INTRINSIC FLEXIBILITY OF WEST NILE VIRUS PROTEASE: IN SOLUTION CHARACTERIZATION

Committee:

_____
Stanley Watowich, Ph.D. Supervisor

_____
Kyung Choi, Ph.D. Chair

_____
Werner Braun, Ph.D.

_____
Alexander Freiberg, Ph.D.

_____
Marcelo López-Lastra, Ph.D.

_____
Juan Olano, MD.

_____
Dean, Graduate School

INTRINSIC FLEXIBILITY OF WEST NILE VIRUS PROTEASE: IN SOLUTION CHARACTERIZATION

by

Andrea D. Garces

Dissertation

Presented to the Faculty of the Graduate School of The University of Texas Medical Branch in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Medical Branch

2013

Dedication

To my parents, because they are the reason of my success: my father, Carlos Garcés Astudillo, who has been both the inspiration and model to follow on my journey to new horizons, and my mother, Teresa Fernández Paganini, who gave me support, strength and love when I most needed.

To my sister and best friend, Katherine Garcés Fernández .

To Rodrigo Diaz-Espinoza, for his encouraging feedback, care and love,

and

To my soon-to-be-born daughter.

Acknowledgements

First, I would like to express my deepest gratitude to my mentor Dr. Stan Watowich. Dr. Watowich was an excellent research advisor during these years, he gave me support when I most needed and allowed me to develop my independence and critical thinking to become what I am today. Also, I want to extend my immense appreciation to present and former members of Watowich lab, Dr. Suzanne Tomlinson, Dr. Usha Viswanathan and Dr. Robert Malstrom.

For their constant guidance and scientific challenges I want to thank to my dissertation committee members, Dr. Kyung Choi, Dr. Werner Braun, Dr. Alexander Freiberg, Juan Olano, all of them from UTMB and Dr. Marcelo López-Lastra from Pontificia Universidad Católica de Chile (Chile).

Special thanks to my program director, Dr. Andres Oberhauser and Dr. James C. Lee, for giving me the chance to extend my scientific horizons out of Chile and be part of the Structural Biology program as graduate student at UTMB.

Thanks to my friends and colleagues, Dr. Keerthi Gottipati, Jeff Chen, Whenzel Lu and my Chilean/Spanish friends in Houston to make this time fun and joyful.

My deepest thanks to the most important person in my life: Rodrigo Diaz-Espinoza, with whom we will raise a family soon, our biggest project together.

INTRINSIC FLEXIBILITY OF WEST NILE VIRUS PROTEASE: IN SOLUTION CHARACTERIZATION

Publication No _____

Andrea Garces, PhD

The University of Texas Medical Branch, 2013

Supervisor: Stanley Watowich

ABSTRACT

West Nile virus (WNV) is a mosquito-borne flavivirus with a rapidly expanding global distribution. Infection can cause severe neurological disease and fatalities in humans. Efforts are ongoing to develop antiviral drugs that inhibit the WNV protease, a viral enzyme required for polyprotein processing. Unfortunately, little is known about the solution structure of recombinant WNV protease (NS2B-NS3pro) used for antiviral drug discovery and development, although X-ray crystal structures and NMR studies have provided valuable insights about the interactions between NS2B-NS3pro and peptide-based inhibitors. To determine the solution structure and dynamics of the substrate-free WNV NS2B-NS3pro and understand the conformational changes due to substrate binding, we performed Circular dichroism spectroscopy, fluorescence experiments, small angle X-ray scattering (BioSAXS) and Fourier transform infrared spectroscopy (FT-IR) experiments. Our results suggested that in absence of a substrate analogous, the NS3pro domain adopted a well-folded compact tertiary structure similar to the observed crystallographic structures. However, while the NS2B cofactor maintained its secondary structure folding elements, it adopted multiple conformations distinct from previously described open and closed structures. In presence of a substrate analogous, WNV NS2B-NS3pro acquires a more compact folding state, which can be represented by the closed state described by X-ray crystallography data, suggesting that the substrate analogue acts as a folding inductor. This detailed description of the folding state and intrinsic flexibility of WNV NS2B-NS3pro can aid structure-guided discovery and optimization of WNV protease inhibitors as well as improve understanding of WNV polyprotein post-translation processing.

TABLE OF CONTENTS

Acknowledgements i

Abstract ii

Table of contents iii

List of Figures vi

List of Tables viii

List of Abbreviations ix

Chapter 1 INTRODUCTION 1

1.1 West Nile Virus: Infectious agent 1

1.2 Molecular nature of West Nile Virus 1

1.3 WNV protease as target for drug discovery..... 2

1.4 Rational of the study 3

1.4.1 Background problem 3

1.4.2 Central hypothesis 4

Chapter 2 MATERIALS AND METHODS 6

2.1 Over-expression and purification of WNV NS2B-NS3pro. 6

2.2 Circular Dichoism spectroscopy. 6

2.3 ANS binding experiments. 7

2.4 Sedimentation velocity. 7

2.5 Protease kinetic assays. 7

2.6 Bio Small angle X-ray Scattering (BioSAXS) 8

2.6.1 Data collection and quality determination. 8

2.6.2 Data analysis. 8

2.7 Fourier Transformed Infrared Spectroscopy (FT-IR). 10

2.8 Intrinsic Fluorescence under denaturating conditions..... 10

2.9 Limited Proteolysis. 10

Chapter 3 STRUCTURAL CHARACTERIZATION OF THE SUBSTRATE-FREE ACTIVE WNV NS2B-NS3..... 11

3.1 Abstract 11

3.2 Results ……..…... 11
3.2.1 Purified WNV NS2B-NS3pro is functionally active and monomeric in solution ...11
3.2.2. Characterization of the folding state of NS2B-NS3pro through Circular Dichroism spectroscopy. ……..…… 12
3.2.2.a Effect of crowding agents. …….. 15
3.2.2.b. Effect of counter ions……..… 15
3.2.2.c. Effect of pH …… 15
3.2.3 Identification of potential hydrophobic cores through ANS binding experiments...17
3.2.3.a ANS binding experiments with WNV NS2B-NS3pro in absence of substrate . ……..….. 17
3.2.3.b Complex WNV NS2B-NS3pro-Aprotinin……... 19
3.2.4 Characterization of the folding state of NS2B-NS3pro through Small Angle X-ray scattering (BioSAXS). …… 21
3.2.5 Solution state models with a flexible NS2B cofactor could reproduce the experimental BioSAXS data. ……..……. 25
3.2.6 Fourier transformed infrared spectroscopy (FT-IR). ……..…..…… 29
3.2.6.a Control proteins……...…31
3.2.6.b FT-IR studies suggested NS2B-NS3pro developed similar secondary structure in solution and in crystalline environments……. 33
3.3 Discussion …...…… 35

Chapter 4 DETERMINATION OF THE SOLUTION STRUCTURE OF WNV PROTEASE IN PRESENCE OF SUBSTRATE ANALOGUE……..……...……. 38
4.1 Abstract ……..…… 38
4.2 Results ...…….... 38
4.2.1 Determination of inhibition constant of peptide analogue by Kinetic assays.…….38
4.2.2 Characterization of substrate-free-, bound to substrate- and denatured states of WNV NS2B-NS3pro by BioSAXS…….. 40
4.2.2.a Effect of substrate analogue Ac-Lys-Lys-Arg-NH_2 on the WNV NS2B-NS3pro folding state …… 40
4.2.2.b Effect of Urea as a denaturating agent…… 40
4.2.3 Comparison of in solution substrate bound- conformation of WNV NS2B-NS3pro

with crystallographic structures. ……..…. 42
4.2.4 Structural stability of WNV NSB-NS3pro followed by Limited Proteolysis ……. 44
4.2.5 Chemical denaturation of WNV NS2B-NS3pro followed by
Intrinsic Fluorescence. …….…… 48
4.2.5.a Effect of the substrate analogue on the protein stability of
WNV NS2B-NS3pro followed by intrinsic fluorescence……..…….. 50
4.2.5.b Effect of Crystallization buffer on the protein stability of
WNV NS2B-NS3pro followed by intrinsic fluorescence. …….…… 51
4.3 Discussion ……...……. 52

Chapter 5 GENERAL CONCLUSIONS ...……..……... 55
Bibliography ……..……....……... 57
Vita ...……..……..……..…… 61

LIST OF FIGURES

Figure 1.1: WNV polyprotein processing. …….…….. 2

Figure 2.1: Schematic representation of structure-based computational screens…….…… 4

Figure 3.1: Test for functionality and oligomerization state of WNV protease ……..…. 13

Figure 3.2: Circular dichroism of WNV NS2B-NS3pro……... 14

Figure 3.3: Secondary structure content of active WNV NS2B-NS3pro followed by CD spectroscopy (Far UV)……...… 16

Figure 3.4: ANS binding experiments of WNV NS2B-NS3pro in absence of substrate …. 18

Figure 3.5: Inhibition effect on Aprotinin on WNV NS2B-NS3pro activity followed by fluorescence signal……...… 20

Figure 3.6: Circular Dichroism spectroscopy of WNV NS2B-NS3pro, Aprotinin and Protease-Aprotinin complex…… 20

Figure 3.7: ANS binding experiments of WNV NS2B-NS3pro-Aprotinin complex…… 22

Figure 3.8: BioSAXS solution characterization of WNV NS2B-NS3pro……..….. 24

Figure 3.9: Comparison of SAXS data collected from solutions of WNV NS2B-NS3pro and calculated from NS2B-NS3pro crystallographic structures……...…. 26

Figure 3.10: Schematic representation of WNV protease models used for EOM……..….. 28

Figure 3.11: BioSAXS analysis of flexible models of WNV NS2B-NS3pro…….…. 29

Figure 3.12: Ensemble of selected conformations of model NS2B-F…….…. 31

Figure 3.13: Determination of secondary structure content of control proteins by FT-IR….. 32

Figure 3.14: WNV NS2B-NS3pro secondary structure determined by FT-IR……..... 34

Figure 4.1: Test for inhibition of WNV protease activity by Ac-KKR-NH_2 as peptide analogue…….…….. 39

Figure 4.2: BioSAXS solution characterization of WNV NS2B-NS3pro in absence, presence of substrate and denatured states. ……...…… 41

Figure 4.3: Comparison of SAXS data collected from WNV NS2B-NS3pro in presence of substrate analogue and from NS2B-NS3pro crystallographic structures…… 45

Figure 4.4: Limited proteolysis on WNV NS2B-NS3pro by proteinase K at different incubation times……..…… 46

Figure 4.5: Limited proteolysis on WNV NS2B-NS3pro by proteinase K……...……. 47

Figure 4.6: Trypthophan residues observed on the crystal structure of WNV NS2B-NS3pro …….. 48
Figure 4.7: Chemical denaturation of WNV NS2B-NS3pro followed by intrinsic fluorescence …… 49
Figure 4.8: Unfolding curves of WNV NS2B-NS3pro followed by intrinsic fluorescence...52

LIST OF TABLES

Table 3.1: Values of BioSAXS parameters of WNV NS2B-NS3pro in absence of substrate …….…….. 30

Table 3.2: Comparison of χ^2 values of EOM obtained from open and closed crystallographic structures…… 31

Table 3.3: Secondary structure percentage estimated by FT-IR experiments: Lysozyme and Trypsine as control proteins……... 33

Table 3.4: Comparison between secondary structure content determined by FT-IR spectroscopy and calculated from model structures ……..…… 34

Table 4.1: BioSAXS parameters of substrate free-, bound to substrate- and denatured states of WNV Protease…… 43

LIST OF ABBREVIATIONS

$(NH_4)^2SO_4$	amonium sulfate
3D	three dimensional
A.U.	arbitrary units
ºC	Celsius degrees
cm	centimeters
Da	Dalton
Dmax	Maximum distance
EOM	Ensemble of conformational model
FT-IR	Fourier transform infrared spectroscopy
hrs	hours
IPTG	isopropyl a-D-thiogalactopyranoside
kDa	kiloDalton
Ki	Inhibition constant
k_{cat}	Catalytic constant
Kd	Dissociation constant
K_2HPO_4	Potassium phosphate buffer
M	Molar
mg	milligrams
min	minutes
mM	millimolar
NaCl	sodium chloride
NaF	sodium fluoride
nm	nanometers
nM	nanomolar
NMR	nuclear magnetic resonance
NSD	normalized spatial discrepancy
OD	optical density
PDB	Protein Data Bank
P(r)	Pair distance distribution plot
PK	proteinase K
Rg	Radio of gyration
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SSE	Secondary Structure Estimation
SI	Substrate Inhibition Mechanism
TMAO	Trimethylamine N-oxide
µg	micrograms
µL	microliter
µM	micromolar
WNV	West Nile virus
WNpro, WNVpro	West Nile protease
x g	g-force

CHAPTER 1

INTRODUCTION

1.1 West Nile Virus: Infectious agent

West Nile Virus (WNV) is a ~11 kb ssRNA virus that belongs to the *Flaviviridae* family *(1)*. WNV is also genetically related to other viruses known to affect the central nervous system in humans such as dengue, tick-borne encephalitis, yellow fever and Japanese encephalitis viruses. The first reported case of WNV infection in USA occurred in New York in 1999 and since then this mosquito-borne virus has been responsible for hundreds of thousands of human infections and tens of thousands of disease cases, becoming the leading cause of epidemic encephalitis in USA (from Center of Disease Control and Prevention) *(2)*. Unfortunately, there are no vaccines or antiviral therapies available to prevent and/or treat human WNV infection, although several veterinarian vaccines have been implemented in horses and other susceptible animals *(3)*. Most of the reported cases of WNV infections in humans are asymptomatic, however, it can cause severe febrile illness, meningitis, encephalitis and acute flaccid paralysis *(4)*. West Nile neuroinvasive disease can be the cause of cognitive dysfunction and movement disorders, limb weakness or paralysis *(4, 5)*.

