الموجودة في الأغشية و المغموسة في طبقتي الدهن، ولا تحتاج إلى كميات كبيره من البروتينات أو إلى وقت كبير كما هو في الطرق السابقة . تم التأكيد في هذا البحث على أن (E protein) الخاص بفيروس السارس هو عبارة عن بروتين لولبي فريد من نوعه ، يوجد منه ست و عشرون حامض أميني مغروسين داخل الغشاء في طبقتي الدهون، وهو عبارة عن بروتين لولبي غير اعتيادي فيه انحنائين داخل الغشاء في الطبقتين الدهنيتين . وكل النتائج تم تأكيدها بتجارب عملية.

CHAPTER I

INTRODUCTION

1.1. The SARS Coronavirus

The coronaviruses (order *Nidovirales*, family *Coronaviridae*, genus *Coronavirus*) are members of a family of large, enveloped, positive sense single-stranded RNA viruses that replicate in the cytoplasm of animal host cells, causing principally respiratory or enteric diseases but in some cases neurological illness or hepatitis. Individual Coronaviruses usually infect their hosts in a species specific manner, mainly via respiratory and fecal-oral routes [1-3].

1.1.1. General overview

In the spring of 2003, when it became clear that a new human Coronavirus was responsible for the Severe Acute Respiratory Syndrome (SARS) [4,5], Coronaviruses became much more recognized. The name Coronavirus, coined in 1968, is derived from the corona (crown) like morphology of these viruses. Based on serological cross reactivity, Coronaviruses are divided into three genera, usually referred to as group I, II, and III. More recent genome sequence analysis has confirmed this grouping, and the identification of the SARS Coronavirus (SCoV) as a new subfamily within the *Coronaviridae* [6, 7]. Coronaviruses are enveloped viruses with round virions of approximately 80 to 120 nm in diameter. The most distinctive feature of this viral family is the genome size. Coronaviruses have the largest genome of all RNA viruses (approximately 29.7 kb). The positive-stranded RNA is complexed with the basic nucleocapsid (N protein) to form a helical capsid found within the viral membrane. The membranes of all Coronaviruses contain at least three viral membrane proteins. These are (i) the Spike protein (S protein), that forms the glycoprotein spike on the virion surface, giving the virus its name; (ii) the membrane (M protein), a protein that spans the membrane three times and has a short N terminal ectodomain and a cytoplasmic tail; and (iii) the small and highly hydrophobic Envelope (E protein).

1.1.2. SARS Coronavirus Envelope (E) protein

The Coronavirus E protein is a small polypeptide, ranging from 8.4 to 12 kDa, which is only a minor constituent of virions. Owing to its small size and limited quantity, the E protein was recognized as a virion component much later than the other structural proteins. As a result, both the structure and function of the protein are currently unknown.

SCoV E protein differs substantially from other Coronavirus E proteins, as these proteins are extremely divergent across the Coronavirus groups (but are relatively conserved among members of a single group). Nevertheless, the same general architecture can be recognized in all E proteins: a short hydrophilic amino terminus (8–12 residues) followed by a large hydrophobic TMD (21–29 residues), and a hydrophilic carboxy terminus (39–76 residues) which constitutes most of the protein. The E protein of the SCoV E virus contributes in the pathogenesis of the virus for its importance in the viral budding. It is a small, 76 amino acid residues, integral protein with an unusually long hydrophobic stretch (25 - 30 amino acid residues), placed in between a hydrophilic N- and C-terminus, ≈ 8 and ≈ 40 residues long, respectively. This protein is very important for virion envelope morphogenesis [8–10], and membrane insertion of the E protein occurs without cleavage of a signal sequence [11]. It was shown that the E protein of the SCoV and other coronaviruses [12, 13] localizes to the Golgi apparatus and ER. The E protein has the ability to dramatically change the morphology of the Golgi apparatus, which could explain, in part, the ability of the E protein to induce apoptosis [11, 14]. Viral morphogenesis and budding has been studied extensively. It was found that in the SCoV [15] and other coronaviruses [11, 12, 16–18] that the expression of M and E proteins are sufficient for production of virus-like particles. This fundamental observation proved that neither the nucleocapsid nor the viral spike are needed for viral budding. Furthermore, expression of Infectious Bronchitis Virus (IBV) or Mouse Hepatitis Virus (MHV) E protein on its own causes the release of vesicles containing E protein, thereby pointing to the importance of coronavirus E protein in the budding process [12, 18]. Finally, mutations in MHV E protein cause marked morphological changes in the resulting viruses [19]. The information currently available about the role of the E protein in the budding process is not complete.

Consequently, the exact function and mechanism of action of this small protein are unknown. Even so, there is some evidence suggesting that the SCoV [20, 21] and the MHV [22, 23] E protein may function as a cation selective ion channel, which is more permeable to cations than to anions. This channel forming property is contained in the amino terminus TMD of the SCoV E protein, which were known by Affinity purified polyclonal antibodies recognizing the N-terminal [20]. Moreover, hexamethylene amiloride inhibited the Human Coronavirus (HCoV) and MHV E protein ion channel conductance in the lipid bi-layer and also inhibited replication of the parent Coronaviruses in cultured cells [23].

The length of the hydrophobic segment of SARS-CoV protein has posed a problem with respect to assigning a topology to that protein. The average length of a transmembrane α -helix is 21 residues [24], far shorter than the hydrophobic stretch of E proteins. In contrast, if a bend is formed by the long hydrophobic stretch, the resulting helices would be much shorter than average [25]. Further investigation is needed to resolve the backbone structure and orientation of the transmembrane domain of the SARS coronavirus E protein in lipid bilayer. An early report suggested a C terminus in the outer site of the membrane and the N terminus in the inner site of the membrane for the Transmissible Gastroenteritis Virus (TGEV) E protein [26]. More extensive investigations of the MHV and IBV E protein concluded that the carboxy terminus of the protein is cytoplasmic (i.e., is situated in the interior of the virion) [11, 12]. More recent results with an epitope-tagged Murine Coronavirus (MCoV) E protein [27] are consistent with a topology model in which the TMD spans the lipid bilayer twice, whereby both termini of the protein reside in the virus lumen. A similar model was suggested for the SCoV E protein using immunofluorescent staining by Q. Yuan *et al*, it was found that both the amino and carboxy terminus of the SCoV E protein are exposed to the viral lumen [28].

1.2. Membrane proteins

Membrane protein is a protein molecule that is attached to, or associated with the membrane of a cell or an organelle. More than half of all proteins interact with membranes.

1.2.1. The biology of membrane proteins

Integral membrane proteins are defined as having part of their structure the transmembrane domain (TMD) embedded in the lipid bilayer. Membrane proteins with at least one TMD account for 20 -30% of the proteins in almost every genome sequenced so far [29, 30]. In contrast to integral membrane proteins, non-integral membrane bound proteins usually bound to the membrane indirectly by interactions with integral membrane proteins or directly by interactions with lipid polar head groups. Although both types of proteins are known as membrane proteins. Transmembrane (TM) proteins are strategically located at the interface between the cytosol and the cell surrounding. As such, they mediate communication between both sides of the membrane and involved in the host biological functions.

One example is the relaying of signals across the cell membrane carried out by the vast family of G protein coupled receptors. These receptors function as molecular switches and transducers in a variety of signaling cascades and processes. Another major role of TM proteins is the transport of solutes across the membrane by pores, channels, and transporters. Examples of such proteins are (i) the Escherichia Coli (*E. coli*) multidrug resistance efflux transporter (EmrE), which confers resistance to a wide variety of toxicants by actively removing them from the cell; (ii) the Na⁺/H⁺ antiporter channel, which exchange one molecule for another, are involved in homeostasis of H⁺ and Na⁺ ions; (iii) the aquaporins (water channel) which conduct water molecules in and out through cell membranes.

Membrane proteins are also involved in processes important to the life of the cell, such as photosynthesis, and cell respiration. The high abundance and strategic location of membrane proteins has made them ideal targets for most of the pharmaceuticals in use today. This, in turn, requires detailed structural and biochemical information about membrane proteins.

Reliable information of this kind can be obtained by performing experiments in the natural environment of membrane proteins the lipid bilayer. This requirement is a major experimental obstacle, especially when it comes to structural biology of membrane proteins [31].

1.2.2. General structure of membrane proteins

Membrane proteins must perform complex signaling and transport functions within the strict confines of the lipid bilayer. But the lipid bilayer is a variable and its surface region is usually composed of charged groups, while the hydrocarbon core is highly hydrophobic. These conditions require a unique structural adaptation of membrane proteins to their environment.

As no hydrogen bonding groups are present in the hydrophobic environment of the membrane, formation of secondary structure hydrogen bonds is highly favorable. It is for this reason that within the lipid-embedded domain of membrane proteins only two major classes of secondary structures have been observed α -helical bundles and β -barrels. By far, the majority of membrane proteins belong to the former class composed of bundled α -helices. β -barrels, in contrast, are found only in the outer membranes of Gram-negative bacteria, mitochondria, and chloroplasts. The most important feature of both these structures is that they satisfy all their potential of creating backbone hydrogen bonding. Hydrogen bonds in α -helical structure are formed mostly between residues 1 and 1+4 of the same helical structure, while within the β -barrel structure; hydrogen bonds are formed between extended anti-parallel peptide strands.

The TMD is optimized to interact favorably with lipid bilayer by exposing a hydrophobic surface that reflects the physicochemical properties of the surrounding hydrophobic lipid bilayer. As a consequence, TM proteins usually have a hydrophobic stretch which is long enough to span the hydrocarbon core of a typical bi-layer. In case of the common α -helical structure, the average length of a TMD was found to be 21-22 amino acids [24], corresponds to the average width of the lipid bi-layer. This value is valid for a canonical α -helical structure, but Transmembrane Domains (TMDs) exhibit more complex architecture as exemplified by the structure of the glycerol /water channel GlpF [32]. Most notable in the structure is the pair of helices that penetrate half way into the bilayer. The structure also contains a highly distorted TM helix, which is a common feature of membrane proteins, about 60% of all TM helices contain significant bends or other distortions, as predicted by Yohannan et al. [33]. Other deviations from ideal α -helical structures are short stretches of π or 3_{10} helix [34, 34]. These distortions lead to a

deviation of the helix axis and/or a change in the periodicity of side chain positions. It is possible that such distortions allow the small structural adjustments needed for precise positioning of functional groups. Non perfect helical structure may also provide weak points in the helical rods that facilitate movements needed for protein function (e.g., helix-breaking properties of Proline residues) [33, 35].

Bitopic membrane proteins have one TMD, while polytopic membrane proteins have more than one α -helical TMD to integrate in the lipid bilayer. Bitopic membrane proteins are further categorized according to the orientation of their TMD. During biogenesis, the N-terminus of a type I integral membrane protein is in the Endoplasmic Reticulum (ER) lumen, whereas in a type II membrane protein the N-terminus is in the cytoplasm. A common structural feature found in many TM proteins is the high affinity of aromatic residues to the lipid-water interface. In polytopic membrane proteins and α -helical bundles these residues often form an aromatic belt a cluster of aromatic residues situated close to the lipid head groups

1.2.3. Structural analysis of membrane proteins

The increasing use of computational methods for structural based drug design, together with the need to better understand the mechanism of action of a protein under question, have created an urgent need to determine the three-dimensional structure of membrane proteins. Despite being such an important and abundant class of proteins, as of October 2006 there are only 116 unique atomic-resolution structures of membrane proteins (or 225 structures if including non- unique structures). This number is negligible compared to the thousands of structures of water soluble proteins. The complete three-dimensional structure of a protein at high resolution can be determined by X-ray crystallography. This technique requires the molecule to form a well ordered crystal which is not possible for all proteins. An alternative to X-ray crystallography is multidimensional nuclear magnetic resonance (NMR) spectroscopy. Using NMR spectroscopy structures of the proteins can be determined in solution. The interpretation of the NMR spectra of large proteins is very complex, so its present application is limited to small proteins (~15-25 kDa) [31]. Hence, elucidating the structure of membrane protein is still one of the biggest challenges facing

the field of structural biology today.

Many alternative approaches have emerged in order to tackle the problem of solving membrane proteins structure. This includes solid-state NMR spectroscopy and low-resolution techniques such as vibrational spectroscopy, electron microscopy, electron paramagnetic resonance, and theoretical calculations.

1.3. Peptide synthesis and purification

In order to study the secondary structure of the target peptide, using any of the methods mentioned above, it is important to synthesize this protein. The most accepted method for building peptides in the lab is the Solid-phase synthesis method. To start working with the synthesized peptide it is important to obtain a highly pure peptide, so purification of the synthesized peptide is the next step; using high performance liquid chromatography which will separate the pure peptide from impurities .

1.3.1. Peptide synthesis

peptide synthesis is the creation of peptides, which are organic compounds in which multiple amino acids bind via peptide bonds which are also known as amide bonds, for the purpose of peptide synthesis, amino acids can be considered as having two main functionalities to manipulate, the amino and carboxyl groups. Functional groupings are also present in the side chains of many of the principle amino acids. These functionalities must be protected so that they do not interfere with the formation of the peptide bond, peptides are synthesized by coupling the carboxyl group or C-terminus of one amino acid to the amino group or N-terminus of another, there are two way for the synthesis , the Liquid-phase synthesis and the Solid-phase synthesis[36].

1.3.1.1. Liquid-phase synthesis

Liquid-phase synthesis is used in large-scale production of peptides for industrial purposes synthesis of well known short peptides. It is a classical approach to peptide synthesis in solution, which has a long cycle time, and it has limitation in number of amino acids to be combined.

So it has been replaced in most labs by solid-phase synthesis.

1.3.1.2. Solid-phase synthesis

Solid-phase peptide synthesis (SPPS), pioneered by Merrifield in 1959, resulted in a paradigm shift within the peptide synthesis community. It is now the accepted method for building peptides and proteins in the lab in a synthetic manner. SPPS allows the synthesis of natural peptides, which are difficult to express in bacteria, the incorporation of unnatural amino acids, peptide/protein backbone modification, and the synthesis of D-proteins, which consist of D-amino acids, and synthesis $^{13}\text{C}=^{18}\text{O}$ amino acids.

Small solid beads, insoluble yet porous, are treated with functional units (linkers) on which peptide chains can be built. The peptide will remain covalently attached to the bead until cleaved from it by a reagent such as trifluoroacetic acid. The peptide is thus immobilized on the solid-phase, and can be retained during a filtration process, whereas liquid-phase reagents and by-products of synthesis are flushed away.

The general principle of SPPS is one of repeated cycles of coupling-deprotection. The free N-terminal amine of a solid-phase attached peptide is coupled to a single N-protected amino acid unit. This unit is then deprotected, revealing a new N-terminal amine to which a further amino acid may be attached (Figure (1)). There are two majorly used forms of SPPS -- Fmoc and Boc. Unlike ribosome protein synthesis, solid-phase peptide synthesis proceeds in a C-terminal to N-terminal fashion. The N-termini of amino acid monomers is protected by these two groups and added onto a deprotected amino acid chain [37].

1.3.1.2.1. Fmoc SPPS

This method was introduced by Carpino in 1972 and further applied by Atherton in 1978. Fmoc stands for Fluorenyl-methoxy-carbonyl which describes the Fmoc protecting group, first described as a protecting group by Carpino in 1970. The crucial link in any polypeptide chain is the amide bond, which is formed by the condensation of an amine group of one amino acid and a carboxyl group of another. Generally, an amino acid consists of a central carbon atom (called the α -carbon) that is attached to four other groups: hydrogen, amino group, carboxyl group, and a side chain group. The side chain group defines the different structures of amino acids. Certain side chains contain functional groups that can interfere with the formation of the amide bond. Therefore, it is important to mask the functional groups of the amino acid side chain, The N-termini of amino acid monomers is protected by Fmoc too. Removal of side-chain protecting groups and peptide from the resin is achieved by incubating in trifluoroacetic acid (TFA) [36].

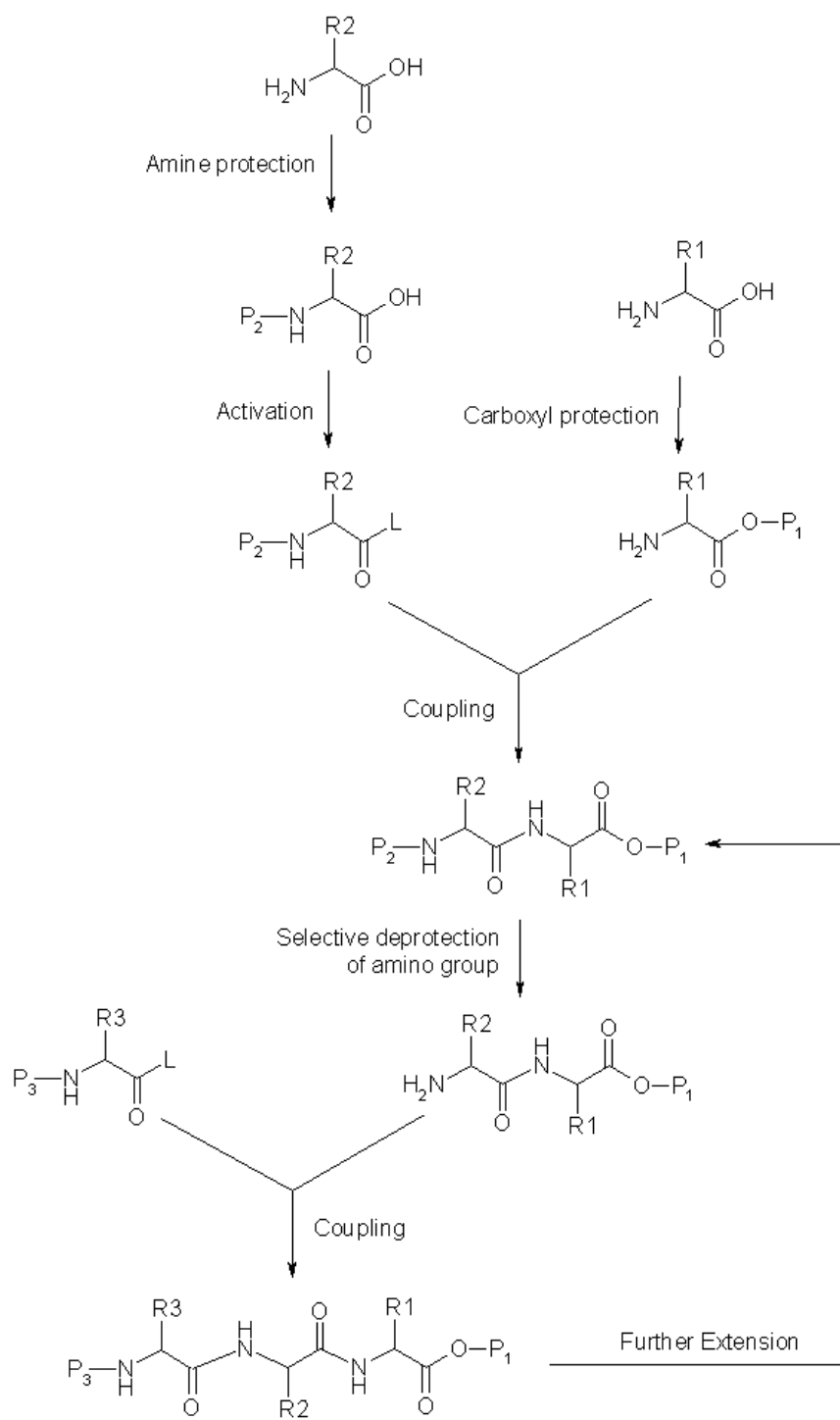


Figure (1.1): peptide synthesis protocol showing the four main steps: Protection activation, coupling and selective de-protection and further extension.

1.3.2. Peptide Purification

Since the purity of the peptide obtained from the synthesis is sequence dependent, major impurities can range from 10% to 50% by weight. They consist of small water soluble molecules, salts and protecting groups from the cleavage reaction, deletion peptides created due to incomplete coupling during synthesis, and modified peptides created during the cleavage. These species can be removed using reverse phase HPLC.

1.3.2.1. Reverse Phase Chromatography

In reverse phase chromatography, the packing is non-polar and the solvent is polar with respect to the sample. Retention is the result of the interaction of the non-polar components of the solutes and the non-polar stationary phase. Typical stationary phases are non-polar hydrocarbons, waxy liquids or bonded hydrocarbons, and the solvents are polar aqueous organic mixtures such as methanol-water or + acetonitrile-water. Silica is a reactive substrate to which various functionalities can be attached or bonded; the functionalities most widely bonded to silica are the alkyl (C_{18} , C_4 and C_8), aromatic phenyl, and cyano and amino groups. The retention time is therefore longer for molecules which are more non-polar in nature, allowing polar molecules to elute more readily. Retention time is increased by the addition of polar solvent to the mobile phase and decreased by the addition of more hydrophobic solvent. Reverse phase chromatography operates on the principle of hydrophobic interactions which result from repulsive forces between a relatively polar solvent, the relatively non-polar analyte, and the non-polar stationary phase. The driving force in the binding of the analyte to the stationary phase is the decrease in the area of the non-polar segment of the analyte molecule exposed to the solvent.

The hydrophobic effect is decreased by adding more non-polar solvent into the mobile phase. This shifts the partition coefficient such that the analyte spends some portion of time moving down the column in the mobile phase, eventually eluting from the column [38].

1.4. Vibrational spectroscopy

Vibrational spectroscopy is a collective term used to describe two analytical techniques – infrared and Raman spectroscopy. These techniques are used to measure the vibrational energy levels in a sample. These energy levels are associated with the chemical bonds in a sample. The spectrum (the output of the spectrometer) of a sample is unique, like a fingerprint.

1.4.1. ATR-FTIR spectroscopy

The Infra-Red (**IR**) range of the electromagnetic spectrum (2.5 μm - 700 nm) corresponds to the energies of vibrational transitions in molecules. These transitions are sensitive to the nature of the chemical bond (e.g., type of atoms, single- or double-bond, resonance), the chemical environment (e.g., hydrogen bonding, electrostatic properties, inter or intra-molecular interactions), and the orientation of the molecule relative to the IR radiation. These, together with the sensitivity of modern spectrometers, make Fourier Transform Infra-Red (FTIR) spectroscopy a popular tool for molecular analysis.

FTIR spectroscopy is particularly useful for membrane protein studies as it allows the analysis of such proteins as it is in their native environment, the lipid bilayer. This is because the lipid environment does not perturb the resolution or sensitivity of the spectra. It is also possible to simultaneously study the lipid bilayer, as there is no overlap between the absorption frequencies of the membrane and the embedded protein. By measuring the frequencies of the protein vibrational modes, mainly that of the amide I mode, one can determine the average secondary structure of the protein from its IR spectrum. Amide I is the most intense absorption band in proteins. It is primarily governed by the stretching vibration of the C=O (70-85%) and C-N groups (10-20%). Its frequency is found in the range between 1600 and 1700 cm^{-1} . The exact band position is determined by the backbone

conformation and the hydrogen bonding pattern.

Amide II is found in the 1510 and 1580 cm^{-1} region and it is more complex than amide I. Amide II derives mainly from in-plane N-H bending (40-60% of the potential energy). The rest of the potential energy arises from the C-N (18-40%) and the C-C (about 10%) stretching vibrations.

Amides III, V are very complex bands dependent on the details of the force field, the nature of side chains and hydrogen bonding. Therefore these bands are of little use. However, on the secondary structure of the protein, since the backbone C=O group is involved in different hydrogen bonding in different secondary structures. As a result, the experimentally observed amide I band can be generally classified as follows: 1662-1645 cm^{-1} for α -helices, 1689-1682, and 1637-1613 cm^{-1} for β -sheets, and 1682-1662 cm^{-1} for turns as shown in Figure (1.2). Among the various IR techniques, Attenuated Total internal Reflection Fourier Transform Infra-Red (**ATR-FTIR**) is one of the most useful methods for obtaining spectroscopic data on biological membranes [39]. In this type of measurements the incoming light beam is completely reflected when it impinges on the surface of the internal reflection element (Figure 1.3). Within this element, a standing wave is established normal to the reflecting surface. Consequently, the standing wave generates an electromagnetic disturbance that exists in the rarer medium beyond the reflecting interface. This so called evanescent wave is characterized by its amplitude, which falls off exponentially with the distance from the interface [40] and is affected (i.e., absorbed) by the sample on the surface of the ATR element.

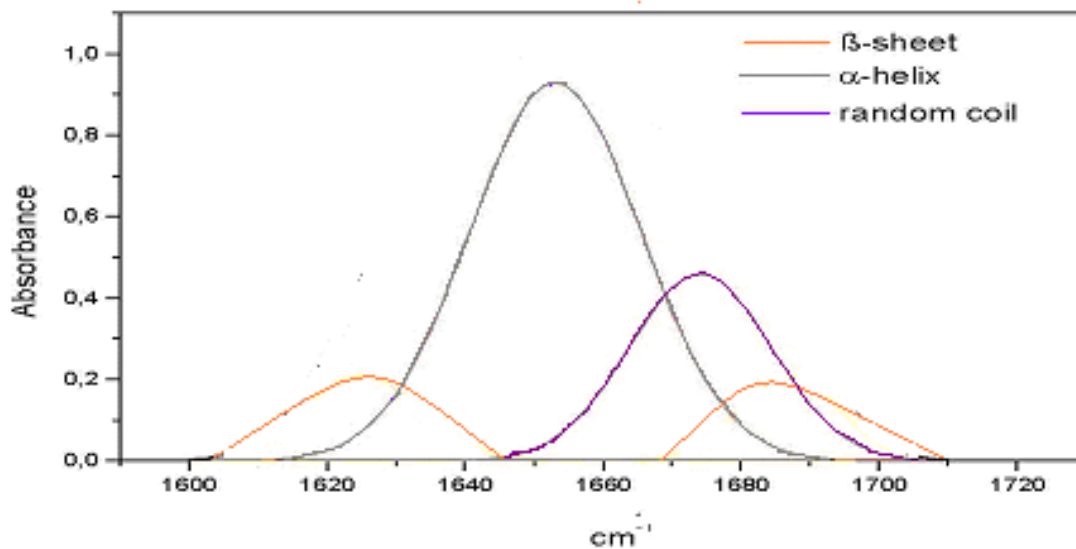


Figure (1.2): ATR-FTIR spectra corresponding to the amide I region shows Peaks of α -helical peptide segment in gray , β -strand peptide segment in orange and random coil peptide segment in violet.

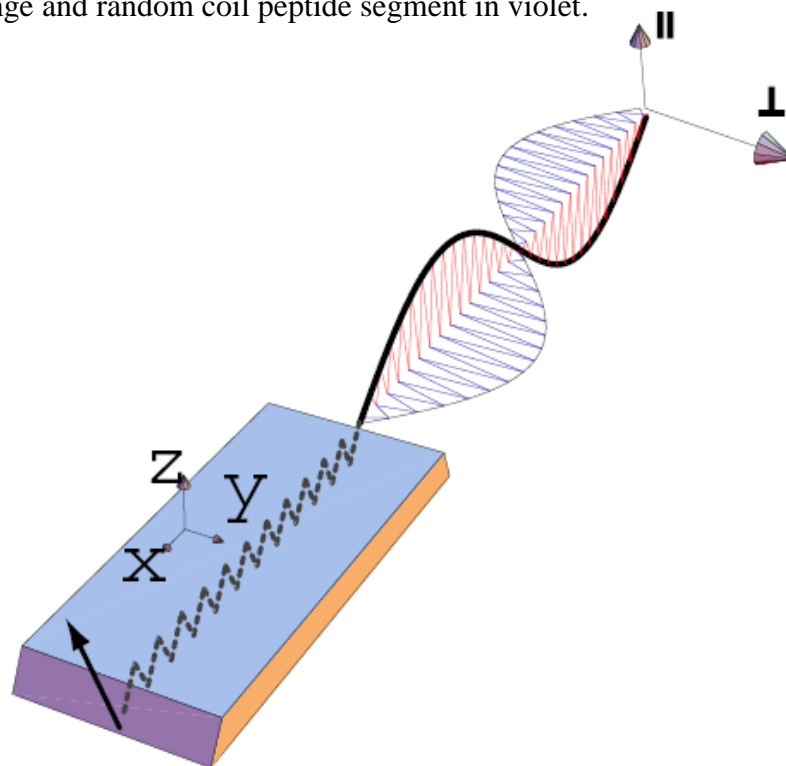


Figure (1.3): The trapezoid ATR element and the incident polarized light. The polarized light (parallel in red or perpendicular in blue) is reflected 22 times within the confines of the internal reflection element. The evanescent wave interacts with the sample deposited on the surface of the crystal.