1.2 Molecular nature of West Nile Virus

The WNV genome encodes a polyprotein that includes three structural (Capsid, pre Membrane and Envelope) and seven non-structural (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) proteins *(6, 7)* (Figure 1.1 (adapted from Assenberg R., et al., 2009 *(7b)*)). Non-structural proteins are functionally related to virus assembly and maturation. NS3 is a multifunctional protein with a C-terminal portion that contains a nucleotide/RNA triphosphatase and helicase domain and a N-terminal domain that encodes a serine protease (NS3pro) *(3)*. Although *in vitro* activity of recombinant NS3pro requires interaction with a hydrophilic segment (~40 amino acids) from the transmembrane protein NS2B *(8, 9)*, NS3pro activity is only slightly modified by interaction with the C-terminal domain of NS3 *(10, 11)*.

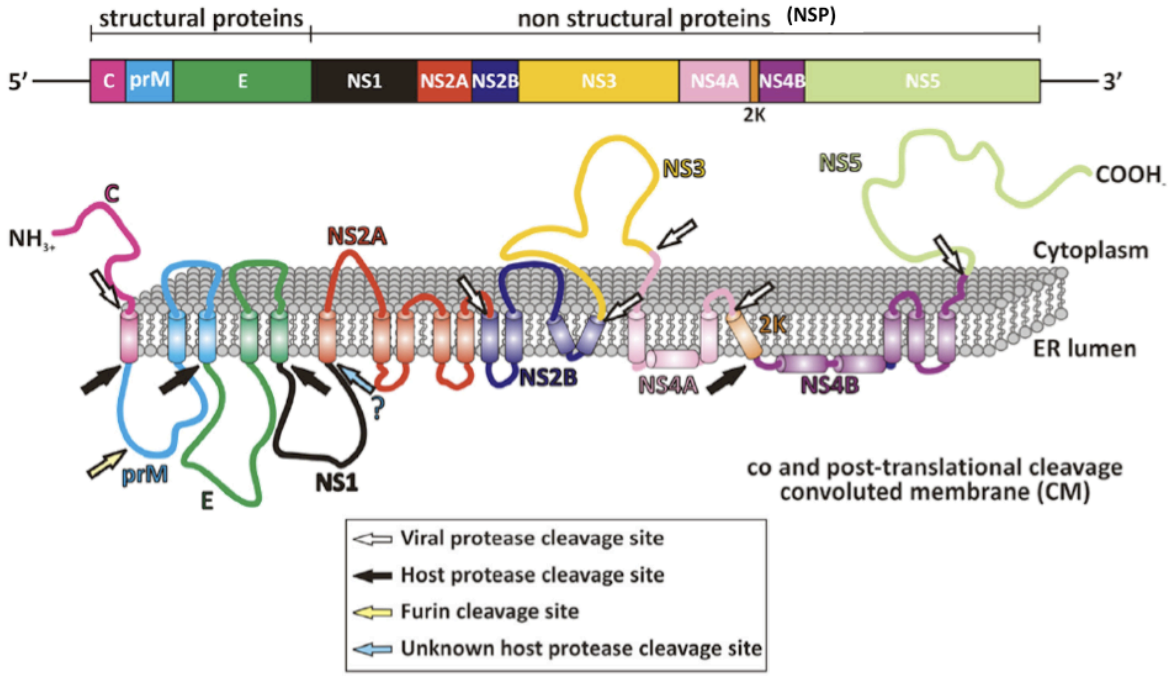

Figure 1.1 WNV polyprotein processing. Top panel. Organization of the viral ssRNA indicating the three structural and the seven non structural proteins. Bottom panel. Spatial distribution of the viral proteins in the endoplasmic reticulum membrane. (Modified from Assenberg R., et al., J Virol. 2009 *(7b)*.

1.3 WNV protease as target for drug discovery.

WNV protease is a promising target for antiviral drugs due to its critical role in cleaving the NS2A/B, NS2B/3, NS3/4A and NS4B/A junctions during post-translational processing of the viral polyprotein *(6, 7)*. High-throughput screening studies directed against the WNV protease have identified peptidic leads as di- and tripeptide aldehydes, tetrapeptide aldehydes, and D-arginine-based inhibitors *(12-14)*. In addition, low molecular weight non-peptidic protease inhibitors have been generated from biochemical high throughput and structure-based computational screens *(15-18)*. Unfortunately, small molecule protease inhibitors have exhibited micromolar binding constants (Ki) to recombinant NS2B-NS3pro. To help develop more potent inhibitors, the protein has been crystallized in absence of substrate, although inactive, *(19)* and co-crystallized with substrate-like peptide (e.g., Naph-KKR-H; *(20)* and protein (BPTI, *(19)*) inhibitors. A portion of the NS2B chain was observed to adopt different conformations in absence

and in presence of substrate: in the "open" structure the NS2B cofactor was ~38 Å from the active site (Cα Ans 84 of NS2B to Cα Ala 51 of NS3pro) whereas in the "closed" structure the NS2B was wrapped around NS3pro and interacted with inhibitors within the active site.

Complementary NMR studies were performed to understand inhibitor-induced NS2B conformational changes *(21, 22)*. These solution state studies concluded that in presence and absence of a substrate analogue, WNV NS2B-NS3pro existed as ensembles consisting predominantly of the closed conformation in equilibrium with small populations of open conformations where the C-terminal region of NS2B (residues 72-90) were likely disordered; inhibitors shifted the equilibrium to a more strongly populated closed state. Although there was general agreement between X-ray and NMR experiments regarding the conformation of WNV NS2B-NS3pro bound to an inhibitor, the substrate-free conformation of WNV NS2B-NS3pro remains ambiguous since X-ray crystallographic and NMR studies suggested the dominant structure corresponded to either an open or closed conformation, respectively. In a recent work, the effect of membrane on the flexibility of WNV protease was also assayed in NMR experiments, however, the intrinsic dynamic that the free protease seems to have, was again observed through a just partial assignment *(23)*.

1.4 Rational of the study

1.4.1 Background problem

Developing drugs that inhibit WNV protease activity using structure-based computational screens (Figure 1.2) has become fruitless mainly due to the discrepancies between the predicted and the experimentally determined affinity constants (Ka), the latest resulting often poorer than the former ones. Among the potential sources of this problem, it is possible to identify 1) the theoretical determination of Ka by the docking software itself due to calculation errors on the predicted binding energy or conformational changes due to ligand binding that the software was not able to incorporate, 2) the small compound library utilized, in which solubility problems are very often observed restricting the compounds to test experimentally, and 3) the definition of the target protein, in this case the use of a crystal structure that provides the 3D coordinates for the virtual docking. In this last scenario, it is assumed that the folding state of the target protein in solution is well represented by the crystal structure and in the case of WNV NS2B-NS3pro there

is some evidence that suggest that a flexibility factor should be considered at the moment of defining a viral protease as a target for drug discovery.

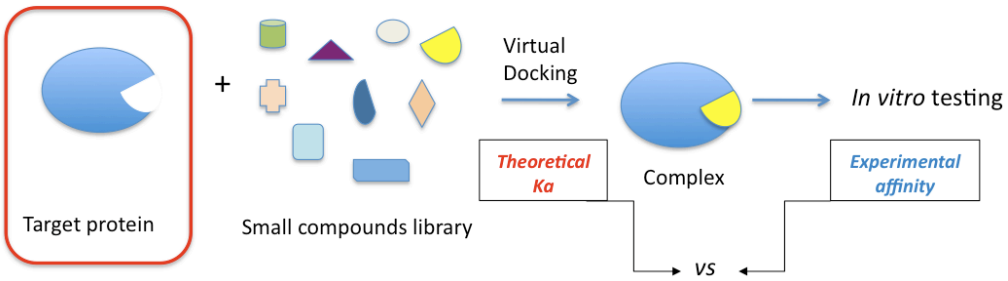

Figure 1.2: Schematic representation of structure-based computational screens.

Much of the ambiguity regarding the NS2B-NS3pro structure in absence of substrate centers on the conformation and flexibility of the NS2B chain and the interactions it makes with NS3pro *(19, 20)*. The discrepancies between in solution NMR data and X-ray crystallography on the characterization of WNV protease in its substrate-free-conformation led me to use a more straightforward approach that overcome problems associated to mutations or excessive protein dynamic, therefore offering a better-defined target for drug discovery that may improve the finding of effective leads. In addition, a detailed characterization of the effect of a substrate analogue on the structural stability and folding state of WNV protease requires to be done, in order to provide a better understanding of its role on the protease activity and highlight the molecular details that may rule the protease inhibition.

Therefore, this study seeks to provide a detailed description of the folding state and structural flexibility that WNV NS2B-NS3 protease adopts in solution, in absence and presence of a substrate analogue, through different experimental approaches. These include Circular dichroism, Small angle X-ray scattering, fluorescence experiments, limited proteolysis and FT-IR experiments on a functionally active protein sample. Since the main focus of this dissertation is the structural characterization of the WNV protease as a target for drug discovery, I will use the same protein construct used for inhibition assays in the field *(20)*.

1.4.2 Central hypothesis

Based on the presented experimental evidence, my working hypothesis about the folding state of WNV protease indicates; ***WNV protease is a partially folded protein that acquires its***

functional structure in the presence of its substrate. In order to test this hypothesis, two specific aims were designed:

Specific Aim 1: *Structural characterization of the substrate-free active WNV NS2B-NS3pro.* A detailed characterization of the WNV protease in absence of substrate analogue requires to be done in order to provide a well-defined target for drug discovery. To this end, I used an experimental protocol that describes its potential structural flexibility and folding state in absence of its substrate analogue, without artifacts caused by excessive dynamics or artificial mutations.

Specific Aim 2: *Determination of the solution structure of WNV protease in presence of substrate analogue*. According to preliminary data, WNV protease would adopt a more compact conformation in presence of peptidic inhibitors. To test this idea I performed structural studies in solution that characterize WNV protease in presence of a substrate analogue. This aim will provide a better understanding of its role on the stabilization of the folding state of the protease and highlight the mechanisms of inhibition.

CHAPTER 2

MATERIALS AND METHODS

2.1 Over-expression and purification of WNV NS2B-NS3pro. A gene fragment coding WNV NS2B-NS3pro (formed from in-frame concatenation of NS2B residues 48-93, a Gly_4-Ser-Gly_3 linker, NS3 residues 1-188, and a hexahistidine purification sequence; identical to the amino acid sequence of protein structure 3E90A *(20)*, was codon-optimized for *E. coli* expression and commercially synthesized (DNA2.0, USA). The resulting gene was subcloned into the expression plasmid pJexpress414 (DNA2.0) and used to transform BL21(DE3) Codon Plus-RIPL *E. coli* cells. Culture was grown in 2xYT medium (50 mg/ml carbenicillin) at 37 °C until OD_{600nm} of 0.6. Protein expression was induced by adding 0.5 mM IPTG and shaking for 4 hours at 37 °C. Cell pellet was resuspended with ice-cold lysis buffer (50 mM HEPES-HCl pH 7.5, 300 mM NaCl, 10 mM imidazole, 5 % glycerol, 300 ug/ml lysozyme, 30 ug/ml DNase, 10 mM $MgCl_2$, 0.05% Triton-X-100) and the clarified supernatant incubated with Nickel resin (GE Healthcare) for 1 hr at 4 °C. Resin was washed with 10 mM and 50 mM imidazole, and bound protein eluted with 500 mM imidazole. Fractions containing NS2B-NS3pro were dialyzed at 4 °C against 20 mM Glycine-NaOH pH 9.5, 5% glycerol. Protein was concentrated by centrifugation at 4 °C and frozen at -80 °C in the sample buffer. Final purification utilized size exclusion chromatography with a S75 column (GE Healthcare Life Science) with 10 mM potassium phosphate pH 8.0. Fractions corresponding to WNV NS2B-NS3pro were collected and concentrated. His-tag was not cleaved off.

2.2 Circular Dichroism spectroscopy. Experiments were performed with an AVIV Model 215 circular dichroism spectrometer using 200 μl sample, 0.1 mm path length, at final protein concentration of 0.1, 0.5 and 10 mg/ml. Measurements by triplicate were done at far (180-260 nm) and near (250 - 350 nm) UV regions for sample and buffer. Chemical denaturation was assayed using 6 M Urea in 10 mM potassium phosphate buffer pH 8.0. Crowding conditions were tested with 0 - 4 M TMAO and 40% Ficoll70. Effect of counter ions was assayed with 0 - 500 mM NaF and 0 - 500 mM $(NH_4)_2SO_4$. pH effect was studied with protein solved in 0.1 M Acetic acid/0.1 M Sodium acetate and 10 mM potassium phosphate buffer. Temperature effect on the

secondary structure content of WNV NS2B-NS3pro was analyzed at 4, 20 and 55 °C. Effect of higher temperatures was impossible to study due to equipment limitations.

2.3 ANS binding experiments. WNV NS2B-NS3pro and NS2B-NS3pro/BPTI complex samples at 0.05 mg/ml final concentrations were incubated with 0-6 M Urea in 10 mM K phosphate buffer pH 8.0 at 25 °C, 50 μM ANS (final concentration) was added to perform fluorescence experiments. Final volume per well was 100 μl. Each condition was done by duplicates. Excitation wavelength was 380 nm while the emission wavelength was 400-600 nm. Proper buffer measurement were performed and subtracted from the sample.

2.4 Sedimentation velocity. Experiments were performed with a Beckman-Coulter XL-A centrifuge using 400 μl of protein sample at final concentrations of 2.5 and 0.3 mg/ml in 10 mM K phosphate pH 8.0. Data acquired from runs at 45,000 rpm, 20 °C were analyzed with SEDFIT *(24)* to generate c(S) versus S plots, where c(S) is the concentration of protein divided by the sedimentation coefficient at the respective S position. Confidence level was set to P=0.95. Determination of theoretical sedimentation coefficient for crystallographic structures was done using HYDROPOD v10 *(25)*. Theoretical frictional ratio (f/f_0) was obtained using the following equations:

$$s_{20,w} = M\,(1-\nu_{20}\rho_{20,w})/Af$$

$$f_0=6\pi\eta(3M\nu/4\pi A)^{1/3} \qquad \text{(Stokes equation)}$$

where M is the molecular weight, ν_{20} is the solute partial specific volume at 20 °C, $\rho_{20,w}$ is the solution density at 20 °C, A corresponds to Avogadro number and f is the frictional coefficient. f_0 is the minimum frictional coefficient η is solvent viscosity.

2.5 Protease kinetic assays. Activity experiments were performed following previously established protocols *(26)*. Briefly, purified WNV NS2B-NS3 protease and a 7-amino-4-methylcoumarin (AMC) fluorophore-linked peptide substrate (Boc-GKR-AMC, Bachem, USA) were incubated at 25 °C in 50 mM glycine-NaOH pH 9.5, 30 % glycerol, 1 mM CHAPS. Release of free AMC was monitored every 10 min using a Fluorolog FL3-22 spectrofluorometer (Horiba Jobin Yvon) at 465 nm after excitation at 380 nm. Protein final concentration was 100 nM, while substrate concentration ranged from 1200 μM to 37.5 μM. Data was analyzed using Dynafit

software *(27)* to determine reaction mechanism and kinetic parameters.

2.6 Bio Small angle X-ray Scattering (BioSAXS)

2.6.1 Data collection and quality determination. Data was collected from a Rigaku (USA) BioSAXS-1000 Kratky camera on a Rigaku FRE+ X-ray source with a Cu target (1.54 Å wavelength). Experiments were performed at 20 °C. Protein samples were filtered through a 100 kDa filter and dialyzed overnight at 4 °C against a) 10 mM potassium phosphate pH 8.0, b) 10 mM potassium phosphate pH 8.0, 1.3 mM substrate analogue Ac-KKR-NH_2, and c) 10 mM potassium phosphate pH 8.0, 6 M Urea. Three different protein concentrations (1.0, 2.5, and 5.6 mg/ml) were used to test the presence of aggregates and/or electrostatic repulsion. BioSAXS data was collected from averaging 3 - 6 experiments, each with a 60 min exposure. No radiation damaged was detected after 60 min of exposure time during 10 hr at 20 ˚C (data not shown). Buffer absorbance spectra were separately collected and subtracted from sample spectra prior to data processing. Data processing was done with SAXSLab software (Rigaku, USA) and data analysis was performed using PRIMUS (ATSAS, suite 2.3) *(28)*. Quality of experimental BioSAXS data and determination of folding states were analyzed on the Guinier region and Kratky plot, respectively *(29)*. The scattering intensity at zero angle (I_0) and the radius of gyration (R_g) were evaluated using the Guinier approximation: $\ln(I) \sim \ln(I_0) - (qR_g)^2/3$, with the limit $qR_g<1.3$ *(30)*. In addition, R_g and maximum distance in the particle (D_{max}) were determined from the pair-distance distribution function P(r), which was obtained by indirect Fourier inversion of the entire scattering curve using GNOM *(28)*. The hydrated volume was obtained from the Porod invariant and the molecular weight was computed by scaling the Porod volume by 0.59 *(30)*.

2.6.2 Data analysis. Theoretical scattering curves were generated using the CRYSOL software package *(31)*. Crystallographic structures of either open (PDB identifier 2GGV; (19)) or closed (PDB identifier 3E90; (20)) WNV NS2B-NS3pro were fit to the experimental scattering data. The agreement between theoretical and experimental scattering curves was evaluated by the discrepancy value χ^2 within a range of $q< 0.5$ *(28)*.

Ab initio modeling was performed by simulated annealing of a single-phase dummy atom model using the program DAMMIF *(32)*. Twenty independent models were calculated with no symmetry restrictions. The program DAMAVER *(33)* was used to align *ab initio* models, select

the most likely models, and build averaged models. Mean and standard deviation of the normalized spatial discrepancy (NSD) was calculated as 0.7 and 0.09, respectively, indicating model reconstructions were reliable. The SUPCOMB program *(33)* was used to superimpose crystallographic structures and the averaged molecule envelope.

The ensemble optimization method *(34)* was used to select conformational ensembles that best represented the experimental BioSAXS data. This software consisted of two parts: RANCH (RANdom CHain) generated a pool of 10,000 random models based upon sequence and structural constraints, and GAJOE (Genetic Algorithm Judging Optimization of Ensembles) used a genetic algorithm to select members from the pool of random models whose theoretical scattering curves best described the experimental BioSAXS data *(34)*. Six ensemble models (Figure 3.10) were constructed to compare NS2B chains with differing degrees of flexibility linked to a rigid NS3pro domain: (1) NS2B-NT assumed residues 51 - 72 of the NS2B chain were flexible while the C-terminus of NS2B and the NS3pro domain were rigid with a structure described by the closed conformation of NS2B-NS3pro (PDB identifier 3E90); (2) NS2B-DDD assumed residues 51-83 of the NS2B chain, including the DDD triad, were flexible and the remainder of the molecule was rigid with a structure described by the closed conformation of NS2B-NS3pro; (3) NS2B-U assumed residues 51 - 97 of the NS2B chain were flexible and the NS3pro domain was rigid as described by the closed conformation of NS2B-NS3pro; (4) NS2B-CT assumed residues 73 - 97 of the NS2B cofactor were flexible and the rest of the protein was rigid; and (5) NS2B-F assumed the Gly-Ser linker between the NS2B and NS3pro segments was flexible and the individual chains adopted the rigid structures described by the closed conformation of NS2B-NS3pro. In models NS2B-CT and NS2B-F the positions of the rigid regions of NS2B were highly variable since they were located upstream of the NS2B flexible regions and thus dependent on the position and orientation of the N-terminus of the flexible region. A sixth model was generated to mimic the conformations suggested by Su *et al.* (2009) *(21)* in which only the NS2B C-term is flexible and the NS2B N-term was kept constrained to the interactions with NS3pro. BioSAXS parameters for the ensemble models were calculated with GNOM and DATPOROD subroutines *(28)*.

The program OLIGOMER *(28)* was used to test if mixtures of NS2B-NS3pro conformations could produce scattering curves consistent with the observed BioSAXS data. This

program could identify the volume fraction of each component in the mixture that best fit the experimental BioSAXS data *(28)*. This analysis was performed with crystal structures corresponding to the open (PDB identifier 2GGV) *(19)* and closed (PDB identified 3E90, chain A) *(20)* conformations of NS2B-NS3pro.

2.7 Fourier Transformed Infrared Spectroscopy (FT-IR). Aliquots of purified WNV NS2B-NS3pro in BioSAXS buffer were lyophilized overnight and resuspended in D_2O. Samples were incubated at 4 °C overnight, filtered through a 100 kDa MWCO filter (Millipore) and concentrated by centrifugation. Control experiments were performed to ensure lyophilization did not affect protease activity (data not shown). FT-IR experiments were conducted using a JASCO FT/IR-4100 spectrometer. Protein sample was added on top of a diamond PRO450-S Attenuated Total Reflectance unit adapted to the FT/IR-4100 system. Experimental parameters included 64 scans per protein sample with a resolution of 4.0 cm^{-1}. Deuterated buffer spectra were collected and subtracted from protein sample spectra. Data fitting and secondary structure estimations were done by multi-component analysis through the Secondary Structure Estimation (SSE) software *(35)*.

2.8 Intrinsic Fluorescence under denaturating conditions. WNV NS2B-NS3pro (final concentration of 0.1 mg/ml) was incubated 30 min at 37 °C in presence of 10 mM K Phosphate buffer pH 8.0 at Urea concentrations ranging from 0 to 8 M as final concentrations. Effect of substrate analogue Ac-KKR-NH_2 was tested at final concentration of 1.3 mM, according to the determined inhibition constant. Effect of crystallization conditions was assayed using 0, 1 and 3.5 M Sodium formate pH 7.2. Excitation wavelength was set at 280 nm, while emission was collected at 300-450 nm in a Fluorolog FL3-22 spectrofluorometer (Horiba Jobin Yvon), slits were set at 2-2 nm. Each condition was tested by duplicate.

2.9 Limited Proteolysis. WNV NS2B-NS3pro (final concentration of 0.5 mg/ml) was incubated 10 min at 37 °C in presence and absence of the substrate analogue Ac-KKR-NH_2 and different concentrations of Proteinase K in potassium phosphate buffer pH 8.0. Reactions were stopped adding PMSF (final concentration 1 mM). Degradation products were analyzed by SDS-PAGE.

CHAPTER 3

STRUCTURAL CHARACTERIZATION OF THE SUBSTRATE-FREE ACTIVE WNV NS2B-NS3PRO

3.1 Abstract

The so far fruitless efforts to develop antiviral drugs that inhibit WNV protease activity during the viral polyprotein processing, raise concern about whether the target being used by the classical structural-based approaches is the optimal one or not. Although X-ray crystal structures and NMR studies have provided valuable insights about the interactions between NS2B-NS3pro and peptide-based inhibitors, little is known about the solution structure of recombinant WNV protease (NS2B-NS3pro) in its substrate-free conformation. In this chapter I determine the solution structure and dynamics of the substrate-free WNV NS2B-NS3pro through a detailed characterization using circular dichroism (CD) spectroscopy, fluorescence experiments, small angle X-ray scattering (BioSAXS) and Fourier transform infrared spectroscopy (FT-IR) experiments. Our results indicate that the tertiary structure of WNV NS2B-NS3pro in absence of substrate and in solution is significantly different from its crystallographic structure, with the NS2B cofactor being highly flexible and occupying different orientations. However, the secondary structure of NS2B-NS3pro was similar in solution and crystal states, which may be stabilized by transient NS2B-NS3pro interactions. These findings highlight a better-defined protein target that may improve the efforts of antiviral drug discovery as well as our understanding of WNV polyprotein processing.

3.2 Results

3.2.1 Purified WNV NS2B-NS3pro is functionally active and monomeric in solution. Histidine-tagged WNV NS2B-NS3pro was over-expressed in *E. coli* cells and purified by Nickel affinity chromatography, followed by size exclusion chromatography according to standard protocols *(36)*. WNV NS2B-NS3pro was tested for protease activity using Boc-GKR-AMC as substrate *(26)*. Purified WNV NS2B-NS3pro had catalytic activity, with substrate inhibition

clearly evident at substrate concentrations > 600 μM (Figure 3.1 A). Kinetic parameters were estimated to be K_d=542 ± 33 μM, K_{d2}=1486 ± 140 μM, and k_{cat}=0.7 ± 0.03 s^{-1} in 50 mM Glycine-NaOH pH 9.5, 30% Glycerol, 1 mM CHAPS, consistent with previous studies *(26)*. Same functional assay was performed in 10 mM K phosphate pH 8.0 (buffer used for following experiments), in which protein followed a substrate inhibition mechanism and kinetics parameters were estimated as K_d=585.4 ± 67 μM, K_{d2}=827 ± 120 μM and k_{cat}=0.074 ± 0.006 s^{-1}. No significant differences on affinity constant were observed and a 10-fold decrease on catalytic constant, maybe due to the absence of Glycerol on the solution. Thus, WNV protease was active under the tested conditions.

Sedimentation velocity experiments determined that purified NS2B-NS3pro was essentially homogeneous in solution with a molecular weight of ~28 kDa for the predominant (>99%) species; this molecular weight was consistent with the molecular weight expected for a monomer of WNV NS2B-NS3pro. Similar results were observed when the protein was analyzed at concentrations of 0.3 and 2.5 mg/ml, suggesting the monomeric state was independent of protein concentration (Figure 3.1 B). From the sedimentation velocity experiment, the frictional ratio of WNV NS2B-NS3pro was calculated as 1.44, which was higher than the 1.2 ratio typically observed for globular spherical proteins *(37)*. The theoretical frictional ratio for the crystallographic structures in closed and open conformations was calculated to be 1.23 and 1.28, respectively. These results suggested that purified WNV NS2B-NS3 protease was active, monomeric, and may not have adopted the compact globular state described by X-ray crystallographic studies *(19, 20)*.

3.2.2. Characterization of the folding state of NS2B-NS3pro through Circular Dichroism spectroscopy. Folding state of WNV NS2B-NS3pro was evaluated through Circular dichroism (CD) spectroscopy. Analysis of the signal observed at the far UV region (180-260 nm) allows to determine the secondary structure content of the protein under study, whereas the profile observed at the near UV region (250-380 nm) gives information about the structural conformation of aromatic residues. The effect of chemical denaturant agent Urea on the folding state was followed by the CD signal and it was compared to the one under native conditions. I measured the CD signal of WNV NS2B-NS3pro at pH 6.0 to 9.5 in native conditions and in presence of 6 M Urea, after proper buffer subtraction, results are shown in Figures 3.2 A and B, respectively.