The polarized Light can be defined as the light that has a dominant direction in on plane. Natural light is not polarized, while laser beam is polarized.

1.4.2. Site-specific isotope labeling

Low spectral resolution is one of the major drawbacks of IR spectroscopy as a tool for structural studies of biological macromolecules. Biological samples usually contain a population of chemical bonds which differs slightly from one another (e.g., the peptide bond). Because of this, the absorption band of the whole population is usually much wider than the absorption band of a single vibrational transition. The structural information obtained from such measurements is therefore an average property of the macromolecule.

A possible solution for the low resolution problem is the use of site-specific isotope labeling [41]. An isotopic label shifts the absorption band of the labeled site to a transparent region of the spectrum, without altering the chemical nature of the molecule. For example, site-specific isotope labeling of a backbone C=O group, together with the sensitivity of the Amide I vibrational mode to the protein secondary structure, allow the measurement of local conformation at the level of a single residue [42].

The isotopic labels that have been used so far are the GlyCD2 [43], the triple C-deuterated methyl group of alanine [44], and most frequently, the carbonyl group of the peptide bond. The carbonyl group can be labeled in two different ways with the $^{13}\text{C}=\text{}^{16}\text{O}$ label [45-48] or with the $^{13}\text{C}=\text{}^{18}\text{O}$ label [45, 49-51]. The $^{13}\text{C}=\text{}^{18}\text{O}$ label has two major advantages over the $^{13}\text{C}=\text{}^{16}\text{O}$ label.

1. The -40 cm^{-1} shift of the absorption band due to $^{13}\text{C}=\text{}^{16}\text{O}$ incorporation yields an isotope-edited mode present only as a shoulder on the unlabeled C=O peak. On the other hand, the frequency shift of -65 cm^{-1} arising from the double label places the $^{13}\text{C}=\text{}^{18}\text{O}$ absorption band so that it is nearly baseline resolved from the much larger unlabeled peak.

2. Because of the high natural abundance of $^{13}\text{C}=\text{}^{16}\text{O}$ (1.1%), in a spectrum of a protein over 45 amino acids long, 50% of the isotope-edited peak is due to naturally-occurring isotope. In contrast, the natural abundance of the $^{13}\text{C}=\text{}^{18}\text{O}$ label (0.0026%) is negligible, so that the IR absorption of the $^{13}\text{C}=\text{}^{18}\text{O}$ amide I normal mode is absolutely unique to the labeled site.

1.4.3. Site-specific infrared dichroism

Site-specific dichroism is a vibrational spectroscopy technique proposed by Arkin *et al.* [45], that utilizes site-specific isotope labeling and polarized IR light. Dichroism (Equation. 1.1) is defined as the ratio between the absorption of parallel and perpendicular polarized light:

$$\mathcal{R} \equiv \frac{A_{\parallel}}{A_{\perp}}$$

Equation (1.1): Dichroism which is defined as the ratio between the absorption of parallel and perpendicular polarized light

This new vibrational spectroscopy technique is based on site-directed isotope labeling and site-specific infrared dichroism (SSID). In SSID not only frequencies of the labeled sites measured, but the dichroism as well, yielding site-specific orientational restraints. The orientation of a single vibrational mode in an α -helix, represented as vector \mathbf{P} in Figure (1.4), can be expressed as a function of several parameters: (i) the helix tilt angle β , (ii) the

angles α and δ relating the helix axis and the vibrational mode, and (iii) the rotation angle ω of the helix about its axis.

Isotope label shifts the frequency of a specific vibrational mode so that its dichroism may be directly measured. This analysis is aided by the fact that the relative rotational pitch angles of the labeled sites are known: in α -helices, consecutive sites are related by a 100° rotation [44]. Subsequently, a structural model is derived by combination of experimental data, obtained from SSID, and implementation of the resulting restraints as energy refinement terms in a molecular dynamics simulation [46]. Site-directed dichroism enables the determination of the helix tilt angle, β , and pitch angle, ω , from two selectively labeled oriented samples.

Isotopically labeled probes are most suitable for the use of exploring the secondary structure of proteins in as much as they do not change the native properties of the protein, and yield interpretable structural information. SSID has already been used successfully in the studies of the transmembrane domain of several proteins [44]. The dichroic ratio of the helix is calculated by using the amide I absorption band of the canonically distributed α -helical C=O transition dipole moment centered at $\approx 1657 \text{ cm}^{-1}$ [39].

The dichroism of the labeled site is calculated by using the shifted vibrational mode (60 cm^{-1} for the $^{13}\text{C}=\text{}^{18}\text{O}$ Amide I label). The measured dichroism yields orientational restraints that are later used as energy refinement terms in molecular dynamics simulations).

Examples of the use of site-specific isotope labeling with ATR-FTIR are the studies on the transmembrane domain of GPA [44], Influenza AM2H+ Channel [47], Human Immunodeficiency Virus type 1 (HIV-1) vpu protein [46], Influenza C virus CM2 protein [48], Phospholamban [49], the T-cell receptor CD3- [38, 43-46, 51, 52], and the transmembrane domain of the trimeric major histocompatibility complex class II-associated invariant chain [45].

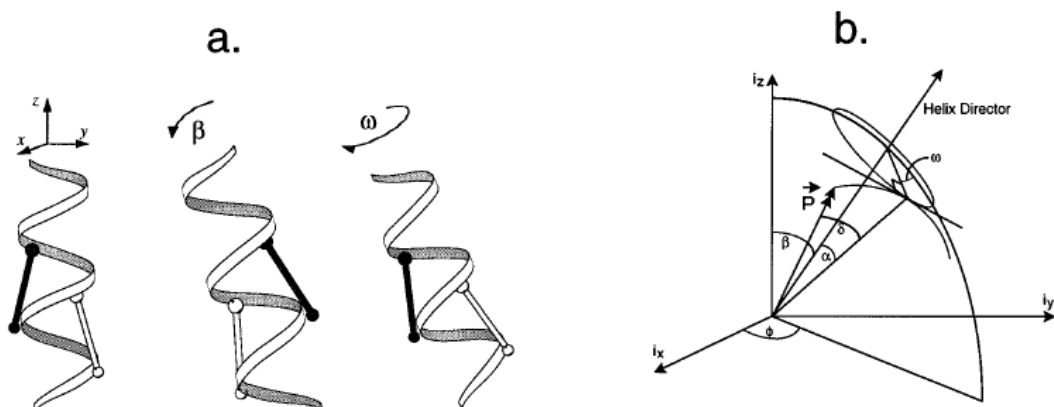


Figure (1.4): (a) A schematic diagram of a helix containing two C=O bonds (exaggerated in length for illustrative purposes). The relative orientation of the two bonds, with respect to the z axis, changes upon tilting the helix by the angle β , as well as changing the rotational pitch angle of the helix about its axis by the angle ω . (b) The explicit diagram used in the mathematical derivations of a helix tilted from the z axis by the angle β , containing a vibrating bond, \vec{P} , related to the helix director by the angles α and δ . The bond is positioned with a rotational pitch angle ω around the helix director.

1.5. Prediction of transmembrane domains

Although the lipid bilayer is an obstacle for structure determination by experimental methods, it has greatly simplified sequence based prediction of TMDs. Hydrophobic residues are a clear majority in transmembrane helices, which makes hydrophobicity an important feature in defining TM helices. The basic assumptions in TMD recognition are

that stretches of about 20 consecutive hydrophobic residues can span the membrane, and that such stretches can be easily recognized in protein sequences.

Many hydrophobicity scales have been developed over the past decade on the basis of solubility measurements of the amino acids in different solvents (i.e., water and hydrophobic solvent), vapor pressure of side chain analogues, analysis of side chain distribution within soluble and TM proteins, and theoretical energy calculations. Two of the most frequently used hydrophobicity scales are those introduced by Kyte and Doolittle in 1982 [53], and by Engelman *et al.* (the GES scale) in 1986 [54]. Another hydrophobicity scale was recently suggested by Zviling *et al.* [55], who used genetic algorithm for calculating an optimized hydrophobicity scale. When used alone, the simple hydrophobicity of TM residues can predict about 80% of TM helices. This is compared to much more accurate algorithms, based on neural-network [56] or hidden Markov model methods [30], which can predict >97% of TM helices. These predictors take advantage of evolutionary data from multiple sequence alignments and consider sequence elements in order to refine TMDs identification.

1.6. Global search molecular dynamics simulation

Molecular interactions between transmembrane α -helices can be explored using global searching molecular dynamics simulations (GSMDS), a method that produces a group of probable low energy structures, limiting the number of possible conformations. Brunger and co-workers [49, 57] have developed a procedure to explore transmembrane helix interactions based on global searching molecular dynamics simulations. In this method, multiple symmetric bundles of helices are constructed, each differing from the other by the rotation of the helices about their axes. These are then used as starting positions for molecular dynamics and energy minimization protocols. The output structures from these simulations are compared and grouped into clusters that contain similar structures. An average of the structures forming a cluster represents a model with characteristic interhelical interactions and helix tilt[58].

The correct model is selected amongst the several different clusters, based on existing experimental data, either from mutagenesis or orientational data from site specific infrared dichroism used as spatial restraints. Alternatively, they have also used a purely computational approach where simulations are performed on close sequence variants that are likely to share the same structure [59].

CHAPTER 2

PURPOSE OF THE THIS WORK

2.1. The main purposes of this study are:

- Analyzing the secondary structure of the transmembrane domain of SARS CoV E protein using a new biophysical method which is called (Attenuated Total Internal Reflection-Fourier Transform Infrared spectroscopy) (ATR-FTIR).
- Further study of the hairpin conformation of the E protein of SARS CoV as postulated by *Arbely et al* [26], in which he showed that the iodinated phenylalanine 23 was displaced about 16.5\AA from the bi-layer center in the lipid head group region.
- Using Multiple site specific infrared dichroism, in which the transmembrane domain had been labeled at multiple positions along the sequence with multi labeled amino acids with ($^{13}\text{C}=^{18}\text{O}$), which allow gathering of information at different points along the helix, and provides accurate information about the backbone of the secondary structure of the transmembrane proteins
- Suggest a model for the structure of the SCoV E protein TMD.

CHAPTER 3

MATERIALS AND METHODS

3. Materials and Methods

3.1. Instrumentation:

- * Systems for the synthesis of the peptide, which represents the transmembrane domain of SARS CoV E protein. That include the Fmocing system, to protect the N-terminus group of the amino acids, and system for amino acids labeling with ^{13}C = ^{18}O .
- * High Performance Liquid Chromatography (HPLC) Waters Breeze HPLC System with binary pump for the purification of the E protein with two detectors UV and fluorescence detectors using Jupiter 5μ C4-300Å column (Phenomenex, Chesire, UK)
- * Mass Spectrometer, Trace Ms, from Thermo Quest, Finnigen.
- * Fourier Transform Infrared Spectroscopy machine, Nicolet Avatar E.S.P. FT-IR Spectrometer from Thermo Electron Corporation, with DTGS detector.
- * Fourier Transform Infrared Spectroscopy machine, Nicolet Magna-560 infrared spectrometer (Nicolet Instrument, Madison) purged with N_2 and equipped with a MCT-A detector cooled with liquid nitrogen.
- * Attenuated Total internal Reflection unit (ATR) a 25 reflections ATR accessory from Graseby Specac (Kent, UK) and a wire grid polarizer (0.25 μM , Graseby Specac).
- * Germanium (Ge) trapezoidal internal reflection element (50×2×10 mm).
- * Integration program origin pro 7.0 (OriginLab Corporation. Northampton, MA) software for the peak fitting and integrating areas under peaks.

- * Freeze Drying machine (Lyophilizer) from Labconco.
- * Rotary Evaporator from Labconco.

- * Bench top pH meter from (WTW).

- * Analytical balance from Sartorius CP.

- * Ultra pure water system from NANO pure Diamond.

3.2. Materials and chemicals

- * 1,2-dimyristoyl-sn-glycero-3-phosphocholine, (DMPC), was obtained from Avanti Polar Lipids in powder form.

- * Hexafluoro-2-propanol, purity $\geq 99.8\%$ HPLC grade (HFIP), was obtained from Merck.

- * Trifluoroacetic Acid, (TFA), for HPLC, $\geq 99.0\%$ (GC), was obtained from Aldrich.

- * Anhydrous dioxane, 99.8%, was obtained from Aldrich.

- * Heavy water, H_2^{18}O , (95% ^{18}O), was obtained from ISOTECH.

- * Argon gas.

- * Sulfuric acid, purity 99.9% was obtained from Aldrich.

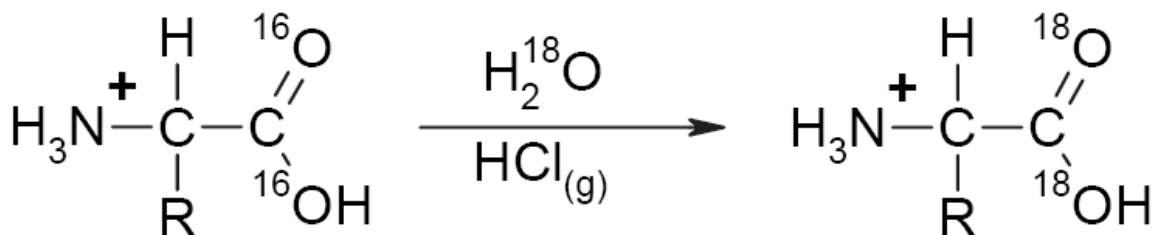
- * Calcium chloride anhydrous, 99.99%, was obtained from Aldrich in powder form.
- * Hydrochloric acid, 37%, AR grade, was obtained from Aldrich.
- * Acetonitrile, Chromasolv grade, for HPLC, $\geq 99.9\%$, was obtained from Sigma-Aldrich.
- * 2-Propanol, Chromasolv grade, for HPLC, $\geq 99.8\%$ (GC), was obtained from Sigma-Aldrich.
- * Potassium Carbonate, anhydrous, 99.99%, was obtained from Aldrich in powder form.
- * 9-fluorenylmethoxycarbonyl-Cl, (Fmoc-Cl), was obtained from Fluka, Buchs, Switzerland.
- * Ether, was obtained from Sigma-Aldrich, St. Louis, Mo, USA.
- * Ethyl acetate, HPLC grade, $\geq 99.8\%$ was obtained from Sigma-Aldrich, St. Louis, Mo, USA.
- * Magnesium sulfate, anhydrous, was obtained from Mallinckrodt Baker, St. Louis, MO, USA.
- * Petroleum ether ACS reagent was obtained from Sigma-Aldrich.

3.3. Methods

3.3.1. $^{13}\text{C}=^{18}\text{O}$ labeling

All labeling reactions were performed in dry environment and under continuous flow of Argon gas to prevent any evaporation from the reaction. In a typical experiment, 3.57×10^{-3} mol of amino acid was dissolved in a mixture of 10 ml anhydrous 1, 4-dioxane (Sigma-Aldrich, St. Louis, Mo, USA) and 1 ml of H_2^{18}O . The mixture was stirred in ice-cold bath and acidified by bubbling dry HCl (g) Equation (3.1). Dry HCl (g) was generated by drop wise addition of 10 ml H_2SO_4 (99.9%) to 10 gr of CaCl_2 , and bubbling the evolved HCl (g) through the reaction mixture for 20 min. The acidified mixture was stirred under reflux for 2 hr as shown in figure (3.1), cooled to room temperature, evaporated under reduced pressure and lyophilized over night.

The labeling was confirmed by mass spectroscopy for the amino acids and also for the heavy water.



Equation (3.1): labeling reactions equation shows labeling of the amino acids with ^{18}O using the heavy water H_2^{18}O .

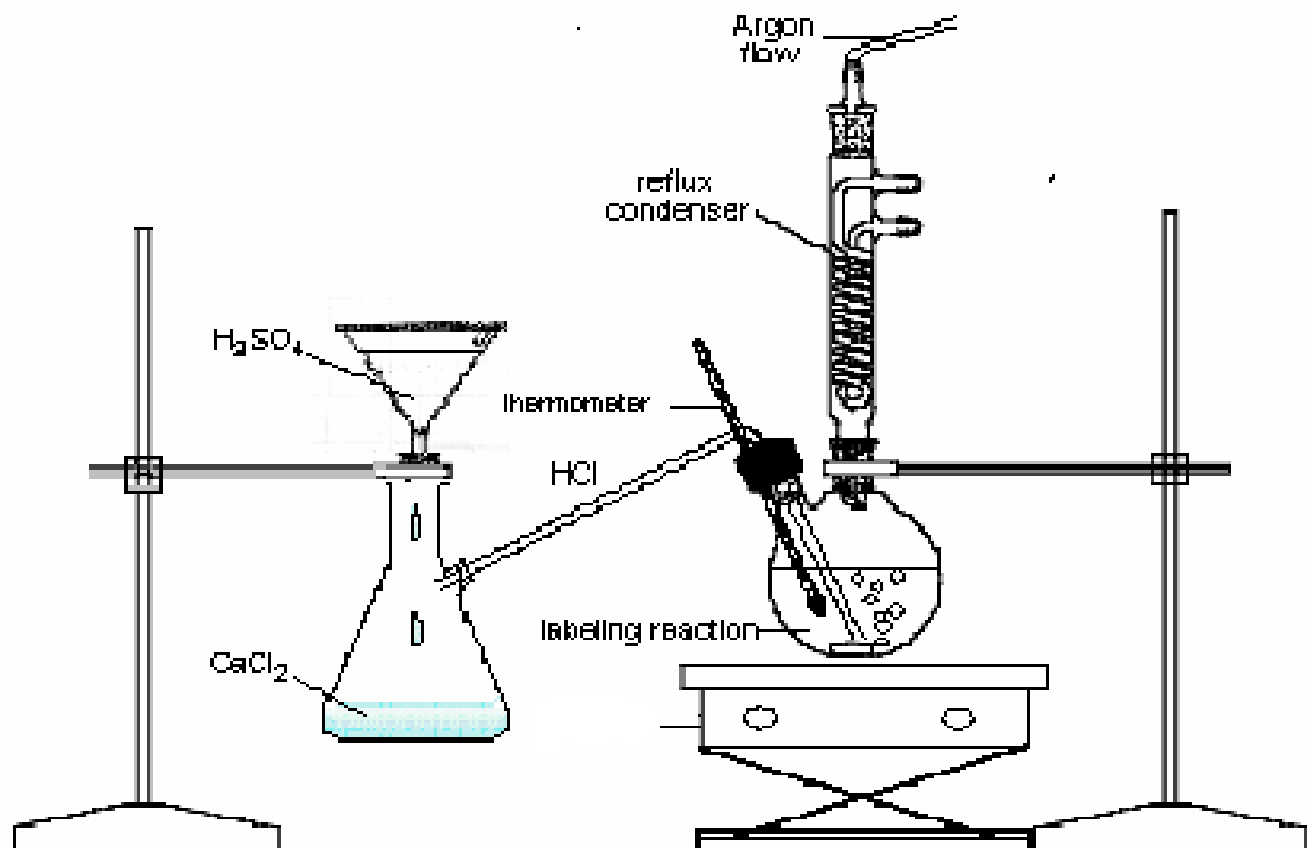
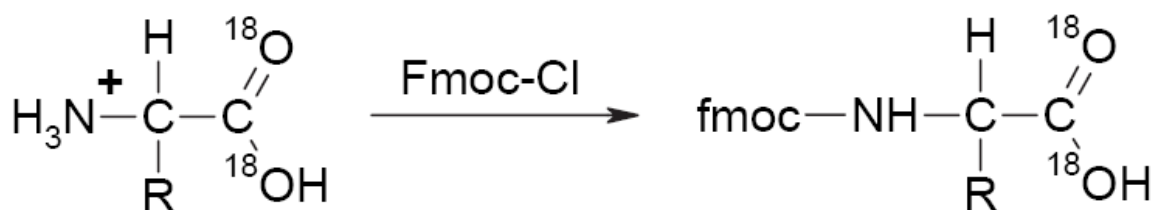


Figure (3.1): schematic diagram shows the labeling system for the amino acids with $^{13}\text{C}=^{18}\text{O}$ labeling.

3.3.2. Fmoc protection

The product of the $^{13}\text{C}=^{18}\text{O}$ labeling reaction (3.57×10^{-3} mol of amino acid) was used for the Fmoc protection reaction Equation (3.2) without any further purification. The amino acid was dissolved in a mixture of 25 ml K_2CO_3 solution (13 gr/100 ml; BDH Laboratory Supplies, Poole, England) and 11 ml 1, 4-dioxane (Sigma-Aldrich, St. Louis, Mo, USA). (14.3 gr) of Fmoc-Cl (Fluka, Buchs, Switzerland) was dissolved in 11 ml 1, 4-

dioxane and the solution was added drop wise to the reaction mixture, stirred in an ice bath. The reaction was allowed to warm to room temperature and stirred over night. The solution was then diluted with water, washed three times with ether (Sigma-Aldrich, St. Louis, Mo, USA), and the combined ether extracts washed twice with brine. The combined water extracts were cooled in an ice-bath and acidified by adding dilute HCl (aq) and washed three times with ethyl acetate (Sigma-Aldrich, St. Louis, Mo, USA). The combined ethyl acetate extracts were washed with brine, dried over MgSO₄(s) (Mallinckrodt Baker, St. Louis, MO, USA), filtered, evaporated under reduced pressure and lyophilized over night. The dry product was recrystallized using the solvent system described in Table (3.1).



Equation (3.2): Fmoc protection reaction Equation to protect the N-terminus group of the amino acids and system for the labeled amino acids with $^{13}\text{C}=^{18}\text{O}$.

Fmoc-amino acid	Solvent system
Fmoc-Ala-OH	EtOAc – Petroleum ether
Fmoc-Gly-OH	CH ₂ Cl ₂ – Petroleum ether
Fmoc-Ile-OH	CH ₂ Cl ₂ – Petroleum ether
Fmoc-Leu-OH	CH ₂ Cl ₂ – Petroleum ether
Fmoc-Phe-OH	EtOAc – Petroleum ether
Fmoc-Val-OH	CH ₂ Cl ₂ – Petroleum ether

Table 3.1: The solvent system used for recrystallization of Fmoc protected and $^{13}\text{C}=^{18}\text{O}$ labeled amino acids. Recrystallized amino acids were filtered

and dried over night under reduced pressure.

3.3.3. Protein synthesis

Peptides encompassing the predicted transmembrane domain of SARS coronavirus E protein (Figure (3.2), residues 7 to 38 (Figure (3.3) were synthesized by standard solid-phase N-(9- fluorenyl) methoxycarbonyl chemistry in Keck Foundation Biotechnology Resource Laboratory (Yale University). The peptides were cleaved from the resin with trifluoroacetic acid (TFA, Aldrich) and lyophilized.

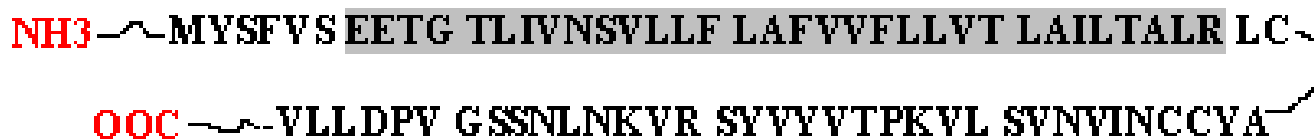


Figure (3.2): diagram of the of SARS coronavirus E protein, outlining the synthesized transmembrane domain region from residue 7 to 38.

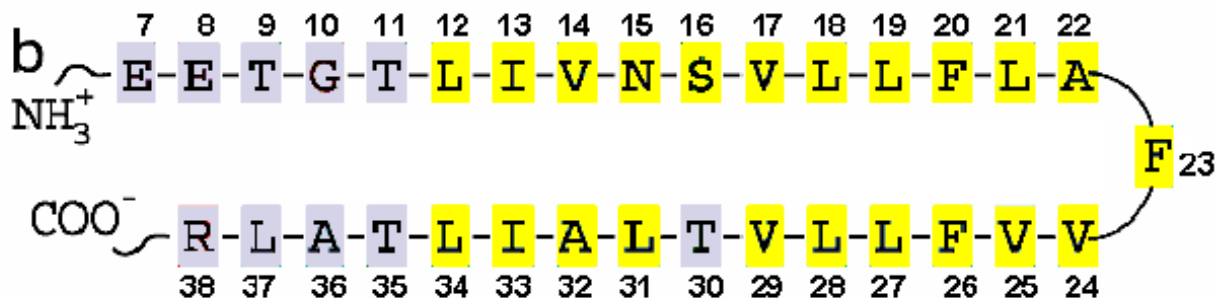


Figure (3.3): diagram of the synthesized transmembrane domain region from residue 7 to 38 of the SCoV E protein outlining the labeled residues by yellow shading.

3.3.4. Protein purification

The crude peptide was dissolved in trifluoroacetic acid (TFA) (final concentration about 5 mg/ml) and immediately injected on Jupiter 5 μ C4 300Å HPLC column phenomenex in order to be purified with the semi-preparative reversed phase high pressure liquid chromatography. The products were equilibrated with 80% H₂O, 8% (v/v) acetonitrile (and 12% (v/v) 2- propanol (Peptide elution was achieved with linear gradient to a final solvent composition of 40% (v/v) acetonitrile, and 60% (v/v) 2-propanol. All solvents contained 0.1% (v/v) TFA. The eluent was collected over a period of 5 minutes from the 20th minute to the 25th minute as shown in Figure (3.4), Table (3.2). In each case, the primary peak which contains the synthesized TMD of SARS CoV E protein was well separated from the other peaks. As seen in Figure (3.5) the peptide peak and the TFA peak were both detected by the UV detector, but in the fluorescence detector the peptide peak was more easily detected because of the presence of Phenylalanine (Phe) which has fluorescence properties at (excitation 257nm, emission 282nm) as shown in Figure (3.6). Peptide purity was confirmed by mass spectrometry.

Time(min)	Hydrophilic phase %	Hydrophobic phase %	Flow rate: ml/min
Initial	80	20	0.1
1	80	20	3.0
9	80	20	3.0
21	0	100	3.0
28	0	100	3.0
30	80	20	3.0
33	80	20	3.0
34	80	20	0.1

Table (3.2): The gradient of the pumped solvents used in the HPLC purification, the hydrophilic phase has the water, and the hydrophobic phase has the acetonitrile and 2-propanol.

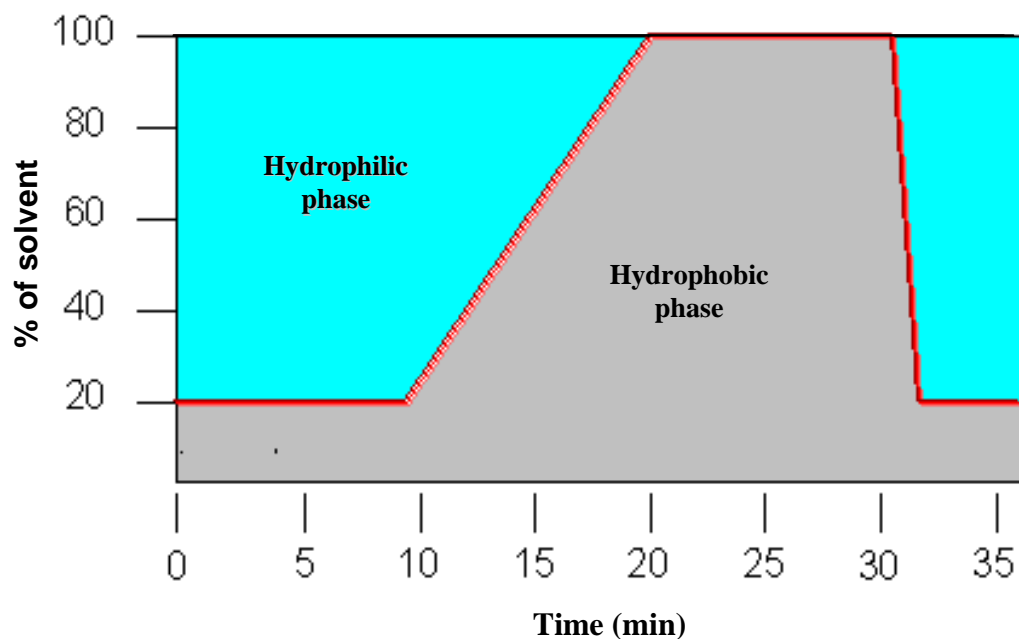


Figure (3.4): Schematic diagram shows the gradient of the solvents used in the HPLC process; the hydrophilic phase has the water, where the hydrophobic phase has the acetonitrile and 2-propanol.

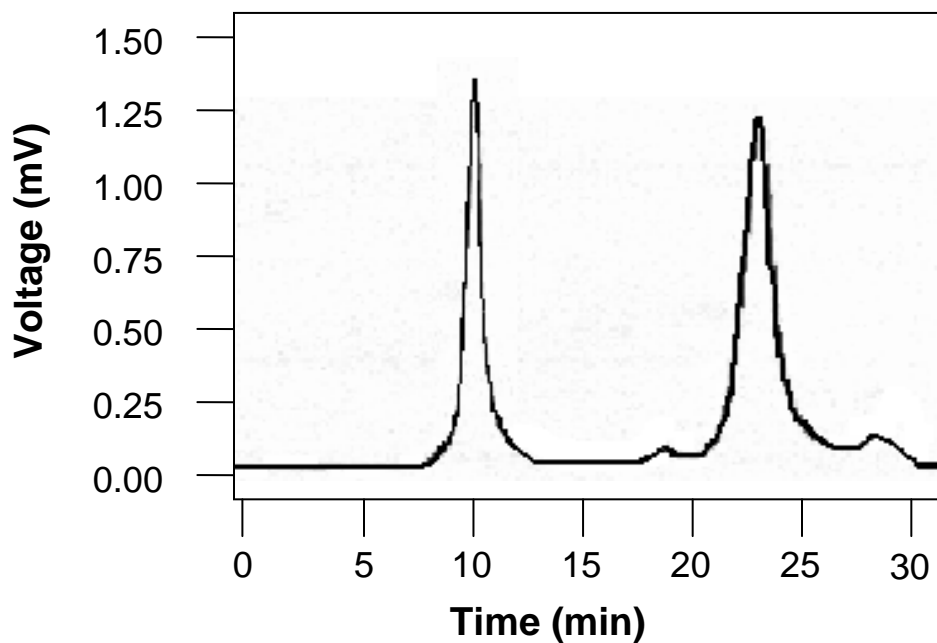


Figure (3.5): HPLC chromatogram of the crude peptide show the peptide peak obtained at the 20-25 min region and the trifluoroacetic acid peak obtained at the 8-13 min region using UV detector.

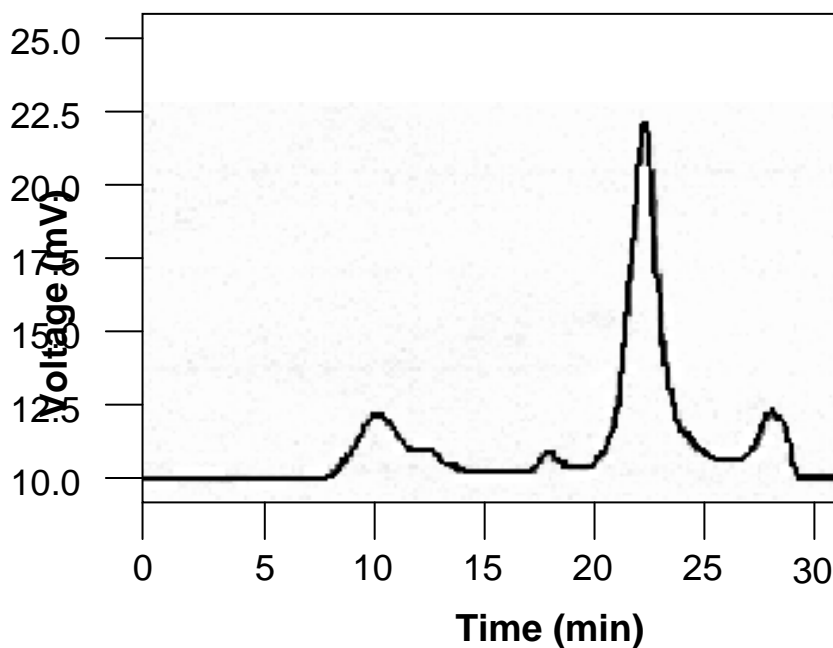


Figure (3.6): HPLC chromatogram of the crude peptides shows the peptide peak at the 20-25 min using fluorescence detector.

3.3.5. Lyophilization (Freeze Drying)

Lyophilization is the process of isolating a solid substance from solution by freezing the solution and evaporating the ice under vacuum, it is carried out using a simple principle of physics called sublimation. Sublimation is the transition of a substance from the solid to the vapor state, without passing through an intermediate liquid phase.

To lyophilize the purified peptide, it was frozen by putting it a in round bottom flask using liquid nitrogen and rotating machine, in the presence of 10 mM HCl (to remove traces of TFA adducts) then it was left over night at 3×10^{-3} m bar in the dry freezing machine.

3.3.6. Reconstitution of peptide in the vesicles

After over night lyophilization of pooled fractions in the presence of 10 mM HCl, 1 mg of the dried peptide was dissolved in a solution that contain, 10 mg dimyristoylphosphocholine (DMPC, Avanti) in 900 ml of 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, Merck). The solvent was evaporated over night under reduced pressure, by means of a rotary evaporator. One milliliter of H₂O was added to the dried product and the solution was mixed for 20 minutes at 30 °C, then the vesicles were spread on the Germanium trapezoidal internal reflection element, and dried for 30 min in dry air flow.

3.3.7. FTIR spectroscopy

Data were recorded on a Nicolet Magna-560 infrared spectrometer (Nicolet Instrument Corporation, USA) purged with dry air and equipped with an MCTA detector, cooled with liquid nitrogen. A total of 1000 interferograms were collected at a resolution of 2 cm⁻¹. Attenuated total reflection (ATR) spectra were measured with a 25 reflections ATR

accessory from Graseby Specac (Kent, UK) and a wire grid polarizer (0.25 μM , Graseby Specac). Three hundred microlitres of sample (1 mg/ml of peptide and 10 mg/ml of lipid) were deposited onto a Ge trapezoidal internal reflection element (50 \times 2 \times 10mm) flowed by removal of bulk solvent. For the purpose of solvent exchange, air was followed over the sample that was bubbled through $^2\text{H}_2\text{O}$ or H_2O . Transmission FTIR spectra were collected by depositing 50 μl of sample (1 mg/ml of peptide and 10 mg/ml of lipid) on a CaF_2 window. The dichroic ratios of the Amide I band were calculated by integrating between 1670 cm^{-1} and 1645 cm^{-1} . All integrations were performed by using a straight baseline that contains points immediately before and after the band. Reported standard deviations values represent a minimum of two data sets measured.

3.4. Protein prediction programs

Computational studies have been successful at predicting helical membrane protein topologies, i.e., identifying helical TM domains and predicting their in/out orientation relative to the membrane. It was based on the hypothesis that hydrophobic protein portions could form stable structures across the bilayer using hydrophobic helices.

Two general observations have been useful for predicting TM regions and their topologies. I. Hydrophobic residues are enriched in TM helical segments where they traverse the hydrophobic region of a membrane. II. Cytoplasmic segments contain significantly more positive charges than periplasmic segments.

Predict protein server is an Internet service for sequence analysis and the prediction of protein structure and function. Users submit protein sequences or alignments. The servers

return multiple sequence alignments predictions, for transmembrane helices, or coiled-coil regions, or predictions of transmembrane strands, via e-mail to the users.

We used three different prediction servers:

1. The PSIPRED Protein Structure Prediction Server.
2. Predict Protein server.
3. Membrane Protein Secondary Structure Prediction Server.

CHAPTER 4

RESULTS AND DISCUSSION

4.1. Results

4.1.1 FTIR spectroscopy evidence that the E protein of SARS CoV is an α -helix

In an effort to better understand the structure of the TMD of SARS CoV E protein, we decided to structurally analyze the transmembrane domain of E protein using FTIR spectroscopy. After over night lyophilization of pooled fractions in the presence of 10 mM HCl to remove traces of TFA adducts. ATR-FTIR spectra were obtained using both parallel and perpendicular polarized light. In the FTIR spectrum (1600– 1700 cm^{-1}) it is important to focus on the amide I vibrational mode as shown in Figure (4.1, 4.2, and 4.3). The amide I vibrational mode comes mostly from the peptidic C=O stretch, which correlates to the secondary structure of the protein. Peaks resonating at 1655 cm^{-1} , 1630 cm^{-1} and 1645 cm^{-1} correspond to α -helical, β -strand and random coil peptide segments, respectively [39]. The locations of both the amide I and amide II bands are sensitive to the secondary structure content of a protein.

The amide I mode of SCoV E protein TMD is centered at 1657 cm^{-1} with a peak width at half height of 23 cm^{-1} , which indicates that the E protein has a very high helical content, and the absence of any other secondary structure component.

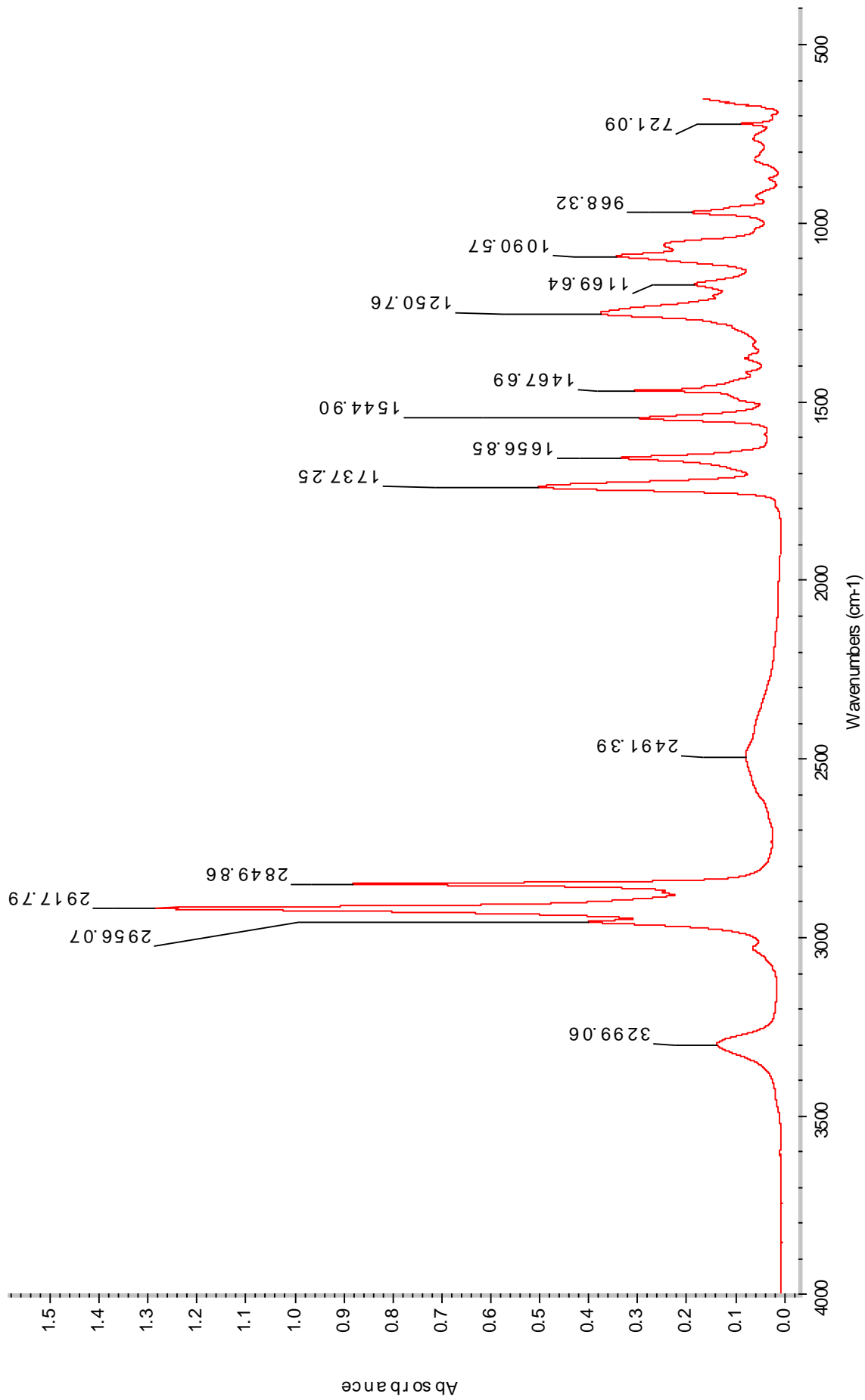


Figure (4.1): ATR- FTIR spectrum of the synthesized SARS coronavirus E protein transmembrane domain peptide reconstituted in lipid bilayer.

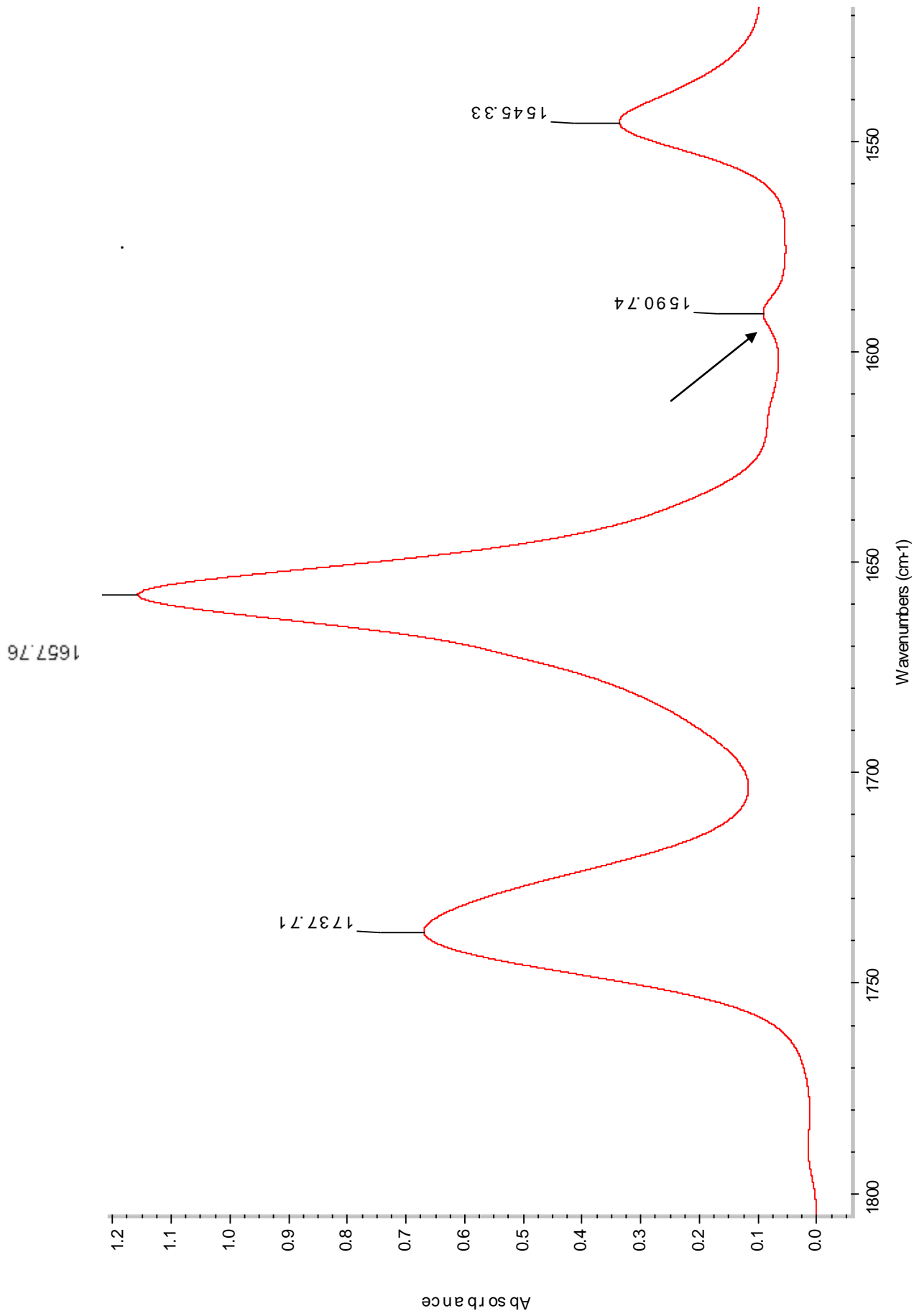


Figure (4.2): ATR- FTIR spectra of the synthesized isotopically labeled site in SARS coronavirus E protein transmembrane domain peptide reconstituted in lipid bilayer focusing on the amide I region. labeled with $^{13}\text{C}=\text{}^{18}\text{O}$ labels at amino acid No. 21 (the pointer indicate the labeled site).

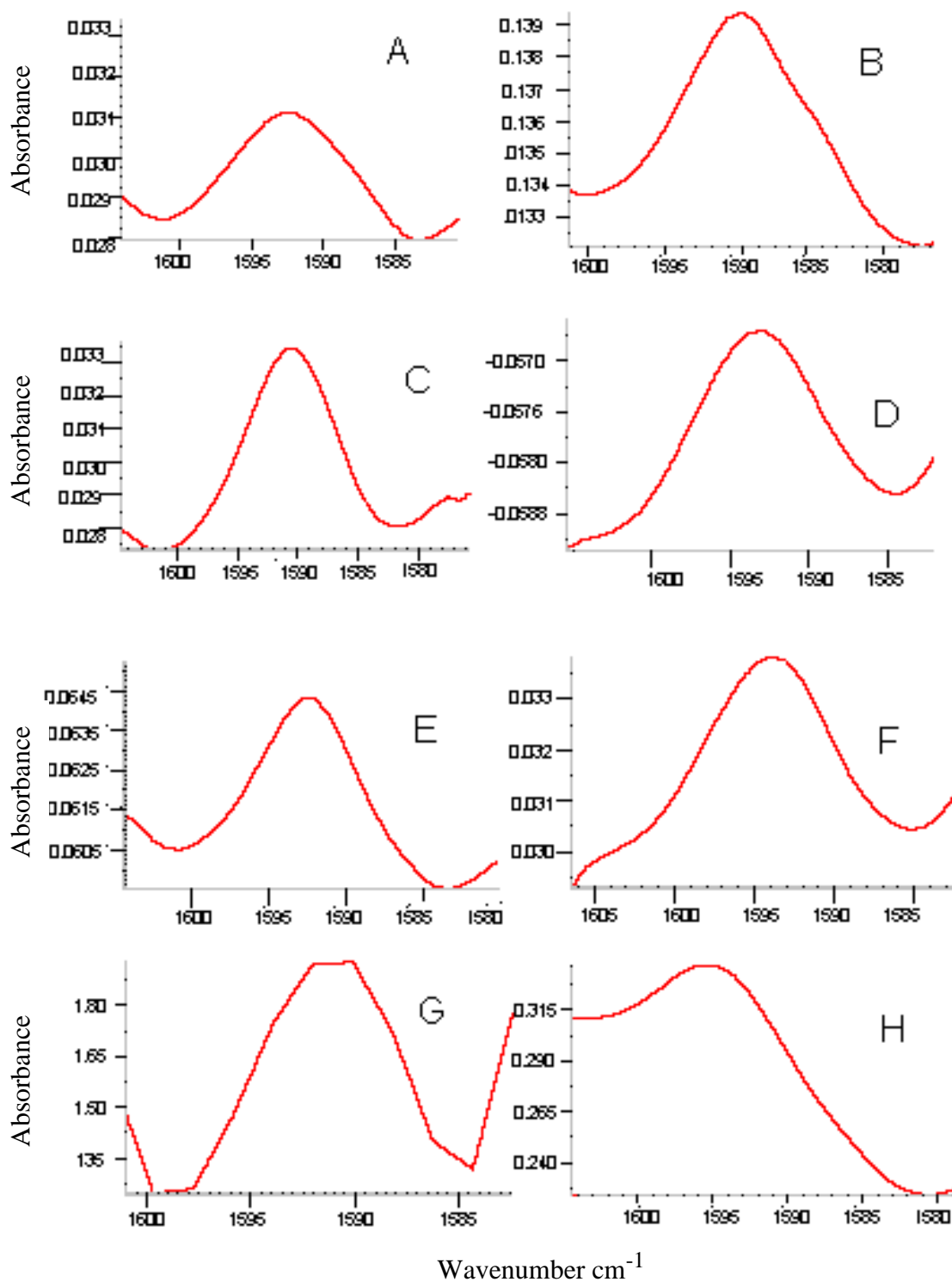


Figure (4.3): ATR- FTIR spectra of isotopically labeled sites in SARS coronavirus E protein transmembrane domain peptide reconstituted in lipid bilayer focusing on the amide I region, with $^{13}\text{C}=^{18}\text{O}$ labels at amino acids A: 13, B: 17, C: 18, D: 21, E: 24, F: 25, G: 28, H: 33.

4.1.2. 26 Residues of the E protein are embedded in the lipid bilayer

Hydrogen isotope exchange is becoming a powerful tool for studying the protein embedded in the lipid bilayer, where the exchange of the hydrogen with the deuterium that present in the heavy water occurs. The deuterium has nearly identical chemical properties to that of the hydrogen [51].

When a polypeptide folds to form the well-defined secondary structure of a protein, many of the amide hydrogens become buried in the lipid bilayer and no longer come into direct contact with the solvent water. This causes a marked decrease in the rate of exchange for these hydrogen.

FTIR spectroscopy is a good method to describe the extent of membrane incorporation of transmembrane proteins. This can be done by observing the reduction in any vibrational mode containing significant contributions from the amide proton (e.g II mode: peptidic N–H deformation) by changing the solvent from H₂O to D₂O. Because the lipid bilayer protects the peptide from exchanging the hydrogen (¹H) with heavy hydrogen (deuterium, ²H), and the exchange occurs only in the areas exposed to the solvent. All the vibrational modes that contain the N–D group will shift elsewhere and H/D exchange can be quantitated by measuring the reduction in the amide II mode directly, upon flushing the membrane with air saturated with D₂O for several hours, the reduction in the amide II peak, centered at 1545 cm⁻¹, is reduced as shown in Figure (4.4), amide I is primarily governed by the stretching vibration of the C=O (70-85%) and C-N groups (10-20%), where the amide II which is more complex derived mainly from N-H bending (40-60% of potential energy), the rest potential energy arises from the C-N (18-40%) and C-C (about 10%) stretching vibrations[52].

We can calculate the extent of exchange by normalizing both spectra (in H₂O and D₂O) on the amide II peak. The resulting exchange rate of 19 ± 2% is indicative of only six out of 32 residues in the peptide undergoing H/D exchange. This means 26 residues of SCoV E protein are protected from exchange as a result of their being embedded in the lipid bilayer.

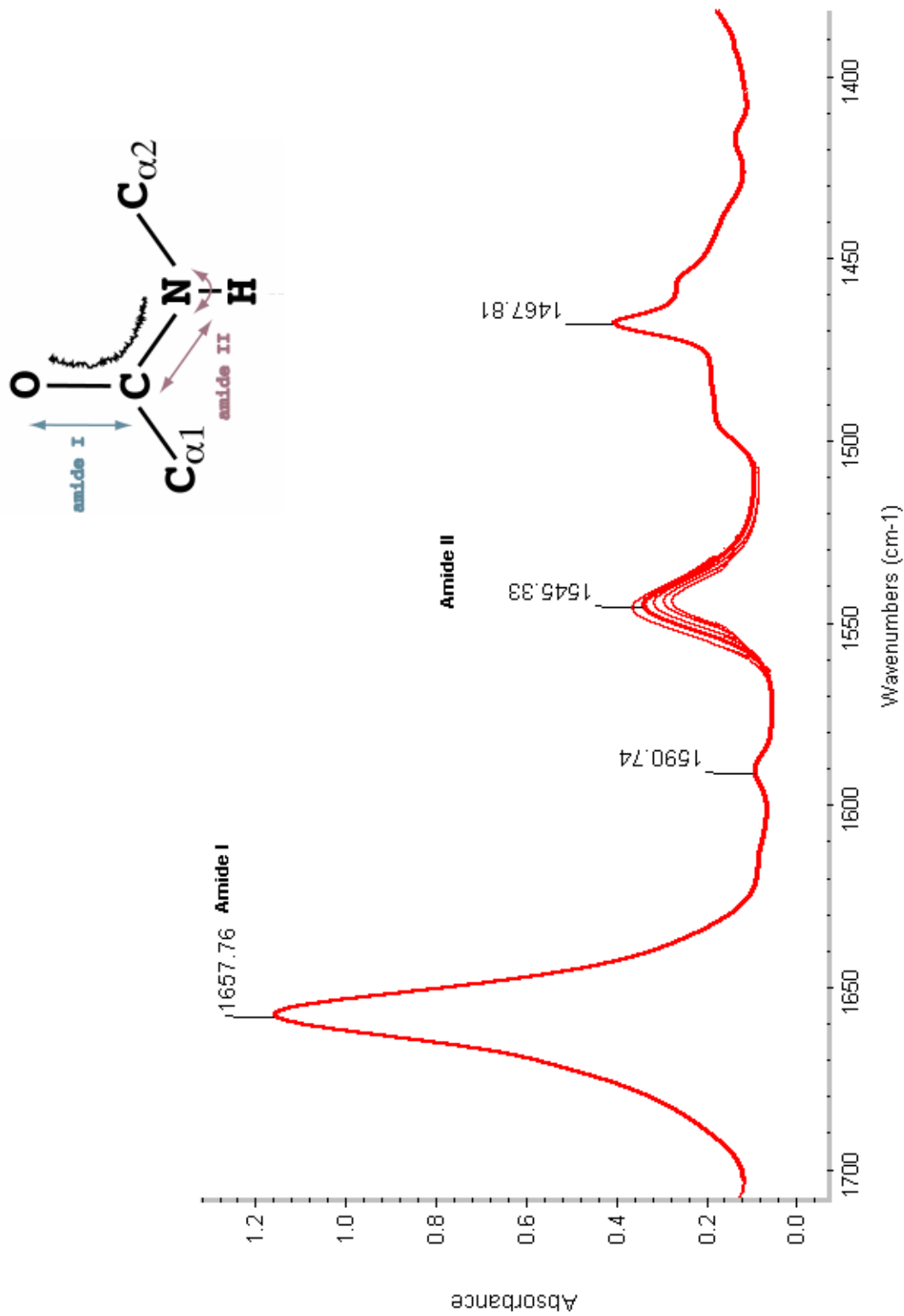


Figure (4.4): ATR- FTIR spectra of isotopically labeled sites in SARS coronavirus E protein transmembrane domain peptide reconstituted in lipid bilayer focusing on the amide I region and amide II region, showing the reduction in the amide II peak, centered at 1545 cm⁻¹, where there is no change in the amide I which centered at 1657 cm⁻¹, after the exposure to D₂O.

4.1.3. Data analysis and calculations

Calculation of the data measured by the ATR- FTIR spectroscopy was analyzed using an advanced computer program (OriginPro 7 (SR2) from OriginLab™). This program can integrate and fit the peaks of interest.

Using this program, we determined the position for each peak center of the helix and for the labeled sites (with $^{13}\text{C}=\text{}^{18}\text{O}$ at residues No. 12 to 14, 17 to 29, 31, 33, and 34 of the SARS E protein), The average value for the center of the peaks was found to be at 1657 cm^{-1} for the helix, and at 1592 for the labeled sites, in saturated air with both H_2O and D_2O . As shown in tables (4.1-4.4).