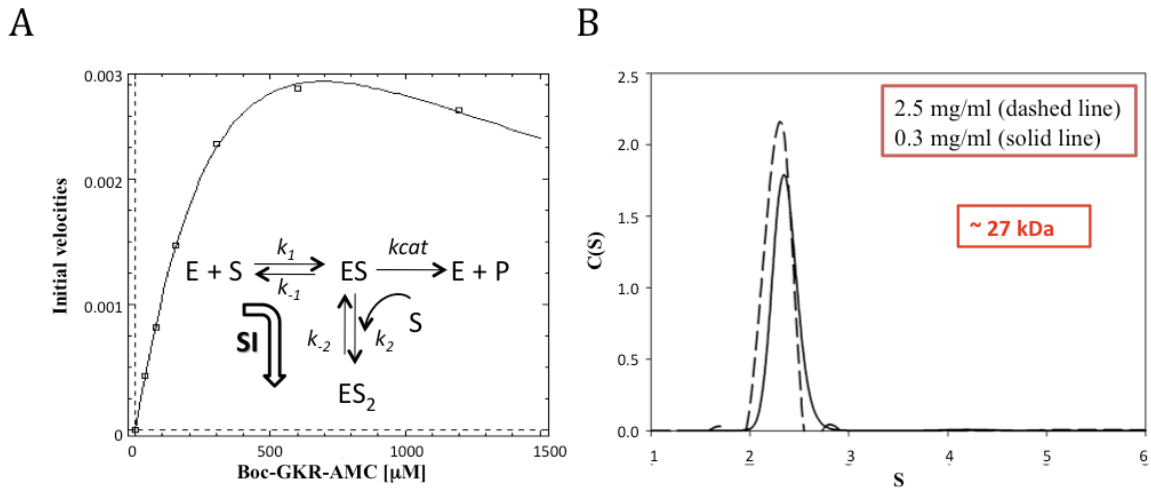

Figure 3.1. Test for functionality and oligomerization state of WNV protease. A. Saturation curve showing WNV NS2B-NS3pro activity with a Boc-GKR-AMC peptide substrate in 10 mM K phosphate pH 8.0. A substrate inhibition model (solid curve) with kinetic parameters K_d=585.4 ± 67 µM, K_{d2}=827 ± 120 µM and k_{cat}=0.074 ± 0.006 s^{-1} best fit the experimental data points (open squares). *B*. Sedimentation velocity experiment with 2.5 mg/ml (dashed line) and 0.3 mg/ml (solid line) WNV protease in 10 mM K phosphate pH 8.0. Absorbance was measured at 298 and 280 nm, respectively.

A typical random coil conformation signature is observed in all the cases where the protein was under native conditions, just a slight difference was noticed around 200 nm, that may be better explained with a more detailed study on pH effect (Section *3.2.2.c)*. In presence of Urea a loss of the CD signal indicates unfolding of the protein. In order to determine whether the secondary structure elements observed on the crystallographic structure (3E90.pdb) were detected by the CD spectroscopy, I processed a theoretical CD signal *(38)*, which indicated the presence of beta strand content as observed on the crystal state (Figures 3.2 C and D). I also measured the CD signal on the near UV region in native (Figure 3.2 E) and denaturing conditions (Figure 3.2 F). A significant change on the environment of aromatic residues was detected indicating a conformational change due to the Urea-induced protein denaturation.

These results suggest that there is an important random coil element representing the solution state of WNV NS2B-NS3pro, which is altered due to the presence of Urea as chemical denaturant agent. Interestingly, the theoretical CD signal obtained from the crystal state seems to

detect its beta content and suggests that in solution WNV NS2B-NS3pro would adopt a different and maybe more flexible conformation than the one represented by the crystallographic structure.

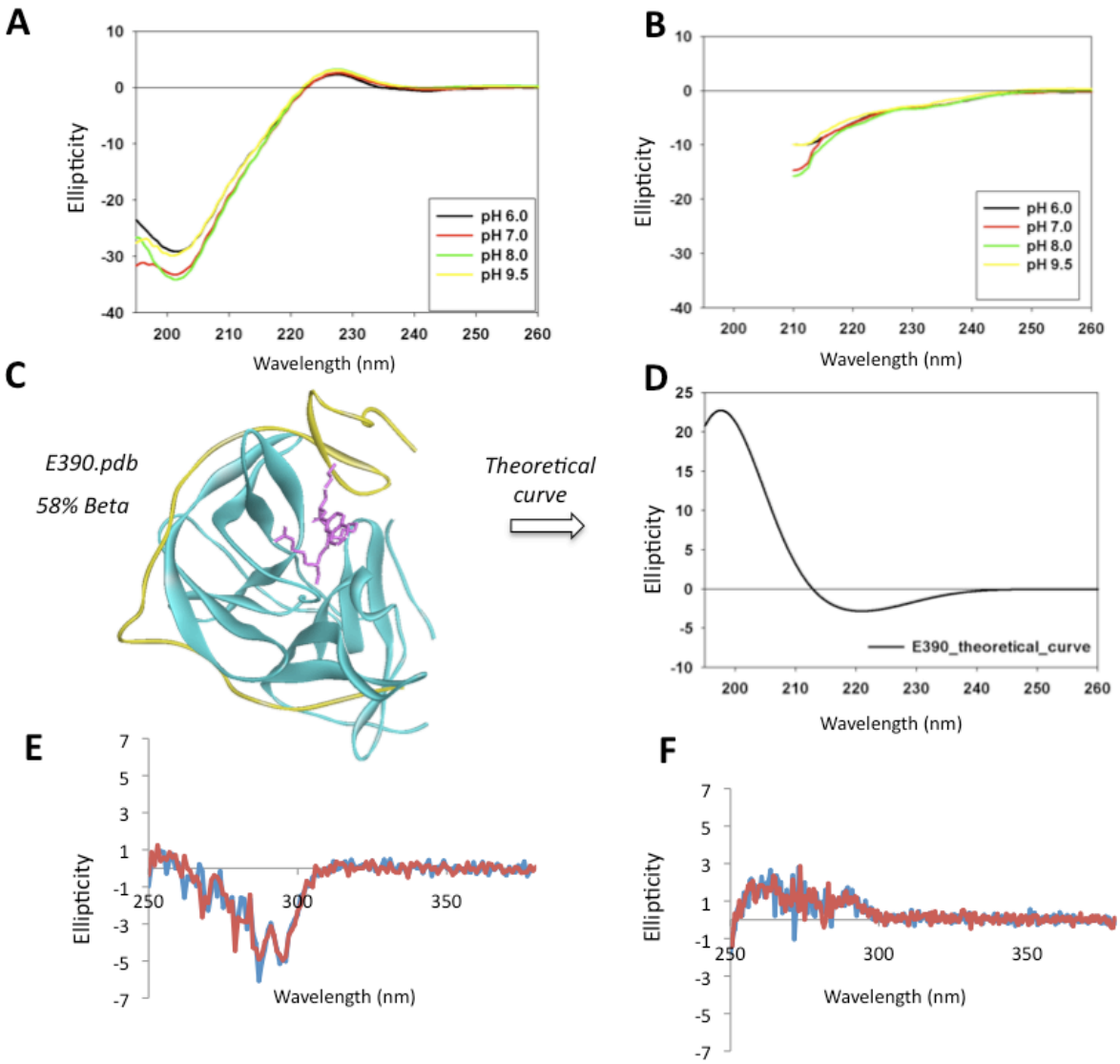

Figure 3.2. Circular dichroism of WNV NS2B-NS3pro. A. At far UV region under native conditions recorded at 0.5 mg/ml protein concentration. *B.* At far UV region under denaturing conditions recorded at 1 mg/ml protein concentration. *C.* Crystallographic structure of WNV NS2B-NS3pro in bound conformation, where NS2B cofactor is depicted in yellow whereas the NS3pro is shown in cyan, peptidic inhibitor is shown as pink sticks. *D.* Theoretical CD curve obtained from 3E90.pdb using DichroCalc software *(38)*. *E.* CD signal at near UV region under native conditions recorded at 10 mg/ml of protein concentration. *F.* CD signal at near UV region under 6 M Urea. Red and blue lines correspond to duplicates.

3.3.2.a. Effect of crowding agents. Since a very flexible protein has the potential to adopt a more compact conformation in presence of crowding agents such as TMAO or Ficoll70 *(39)*, I decided to test if these conditions would alter the CD signal observed under native conditions (Potassium phosphate pH 8.0). Figure 3.3 A shows the effect of working with 0 to 3.19 M TMAO, from which no significant differences were observed on the CD signal (final protein concentration of 0.1 mg/ml). I performed a parallel experiment to check if the presence of TMAO would affect the protease activity using fluorescence signal and as a result no important differences were detected (Figure 3.3 B). A second crowding agent was used, 40% FICOLL70, but unfortunately I observed protein aggregation at very low concentration of the agent, so that no data was collected.

3.2.2.b. Effect of counter ions. It is known that flexible proteins have propensity to exhibit higher percentage of charged amino-acids in its primary sequence *(39)* than protein with a well compact folding state. WNV NS2B-NS3pro has an isoelectric point of 5.45 which is an indication of that electrostatic repulsion may occur between the side chain interactions promoting a less ordered state. In order to test whether the presence of counter ions would affect the secondary structure of WNV NS2B-NS3pro, I choose two different salts from the Hofmeister series according to their known effect on protein solubility/stability and propensity to induce denaturation. I performed CD experiments in presence of 0 to 500 mM NaF and 0 – 500 mM $(NH_4)_2SO_4$, with the latter being described as having a stronger effect on inducing protein stability, decreasing solubility through a salting out effect and decreasing protein denaturation. Figures 3.3 C shows an increasing change on the CD signal at the characteristic minimum at 200 nm for random coil proteins, that may indicate a less disordered folding state due to increasing NaF concentrations. Figure 3.3 D shows that no significant differences were detected when $(NH_4)_2SO_4$ was present at 0 to 250 mM final concentrations, and just a slight change was observed at 200 nm when 500 mM $(NH_4)_2SO_4$ was used.

3.2.2.c. Effect of pH. Following the above concepts, I did a more detailed characterization to determine the effect of different pH on the secondary structure content of WNV NS2B-NS3pro. I used pH from 1.0 to 9.5 with different buffers (see materials and methods) to observe whether the predominant random coil conformation signal was altered. Figure 3.3 E clearly shows, after proper buffer subtraction, that the classical shoulder of random coil signal observed at 228 nm is affected as the pH becomes more acidic. The same phenomenon is observed at ~220 nm. The

effect of pH on the CD signal is observed more clearly in Figure 3.3 F, where the signal at 228 nm was plotted as function of pH. A sigmoidal curve suggests that decreasing pH would favor the formation of a more ordered folding state on WNV NS2B-NS3pro with a midpoint close to pH 4.0.

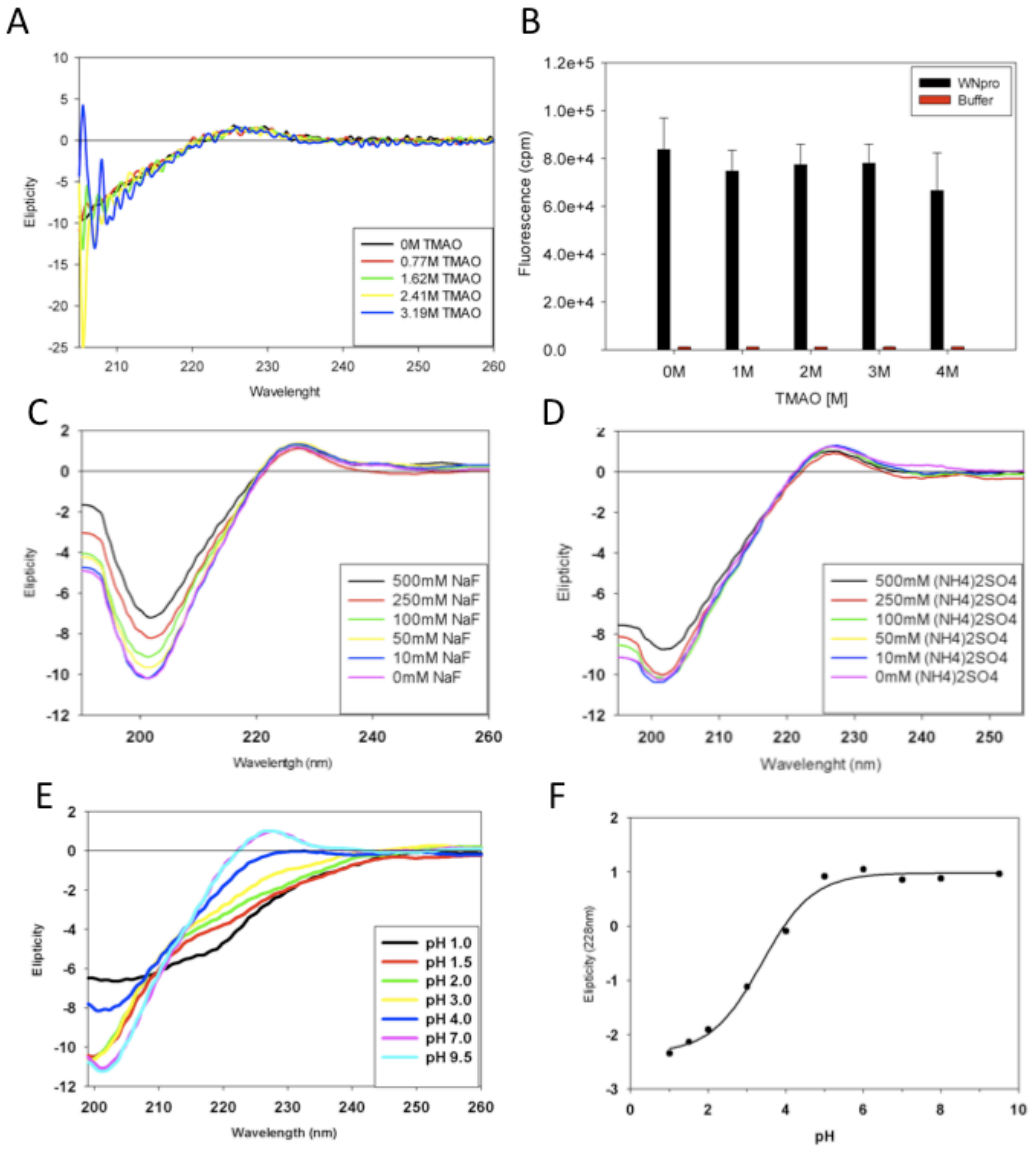

Figure 3.3. Secondary structure content of active WNV NS2B-NS3pro followed by CD spectroscopy (Far UV). A. Effect of TMAO as a crowding agent. *B.* Effect of TMAO on the WNV NS2B-NS3pro activity followed by fluorescence. *C.* Effect of NaF as a counter ion *D.* Effect of $(NH_4)_2SO_4$ as a counter ion *E.* and *F.* Effect of pH.