The area under the helix and the sites labeled with $^{13}\text{C}=\text{}^{18}\text{O}$ were determined in the synthesized TMD of the SARS Coronavirus E protein using the same program. Integrations of the area under the peaks of the helix and the site started from 1702 cm^{-1} to 1620 cm^{-1} for the helix peak, and from 1602 cm^{-1} to 1582 cm^{-1} for the labeled sites peaks, in saturated air with H_2O or D_2O . As shown in tables (4.5, 4.6, 4.8, 4.9).

Amino acid	position	Labeled Site peak center in cm^{-1}	Helix peak center in cm^{-1}
Leu	12	--	1657
Ile	13	1592	1658
Val	14	1596	1659
Val	17	1590	1658
Leu	18	1591	1658
Leu	19	1593	1659
Phe	20	1596	1659
Leu	21	1593	1657
Ala	22	1595	1658
Phe	23	1596	1659
Val	24	1592	1659
Val	25	1593	1657
Phe	26	1585	1657
Leu	27	1590	1660
Leu	28	1591	1658
Val	29	1596	1657
Leu	31	1581	1659
Ile	33	1597	1658
Leu	34	1607	1659

Table (4.1): Determination of the Position of the peak's centers of the helix and the sites labeled with $^{13}\text{C}=^{18}\text{O}$ labels in H_2O saturated samples (A).(residue No. 12 has very weak label, and did not has peak for the labeled site)

Amino acid	position	Labeled Site peak center in cm ⁻¹	Helix peak center in cm ⁻¹
Leu	12	--	1659
Ile	13	1592	1658
Val	14	1596	1659
Val	17	1590	1658
Leu	18	1591	1657
Leu	19	1593	1659
Phe	20	1596	1659
Leu	21	1593	1657
Ala	22	1595	1658
Phe	23	1596	1659
Val	24	1592	1658
Val	25	1593	1657
Phe	26	1584	1657
Leu	27	1590	1659
Leu	28	1591	1658
Val	29	1597	1657
Leu	31	1581	1658
Ile	33	1597	1658
Leu	34	1605	1658

Table (4.2): Determination of the Position of the peak's centers of the helix and the sites labeled with ¹³C=¹⁸O labels in H₂O saturated samples (B). (residue No. 12 has very weak label, and did not has peak for the labeled site)

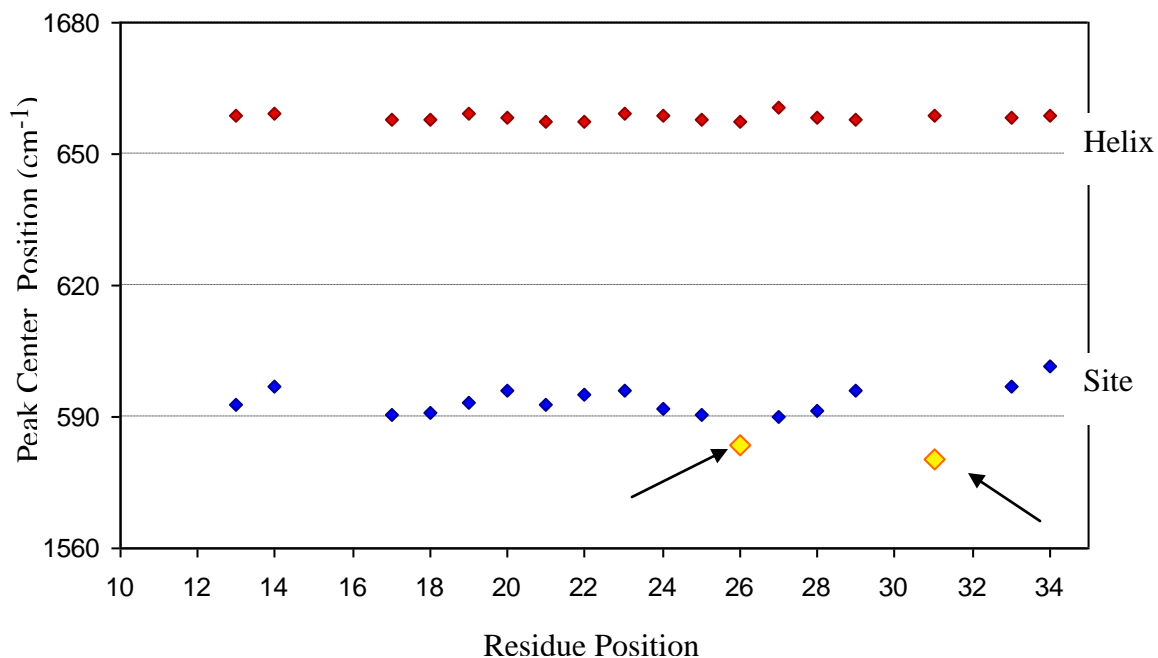


Figure (4.5.a): Schematic diagram showing the Position of peak's centers of the helix and the labeled site with $^{13}\text{C}=^{18}\text{O}$ labels of the synthesized TMD of the SCoV E protein in ATR-FTIR spectrum in H_2O saturated air sample table (4.1), the two arrows pointed at residues No. 26 and 31 Phe and Leu respectively.

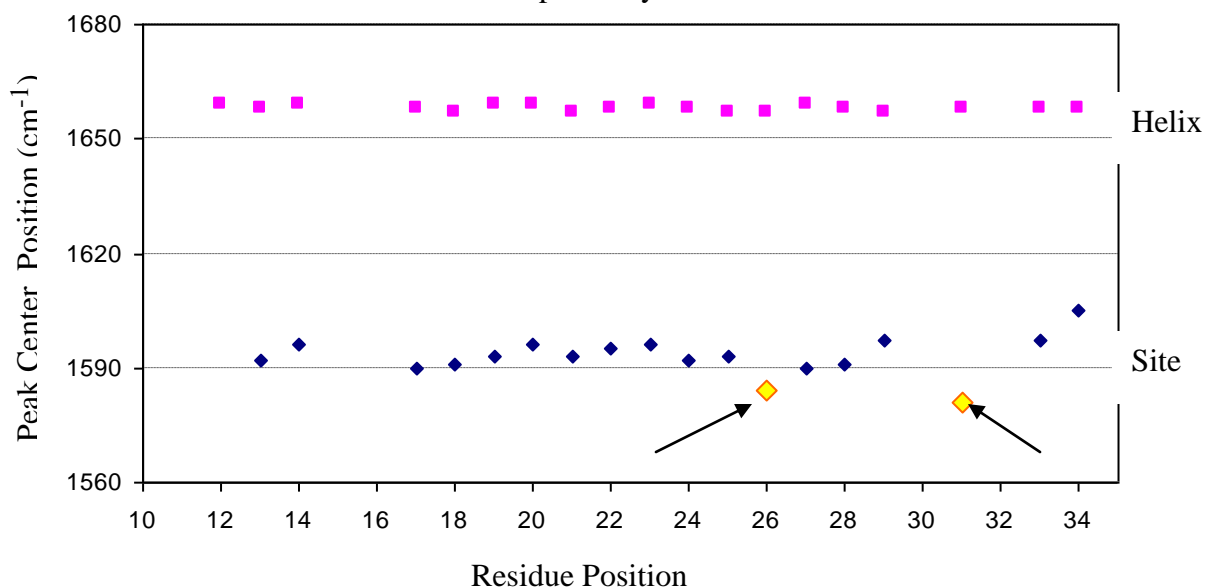


Figure (4.5.b): Schematic diagram showing the Position of peak's centers of the helix and the labeled site with $^{13}\text{C}=^{18}\text{O}$ labels of the synthesized TMD of the SCoV E protein in ATR-FTIR spectrum in H_2O saturated air samples table (4.2), the two arrows pointed at residues No. 26 and 31 Phe and Leu respectively.

Amino acid	position	Labeled Site peak center in cm⁻¹	Helix peak center in cm⁻¹
Leu	12	---	1656
Ile	13	1593	1659
Val	14	1597	1659
Val	17	1591	1658
Leu	18	1591	1658
Leu	19	1593	1659
Phe	20	1596	1658
Leu	21	1593	1657
Phe	23	1596	1659
Val	24	1592	1659
Val	25	1591	1658
Phe	26	1584	1657
Leu	27	1590	1660
Leu	28	1591	1658
Val	29	1596	1658
Leu	31	1580	1659
Ile	33	1597	1658
Leu	34	1602	1659
Average		1592	1658

Table (4.3): Determination of the Position of the peak's centers of the helix and the sites labeled with $^{13}\text{C}=^{18}\text{O}$ labels in D_2O saturated samples (A). (Residue No. 12 has very weak label, and did not has peak for the labeled site)

Amino acid	position	Labeled Site peak center in cm⁻¹	Helix peak center in cm⁻¹
Leu	12	---	1658
Ile	13	1593	1659
Val	14	1597	1659
Val	17	1590	1658
Leu	18	1591	1658
Leu	19	1593	1659
Phe	20	1596	1658
Leu	21	1593	1658
Ala	22	1595	1657
Phe	23	1596	1659
Val	24	1592	1659
Val	25	1594	1657
Phe	26	1583	1657
Leu	27	1590	1659
Leu	28	1592	1658
Val	29	1597	1657
Leu	31	1581	1658
Ile	33	1597	1658
Leu	34	1602	1659
Average		1592	1658

Table (4.4): Determination of the Position of the peak's centers of the helix and the sites labeled with $^{13}\text{C}=^{18}\text{O}$ labels in D_2O saturated samples (B). (residue No. 12 has very weak label, and did not has peak for the labeled site)

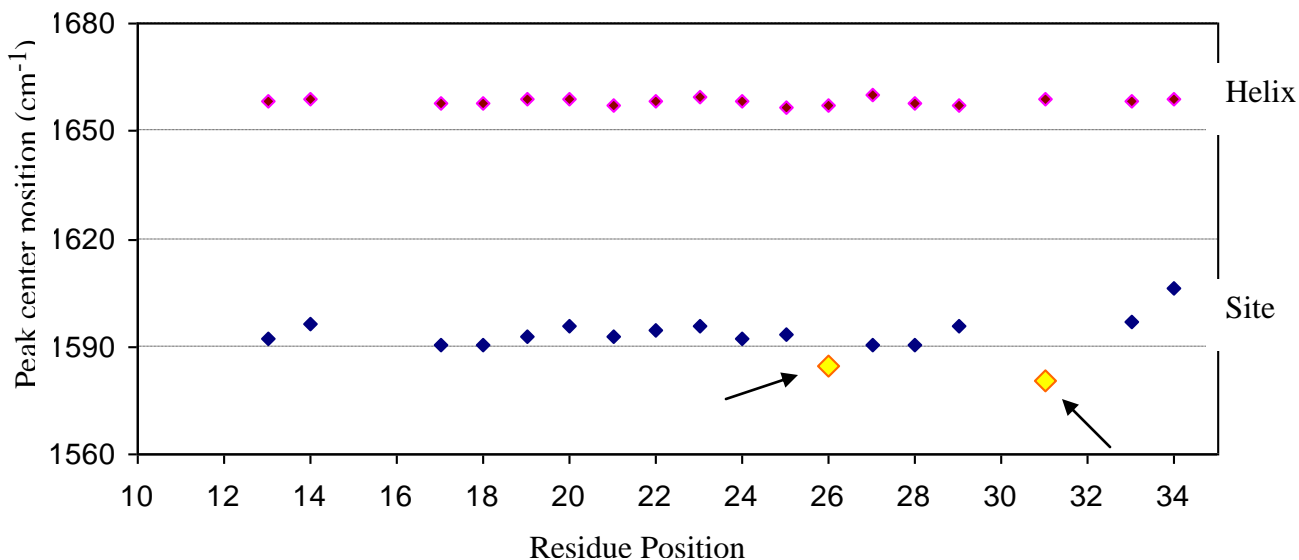


Figure (4.6.a): Schematic diagram that shows the Position of peak's centers of the helix and the site labeled with $^{13}\text{C}=^{18}\text{O}$ labels of the synthesized TMD of the SCoV E protein in ATR-FTIR spectrum in D_2O saturated air sample table (4.3), the two arrows pointed at residues No. 26 and 31 Phe and Leu respectively.

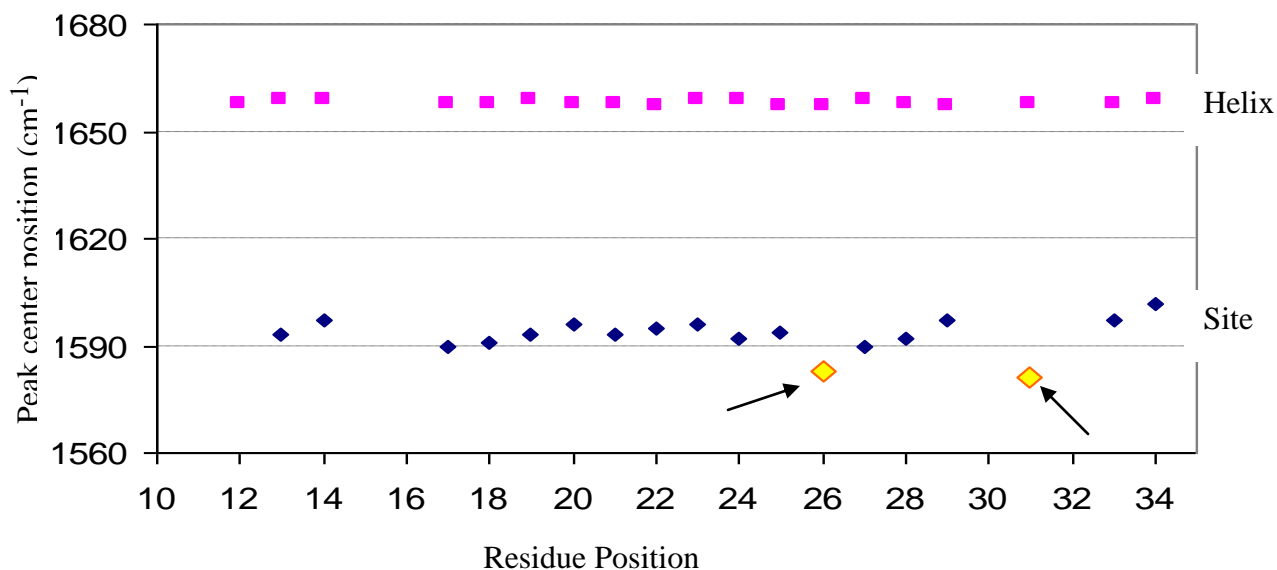


Figure (4.6.b): Schematic diagram that shows the Position of peak's centers of the helix and the site labeled with $^{13}\text{C}=^{18}\text{O}$ labels of the synthesized TMD of the SCoV E protein in ATR-FTIR spectrum in D_2O saturated air sample table (4.4), the two arrows pointed at residues No. 26 and 31 Phe and Leu respectively.

4.1.4. SARS CoV E protein has a unique transmembrane helical hairpin model with two kinks within the lipid bilayer.

From the data obtained we may postulate two possibilities to describe the structure of the E protein: either as a single long transmembrane helix, or as two short helices forming a transmembrane helical hairpin.

If it is a single helix, the hydrophobic stretch is too long relative to the bilayer thickness; on the other hand, the helical hairpin would have to be comprised of two very short helices in order to enable the 26 residues to traverse the membrane twice. Since the smallest possible loop (turn) is of five residues, each helix cannot contain more than 11–12 hydrophobic residues. This value is much smaller than the average of 21 residues per transmembrane helix.

In light of these results, one may consider possible molecular conformations. From previous FTIR results [25], we know that the protein helices are oriented perpendicular. From the density profiles of the iodinated protein, we know that the Phe-23 group is located at the hydrophilic/hydrophobic interface of the bilayer; at least at high hydration. This has led us to conclude that the SARS CoV E protein forms a small alpha-helical hairpin [25].

The position of peak's center for the helix and the site labeled with $^{13}\text{C}=^{18}\text{O}$ labels, for the synthesized TMD of the SCoV E protein, gave precious information about the structure of the TMD of the SCoV E protein. The center of the helix peaks position falls around the same region ($1557\text{-}1559\text{ cm}^{-1}$), and there were no difference between H_2O and D_2O saturated samples. this may indicate that all the synthesized samples of the TMD of E protein of the SARS Coronavirus have the same helical structure, while in the sites labeled with $^{13}\text{C}=^{18}\text{O}$ labels, the peak's centers falls around the same region ($1590\text{-}1597\text{ cm}^{-1}$) except for two residues (26 and 31). These two residues fall at a lower region than the rest of the residues ($1581\text{-}1584\text{ cm}^{-1}$) as shown in Figure (4.5, 4.6). The most likely explanation for this difference is the presence of kinks at these two residues.

It is important to indicate that these two kinks in the TMD of the SCoV E protein proposed for the first time in this work. And it is very important information that will help us in solving the secondary structure of the SARS E protein.

These two residues at position 26 and 31 falls in a lower region than the other residues. This may be due to the possibility that they are sites of a possible turns in TMD of the SCoV E protein, as shown in Figure (4.7). In residue No. 34 the position of the peak's center of the labeled site falls at higher region than the other residue (1605 cm^{-1}). This maybe due to the closeness of the residue to the membrane surface.

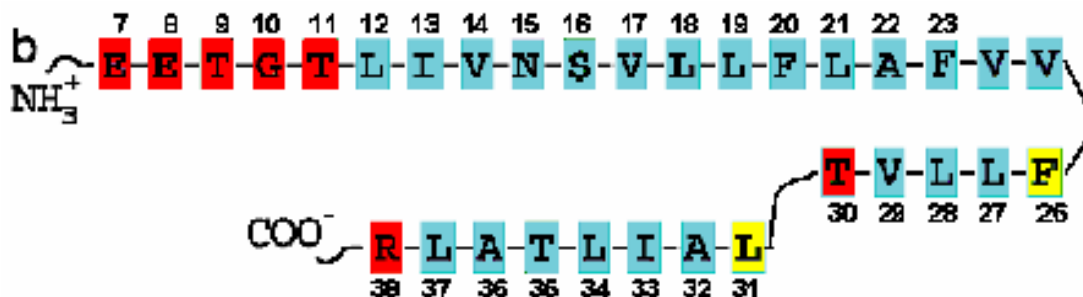


Figure (4.7): Diagram of the SCoV E protein hydrophobic region outlining the two points with expected turn. Hydrophobic amino acid residues are indicated by blue shading and hydrophilic residues are indicated by red shading, the bents site indicated by yellow shading Phe and Leu.

By determining the area under the peaks of the site labeled with $^{13}\text{C}=^{18}\text{O}$ and the helix, we can calculate the dichroism ratio (\mathcal{R}) for the labeled site and for the helix as shown in Table (4.7, 4.10) in both H₂O and D₂O saturated air samples.

The dichroism is equal to the area under the parallel peak (absorption obtained with the parallel polarized light) of the helix or the labeled site \parallel (0°) divided by the area under the perpendicular peak (absorption obtained with the perpendicular polarized light) of the helix or the labeled site \perp (90°).

$$\mathcal{R} \equiv \frac{A_{\parallel}}{A_{\perp}}$$

Equation (4.1): Dichroism which is defined as the ratio between the absorption of parallel and perpendicular polarized light.

Amino acid	position	Area under the labeled site peak		Area under helix peak	
		0°	90°	0°	90°
Leu	12	--	--	29.75573	7.288602
Ile	13	0.127398	0.027571	18.43299	5.212362
Val	14	0.131220	0.009990	11.43141	3.496748
Val	17	0.272301	0.056114	13.37303	4.349404
Leu	18	0.253021	0.055545	15.56820	4.305266
Leu	19	0.197823	0.045196	16.43562	4.154623
Phe	20	0.187070	0.030762	13.75650	3.387700
Leu	21	0.179873	0.031975	10.66120	3.597235
Ala	22	0.170353	0.034120	10.92228	3.038877
Phe	23	0.215216	0.040581	15.73607	3.946313
Val	24	0.259516	0.050628	18.60679	4.704165
Val	25	0.208910	0.021742	15.43259	4.171179
Phe	26	0.077874	0.067887	10.33257	2.507334
Leu	27	0.116225	0.021275	12.79864	3.368135
Leu	28	0.076212	0.012464	16.69371	4.057269
Val	29	0.252402	0.022970	16.42986	4.129005
Leu	31	0.114259	0.074300	14.81473	3.848605
Ile	33	0.343566	0.052944	21.92058	5.801188
Leu	34	0.354527	0.077233	20.63195	5.395729

Table (4.5): Calculation of the area under the peaks of the helix and the sites labeled with $^{13}\text{C}=^{18}\text{O}$ labels in H_2O saturated samples (A).

Amino acid	position	Area under the labeled site peak		Area under Helix peak	
		0°	90°	0°	90°
Leu	12	--	--	35.1510	7.6043
Ile	13	0.1393	0.0178	16.5203	4.6403
Val	14	0.0915	0.0133	5.3695	1.1499
Val	17	0.2892	0.0403	13.1283	3.6783
Leu	18	0.2598	0.0380	14.9701	4.0614
Leu	19	0.1925	0.0506	16.3351	4.4060
Phe	20	0.2010	0.0429	12.2920	2.9925
Leu	21	0.1918	0.0296	9.86590	3.2673
Ala	22	0.3538	0.0503	12.3788	3.6955
Phe	23	0.2526	0.0551	15.0957	3.5085
Val	24	0.2073	0.0381	11.1034	2.8155
Val	25	0.3411	0.0421	17.9745	4.3144
Phe	26	0.1035	0.0537	9.08420	2.3839
Leu	27	0.1618	0.0332	14.4372	3.3576
Leu	28	0.0763	0.0163	16.4590	4.0634
Val	29	0.3044	0.0461	17.7977	4.0909
Leu	31	0.1639	0.1221	16.3942	3.7881
Ile	33	0.4073	0.0756	23.1024	5.5872
Leu	34	0.3668	0.0928	22.7793	6.0316

Table (4.6): Calculation of the area under the peaks of the helix and the sites labeled with $^{13}\text{C}=^{18}\text{O}$ labels in D_2O saturated samples (A).

Amino acid	position	H2O		D2O	
		R Helix	R Site	R Helix	R Site
Leu	12	4.1	—	4.6	—
Ile	13	3.5	4.6	3.6	7.8
Val	14	3.3	13.1	4.7	6.9
Val	17	3.1	4.9	3.6	7.2
Leu	18	3.6	4.6	3.7	6.8
Leu	19	4.0	4.4	3.7	3.8
Phe	20	4.1	6.1	4.1	4.7
Leu	21	3.0	5.6	3.0	6.5
Ala	22	3.6	5.0	3,4	7.0
Phe	23	4.0	5.3	4.3	4.6
Val	24	4.0	5.1	3.9	5.4
Val	25	3.7	5.1	4.2	8.1
Phe	26	4.1	1.1	3.8	1.9
Leu	27	3.8	5.5	4.3	4.9
Leu	28	4.1	6.1	4.1	4.7
Val	29	4.0	11	4.4	6.6
Leu	31	3.8	1.5	4.3	1.3
Ile	33	3.8	6.5	4.1	5.4
Leu	34	3.8	4.6	3.8	4.0

Table (4.7): The dichoisism ratio (R) for the site labeled with $^{13}\text{C}=^{18}\text{O}$ labels and for the helix in both H_2O and D_2O saturated air samples (A).

Amino acid	position	Area under the labeled site peak		Area under helix peak	
		0°	90°	0°	90°
Leu	12	--	--	13.46640	3.401761
Ile	13	0.119680	0.028991	17.41518	5.078993
Val	14	0.081400	0.008666	5.458285	1.420653
Val	17	0.296281	0.053072	14.15910	4.548433
Leu	18	0.236464	0.061364	14.86435	4.473001
Leu	19	0.180174	0.039303	15.68670	3.670170
Phe	20	0.156377	0.028095	12.39292	3.263825
Leu	21	0.183251	0.029844	11.36454	3.359273
Ala	22	0.157273	0.032392	10.07528	3.005962
Phe	23	0.200536	0.036168	15.06762	3.862930
Val	24	0.307432	0.063343	20.17228	5.399378
Val	25	0.196163	0.020026	14.17995	4.040678
Phe	26	0.072359	0.042802	10.08087	2.658089
Leu	27	0.110958	0.022639	11.57420	2.916932
Leu	28	0.082443	0.014867	18.79954	4.727484
Val	29	0.270462	0.028811	16.08342	4.494517
Leu	31	0.133665	0.073168	16.84216	4.292812
Ile	33	0.352917	0.055369	22.65978	5.859195
Leu	34	0.315501	0.074127	19.31885	4.514029

Table (4.8): Calculation of the area under the peaks of the helix and the sites labeled with $^{13}\text{C}=^{18}\text{O}$ labels in H_2O saturated samples (B).

Amino acid	position	Area under the labeled site peak		Area under helix peak	
		0°	90°	0°	90°
Leu	12	--	--	12.2565	3.2619
Ile	13	0.1440	0.0202	19.2936	4.7834
Val	14	0.1646	0.0261	11.4232	3.2393
Val	17	0.3465	0.0497	15.3726	4.0772
Leu	18	0.2874	0.0425	16.1109	3.9108
Leu	19	0.1866	0.0434	15.2248	3.6588
Phe	20	0.1920	0.0460	12.6132	2.9447
Leu	21	0.2073	0.0381	11.1034	2.8155
Ala	22	0.1035	0.0152	8.8474	2.2738
Phe	23	0.2397	0.0509	14.2460	3.6320
Val	24	0.2103	0.0433	14.2886	3.5999
Val	25	0.3388	0.0267	21.2583	5.0233
Phe	26	0.0480	0.0357	10.2655	2.7062
Leu	27	0.1317	0.0264	11.6150	2.9337
Leu	28	0.1164	0.0138	18.6826	4.8121
Val	29	0.2719	0.0289	16.3566	4.3694
Leu	31	0.1359	0.0796	16.8152	4.2128
Ile	33	0.4405	0.0858	24.2601	5.8277
Leu	34	0.4447	0.1154	18.9193	5.3578

Table (4.9): Calculation of the area under the peaks of the helix and the sites labeled with $^{13}\text{C}=^{18}\text{O}$ labels in D_2O saturated samples (B).

Amino acid	position	H2O		D2O	
		R Helix	R Site	R Helix	R Site
Leu	12	4.0	—	3.8	—
Ile	13	3.4	4.1	4.0	7.1
Val	14	3.8	9.4	3.5	6.3
Val	17	3.1	5.6	3.8	7.0
Leu	18	3.3	3.9	4.1	6.8
Leu	19	4.3	4.6	4.2	4.3
Phe	20	3.8	5.6	4.3	4.2
Leu	21	3.4	6.1	3.9	5.4
Ala	22	3.4	4.9	3.9	6.8
Phe	23	3.9	5.5	3.9	4.7
Val	24	3.7	4.9	4.0	4.9
Val	25	3.5	9.8	4.2	12.7
Phe	26	3.8	1.7	3.8	1.3
Leu	27	4.0	4.9	4.0	5.0
Leu	28	4.0	5.5	3.9	8.4
Val	29	3.6	9.4	3.7	9.4
Leu	31	3.9	1.8	4.0	1.7
Ile	33	3.9	6.4	4.2	5.1
Leu	34	4.3	4.3	3.5	3.9

Table (4.10): The dichotomy ratio (R) for the site labeled with $^{13}\text{C}=^{18}\text{O}$ labels and for the helix in both H_2O and D_2O saturated air samples (B).

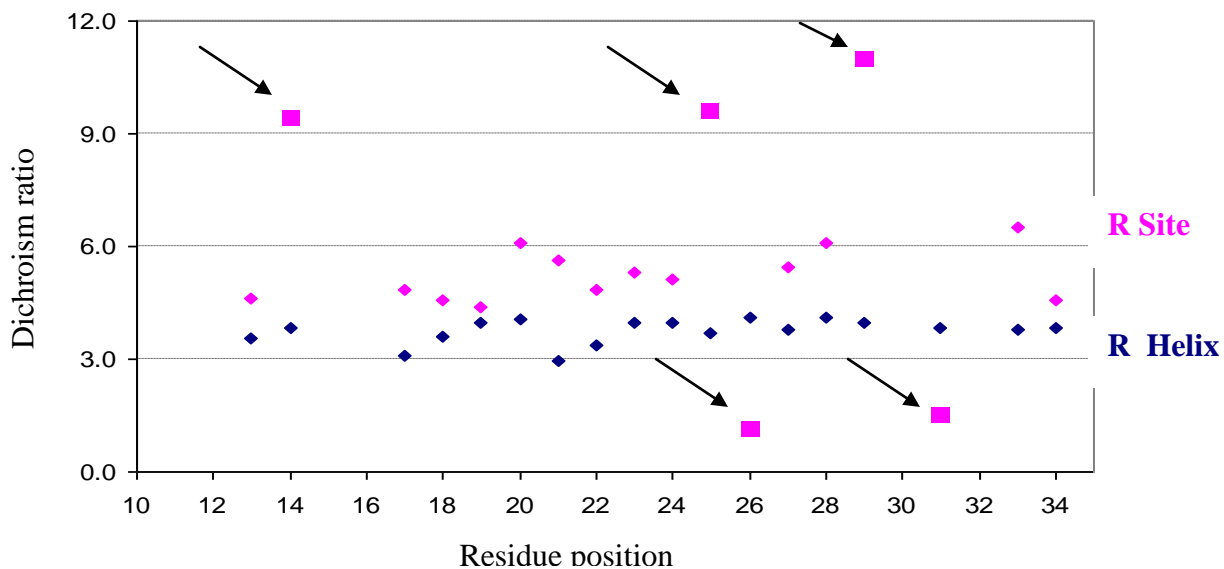


Figure (4.8.a): Schematic diagram shows the dichroism ratio (R) for the labeled site and for the helix in H₂O saturated air samples table (4.7), the upper arrows pointed at residue No. 14, 25 and 29, the lower arrows pointed at residue No. 26 and residue No. 31.

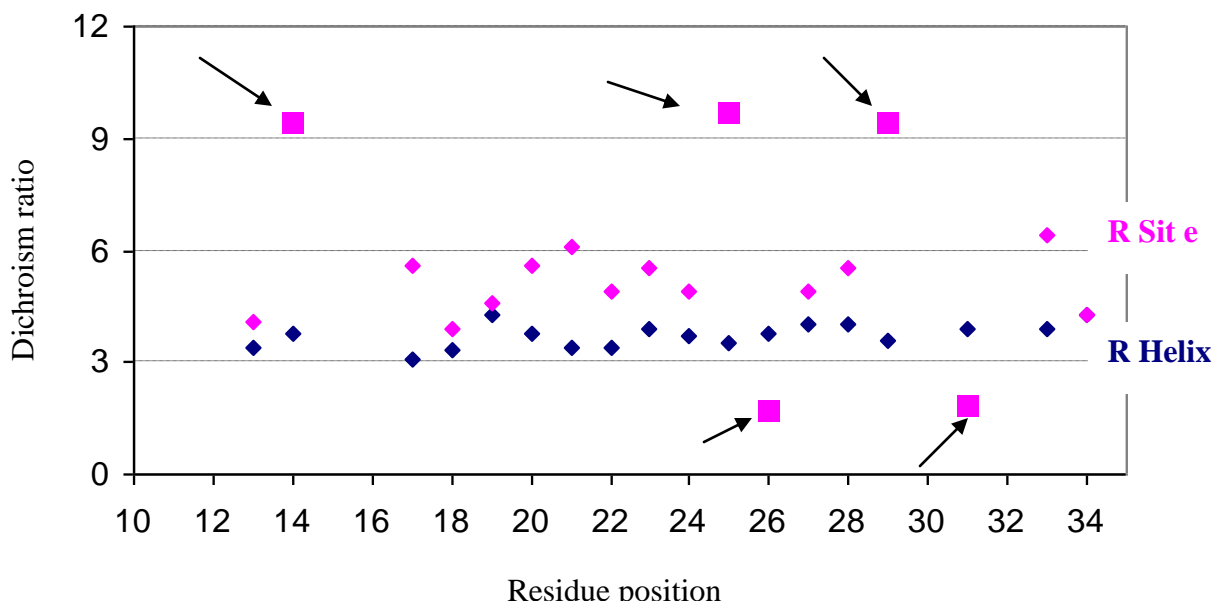


Figure (4.8.b): Schematic diagram shows the dichroism ratio (R) for the labeled site and for the helix in H₂O saturated air samples (4.10), the upper arrows pointed at residue No. 14, 25 and 29, the lower arrows pointed at residue No. 26 and 31.

4.1.6.3. Membrane Protein Secondary Structure Prediction Server

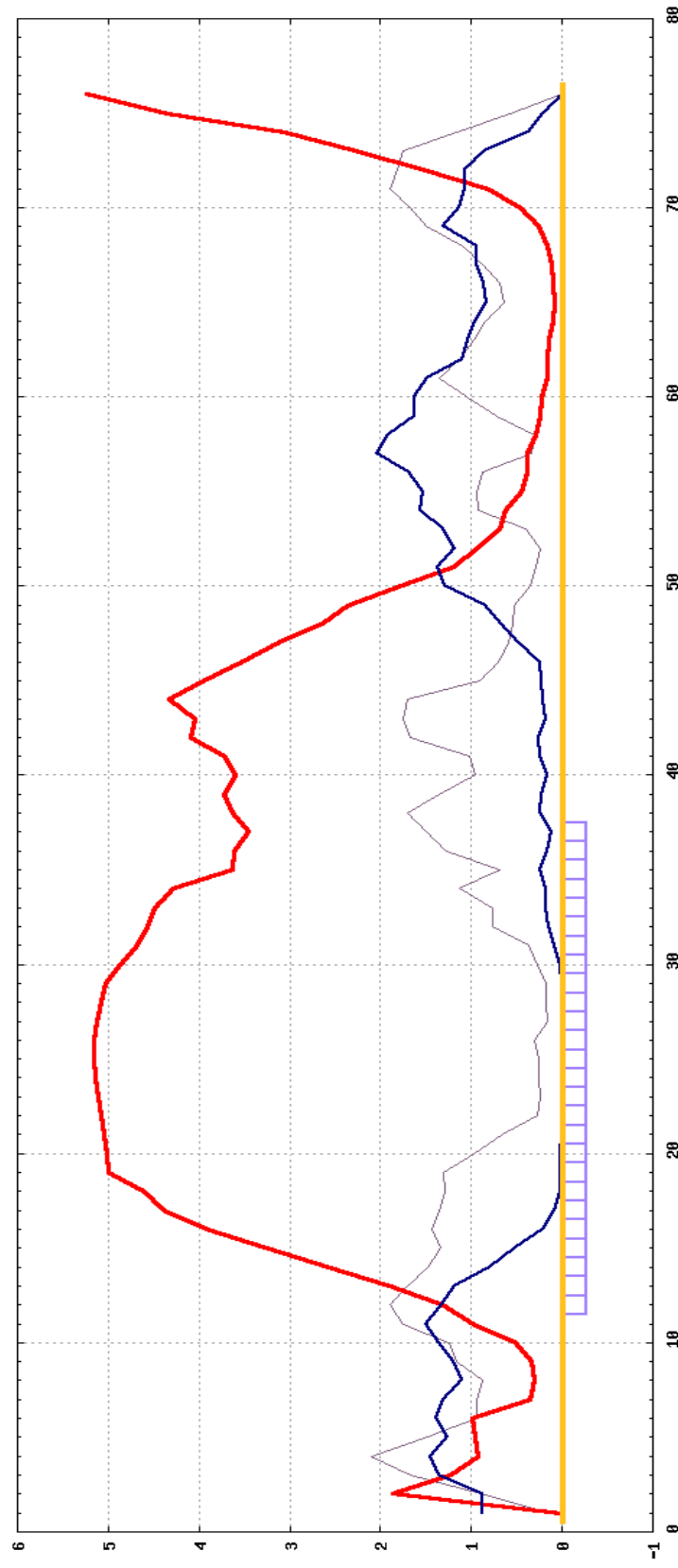


Figure (4.12): Membrane Protein Secondary Structure Prediction Server results showing the protein sequence prediction, the helix prediction starts from the 17th residue of the SARS E protein
(Red line: Transmembrane helix, Blue line: Beta sheets, Gray line: hydrophobic moment index, Violet boxes (below): Predicted transmembrane helix position).

4.2. Discussion

The SCoV life cycle involves two distinct and important stages, the first is attachment and entry to target cells, and the second is budding and release from infected cells. In this research, we decided to study the SCoV E integral membrane protein, involved in these two stages. The E protein is an important factor in the budding process of the virus, which plays an important role in viral assembly along with the M protein. Where it causes major deformation of the ER membrane. We decided to do further study on secondary structure of SCoV E protein and analyzing the structure of the protein and its affect on the lipid bilayer.

The evidence obtained in this research points to the fact that the SCoV E protein TMD forms an unusually short TM helical hairpin [25] with two kinks. And the secondary structure of the transmembrane domain is highly helical, and more than 80% of the protein is embedded in the lipid bilayer.

Using FTIR spectroscopy we analyzed the TMD of the SCoV E protein using synthesized peptide with labeled residues with $^{13}\text{C}=^{18}\text{O}$ which is capable of yielding accurate information about the position of each individual residue site in the TM helix bundle. We proposed a model structure for the protein that improves the earlier model proposed by *Arbely E et al* in 2004 [25], where they used unlabeled synthesized peptide, which gave them an approximate information about the position of each residue in the TM helix bundle because of the absence of the labeled residues with $^{13}\text{C}=^{18}\text{O}$.

The entire hairpin structure containing the loop was subsequently subjected to energy minimization using global searching molecular dynamics simulations, their resulting model was a hairpin structure which contains a center of symmetry upon which it can be inverted as shown in Figure (4.13).

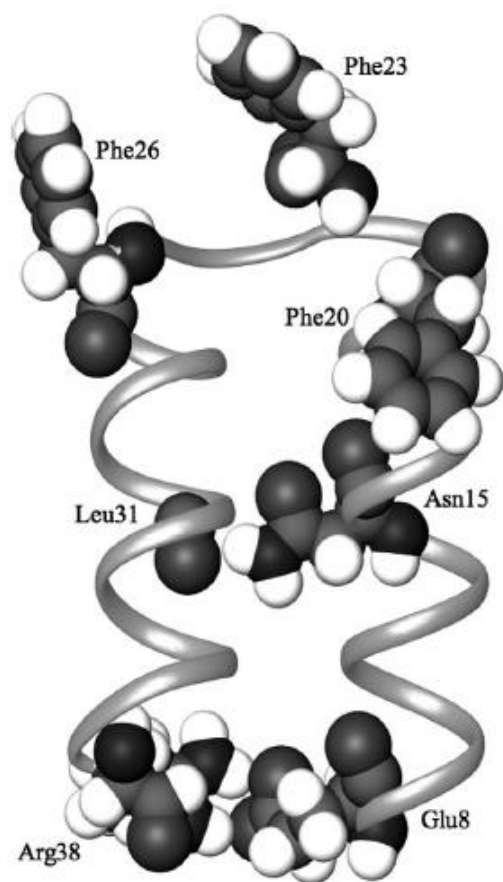


Figure (4.13): Ribbon diagram of the structural model derived for the transmembrane domain of SCoV E protein that proposed by *Arbely et al* in 2004, which shows a hairpin structure contains a center of symmetry upon which it can be inverted.

From the results of the position of the peak's center (Table 4.1-4.4) and the dichroism (Table 4.7, 4.10), we have found two points where we can expect turns in the transmembrane domain of SCoV E protein, these two residues present in the position 26 and 31. It is important to indicate that this was reported for the first time in this work. We are proposing structural models for the transmembrane domain of SCoV E protein, which contain two kinks in the position 26 and 31 Phe and Leu respectively as shown in Figure 4.14.

At residue No. 15 there is a possibility that we may have also a possible kink. The dichroism ratio of residue No. 14 was high just like the dichroism ratio of residues 25 and

29. these two residues followed with an expected turn. Since the residues No. 15 is not labeled with $^{13}\text{C}=^{18}\text{O}$ labels to confirm this result. This must be done. We are planning in the future to label residue No. 15 and 16 to confirm our results.

There is other possibility that the dichroism ratio is high as a result of the weak absorbance intensity of the residue No. 14 spectrum, and this was due to the weak label, and to the low area under the site peaks. As shown in Figure 4.15

The schematic diagram of the structural models (Figure 4.14 and 4.15) designed for the transmembrane domain of SCoV E protein explaining models. Based on the results of the calculation for the dichroism and the peak's center for the synthesized peptide labeled with $^{13}\text{C}=^{18}\text{O}$ labels. It is important to indicate that the angels between the helical segments are not significant in those models. And all the schematic models were designed with respect to a previous study on the SCoV E protein that there is a helical hairpin [25].

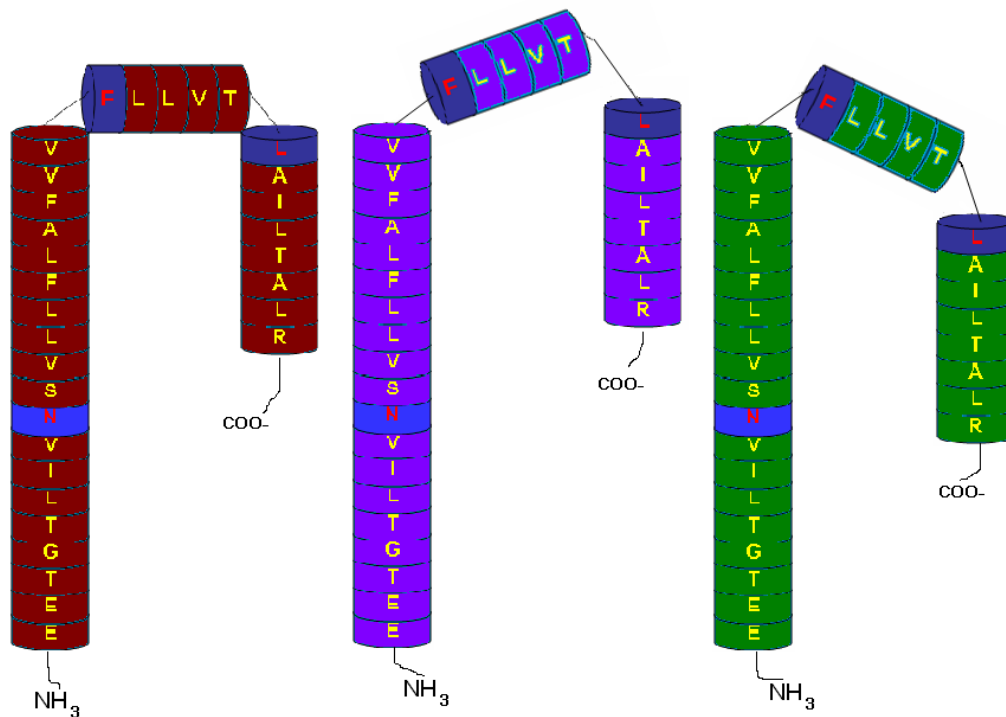


Figure (4.14): Schematic diagram of the structural model derived for the transmembrane domain of SCoV E protein which contain two expected kinks in the position 26 and 31 Phe and Leu respectively.

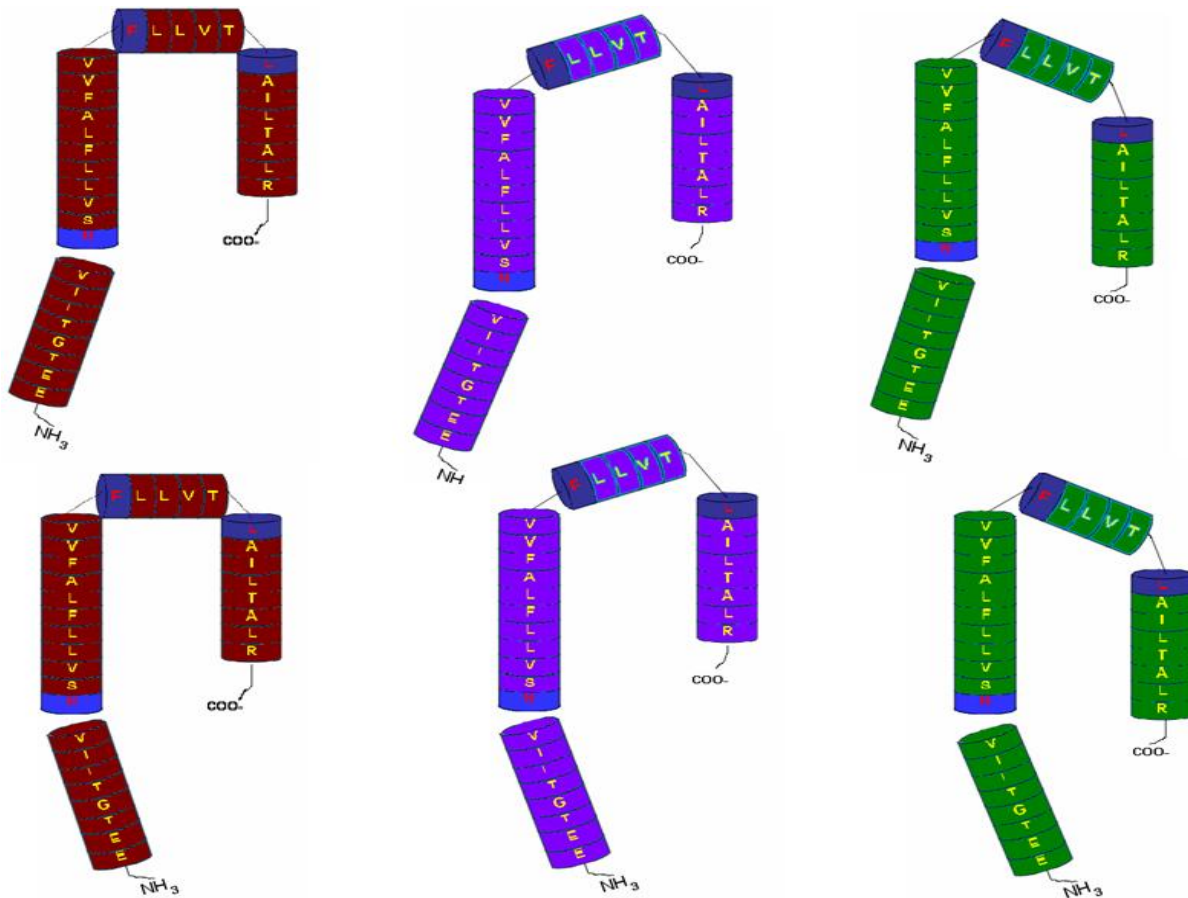


Figure (4.15): Schematic diagram of the structural model derived for the transmembrane domain of SCoV E protein which contain three expected kinks in the position No. 14, 26 and 31 Asn, Phe and Leu respectively.

In our model the shorter arm of the helical hairpin (C terminus) in our model (Figure 4.14 and 4.15) will deform the membrane while it will come close to this membrane, that is because it may expose the hydrophilic residues (non TMD residues) of the SCoV E protein [25].

The transmembrane hydrophobic segment of the SCoV E protein contains the shortest helical hairpin known. It contains 26 hydrophobic amino acid residues that traverse the lipid bilayer as two helices, and form a connecting two turn between the helices. Using negative staining electron microscopy Arbely showed the change in the structure and morphology of the lipid bilayer upon incorporating SCoV E protein transmembrane domain. While vesicles without protein (negative control) Figure (4.16 a) exhibited normal globular structure, SCoV E protein containing vesicles were markedly different Figure

(4.16 c), exhibiting extensive tubulation and deformation. Vesicles containing a positive control transmembrane protein, the MHC class II-associated invariant chain transmembrane domain were similar to vesicles without protein and did not exhibit any tubulation Figure (4.16 b).

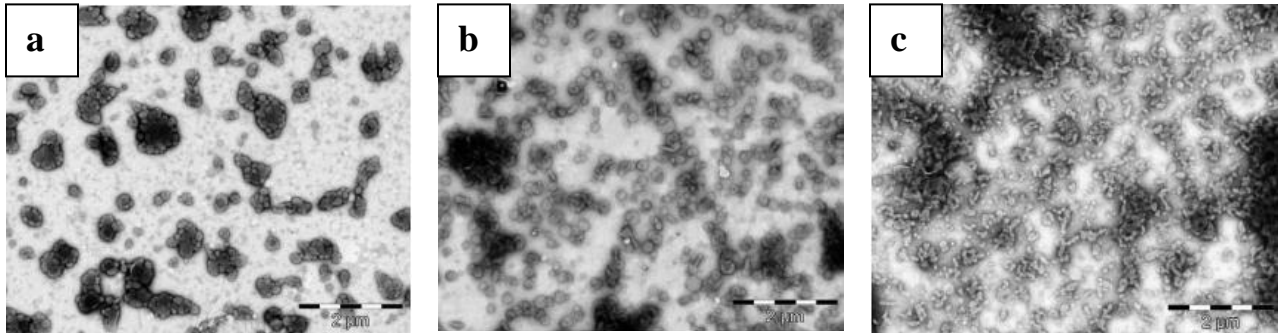


Figure (4.16): Negative staining electron micro-graphs of lipid DMPC vesicles (a), DMPC vesicles contain MHC class II-associated invariant chain transmembrane domain (b) and DMPC vesicles containing SCoV E protein transmembrane domain (c). All vesicles were extruded through a 200 nm filter. (Arbely E et al)

Protein structure prediction is a cornerstone of bioinformatics research. Membrane proteins require their own prediction methods due to their intrinsically different composition. A variety of tools exist for topology prediction of membrane proteins, many of them available on the Internet [61]. Membrane proteins are, in many respects, easier to investigate computationally than experimentally, due to the uniformity of their structure and interactions (e.g. consisting predominately of nearly parallel helices packed together). Using the prediction servers (which are helpful in giving a good idea about the secondary structure of transmembran domain of the protein) we predicted our SARS E protein. The PSIPRED Protein Structure Prediction Server results and predict protein server results Membrane Protein Secondary Structure Prediction Server results was almost the same in predicting the E protein, they all showed that the TM segment is a α -helix but the deference was only in the beginning of the α -helix segment, so it fully agree with our results that the TM domain of the E protein is α -helix, but it did not showed the tuirns that we proposed.

In a bioinformatics study for SARS E protein by Reddy et al [61], they showed also that the E protein have TM segment using different prediction servers and they made molecular

modeling for the α -helix region using molecular dynamics simulation, which also agree with our results that the TM domain of the E protein is α -helix, as shown in Figure (4.17, 4.18).

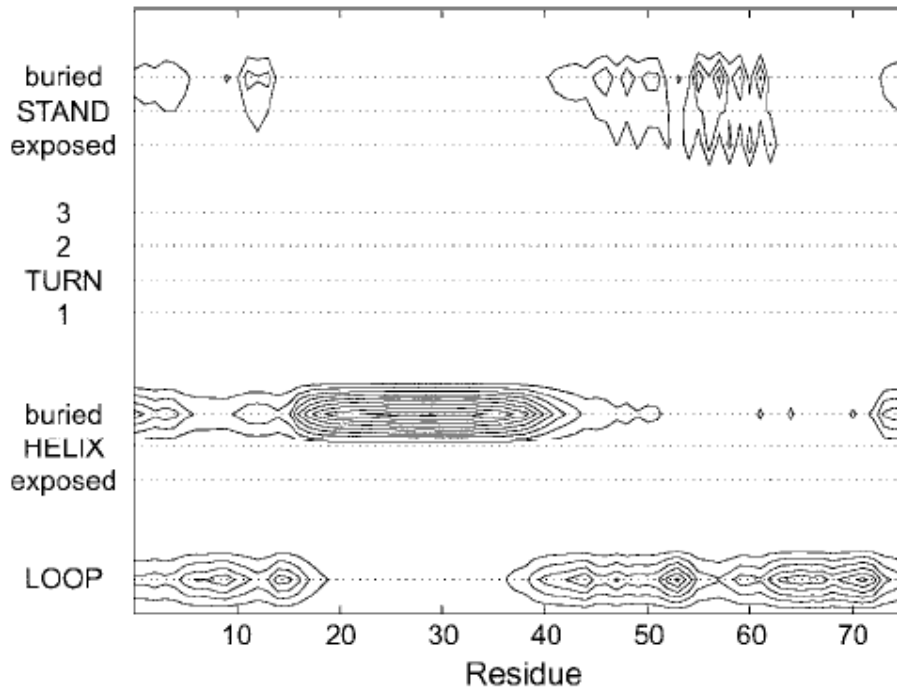


Figure (4.17): Secondary structure probabilities for SARS-CoV E protein from Protein structure prediction, the prediction was: α -helix (aa: 19-38) and b-sheets (aa: 45-52 and 55-63).(Reddy et al)

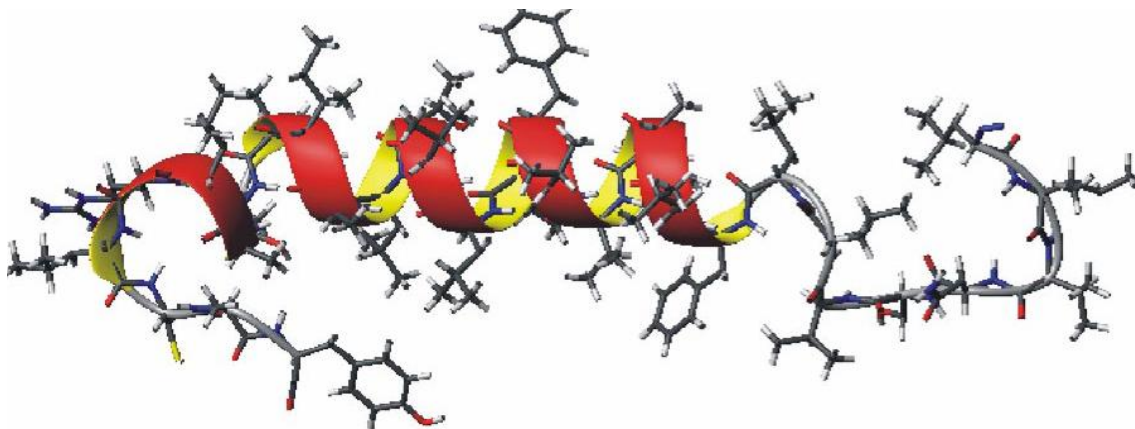


Figure (4.18): Simulated structure of α -helix (aa: 12-40) in SARS. Note that Amino acids: 19-38 is an α -helix. (Reddy et al)

Recently, Torres et al [62]. suggested a pentameric model for the transmembrane domain of the SCoV E protein. The possibility of forming a pentameric helical bundle was first suggested by Torres et al. after performing in vacuo molecular dynamics simulation of the putative TMD. The proposed model was further tested for evolutionary conservation, using 13 coronavirus E protein homologous.

In our point of view there are two basic problems in the way this proposed pentameric model was calculated. First, the authors performed only global search molecular dynamics simulation of canonical (general, basic) helices, ignoring the possibility of a helical hairpin structure. In that case, the simulation will give rise only to helical bundle models composed of canonical _ helices. Second, an evolutionary conservation analysis requires some sequence similarity between the homologue proteins that are used for the analysis. Unfortunately, there is no sequence similarity between the TMD of the SCoV E protein and the TMDs of other coronaviruses's E protein. Later, Torres et al. supported this model by SSID measurements of multiple labeled peptides in oriented lipid bilayers. We suggest that the authors are unaware of the possible implications that a modified peptide sequence might have on a structural study. First, the authors decided to use a shorter 5 residues peptide, which corresponds to residues 9–35 instead of residues 7–38. Second, the authors added two Lysine residues to both the amino and carboxy terminus of the peptide. In our opinion, these two changes have major impact on the structure of the TMD and its interactions with the membrane. One of the most prominent features of the SCoV E protein TMD is the unusually long hydrophobic stretch.

By using a shorter peptide, the authors obviously affected any characteristic of the peptide that this long hydrophobic stretch is responsible to. Furthermore, the addition of four positively charged Lysine residues to a peptide, which is only 27 amino acid residues long, might have a major impact on its structure. Obviously, the positively charged residues may interact with the negatively charged lipid head groups. This interaction might overcome the favored interaction of the aromatic belt with the lipid-water interface. It should be stressed that at the amino terminus of this modified peptide; there are two Lysine residues instead of two Glutamate residues.