3.2.3 Identification of potential hydrophobic cores through ANS binding experiments. Since circular dichroism was suggesting the presence of a major random coil component in solution at the conditions analyzed, I decided to explore a new technique which allowed me to identify the existence of intermediate states of the unfolding process that could unstable enough to be hidden by the usually observed folded and unfolded states. Based on the 1-Anilino-8-naphthalene-sulfonate (ANS) affinity by hydrophobic cores in protein structures and the consequent increasing fluorescence signal due to that binding (10-40 folds), I decided to explore the chances of WNV NS2B-NS3pro having a folded core either in absence of a substrate analogue or due to the presence of Aprotinin (BPTI) as folding stabilizer molecule *(20)*. The existence of an intermediate partially folded state, also called "molten globule"-like state, which by definition is formed by the hydrophobic collapse of secondary structure elements that lack of tertiary interactions, was studied. I incubated WNV NS2B-NS3pro and the protease-BPTI complex with different Urea concentrations to examine the existence of any intermediate folding core or molten globule conformation.

3.2.3.a ANS binding experiments of WNV NS2B-NS3pro in absence of substrate. In order to determine whether an ANS-accessible hydrophobic core is exposed during WNV NS2B-NS3pro unfolding I measured fluorescence signal of protein sample and buffer at different Urea concentrations (Figure 3.4 A). Differences in fluorescence at 500 nm (maximum intensity) between protein and buffer were plotted according to Urea concentrations (Figure 3.4 B). The identification of molten globule state or transient hydrophobic core during unfolding process is often observed as a maximum on the fluorescence curve, typically of 10-40 fold over the base line *(40)*. As is observed on the reference Figure 3.4 C, GlnRS exhibits a molten globule conformation stable enough at 2 M Urea *(40)*.

WNV NS2B-NS3pro characterization suggests an absence of ANS binding due to lack of increase fluorescence signal, therefore no hydrophobic core could be identified at the assayed experimental conditions. This can be either due to absence of a hydrophobic core at the assayed experimental conditions or due to interference with ANS binding affinity as a consequence of the disorder that seems to predominate on WNV NS2B-NS3pro conformation.

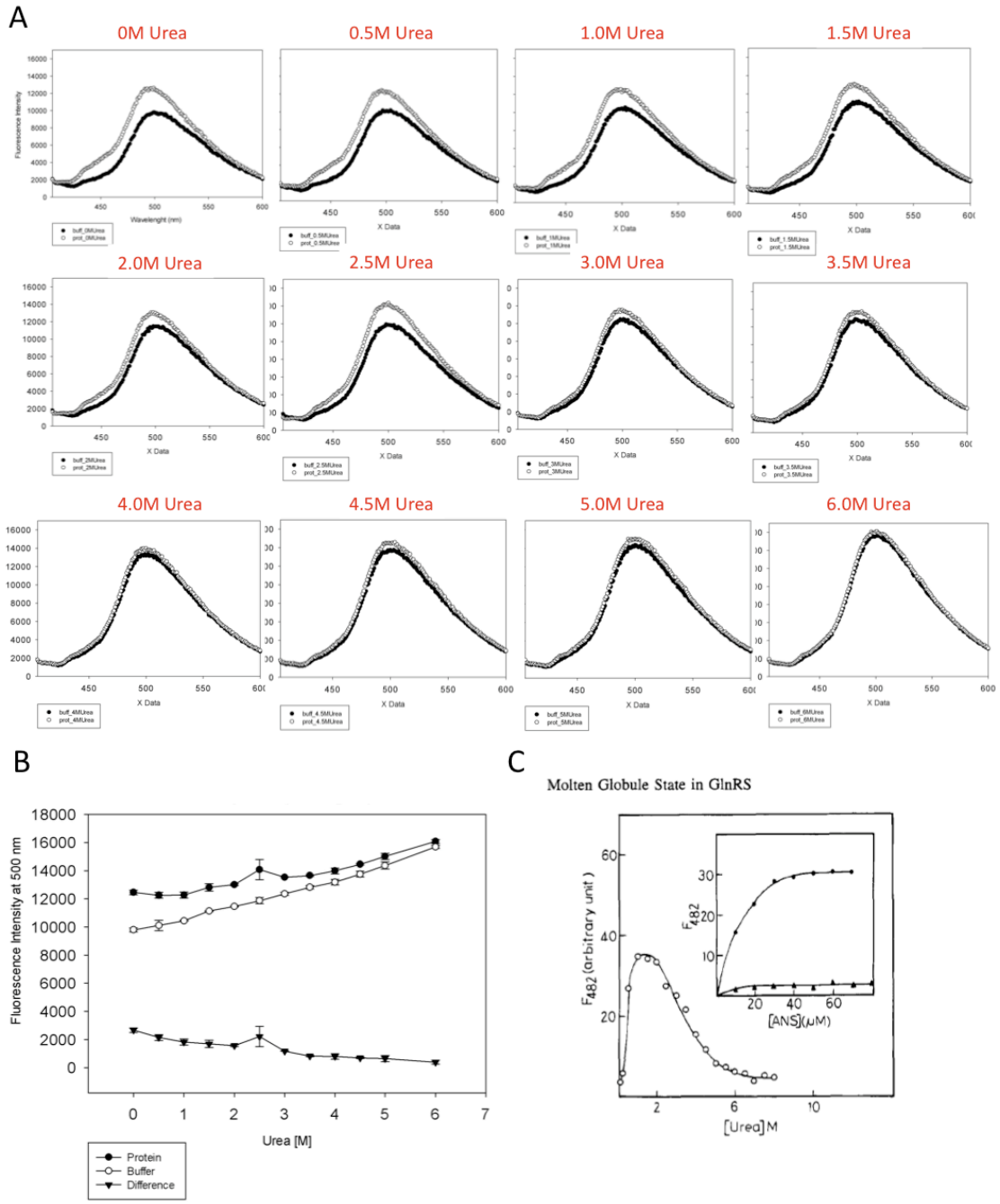

Figure 3.4. ANS binding experiments of WNV NS2B-NS3pro in absence of substrate. A. Emission spectra of ANS fluorescence at 0 – 6 M Urea. White circle curve represent protein sample, while black circle curve corresponds to ANS buffer. *B.* Maximum fluorescence intensities at 500 nm of protein (black circle) and buffer (white circle) samples, and the corresponding difference (black triangle) at each Urea concentration. Error bars are shown in each case. *C.* Reference figure of the identification of a molten globule intermediate state in Glutaminyl-tRNA synthetase *(40)*.

3.2.3.b Complex WNV NS2B-NS3pro-Aprotinin. The complex WNV NS2B-NS3pro-Aprotinin has been solved by X-ray crystallography revealing a new kind of protease-inhibitor interaction *(19)*. In order to explore the possibility of Aprotinin acting as a folding stabilizer of WNV protease I decided to include it on CD experiments. However, a preliminary functional assay was performed to test the inhibition capacity of Aprotinin on the protease function. I performed activity assays that allow quantifying the protease cleaving efficiency following the fluorescence signal emitted by the fluorescence probe AMC once it is cleaved from the substrate provided to the reaction. Figure 3.5 shows the effect of Aprotinin interaction on the protease activity. Increasing Aprotinin concentrations reduce the cleavage activity suggesting that there is interaction between both molecules.

In order to test if the Aprotinin interaction causes any effect on the WNV NS2B-NS3pro folding state I performed CD spectroscopy experiments. Changes on the secondary structure content would suggest whether or not there is an induction of a more ordered conformation of the WNV NS2B-NS3pro. Figure 3.6 A and B show the CD spectra from Aprotinin and WNV NS2B-NS3pro, respectively. Figure 3.6 C shows the CD signal of the complex protease-BPTI at different molar ratios. The observed changes on the CD spectra seems to be due to increasing concentration of BPTI instead of a real induction of a more structured protease fold. Thus, no evident folding induction was observed through this technique.

I used ANS binding experiments as a parallel approach to test the possibility of Aprotinin acting as a folding inductor of WNV NS2B-NS3pro. Following the methodology previously explained I performed fluorescence experiments in which ANS signal is measured under different conditions. Increasing emission signal (10-40 fold) would be an indication of ANS binding to any hydrophobic pocket accessible on the complex. Figure 3.7 A shows the ANS emission curves of the complex protease-BPTI in solution under a range of Urea concentrations (black circle). In addition, the curves of WNV NS2B-NS3pro, Aprotinin and corresponding buffer are indicated with green triangle, yellow triangle and red circle, respectively. I determined the maximum intensity of each curve as 500 nm, and they were plotted according to Urea concentrations as shown in Figure 3.7 B.

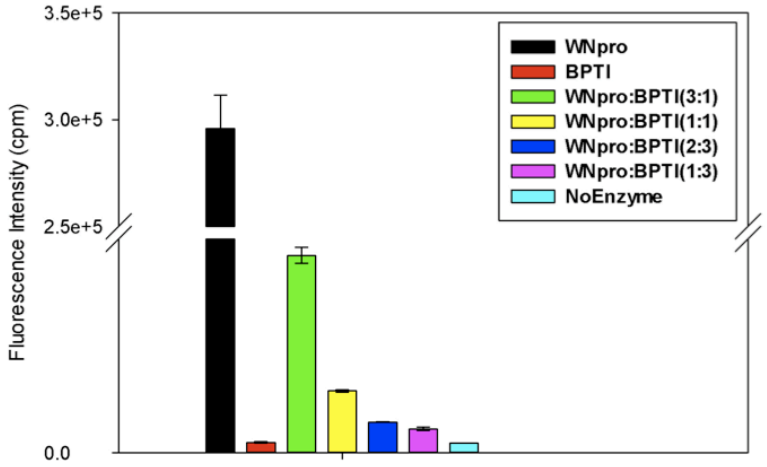

Figure 3.5. Inhibition effect on Aprotinin on WNV NS2B-NS3pro activity followed by fluorescence signal. Emmision was recorded at 465 nm after excitation at 380 nm. Error bars correspond to standard deviation from triplicate samples.

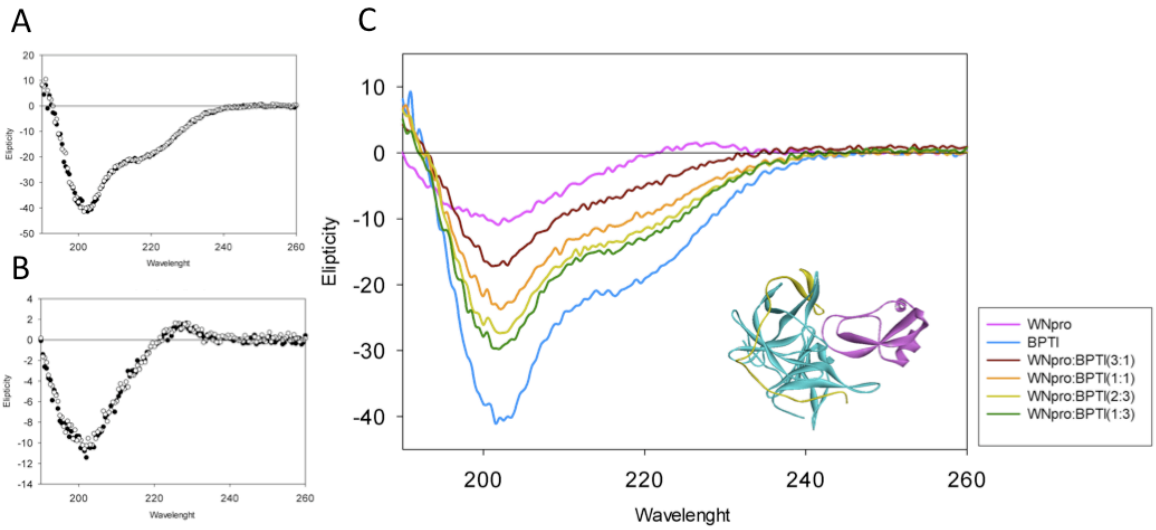

Figure 3.6. Circular Dichroism spectroscopy of WNV NS2B-NS3pro, Aprotinin and Protease-Aprotinin complex. A. CD spectra of Aprotinin (0.1 mg/ml), closed and open circles correspond to duplicates. *B.* CD spectra of WNV NS2B-NS3pro (0.1 mg/ml), closed and open circles correspond to duplicates. *C.* CD spectra of complex protease-BPTI at different molar ratios. Insert shows the crystallographic structure of the complex, where NS3pro is depicted in cyan, yellow represents NS2B cofactor and pink indicated BPTI. PDB code 2IJO *(19)*.

The difference on intensity plots does not reveal the expected increment, suggesting that under the experimental conditions it is not possible to identify a molten globule state stable enough, if indeed there is one.

Therefore, the folding state of WNV NS2B-NS3pro as it is depicted on the crystallographic structure does not seem to be present in solution under the tested experimental conditions, even when the protease was in presence of BPTI. Possible explanations can be associated to the fact that the potential "molten globule" state is not stable enough to be sensed or that the complex protease-BPTI is not populated enough even at very low Urea concentrations. A different explanation may be the low accessibility of hydrophobic pockets for ANS binding, or simply the potential role of crystallization conditions on the stabilization of the complex.

As an alternative approach, it was thought to perform quenching fluorescence experiments, in which acrylamide is expected to interact with residues exposed to the solvent altering its polar environment. Thus, based on the Stern-Volmer approximation, changes on the tryptophan fluorescence can reveal the presence of unfolding intermediates as "molten globule" states indicating how buried or exposed they are *(41)*. However, the reliability of this approach was questioned due to three out of the seven tryptophan residues that WNV protease has, seem to be exposed to the solvent (% area of solvent accessibility > 20%) even in native conditions, affecting the interpretation of the results.