There is no doubt that addition of two negatively charged residues to two positively charged residues, will affect the overall structure of the peptide.

Not only that the pentameric model is based on misleading assumptions, recently published another work for the same author that contradicts this model and supports the helical hairpin model. Studied the membrane topology of the SCoV E protein by immunofluorescent staining and proteinase K protection assay. They found that both the amino and carboxy terminus of the SCoV E protein are exposed to the cytoplasmic side of the membrane.

In this research work the obtained dichroic measurements of isotopically labeled TMDs which were performed by utilizing ATR-FTIR spectroscopy supposed to be used in a special computer program that can analyze the structure of regular (normal) helix, which give the measurements (ω , β , F_a and F_b). Thos measurements can be later used in the global searching molecular dynamics simulations, which gave the exact structural model derived for the transmembrane domain of SCoV E protein. But in our research work the E protein that we studied has TM helix bundle with more than one turn.

REFERENCES

1. Peiris, J. *et al.* 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361, p. 1319-1325.
2. Drosten, C. *et al.* 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348, p.1967-1976.
3. Ksiazek, T. G. *et al.* 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* p.1953-1966.
4. Poutanen, S. M. *et al.* 2003. Identification of severe acute respiratory syndrome in canada, *N Engl J Med.* 15;348(20), p.1995-2005.
5. Lee, N. *et al.* 2003. A major outbreak of severe acute respiratory syndrome in hong kong, *N Engl J Med.* 15;348(20), p. 1986-1994.
6. Rota, P. A., *et al.* 2003, Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300(5624), p. 1394–1399.
7. Marra, M. *et al.* 2003. The Genome sequence of the SARS-associated coronavirus. *Science* 300(5624), p. 1399–1404.
8. Nal, B., Chan, C., Kien, F., Siu, L., Tse, J., Chu, K., Kam, J., Staropoli, I., Crescenzo-Chaigne, B., Escriou, N., van der Werf, S., Yuen, K. Y., and Altmeyer, R. 2005. Differential maturation and subcellular localization of severe acute respiratory syndrome coronavirus surface proteins S, M and E. *J Gen Virol* 86(Pt 5), p.1423–1434.

9. Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J., and Rottier, P. J. 1996. Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J* 15(8), p. 2020–2028.
10. Liao, Y., Yuan, Q., Torres, J., Tam, J. P., and Liu, D. X. 2006. Biochemical and functional characterization of the membrane association and membrane permeabilizing activity of the severe acute respiratory syndrome coronavirus envelope protein. *Virology* 349(2), p. 264–275.
11. Raamsman, M. J., Locker, J. K., de Hooge, A., de Vries, A. A., Griffiths, G., Vennema, H., and Rottier, P. J. 2000. Characterization of the coronavirus mouse hepatitis virus strain A59 small membrane protein E. *J Virol* 74(5), p. 2333–2342.
12. Corse, E. and Machamer, C. E. 2000. Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles. *J Virol* 74(9), p. 4319–4326.
13. Lim, K. P. and Liu, D. X. 2001. The missing link in coronavirus assembly. Retention of the avian coronavirus infectious bronchitis virus envelope protein in the pre-Golgi compartments and physical interaction between the envelope and membrane proteins. *J Biol Chem* 276(20), p. 17515–17523.
14. An, S., Chen, C. J., Yu, X., Leibowitz, J. L., and Makino, S. 1999. Induction of apoptosis in murine coronavirus-infected cultured cells and demonstration of E protein as an apoptosis inducer. *J Virol* 73(9), p. 7853–7859.
15. Ho, Y., Lin, P. H., Liu, C. Y., Lee, S. P., and Chao, Y. C. 2004. Assembly of human severe acute respiratory syndrome coronavirus-like particles. *Biochem Biophys Res Commun* 318(4), p. 833–838.

16. Bos, E. C., Luytjes, W., van der Meulen, H. V., Koerten, H. K., and Spaan, W. J. 1996. The production of recombinant infectious DI-particles of a murine coronavirus in the absence of helper virus. *Virology* 218(1), p. 52–60.
17. Baudoux, P., Carrat, C., Besnardeau, L., Charley, B., and Laude, H. 1998. Coronavirus pseudoparticles formed with recombinant M and E proteins induce alpha interferon synthesis by leukocytes. *J Virol* 72(11), p. 8636–8643.
18. Maeda, J., Maeda, A., and Makino, S. 1999. Release of coronavirus E protein in membrane vesicles from virus-infected cells and E protein-expressing cells. *Virology* 263(2), p. 265–272.
19. Fischer, F., Stegen, C. F., Masters, P. S., and Samsonoff, W. A. 1998. Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. *J Virol* 72(10), p. 7885–7894.
20. Wilson, L., McKinlay, C., Gage, P., and Ewart, G. 2004. SARS coronavirus E protein forms cation-selective ion channels. *Virology* 330(1), p. 322–331.
21. Liao, Y., Lescar, J., Tam, J. P., and Liu, D. X. 2004. Expression of SARS-coronavirus envelope protein in Escherichia coli cells alters membrane permeability. *Biochem Biophys Res Commun* 325(1), p. 374–380.
22. Madan, V., Garcia Mde, J., Sanz, M. A., and Carrasco, L. 2005. Viroporin activity of murine hepatitis virus E protein. *FEBS Lett* 579(17), p. 3607–3612.
23. Wilson, L., Gage, P., and Ewart, G. 2006. Hexamethylene amiloride blocks E protein ion channels and inhibits coronavirus replication. *Virology* 353(2), p. 294–306.
24. Arkin, I. T. and Brunger, A. T. 1998. Statistical analysis of predicted transmembrane alpha-helices. *Biochim Biophys Acta* 1429(1), p. 113–128 .

25. Arbely, A., Z. Khattari, G. Brotons, M. Akkawi, T. Salditt, and I. T. Arkin. 2004. A highly unusual palindromic transmembrane helical hairpin formed by SARS coronavirus E protein. *J. Mol. Biol.* 341, p. 769-779.
26. Godet, M., L'Haridon, R. Vautherot, J. F., and Laude, H. 1992. TGEV corona virus ORF4 encodes a membrane protein that is incorporated into virions. *Virology* 188(2), p. 666– 675 .
27. Maeda, J., Repass, J. F., Maeda, A., and Makino, S. 2001. Membrane topology of coronavirus E protein. *Virology* 281(2), p. 163–169.
28. Yuan, Q., Liao, Y., Torres, J., Tam, J. P., and Liu, D. X. 2006. Biochemical evidence for the presence of mixed membrane topologies of the severe acute respiratory syndrome coronavirus envelope protein expressed in mammalian cells. *FEBS Lett* 580(13), p. 3192– 3200.
29. Stevens, T. J. and Arkin, I. T. 2000. Do more complex organisms have a greater proportion of membrane proteins in their genomes? *Proteins* 39(4), p. 417–420.
30. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305(3), p. 567–580.
31. Branden, C, Tooze, J, 1991. Introduction to protein structure, Garland publishing, Inc, first edition.
32. Fu, D., Libson, A., Miercke, L. J., Weitzman, C., Nollert, P., Krucinski, J., and Stroud, R. M. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290(5491), p. 481–486.

33. Yohannan, S., Faham, S., Yang, D., Whitelegge, J. P., and Bowie, J. U. 2004. The evolution of transmembrane helix kinks and the structural diversity of G protein-coupled receptors. *Proc Natl Acad Sci U S A* 101(4), p. 959–963.
34. Cordes, F. S., Bright, J. N., and Sansom, M. S. 2002. Proline-induced distortions of transmembrane helices. *J Mol Biol* 323(5), p. 951–960.
35. Riek, R. P., Rigoutsos, I., Novotny, J., and Graham, R. M. 2001. Non-alpha-helical elements modulate polytopic membrane protein architecture. *J Mol Bio.* 306(2), p. 349–362.
36. Thomas J. Lukas, Michael B. Prystowsky, AND Bruce W. Erickson. 1981. Solid-phase peptide synthesis under continuous-flow conditions. *Biochemistry j*, Vol. 78, (5) . p. 2791-2795.
37. Angus R. Brown, Pedro H.H. Hermkens, Harry C.J. Ottenheijm, David C. Rees. 1998. Solid Phase Synthesis. *NEW TOOLS IN SYNTHESIS*, p. 817-827.
38. Walter C. Mahoney and Mark A. Hermodson. 1980. Separation of Large Denatured Peptides by Reverse Phase High Performance Liquid Chromatography. *Biological Chemistry j*. p 11199-11203.
39. Nilsson, I., Saaf, A., Whitley, P., Gafvelin, G., Waller, C., and von Heijne, G. 1998. Proline-induced disruption of a transmembrane alpha-helix in its natural environment. *J Mol Biol* 284(4), p. 1165–1175.
40. Vigano, C., Manciu, L., Buyse, F., Goormaghtigh, E., and Ruyschaert, J. M. 2000. Attenuated total reflection IR spectroscopy as a tool to investigate the structure, orientation and tertiary structure changes in peptides and membrane proteins. *Biopolymers* 55(5), p. 373–380.

41. Harrick, N. *Internal Reflection Spectroscopy*. 1967. Interscience Publishers, New York, first edition.
42. Arkin, I. T. 2006. Isotope-edited IR spectroscopy for the study of membrane proteins. *Curr Opin Chem Biol* 10(5), p. 394–401.
43. Tadesse, L., Nazarbaghi, R., and Walters, L. 1991. Isotopically enhanced infrared spectroscopy: a novel method for examining the secondary structure at specific sites in conformationally heterogeneous peptides. *J Am Chem Soc* 113(18), p. 7036–7037.
44. Torres, J., Kukol, A., and Arkin, I. T. 2000. Use of a single glycine residue to determine the tilt and orientation of a transmembrane helix. A new structural label for infrared spectroscopy. *Biophys J* 79(6), p. 3139–3143.
45. Torres, J. and Arkin, I. T. 2002. C-deuterated alanine: a new label to study membrane protein structure using site-specific infrared dichroism. *Biophys J* 82(2), p. 1068–1075.
46. Arkin, I. T., MacKenzie, K. R., and Brunger, A. 1999. Site-directed dichroism as a method for obtaining rotational and orientational constraints for oriented polymers. *J Amer Chem Soc* 119(38), p. 8973–8980.
47. Kukol, A. and Arkin, I. T. vpu transmembrane peptide structure obtained by site specific fourier transform infrared dichroism and global molecular dynamics searching. *Biophys J* 77(3), 1594–1601, Sep.
48. Kukol, A., Adams, P. D., Rice, L. M., Brunger, A. T., and Arkin, I. T. 1999. Experimentally based orientational refinement of membrane protein models: A structure for the Influenza A M2 H⁺ channel. *J Mol Biol* 286(3), p. 951–962.

49. Kukol, A. and Arkin, I. T. 2000. Structure of the influenza C virus CM2 protein transmembrane domain obtained by site-specific infrared dichroism and global molecular dynamics searching. *J Biol Chem* 275(6), p. 4225–4229.
50. Torres, J., Kukol, A., Goodman, J. M., and Arkin, I. 2001. T. Site-specific examination of secondary structure and orientation determination in membrane proteins: the peptidic $^{13}\text{C}=\text{^{18}\text{O}}$ group as a novel infrared probe. *Biopolymers* 59(6), p. 396–401.
51. Torres, J., Briggs, J. A., and Arkin, I. T. 2002. Multiple site-specific infrared dichroism of CD3- ζ , a transmembrane helix bundle. *J Mol Biol* 316(2), p. 365–374.
52. Kukol, A., Torres, J., and Arkin, I. T. 2002. A structure for the trimeric MHC class II-associated invariant chain transmembrane domain. *J Mol Bio, lvol.* 320(5), p.1109–1117.
53. Torres, J., Briggs, J. A., 2002. and Arkin, I. T. Convergence of experimental, computation and evolutionary approaches predicts the presence of a tetrameric form for CD3-zeta. *J Mol Biol ,vol.*316(2),p. 375–384.
54. Kyte, J. and Doolittle, R. F. 1982. A simple method for displaying the hydrophobic character of a protein. *J Mol Biol,vol.* 157(1), 105–132.
55. Engelman, D. M., Steitz, T. A., and Goldman, A. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. *Annu Rev Biophys Chem,vol.* 15,p. 321–353.
56. Zvilin, M., Leonov, H., and Arkin, I. T. 2005. Genetic algorithm-based optimization of hydrophobicity tables. *Bioinformatics,vol.* 21(11), p.2651–2656.

57. Rost, B., Casadio, R., Fariselli, P., and Sander, C. 1995. Transmembrane helices predicted at 95% accuracy. *Protein Sci*, vol. 4(3), p.521–533.
58. Treutlein, H. R., M. A. Lemmon, D. M. Engelman, and A. T. Brunger. 1992. The glycophorin A transmembrane domain dimer: sequence-specific propensity for a right-handed supercoil of helices. *Biochemistry*, vol 31, p.12726-12732.
59. Torres, Jaume, Briggs, John A G, Arkin, Isaiah T. 2002. Contribution of energy values to the analysis of global searching molecular dynamics simulations of transmembrane helical bundles. *Biophys J*, p. 3063-3071, Vol. 82, No. 6
60. Chien Peter Chen, Burkhard Rost, 2002, State-of-the-art in membrane protein prediction, *Appl Bioinformatics*. 1(1), p. 21-35.
61. Amarender D Reddy, Seung Bum Suh, Reza Ghaffari, N. Jiten Singh, Dae-Jin Kim, Joon Hee Han, and Kwang S. Kim, 2003. Bioinformatics Analysis of SARS Proteins and Molecular Dynamics Simulated Structure of an Alpha-helix Motif, *Bull. Korean Chem. Soc.* Vol. 24, p. 899-900.
62. Torres j, Parthasarathy k, Lin1 x, Saravanan1 R, Kuko K , Xiang Liu1. 2006. Model of a putative pore: the pentameric α -helical bundle of SARS Coronavirus E protein in lipid bilayers. *Biophys J*, vol. 91, p.938-947.

APPENDICES

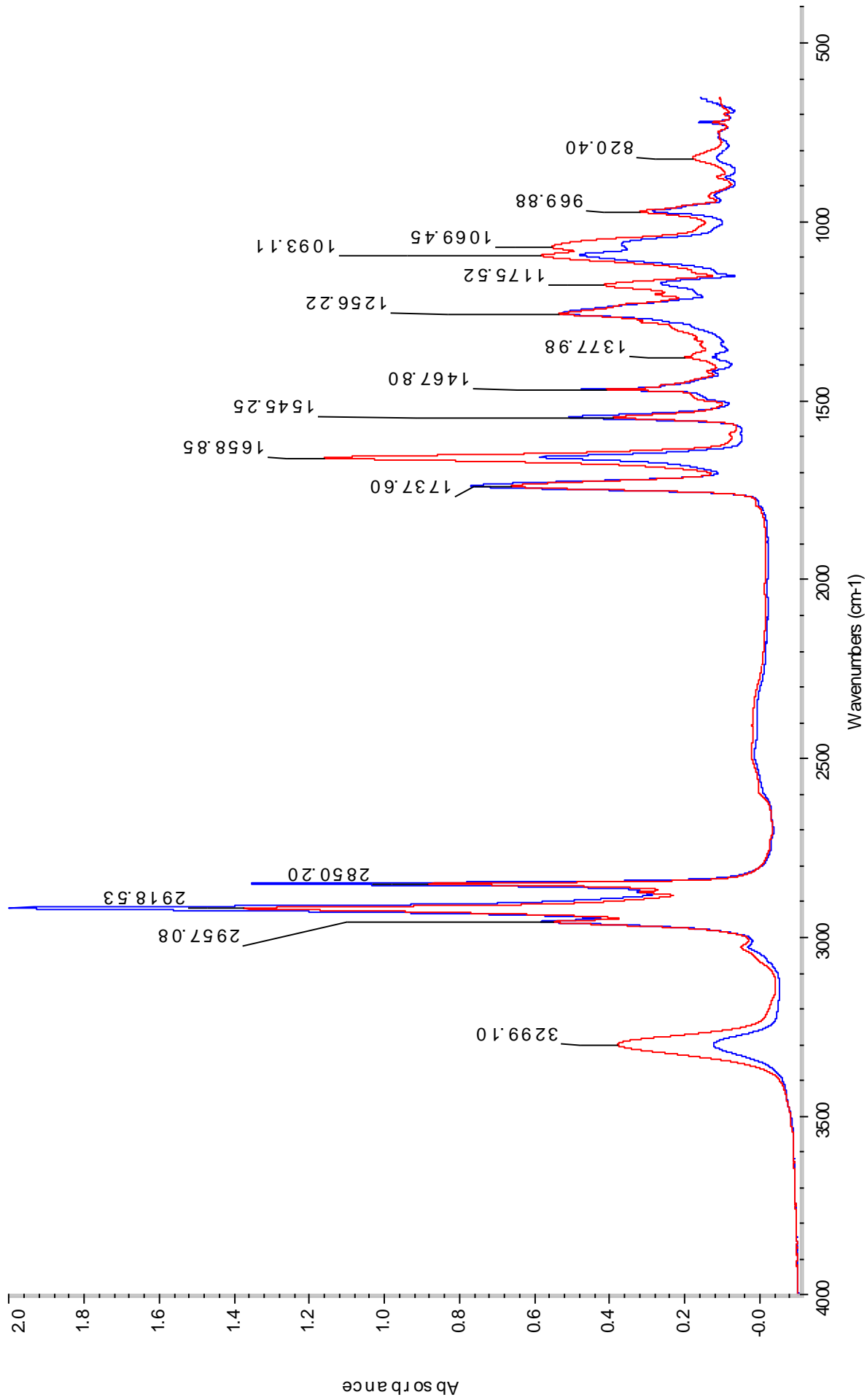


Figure (1): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 13 labeled with ($^{13}\text{C}=\text{^{18}\text{O}}$) (Absorption of parallel in red and perpendicular in blue)

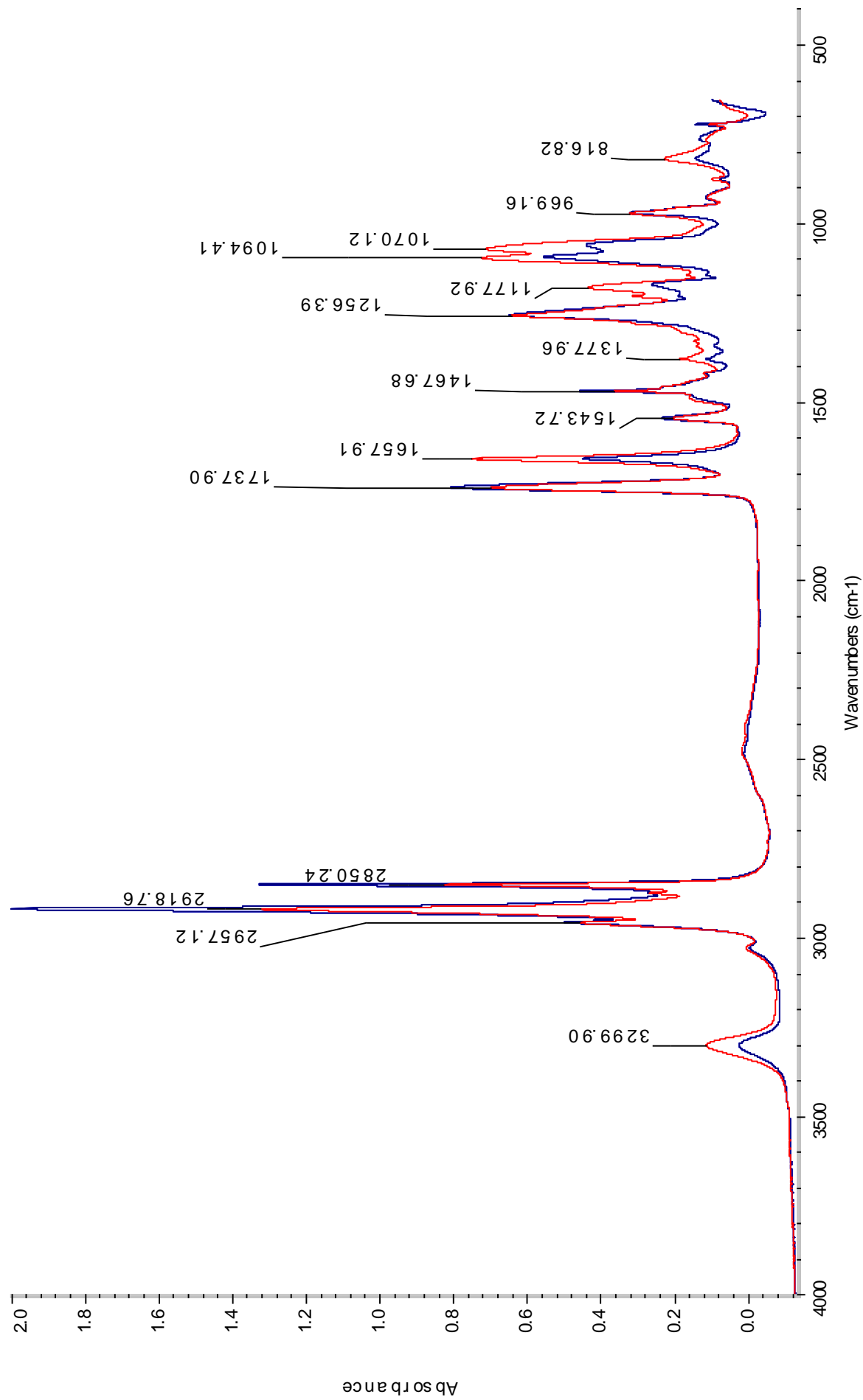


Figure (2): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 14 labeled with (¹³C=¹⁸O) (Absorption of parallel in red and perpendicular in blue)

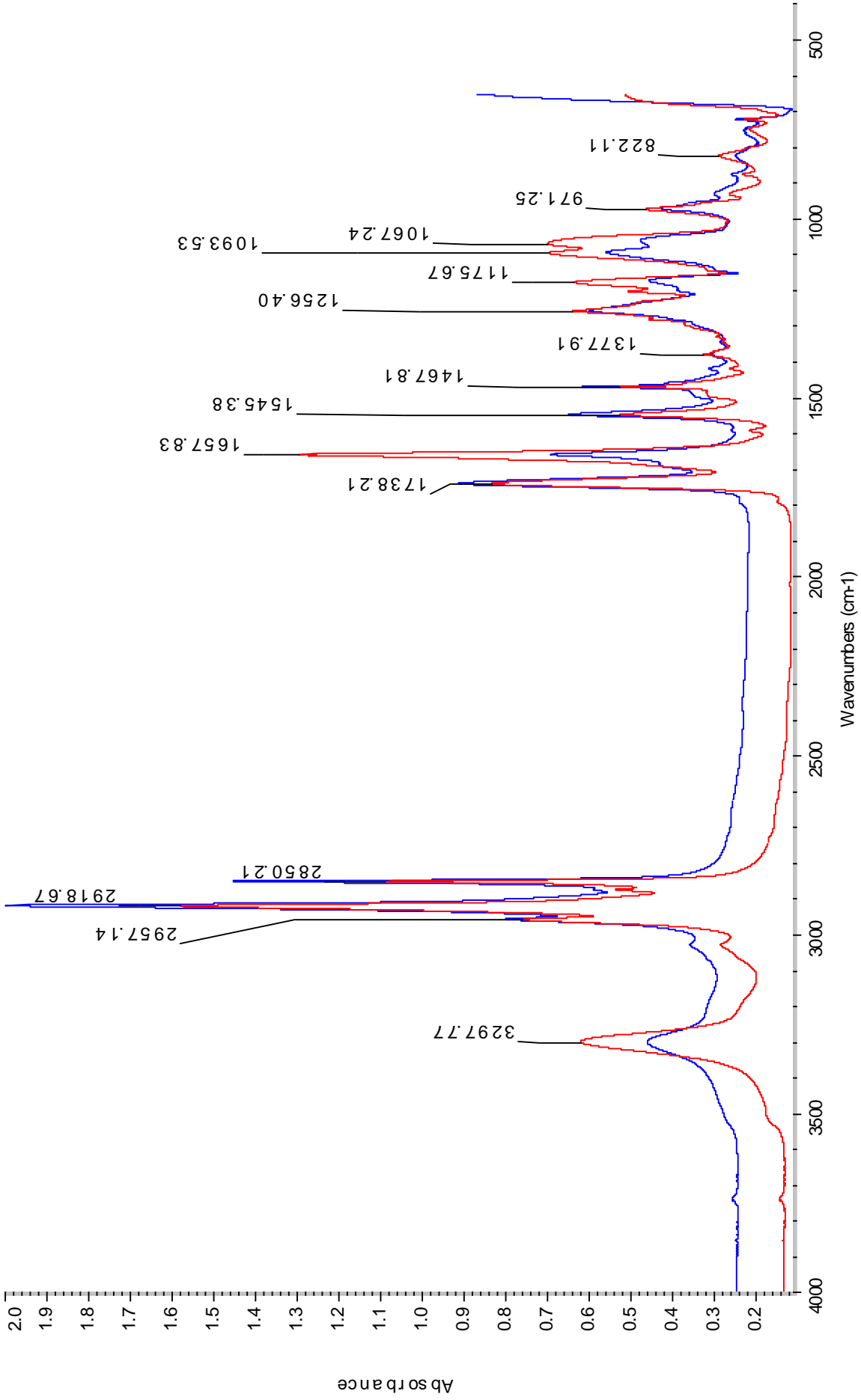


Figure (3): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 17 labeled with (¹³C=¹⁸O) (Absorption of parallel in red and perpendicular in blue)

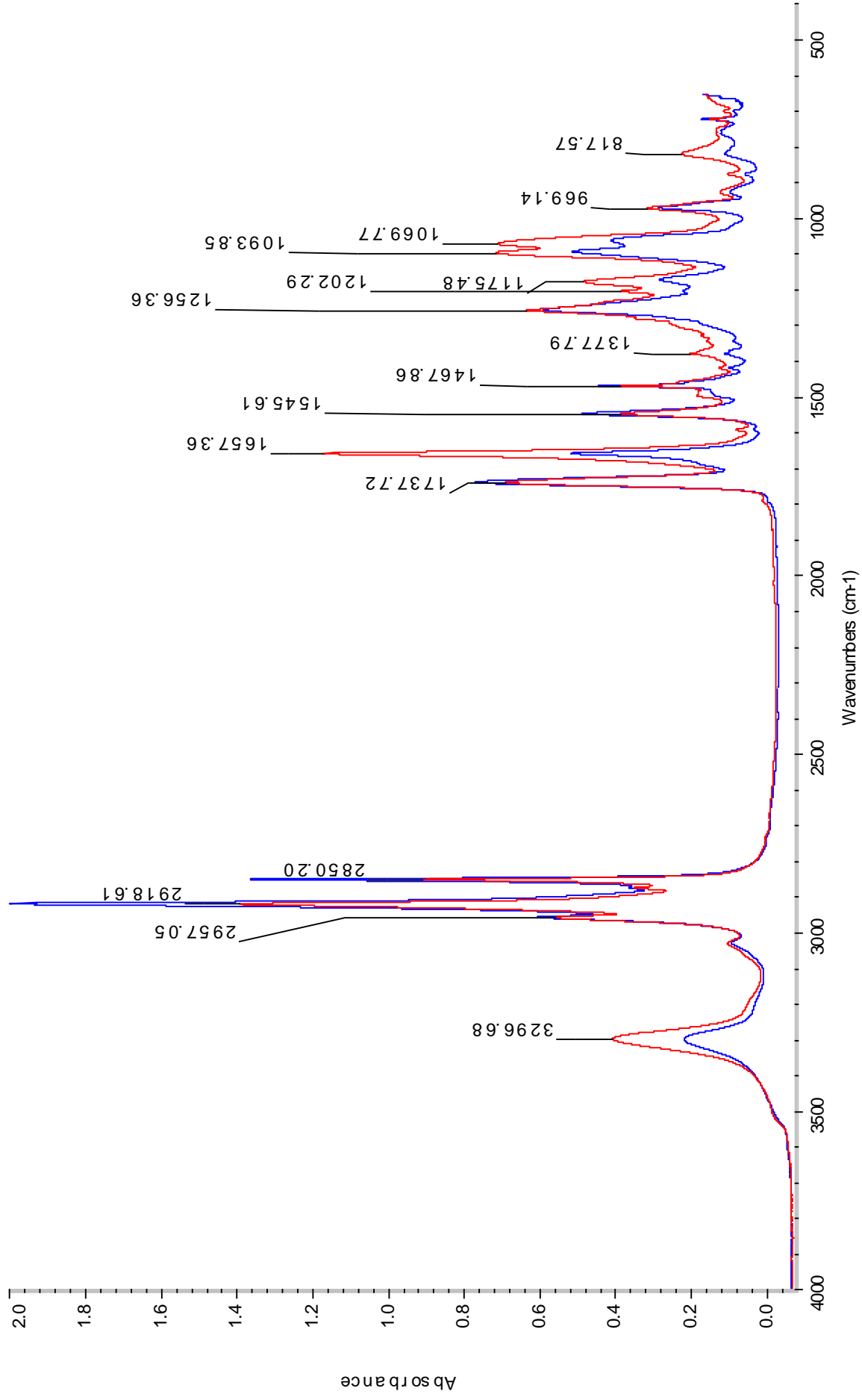


Figure (4): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 18 labeled with ($^{13}\text{C}=\text{18O}$) (Absorption of parallel in red and perpendicular in blue)