3.2.4 Characterization of the folding state of NS2B-NS3pro through Small Angle X-ray scattering (BioSAXS). BioSAXS is a powerful technique used to characterize the folding and oligomerization state of proteins in solution. This approach is especially useful when the structural disorder of the protein under study requires to be evaluated. Thus, I performed BioSAXS experiments to characterize the solution-state conformation of WNV NS2B-NS3pro. No radiation damage was detected during data collection at 20 °C (data not shown). Data collection for 60 minutes from solutions of 2.5 mg/ml samples NS2B-NS3pro produced reproducible scattering curves up to q values > 0.2 (Figure 3.8 A). Linearity of the Guinier plot at $qRg < 1.3$ ($r=0.98$) (Figure 3.8 B) was an indication of good data quality with no evidence of aggregation or inter-particle repulsion *(28, 30)*. Kratky plots (Iq^2 *vs* q) *(29)* of SAXS data have been used to estimate the folding state of proteins in solution, in which compact globular proteins

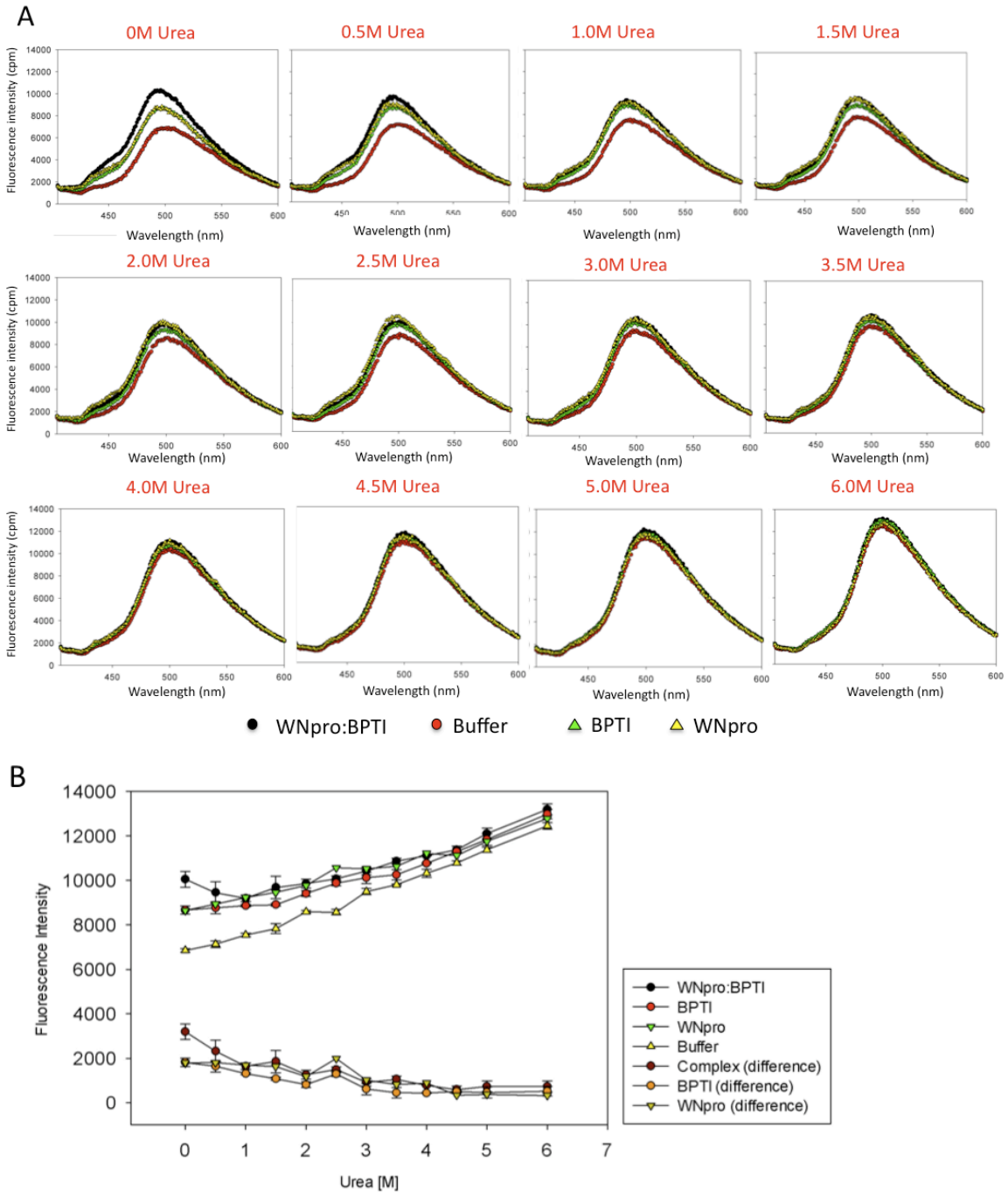

Figure 3.7. ANS binding experiments of WNV NS2B-NS3pro-Aprotinin complex. A. Emission spectra of ANS fluorescence at 0 to 6 M Urea. Curve corresponding to the complex protease-BPTI is shown in black circle, while WNV NS2B-NS3pro, Aprotinin and corresponding buffer are indicated with green triangle, yellow triangle and red circle, respectively. *B.* Maximum fluorescence intensities at 500 nm of each protein, buffer and complex are plotted versus Urea concentration. Difference plot are also shown. Error bars are indicated in each case from duplicated samples.

exhibited a defined peak and conformationally-disordered or elongated proteins showed a tendency for increased Iq^2 values at high q values *(30)*.

In addition, the Iq^2 value trended upwards at q values up to ~0.2. An alternative method to characterize the shape of a particle in solution was a pair distance distribution [P(r)] plot, which represented distances between all electrons within a macromolecule *(29, 30)*. The asymmetrical shape and "tail" of P(r) at high radii (Figure 3.8 D) suggested that WNV NS2B-NS3pro was not folded into a compact and symmetric conformation. Moreover, there was no detectable evidence of oligomerization by BioSAXS. These results were consistent with the hypothesis that WNV NS2B-NS3pro in solution would exhibit features of a partially folded protein with highly flexible regions.

I performed a detailed analysis of WNV NS2B-NS3pro BioSAXS data produced radii of gyration (Rg) of 22.7 ± 0.1 Å and 23.2 ± 0.1 Å, from the Guinier approximation and the Pair-distance distribution function P(r) *(30)*, respectively (Table 3.1). The maximal distance (Dmax) of WNV NS2B-NS3pro was found to be 80 ± 1 Å (Figure 3.8 D). The apparent molecular weight of WNV NS2B-NS3pro in solution was estimated by its hydrodynamic volume using the Porod approximation *(30)*; these calculations yielded a Porod volume of 45,958 ± 5 A^3 and an apparent molecular weight of 27.0 ± 2.7 kDa (Table 3.1). The apparent molecular weight calculated from the Porod volume was consistent with the 26.6 kDa molecular weight calculated from the amino acidic sequence of monomeric WNV NS2B-NS3pro. To determine if the conformation of WNV NS2B-NS3pro in solution was represented by the open conformation (PDB identifier 2GGV) or closed conformation (PDB identifier 3E90) NS2B-NS3pro X-ray crystal structures *(19, 20)* I generated theoretical scattering curves from each crystal structure and I compared them to the experimental solution scattering data. The level of agreement was quantitatively evaluated using χ^2 values, where and χ^2=1 indicates an exact fit between model and experimental data and higher values would represent a poor fit between model and experimental data *(30)*. Scattering curves for the open and closed conformations did not fit the experimental BioSAXS data (Figure 3.9 A, C) and had χ^2 values of 5.04 and 4.24, respectively (Table 3.1). These results suggest that the X-ray structures of the individual open and closed conformations were not accurate representations of the WNV NS2B-NS3pro structure in solution.

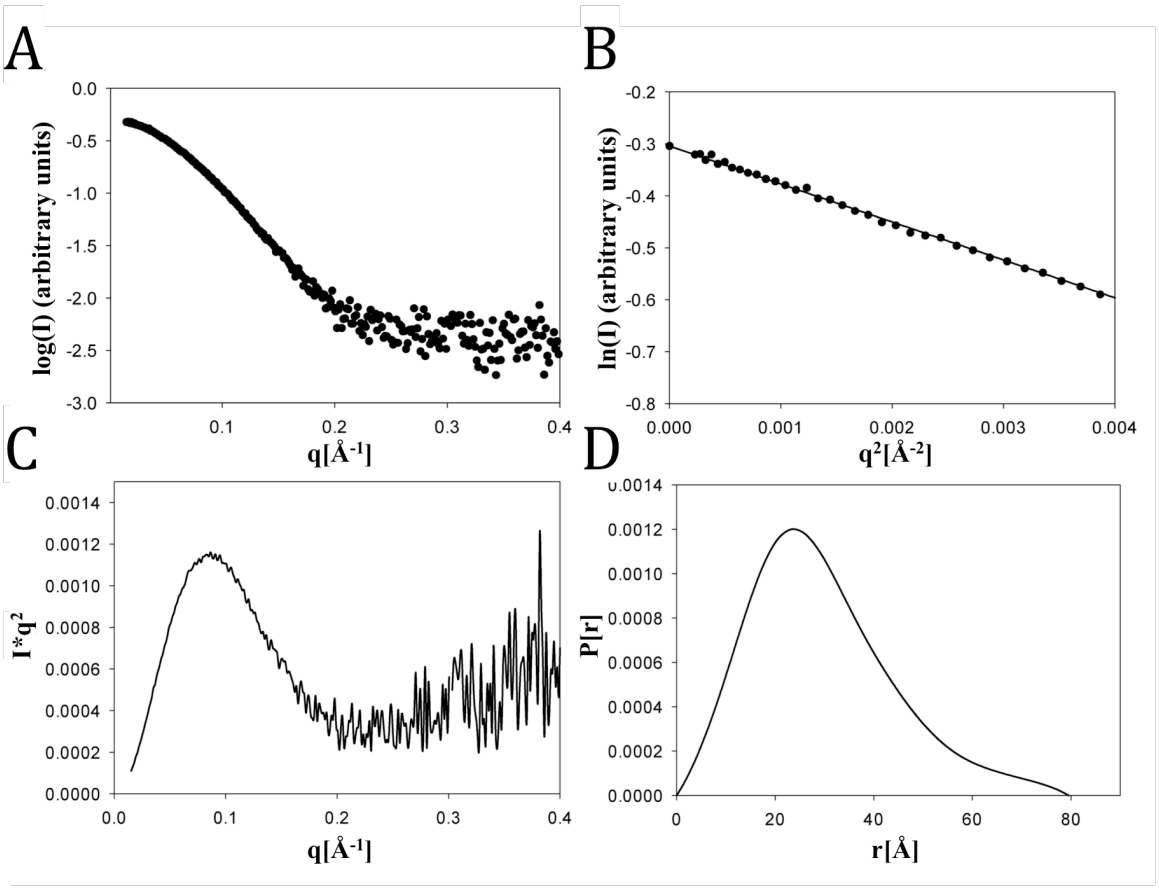

Figure 3.8. BioSAXS solution characterization of WNV NS2B-NS3pro. Small angle X-ray scattering (SAXS) data collected from 2.5 mg/ml WNV protease after overnight dialysis against BioSAXS buffer. *A*. Experimental SAXS data from WNV protease (average of 6 scans). *B*. Guinier plot. The observed linearity at $qR_g<1.3$ indicated no aggregation or electrostatic repulsion was present in the sample (correlation coefficient > 0.98). *C*. Kratky plot. Scattering pattern as q^2I *vs* q. *D*. Pair distance distribution plot from experimental WNV protease data obtained using GNOM software *(28)*.

I next examined if the solution state might contain an equilibrium mixture of open and closed conformations, as suggested by NMR studies *(21, 22)*. Calculations with the OLIGOMER program *(28)* produced scattering curves with a χ^2 value of 5.47 (data not shown), suggesting that the presence of only these two crystallographic conformations in equilibrium does not fit the experimental data.

I compared the structural parameters extrapolated from the modeled and experimental scattering data to further determine if a crystallographic structure might be consistent with the

experimental data (Table 3.1). The open and closed conformations had similar Rg values of 19.8 and 20.0 Å, respectively. In contrast, the open and closed conformations had different Dmax values of 63.6 Å and 57.5 Å, respectively. The Rg and Dmax parameters for the crystallographic structures were significantly smaller than values obtained from experimental BioSAXS data for WNV protease in solution (Table 3.1). Similarly, Porod volumes calculated for the open and closed crystallographic structures suggested they had more compact folding states than the WNV protease in solution (Table 3.1).

The shape and compactness of WNV NS2B-NS3pro in solution can be approximated with *ab initio* models predicted to produce small angle X-ray scattering curves that match the experimental curves. Using the program DAMMIF *(32)* twenty independent models were produced with no symmetry restrictions. The mean value of the normalized spatial discrepancy (NSD) was 0.7 ± 0.09, indicating a high degree of similarity between all models. The χ^2 between the scattering curve from a selected reference model and the experimental scattering data was 1.12, indicating the *ab initio* model could be a reasonable representation of the solution state. The averaged envelope for the *ab initio* models exhibited a roughly elongated ovoid shape with approximate dimensions of 27 Å x 37 Å x 80 Å (Figures 3.9 B, D); the longest dimension was similar to Dmax determined from the pair distance distribution plot. Superimposition of NS2B-NS3pro crystallographic structures with the *ab initio* envelope produced NSD values of 2.28 and 2.24 for the open and closed conformations, respectively. These results provide further evidence that WNV NS2B-NS3pro X-ray structures were not representative of the protease's folded state in solution.

3.2.5 Solution state models with a flexible NS2B cofactor could reproduce the experimental BioSAXS data. Since the crystallographic structures of WNV protease were unable to provide reasonable models for the WNV protease solution state, I used the ensemble optimization method (EOM) (30) to generate ensemble models that were consistent with the experimental SAXS data. A series of ensemble models with varying degrees of NS2B flexibility and movement were constructed (see Methods section for detailed descriptions), (Figure 3.10).

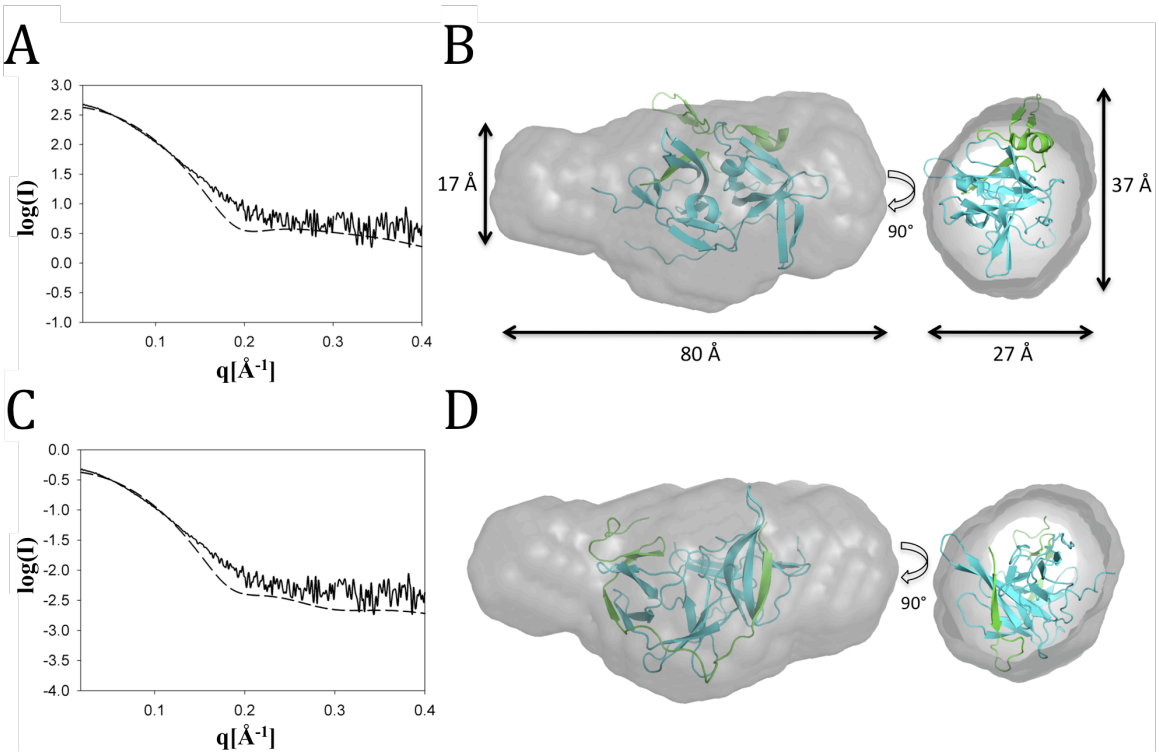

Figure 3.9. Comparison of SAXS data collected from solutions of WNV NS2B-NS3pro and calculated from NS2B-NS3pro crystallographic structures. A. Experimental scattering curve from WNV protease in solution (solid line) and theoretical scattering curve calculated from open crystal structure (dashed line). *B.* WNV crystal structure (open conformation) superimposed into the envelope produced from averaged *ab initio* models that best fit experimental BioSAXS data of WNV protease. *C.* Experimental scattering curve from WNV protease in solution (solid line) and theoretical scattering curve calculated from closed conformation crystal structure (dashed lines). *D.* WNV crystal structure (closed conformation) superimposed into the envelope produced from averaged *ab initio* models that best fit experimental BioSAXS data of WNV protease. Crystallographic structures are depicted as ribbons with the NS2B and NS3pro chains colored green and cyan, respectively.

Theoretical SAXS curves were calculated for each model and compared to the experimental scattering data (Figure 3.11, Table 3.1). Among the tested ensemble models, NS2B-CTf and NS2B-NT showed the poorest agreement between calculated and experimental scattering curves with χ^2 of 3.0 and 2.94, respectively. The NS2B-DDD ensemble model did not provide good agreement to the experimental data with a χ^2 of 1.80 (Table 3.1).

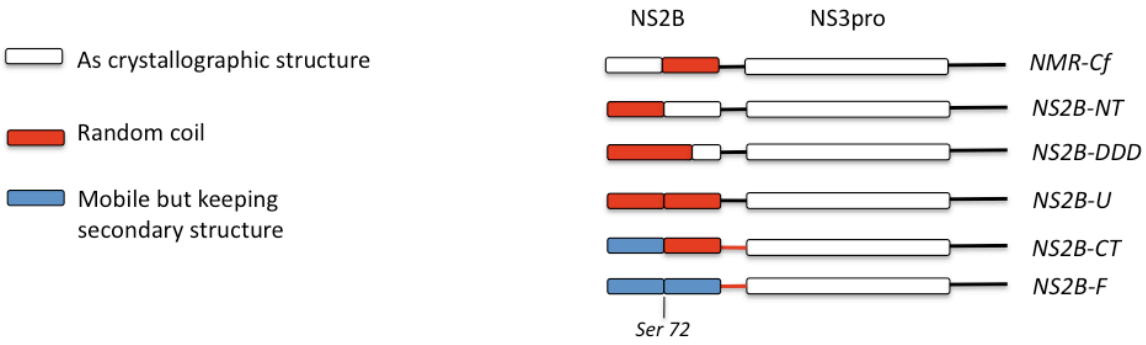

Figure 3.10. Schematic representation of WNV protease models used for EOM. Graphical representation of models NS2B-CTf, NS2B-NT, NS2B-DDD, NS2B-U, NS2B-CT, and NS2B-F. Regions considered folded as in the crystallographic structure are shown in white, flexible regions are depicted in red, and rigid but mobile regions are indicated in blue.

These results suggest that the solution state of WNV protease could not be represented as an ensemble of structures if these conformers retained a significant portion of the NS2B crystallographic structure. The above calculations were performed with rigid regions of the ensemble models corresponding to the closed crystallographic structure; similar χ^2 values (and conclusions) were obtained when the open crystallographic structure was used to model the rigid regions of the ensembles.

An ensemble model (termed NS2B-U) wherein the entire NS2B cofactor was assumed unfolded and fully flexible produced a scattering curve that was in very good agreement (χ^2 = 1.13, Table 3.1) with the experimental data. Essentially equivalent representations of the solution state of WNV protease, as judged by χ^2 values (χ^2 = 1.06, Table 3.1), were produced using ensemble models that combined flexible NS2B C-terminal regions with conformationally rigid N-terminal regions. In the NS2B-CT and NS2B-F ensemble models the NS2B chain, though conformationally rigid, was mobile and occupied highly variable spatial positions since it was attached to the flexible linker. The relative distribution of the radius of gyration (Rg) for randomly generated structures and the best-fit ensemble structures for each of the above models provided a powerful method to characterize the solution state of WNV protease.

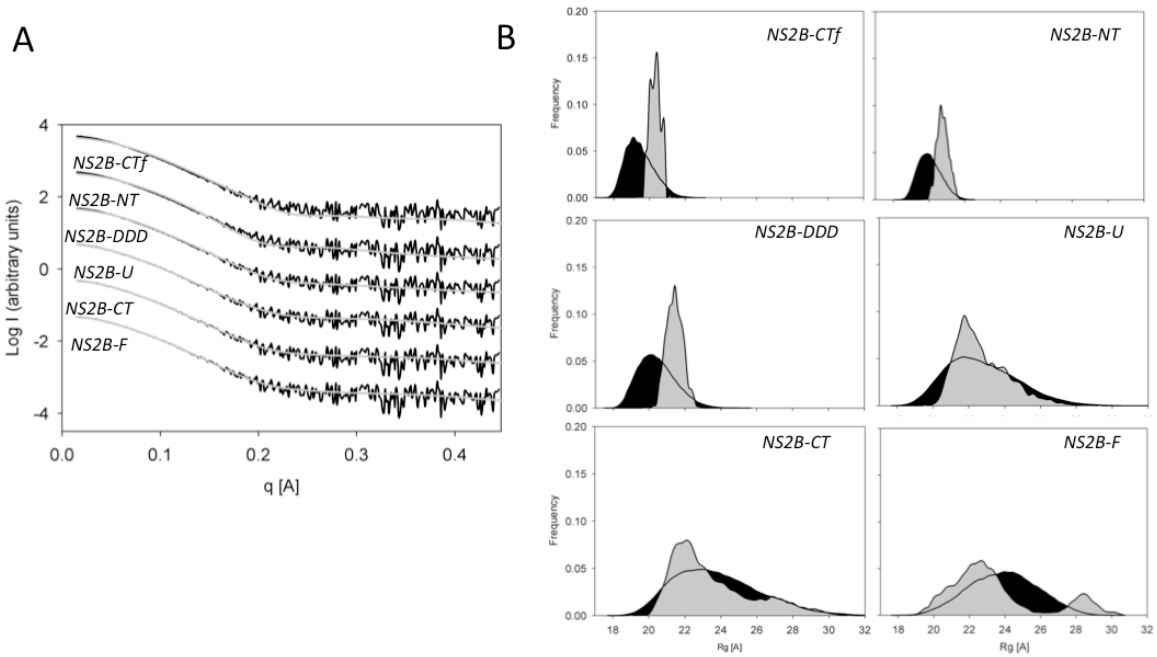

Figure 3.11. BioSAXS analysis of flexible models of WNV NS2B-NS3pro. A. Fitting curves of the theoretical scattering models (gray dashed solid lines) and experimental SAXS data (black solid line). *B.* Distribution of ratio of gyration (Rg) of random (black) and selected (gray) pools for all the models. Ensembles selected for models NS2B-CTf, NS2B-NT, NS2B-DDD, and NS2B-U had average Rg of 20.4 Å, 20.7 Å, 21.4 Å, and 22.9 Å, respectively. Ensembles selected for models NS2B-CT and NS2B-F had bimodal Rg distributions with peaks around 22 Å and 28 Å.

The set of randomly generated structures for each model followed a normal Rg distribution that ranged from very compact models (Rg < 19 Å) to very extended conformations (Figure 3.11 B). The complete set of structures for the NS2B-NT model had Rg distributions that ranged from ~18 Å to ~22 Å, whereas the best-fit ensemble structures for NS2B-NT displayed a narrow Rg distribution with an average value of 20.7 Å. NS2B-CTf model exhibited an average Rg value of 20.4 Å. (Figure 3.11 B; Table 3.1). In the NS2B-DDD model, the best-fit ensemble structures had an average Rg of 21.4 Å (Figure 3.11; Table 3.1). The best-fit ensemble structures for the NS2B-U, NS2B-CT, and NS2B-F models had comparable average Rg of ~22.9 Å (Table 3.1); these values were similar to Rg calculated from Guinier and pair-wise distribution plots for WNV protease (Table 3.1). In addition, the best-fit ensemble models for NS2B-CT and NS2B-F had a bimodal Rg distribution with peaks observed at Rg of ~22 Å and ~28 Å (Figure 3.11 B). This observation suggested that two sub-ensembles were present in solution, with the flexible NS2B region adopting different conformational states in each ensemble. The more compact

conformation represented ~85% of the population, while the extended conformation represented ~15 % of the total ensemble structures. Protease models where the entire NS2B cofactor was mobile and not bound to the NS3pro domain most closely reproduced the SAXS experimental data. Interesting, the theoretical scattering curves (and χ^2 values) were similar for ensemble models where the NS2B chain was treated as fully flexible (e.g., model NS2B-U) or as a rigid segment attached to a short flexible linker at the C-terminus of NS2B (e.g., models NS2B-CT, NS2B-F). The major requirement for reproducing the SAXS data was the ability of the entire NS2B chain to sample numerous conformations including a significant number of highly extended conformations. The open or closed crystallographic structures of NS2B were not obvious from examination of the structures present in ensemble model NS2B-F (Figure 3.12).

Since the open and closed crystal structures are mainly different on the NS2B cofactor conformation, I wanted to test whether similar results were obtained using as a template for the structured regions, the crystallographic open conformation of WNV NS2B-NS3pro instead of the closed state used previously. The results for the models in which NS2B conformation was different in both states are shown in the Table 3.2. No significant differences were found for the NS2B-NT and NS2B-DDD models. Model NS2B-F shown a χ^2 of 1.06 when the closed conformation was used as a template and 1.5 when the open conformation was used. This difference is because the open structure is missing more residues on the linker region (18 residues) compared to the closed structure (4 residues), so flexibility is restricted and that is why the higher χ^2 when compared with the experimental data. When I increased the length of the linker to the same one observed on the closed state, the result was a $\chi^2 = 1.12$.

3.2.6 Fourier transformed infrared spectroscopy (FT-IR). This technique has been widely used to accurately determine secondary structure content in proteins, being at the same time, a better method to analyze beta-sheet rich proteins compared to CD spectroscopy *(42-45)*. Therefore, I decided to characterize WNV NS2B-NS3pro in solution and compare these results with the theoretical secondary structure contents observed on the flexible models that resulted well evaluated by the previous BioSAXS results. The integration of these results will give me a more reliable and consistent determination of the folding state of WNV NS2B-NS3pro in solution.

Table 3.1. Values of BioSAXS parameters of WNV NS2B-NS3pro in absence of substrate.