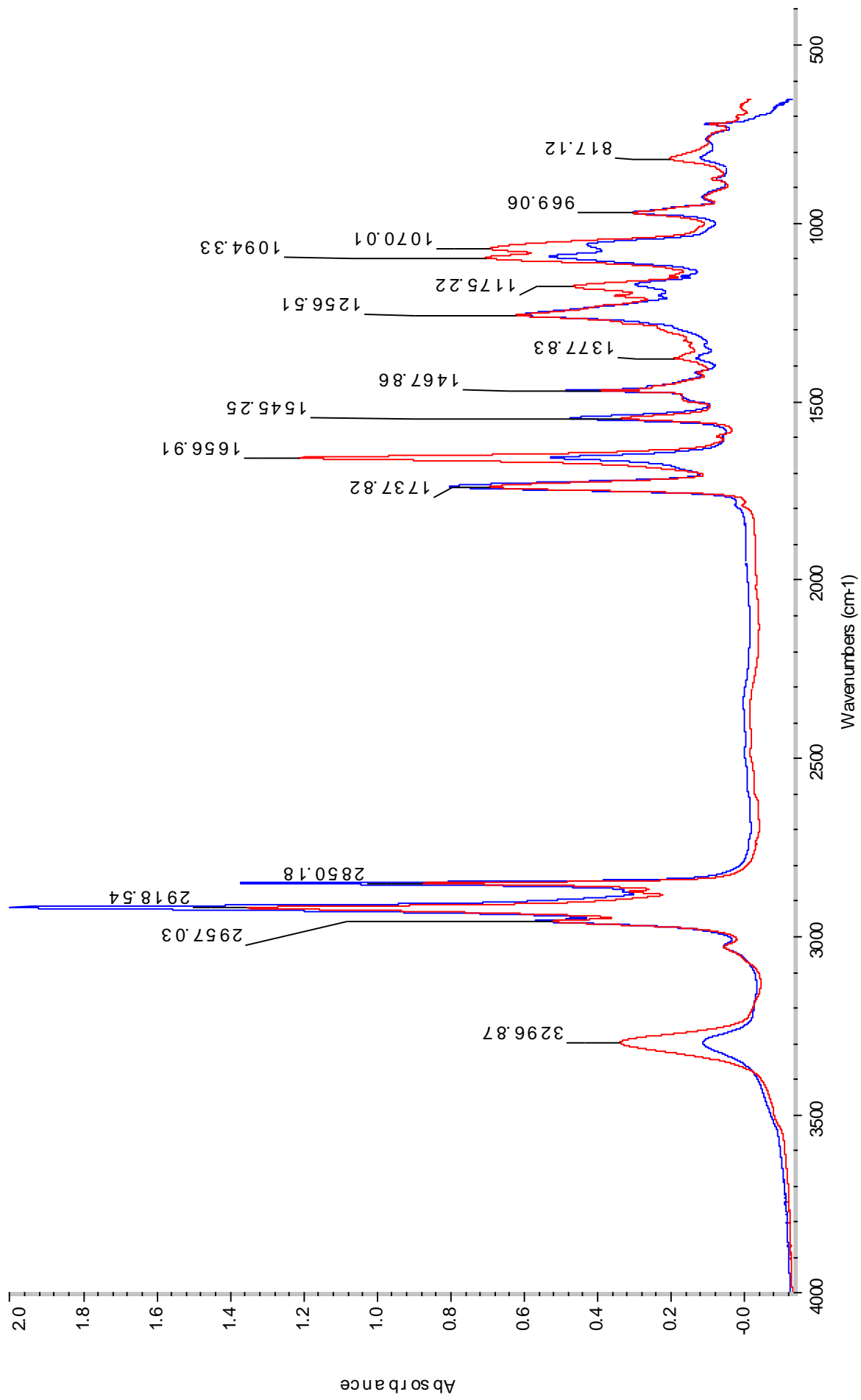


Figure (5): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 19 labeled with ($^{13}\text{C}=^{18}\text{O}$) (Absorption of parallel in red and perpendicular in blue)

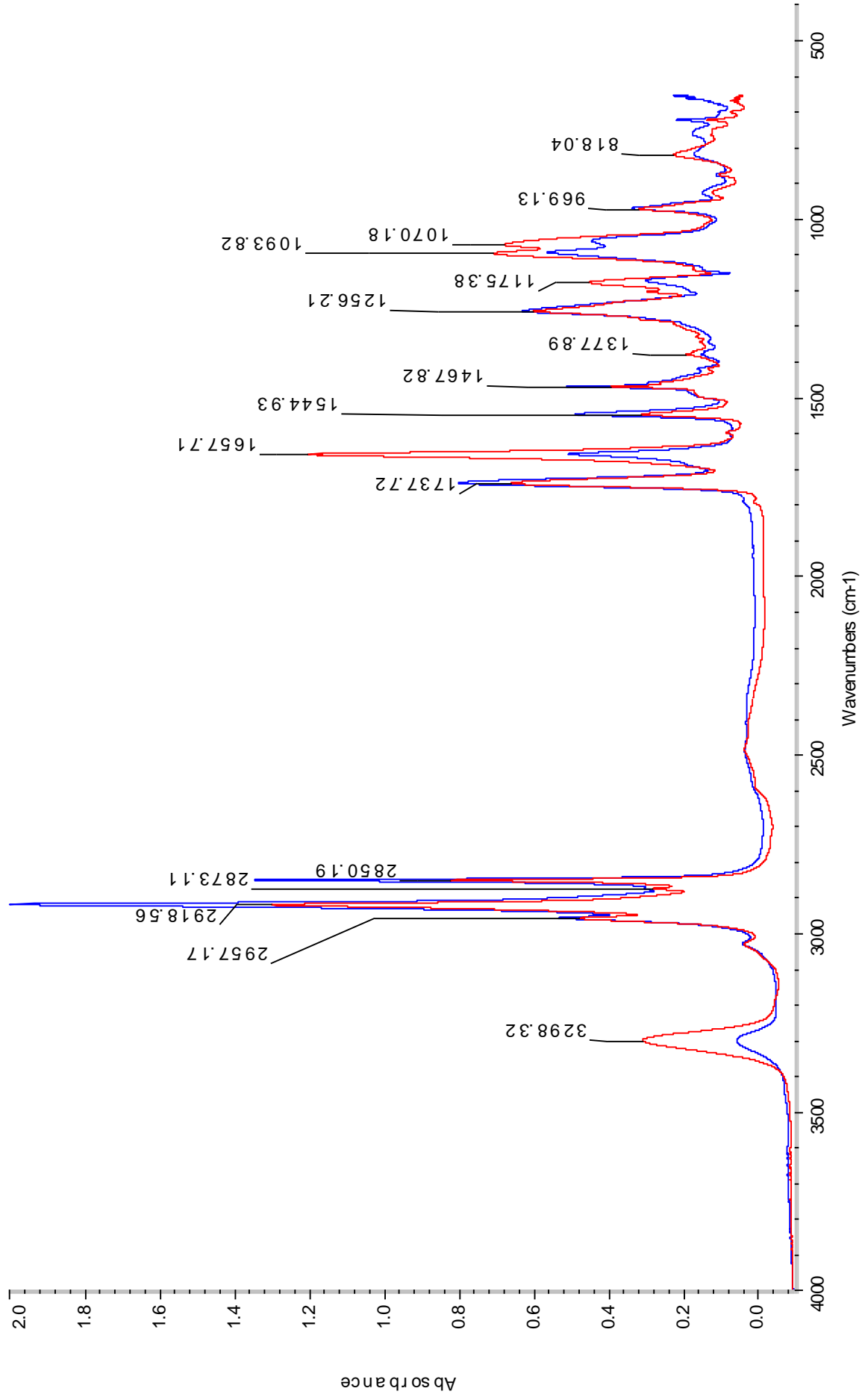


Figure (6): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 20 labeled with ($^{13}\text{C}=\text{18O}$) (Absorption of parallel in red and perpendicular in blue)

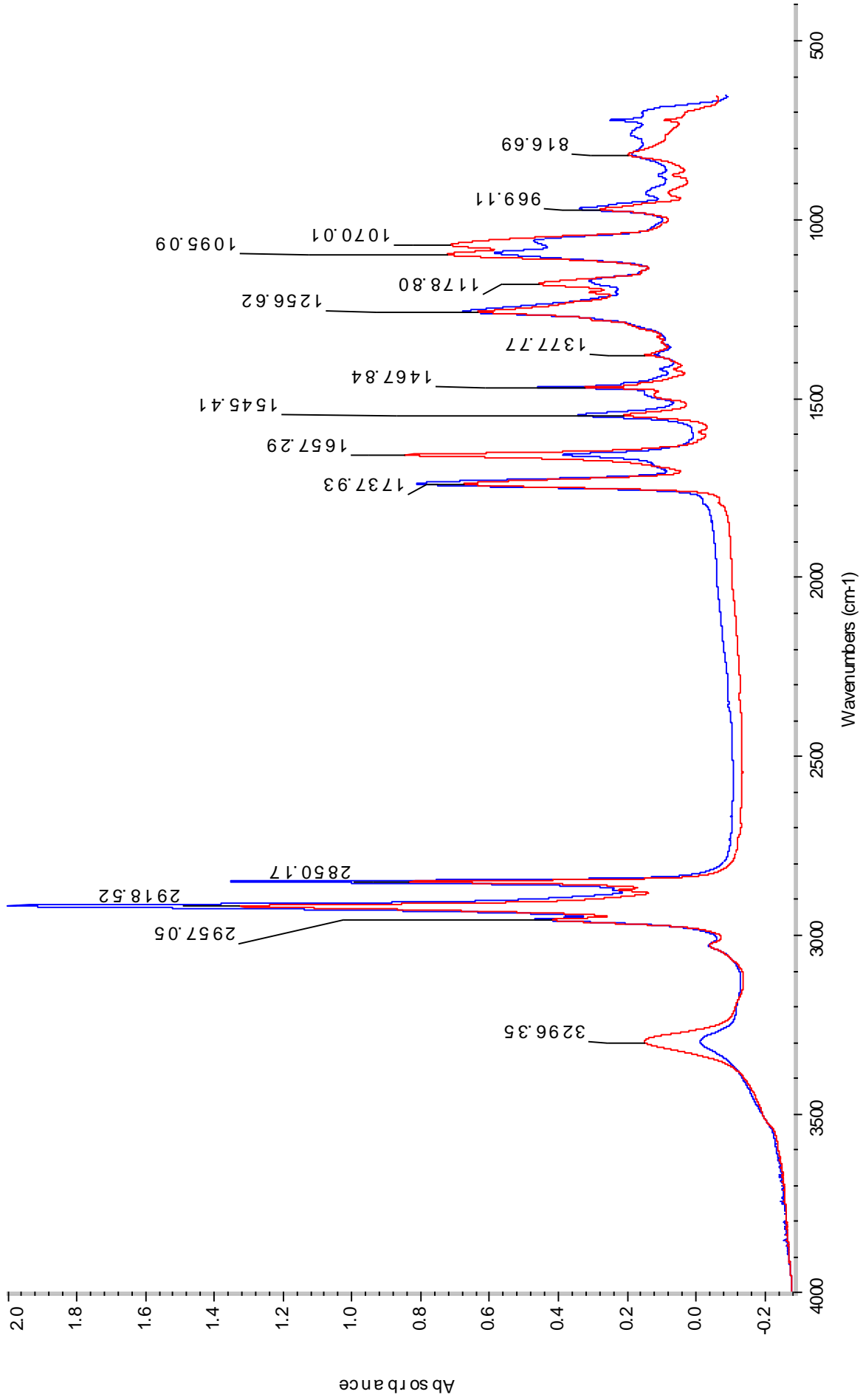


Figure (7): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 21 labeled with (¹³C=¹⁸O) (Absorption of parallel in red and perpendicular in blue)

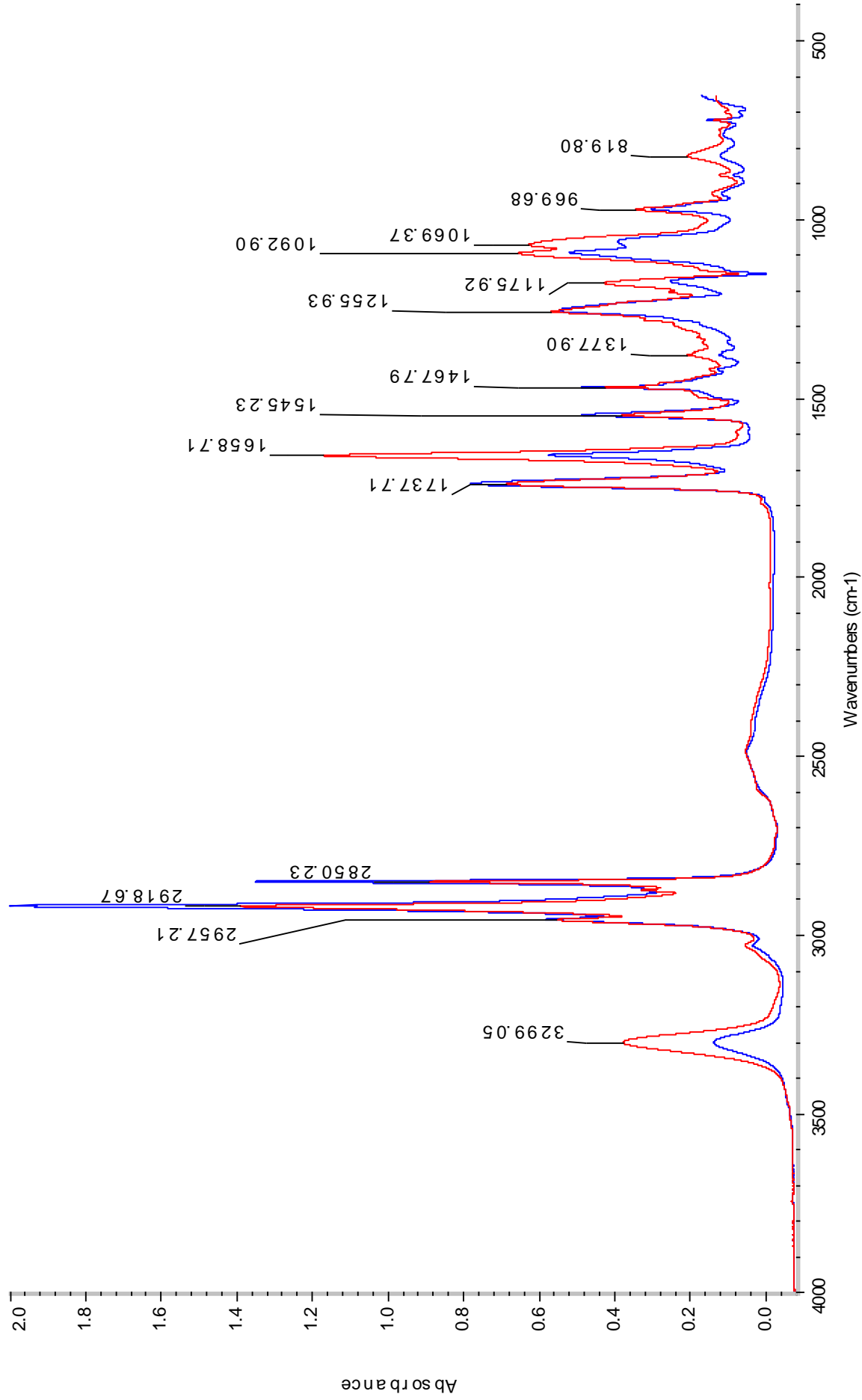


Figure (8): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 22 labeled with ($^{13}\text{C}=\text{O}^{18}\text{O}$) (Absorption of parallel in red and perpendicular in blue)

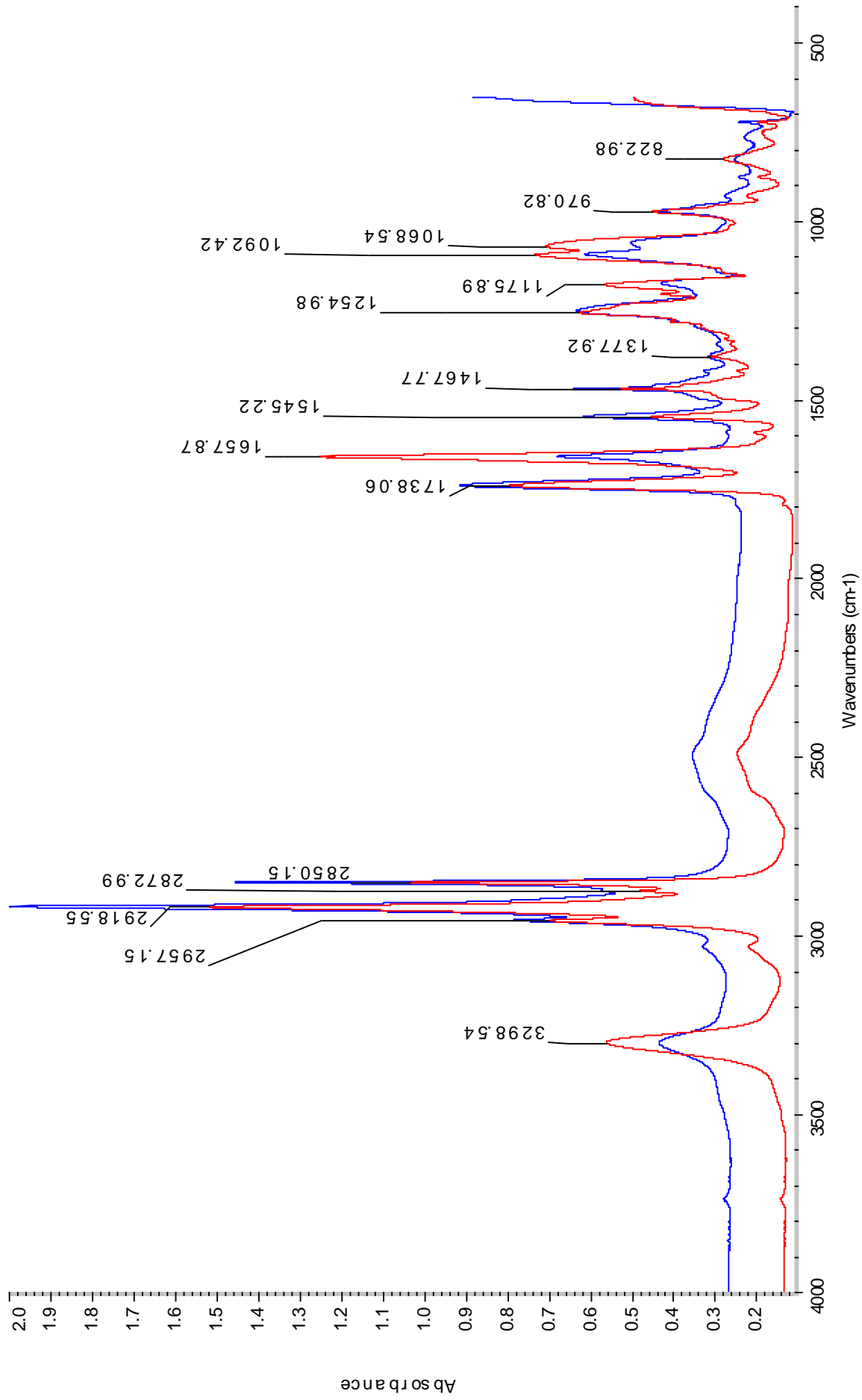


Figure (9): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 23 labeled with ($^{13}\text{C}=\text{O}$) (Absorption of parallel in red and perpendicular in blue)

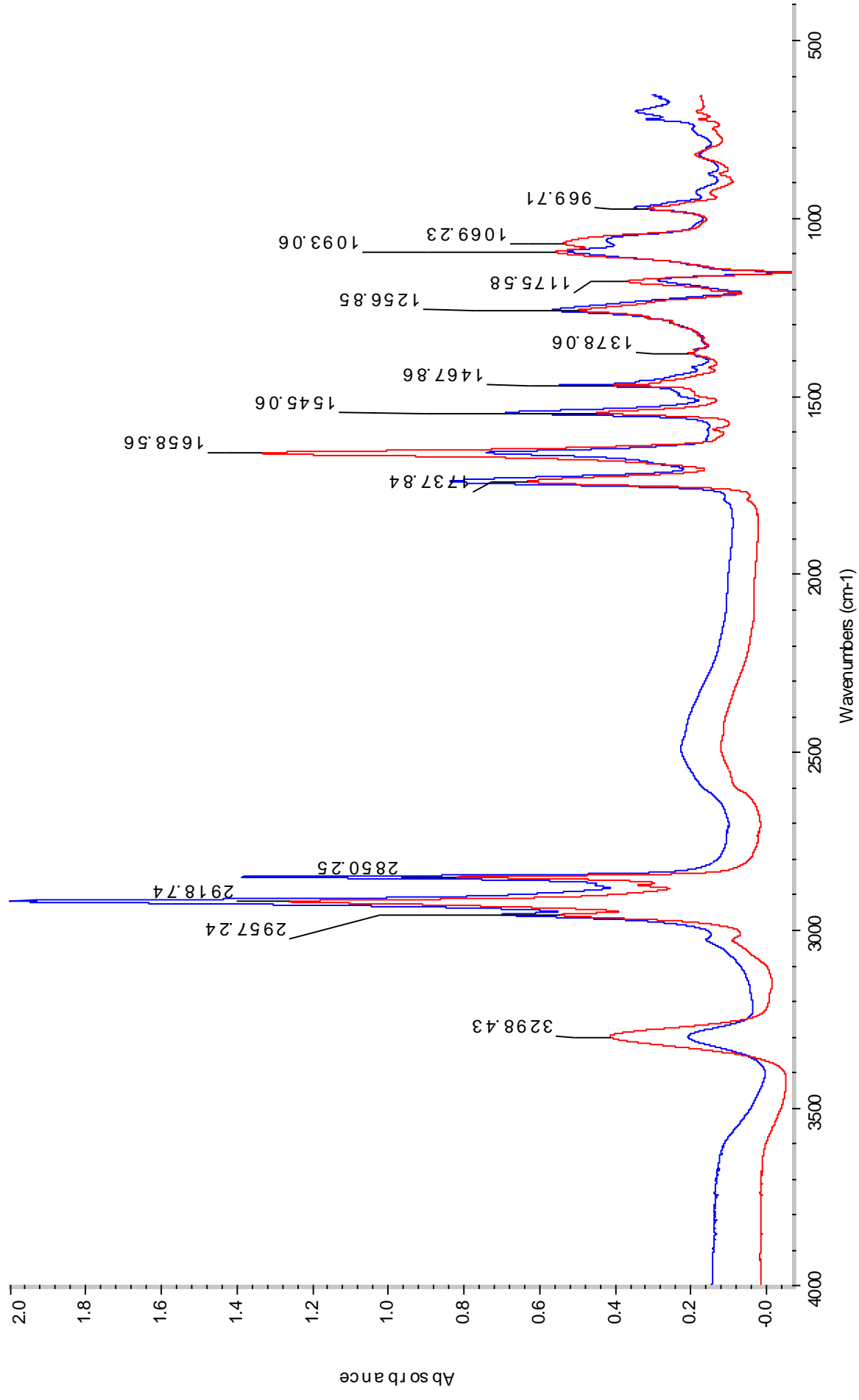


Figure (10): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 24 labeled with ($^{13}\text{C}=\text{18}\text{O}$) (Absorption of parallel in red and perpendicular in blue)

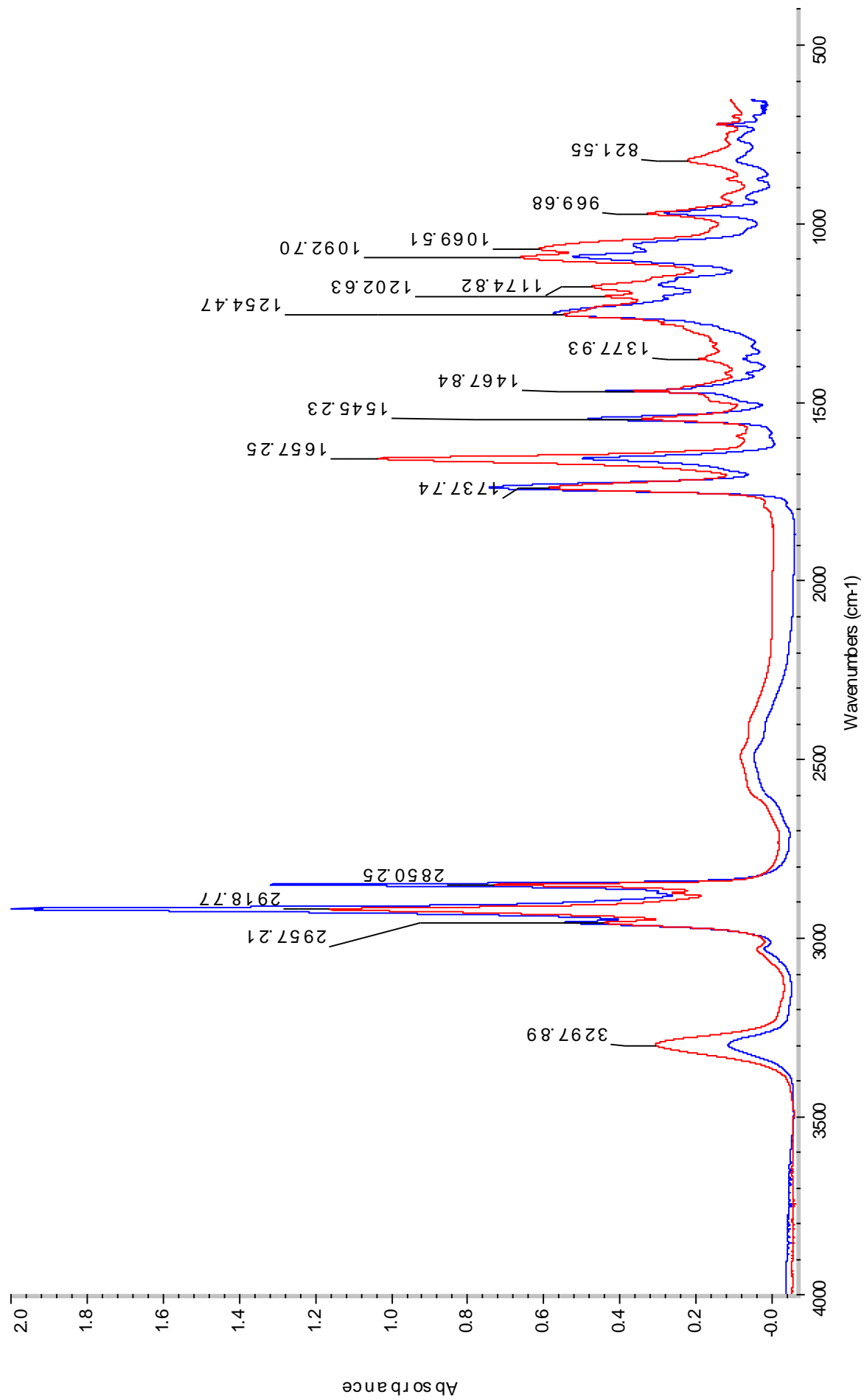


Figure (11): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 25 labeled with (¹³C=¹⁸O) (Absorption of parallel in red and perpendicular in blue)

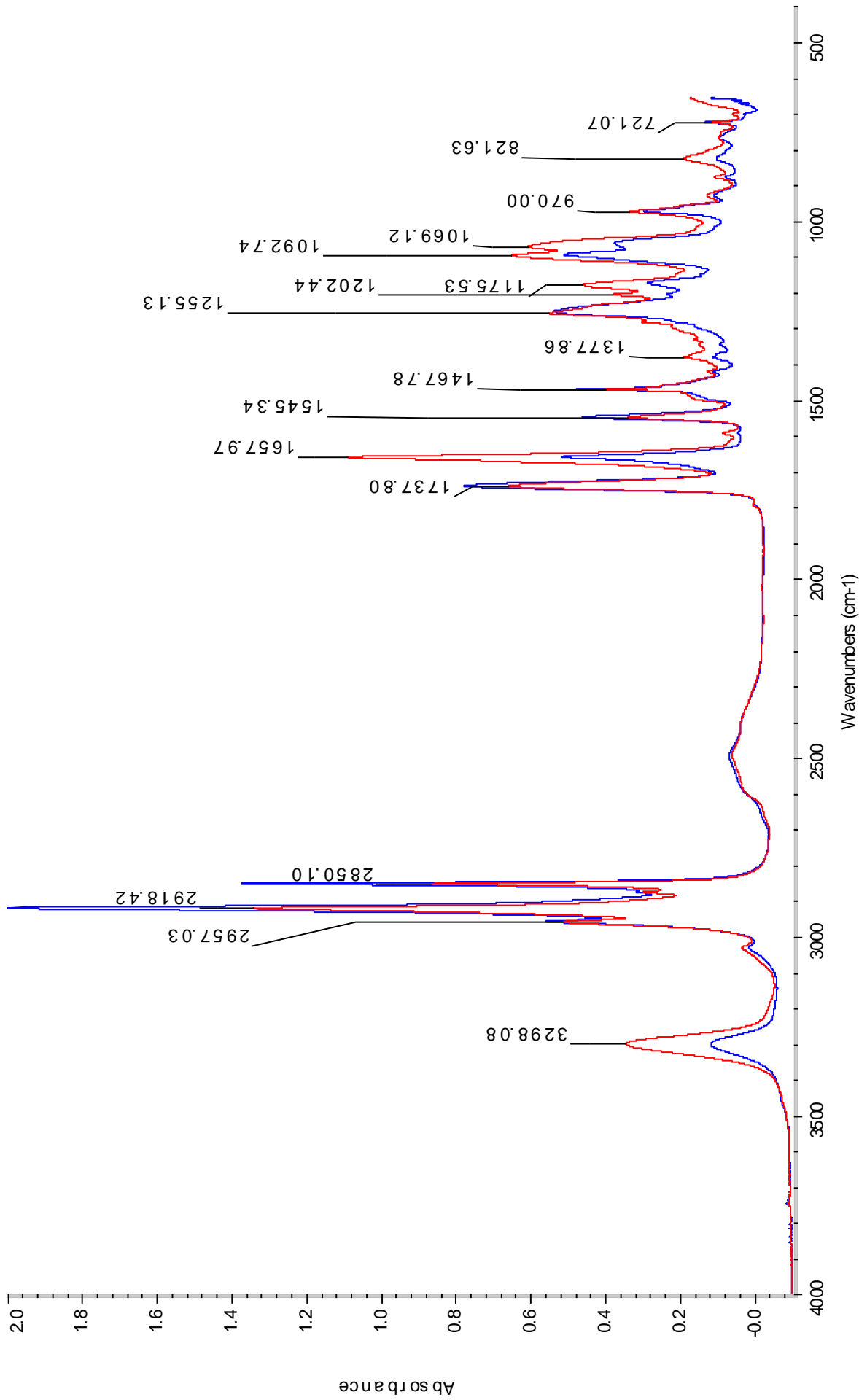


Figure (12): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 26 labeled with (¹³C=1⁸O) (Absorption of parallel in red and perpendicular in blue)

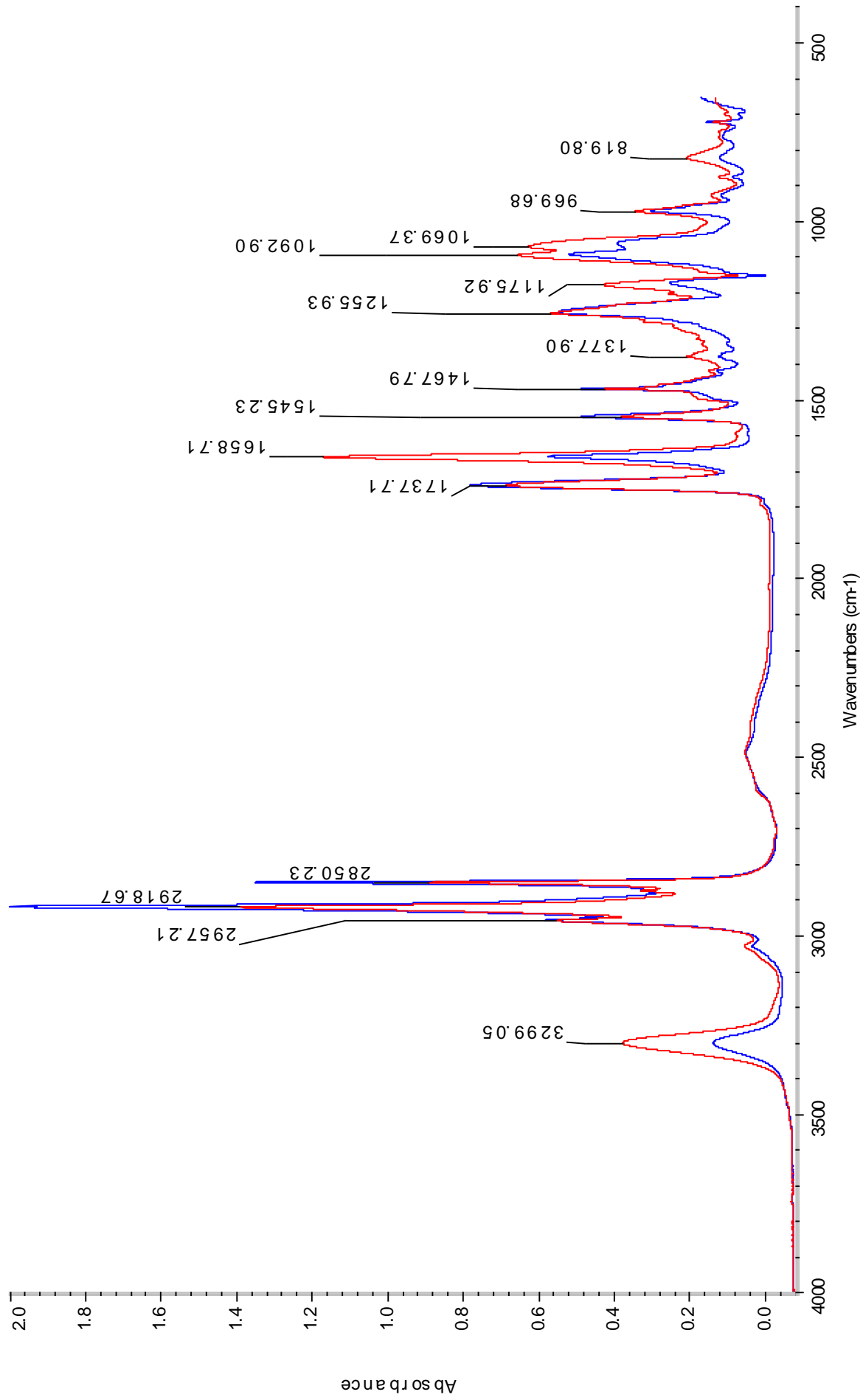


Figure (13): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 27 labeled with ($^{13}\text{C}=\text{^{18}\text{O}}$) (Absorption of parallel in red and perpendicular in blue)

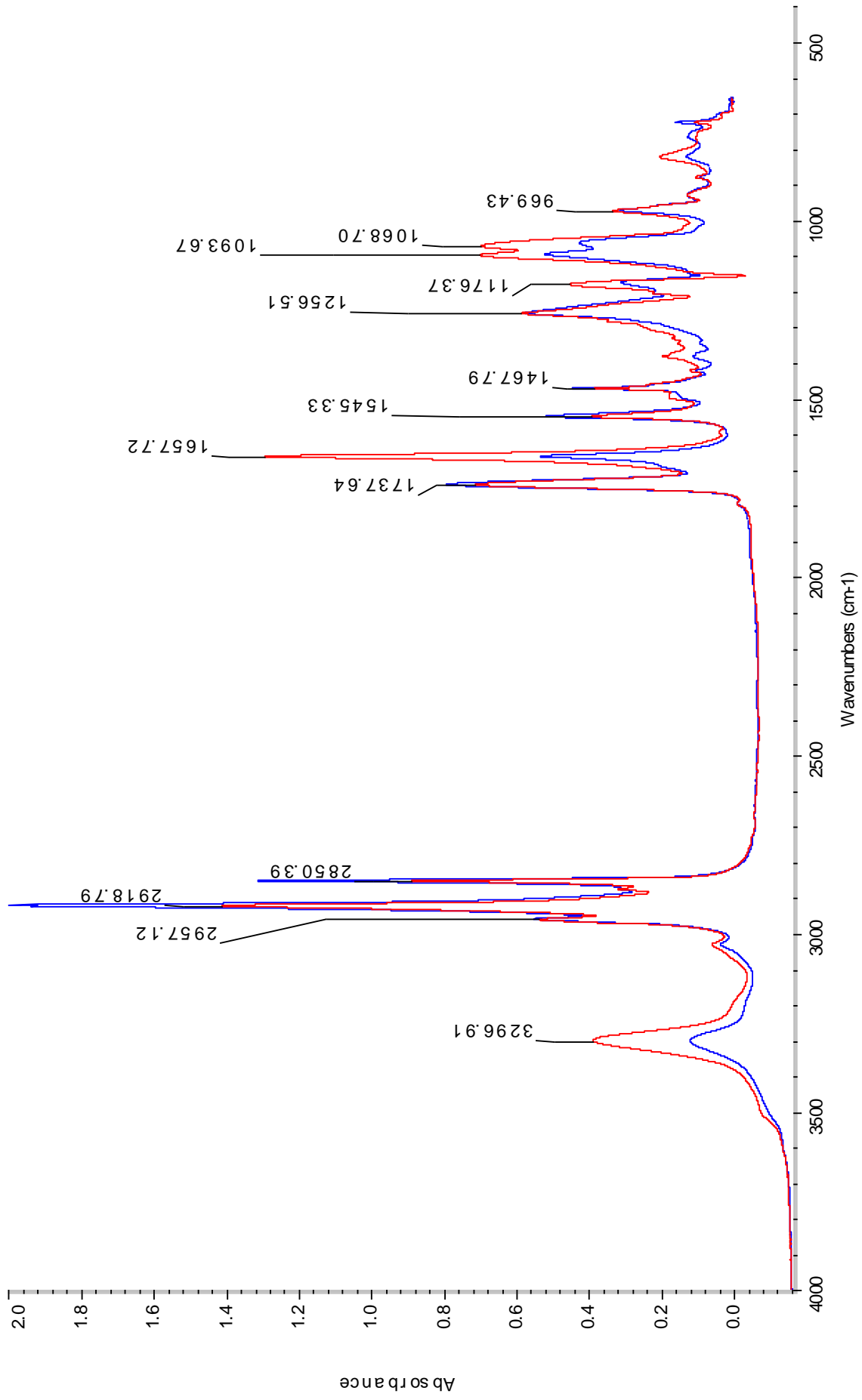


Figure (14): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 28 labeled with ($^{13}\text{C}=\text{^{18}\text{O}}$) (Absorption of parallel in red and perpendicular in blue)

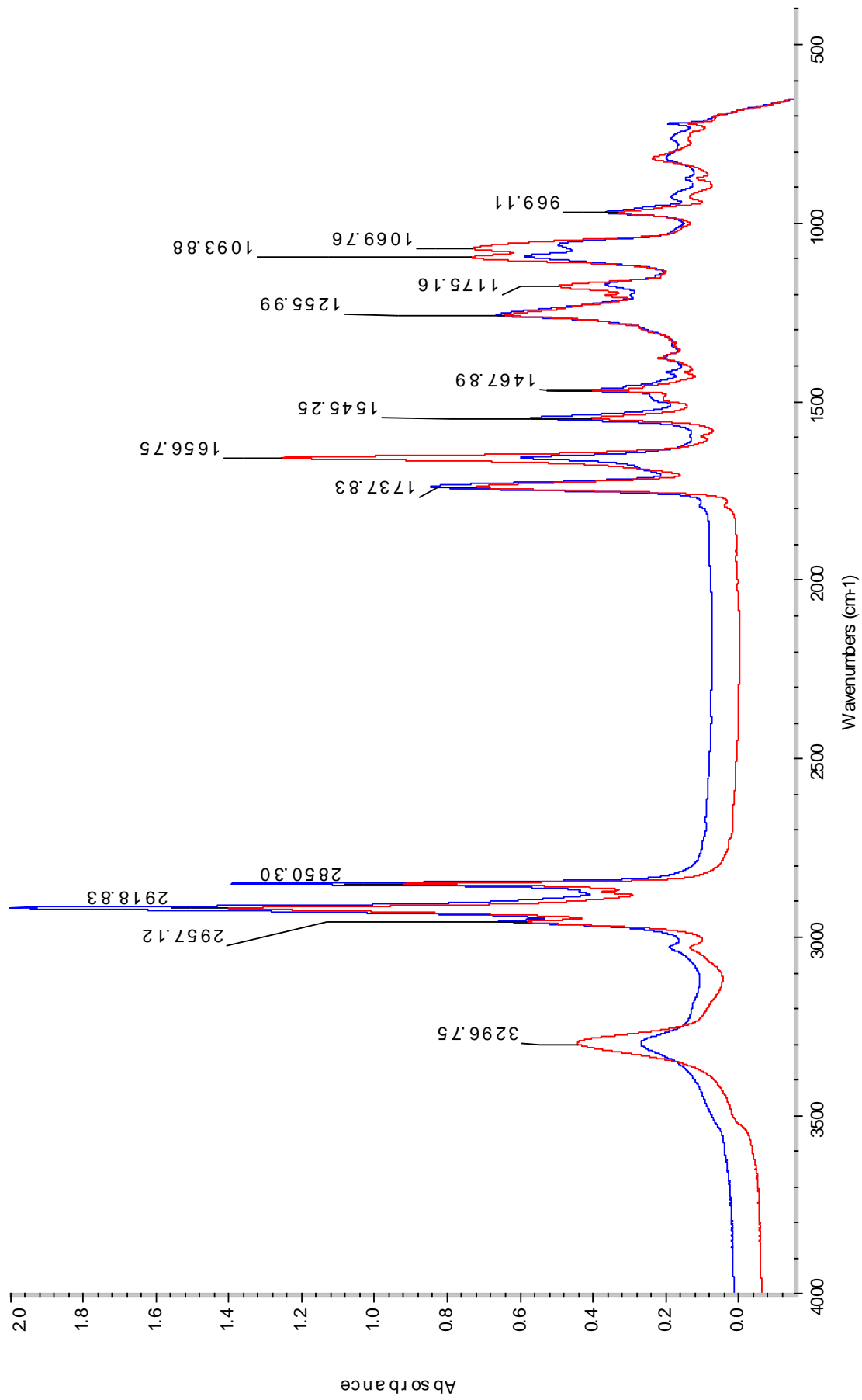


Figure (15): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 29 labeled with (¹³C=¹⁸O) (Absorption of parallel in red and perpendicular in blue)

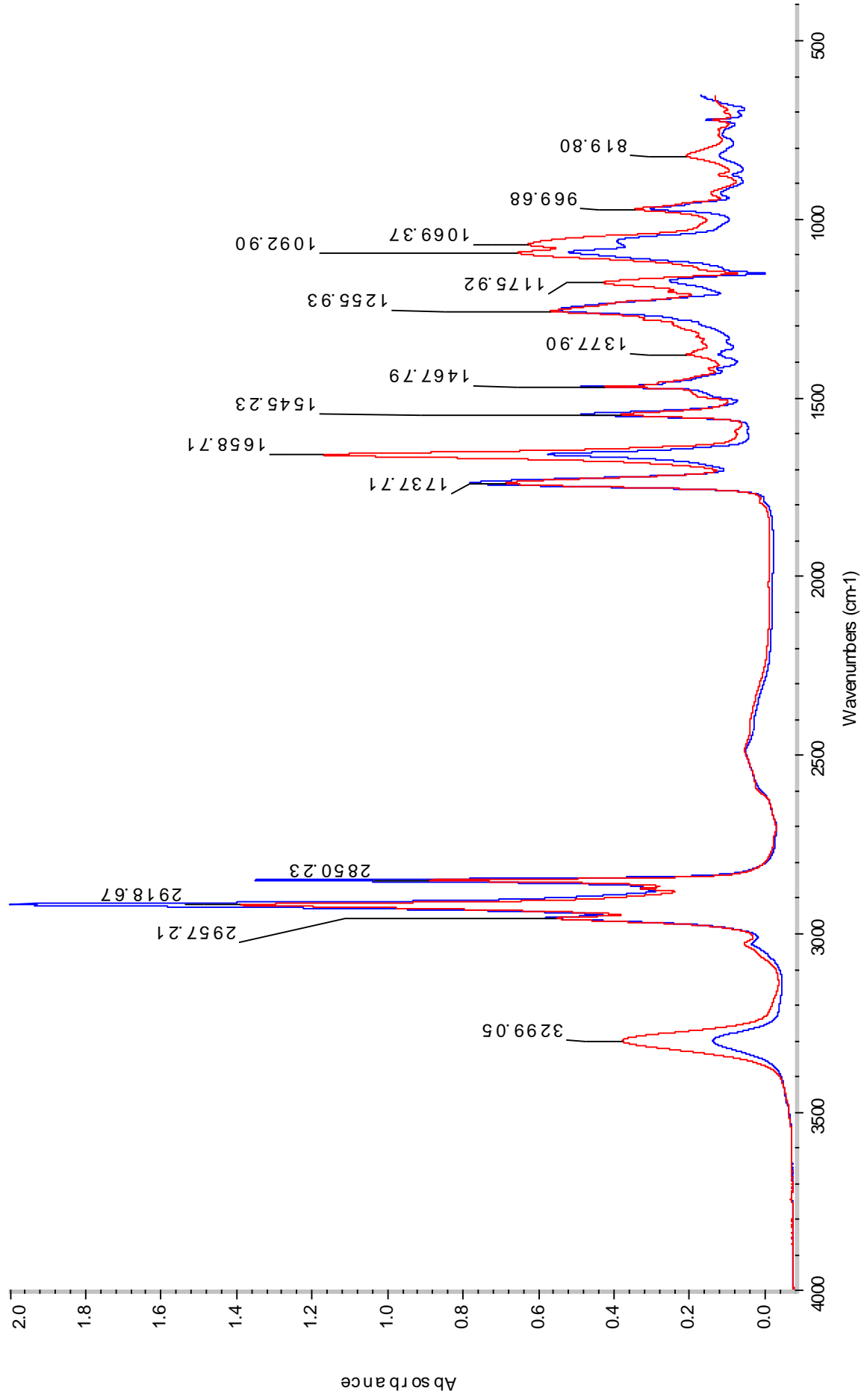


Figure (16): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 31 labeled with (¹³C=¹⁸O) (Absorption of parallel in red and perpendicular in blue)

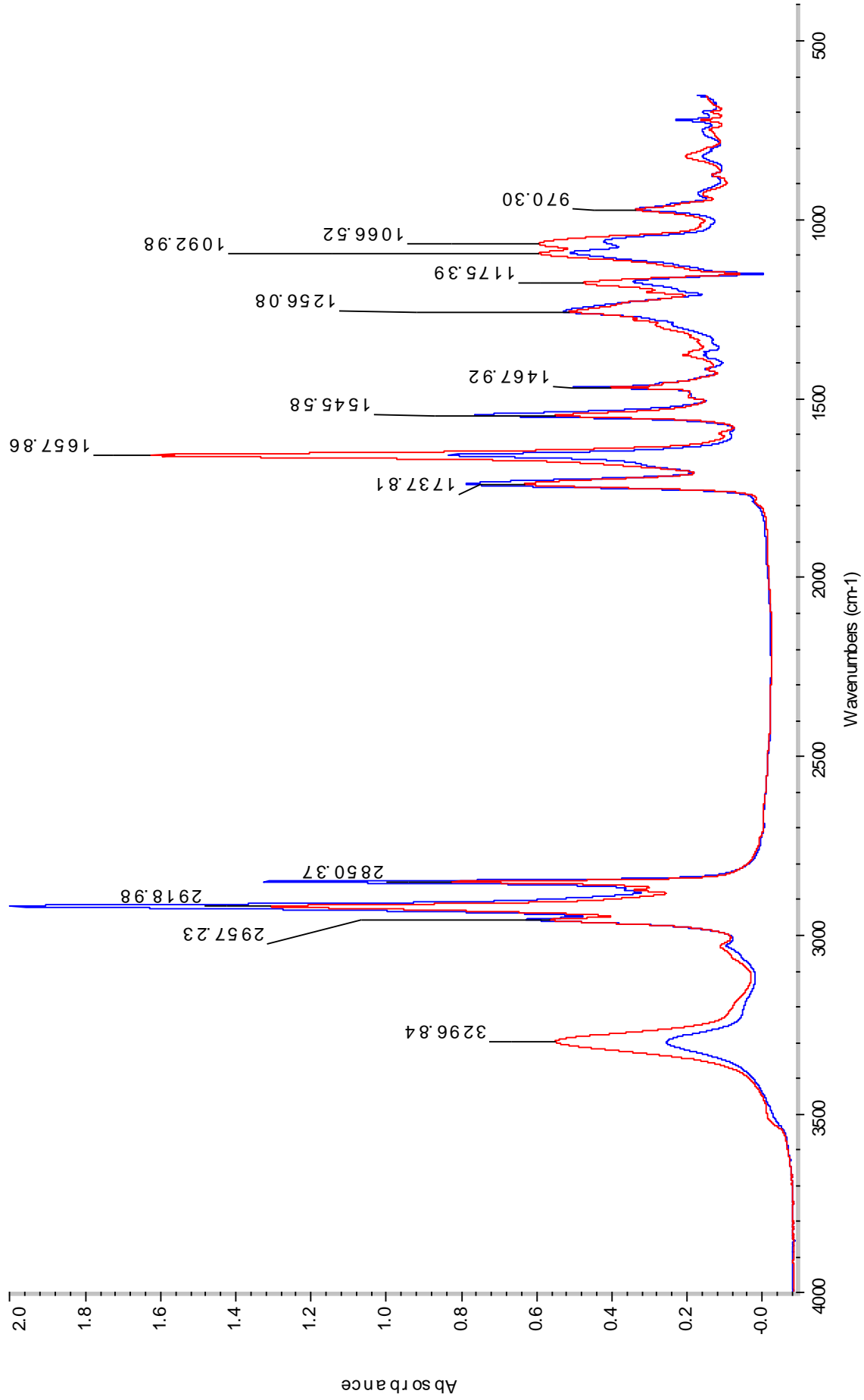


Figure (17): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 33 labeled with (¹³C=1⁸O) (Absorption of parallel in red and perpendicular in blue)

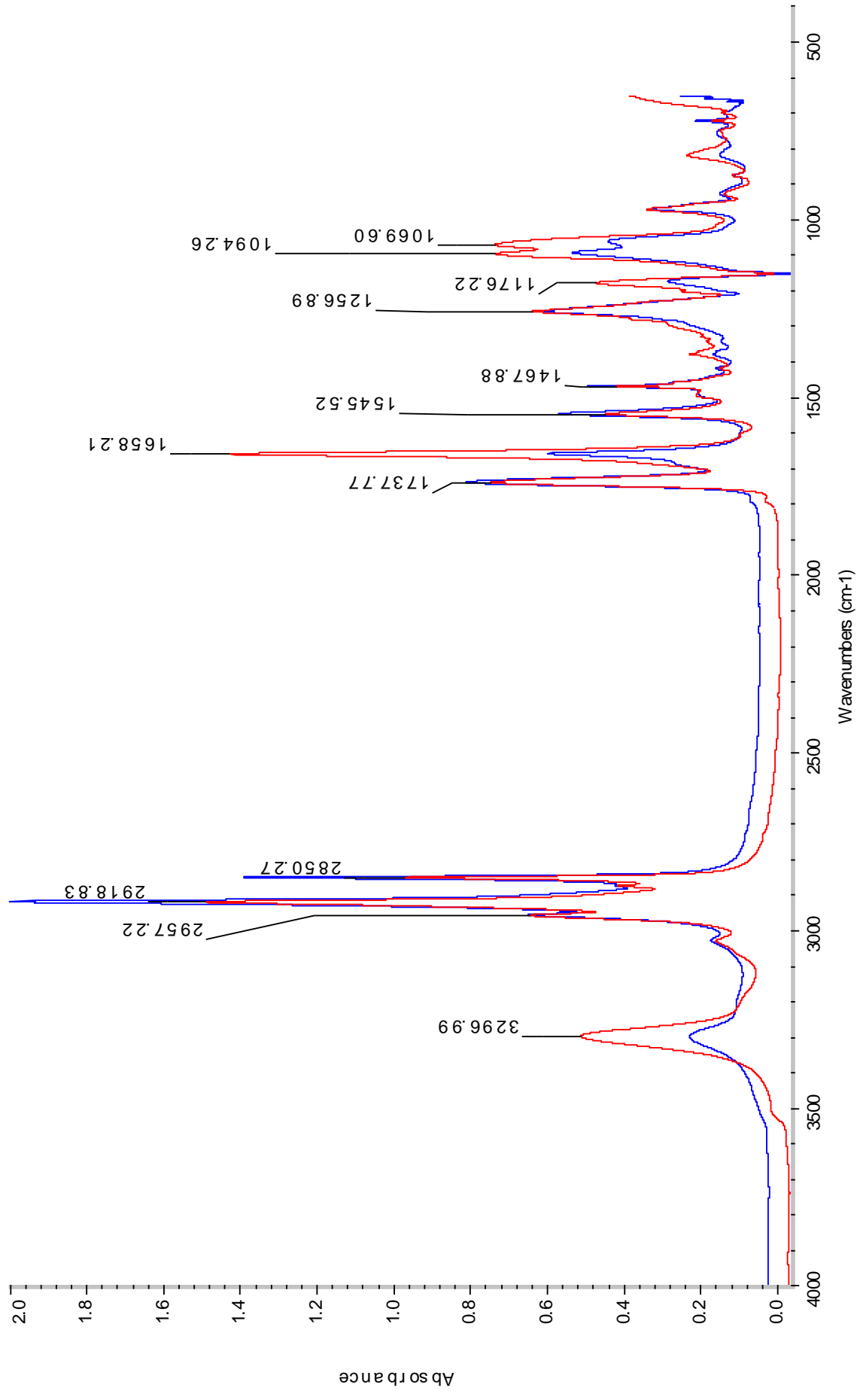


Figure (18): ATR- FTIR spectra of E protein TMD peptide reconstituted in lipid bilayer with residue No. 34 labeled with (¹³C=¹⁸O) (Absorption of parallel in red and perpendicular in blue)