Structural Parameters	Experimental Data	Theoretical Data							
		Crystallographic structures [a]		Ensemble of conformations[c]					
		Closed state	Open state	NS2B-CTf	NS2B-NT	NS2B-DDD	NS2B-U	NS2B-CT	NS2B-F
χ^2	-	4.24	5.04	3.0	2.94	1.80	1.13	1.06	1.06
R_g (Å) Guinier	22.7 ± 0.1	19.8	20.0	20.4	20.7	21.4	22.8	23.0	22.9
R_g (Å) P(r)	23.2 ± 0.1								
D_{max} (Å)	80 ± 1	57.5	63.6	70	75	72	80	78	78
Porod volume estimate (A^3)	45,958 ± 5	32,980	32,430	43,343	41,870	44,486	43,706	47,842	47,228
MW (kDa) [b]	27.0 ± 2.7	23.9	23.3	25.5	24.6	26.1	25.7	28.1	27.7

[a] Closed conformation taken from 3E90 (chain A); open conformation taken from 2GGV. Data calculated with the program CRYSOL.

[b] MW was determined from Porod volume. Error ~10%

[c] SAXS parameters for ensemble models were calculated using the programs GNOM and DATPOROD.

Table 3.2. Comparison of χ^2 values of EOM obtained from open and closed crystallographic structures.

Model	χ^2 from closed conformation	χ^2 from open conformation
NS2B-NT	2.9	2.6
NS2B-DDD	1.8	1.7
NS2B-F	1.06	1.5

*Increasing the length of the linker region on that model resulted in a χ^2 =1.12

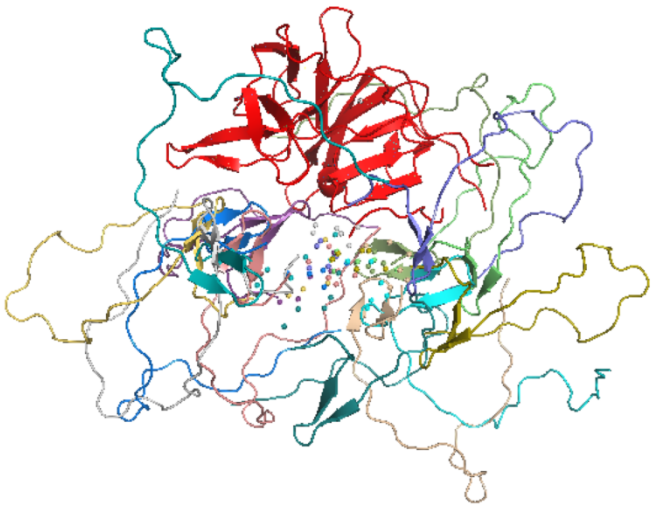

Figure 3.12. Ensemble of selected conformations of model NS2B-F. NS3pro is depicted in red while NS2B cofactor is shown in different orientations while retaining secondary structure elements. Flexible linker is represented with spheres. Figure was produced using Pymol software.

3.2.6.a Control proteins. In order to test my ability on the usage of this instrument I decided to work with two *control* proteins, from which the folding state has been previously determined by FT-IR, and also there is crystallographic data available to compare results between the two techniques. The control proteins to analyze were Lysozyme, due to the high alpha-helix content and Trypsine, based on its high beta-sheet content. Both proteins received the same treatment as the WNV NS2B-NS3pro previous the measurements (See Materials and Methods). Figure 3.13 A shows the spectral deconvolution of the amide-I band of Lysozyme, where the high alpha-helix content is represented by the green line indicating the agreement between previously reported data and the data I collected. Figure 3.13.B shows the analysis of Trypsine, with a strong signal

corresponding to the Beta-sheet content (yellow line), also suggesting an agreement with the reported data and the data I collected.

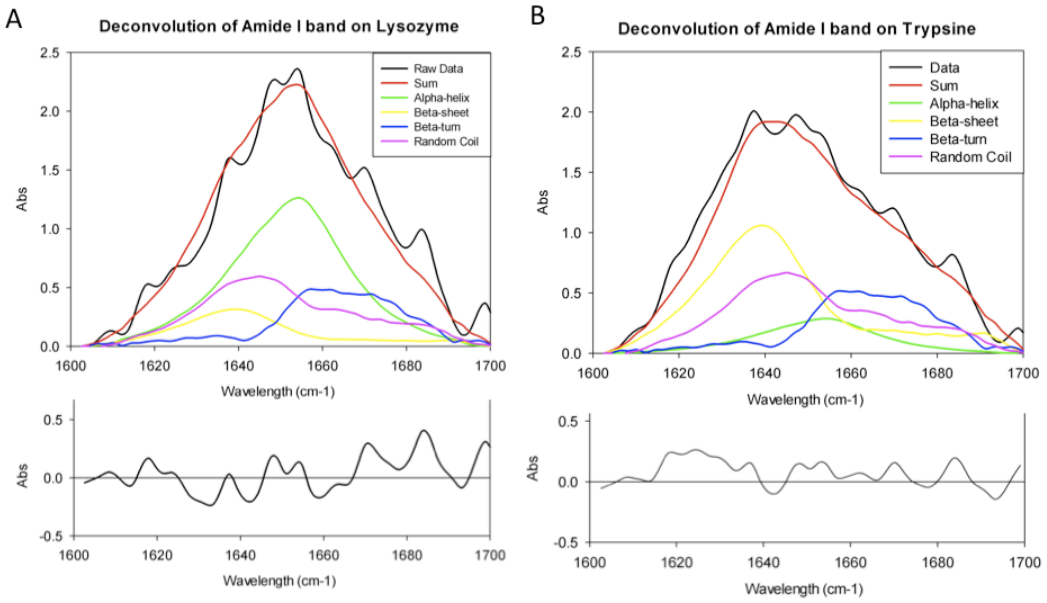

Figure 3.13. Determination of secondary structure content of control proteins by FT-IR. Spectral deconvolution of the amide-I band of a) Lysozyme and b) Trypsine.

After deconvolution into Gaussian components it is possible to describe each secondary structure element and estimate their contribution on the folding state of the studied protein. Table 3.3 shows the previously reported percentages for the Lysozyme and Trypsine determined by FT-IR (Rep FT-IR) and X-ray crystallography (Rep X-ray) *(45)*. FT-IR results from the data I collected are highlighted in green column. Alpha-helix content determined by my measurements was very similar to the reported values in both control proteins. Thus, Lysozyme exhibited a 41%, while the reported FT-IR and X-ray data showed 40 % and 45 %, respectively. On the other hand Trypsine resulted to have 9 % of alpha content, same value observed on previous FT-IR and X-ray crystallography. A similar tendency was observed with the estimation of beta content, although slight differences were noticeable between the reported FT-IR and the reported X-ray data. The data I collected resulted to be on the range reported previously, in which Trypsine resulted to have high beta content as I expected. An increase on the estimated random coil element compared to the reported data was observed, probably due to the software processing which in my study considered as a random coil contribution not just disordered regions but everything else that is not alpha or beta structured.

Table 3.3 Secondary structure percentage estimated by FT-IR experiments: Lysozyme and Trypsine as control proteins.

Protein	Secondary structure (%)											
	Alpha-Helix			Beta-sheet			Beta-turn			Random-coil		
	Rep FT-IR	Rep X-ray	FT-IR	rep FT-IR	rep X-ray	FT-IR	rep FT-IR	rep X-ray	FT-IR	Rep FT-IR	Rep X-ray	FT-IR
Lysozyme	40	45	41	19	19	12	27	23	23	14	13	24
Trypsine	9	9	9	44	56	40	38	24	24	9	11	27

Rep: Acta Biochimica et Biophysica Sinica 2007, 39(8): 549–559

3.2.6.b FT-IR studies suggested NS2B-NS3pro developed similar secondary structure in solution and in crystalline environments. The extent of secondary structure predicted for the NS2B cofactor was a major structural difference between the NS2B-U, NS2B-CT, and NS2B-F ensemble models that best fit the SAXS data. FT-IR experiments were performed to determine the secondary structure content of WNV NS2B-NS3pro in solution. The amide-I band spectra was collected and deconvoluted into Gaussian components describing principal secondary structure elements (Figure 3.14) *(43, 44)*. This analysis suggested the secondary structure content of WNV protease in solution had approximately 3% α-helix, 68% β-strand and β-turn, and 29% random coil (Table 3.4). The FTIR secondary structure content was compared with secondary structure elements present in the open and closed crystallographic structures and the ensemble models (Table 3.4). The NS2B chain in the open and closed crystallographic structures contained a β-hairpin turn towards the C-terminus (residues 78-86) and two β-strands totaling ~16 residues in length.

These elements were used to calculate the secondary structure content of the different ensemble models. The FT-IR results (Table 3.4) were most consistent with an NS2B chain that contained significant secondary structure. The percentage of β-structure measured for the solution state of NS2B-NS3pro was most similar to the secondary structure content predicted if NS2B was structured as in the crystallographic structures or the NS2B-F ensemble model. Ensemble model NS2B-F provided the best fit to both the SAXS and FT-IR data collected for the solution state of NS2B-NS3pro. The NS2B chain in this model retained significant α-helix, β-strand, and β-turn secondary structural elements (similar to those observed in both the open and closed crystallographic structures), yet was flexible in the regions between the secondary structural elements. Moreover, this model assumed the linker between the NS2B and NS3pro chains to be

flexible, allowing the NS2B chain to sample multiple spatial configurations and explore different conformations with respect to NS3pro. Transient interactions between the NS2B chain and the NS3pro domain could occur, and these interactions may stabilize the distinct sub-populations observed in the structures that comprise ensemble model NS2B-F (Figure 3.12).

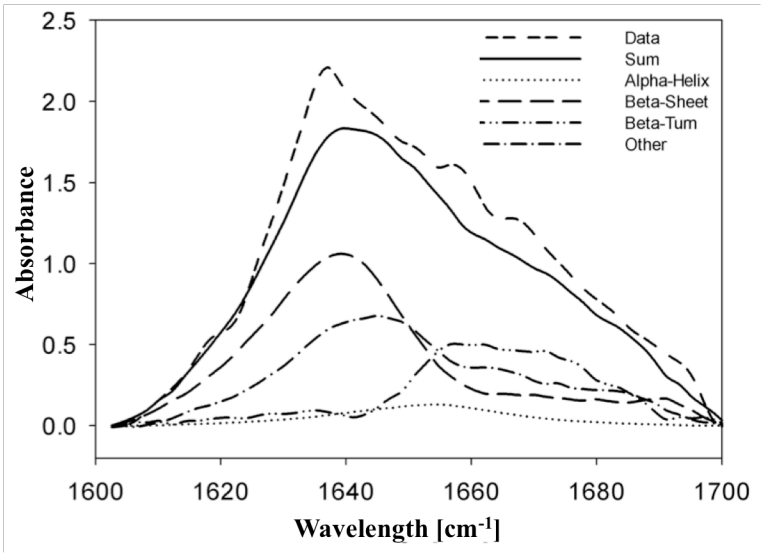

Figure 3.14. WNV NS2B-NS3pro secondary structure determined by FT-IR. Spectral deconvolution of the amide-I band of WNV NS2B-NS3pro in 10 mM K phosphate pD 8.0.

Table 3.4. Comparison between secondary structure content determined by FT-IR spectroscopy and calculated from model structures.

	Alpha content	**Beta content** [b]	**Other**
Experimental FT-IR data [a]	0.03	0.68	0.29
Bound conformation (E390)	0.02	0.51	0.47
Open conformation (2GGV)	0.06	0.50	0.44
Model NS2B-CTf	0.02	0.45	0.53
Model NS2B-NT	0.02	0.46	0.52
Model NS2B-DDD	0.02	0.43	0.55
Model NS2B-U	0.02	0.40	0.58
Model NS2B-CT	0.02	0.45	0.53
Model NS2B-F	0.02	0.51	0.47

[a] Average of 10 measurements of 64 scans each one

[b] Sum of beta turn and beta sheet content.

3.4 Discussion

In an effort to contribute to a more accurate and efficient development of inhibitors against WNV protease, this work provided a detailed description of the folding state of WNV NS2B-NS3pro in solution. Specifically, this chapter was focused on its state in absence of a substrate analogue, since this would be the predominant conformation that the WNV protease would adopt in solution while inhibition assays are performed, so that it should be also considered at the moment of theoretically design and development of new drugs using structure-based computational screenings.

My preliminary concern was to make sure that the purified WNV protease was enzymatically active and monomeric in solution. For this reason I performed kinetic assays following previously established protocols and sedimentation velocity experiments which clearly showed that WNV NS2B-NS3pro was active and monomeric in solution under different protein concentrations. Then, the next step was to determine the secondary structure content of WNV protease in solution, to do so I used Circular dichroism (CD) spectroscopy under a series of different experimental conditions. WNV NS2B-NS3pro exhibited a preliminary random coil component under most of the tested conditions, which included a wide range of pH, crowding environment and the presence of counter ions. This interesting finding was in disagreement with the theoretical CD curve obtained from the crystal structure, which indicated a predominant beta-strand content, suggesting then that what is found in solution is not well represented by the crystal structure. Now, it is important to consider that it has been previously reported that the presence of Tryptophan residues in Aprotinin may induce differences on the CD signal that lead to contradictory secondary structure contents when compared with the crystal structure *(47)*. So that considering the high content of Tryptophans present on the WNV protease sequence, this potential explanation must be considered. CD spectroscopy was also done at near UV region to analyze changes on the aromatic residue's environment due to chemical denaturing agent. The results revealed that Urea altered the CD signal suggesting that an unfolding process maybe occurring, so some structural order should be present on the WNV protease's fold when it was under native conditions. I used ANS binding experiments as an alternative approach to examine any potential transient molten globule state that may be present in solution. Results revealed no significant ANS binding to WNV protease, even when it was in presence of BPTI, what can be

interpreted as no observable transition state was detected with this technique.

Known as a very powerful approach to determine the folding state of proteins in solution, small angle X-ray scattering (SAXS) experiments have been widely used. Since I had a very monomeric and monodisperse protein sample, it was possible to analyze WNV protease with this technique. Although SAXS cannot provide a high resolution structure as X-ray crystallography, SAXS has the great advantage of characterizing the oligomerization state of the sample and the folding state in terms of flexibility and shape of the molecule while it is in solution, avoiding potential artifacts caused by saturated salt and crowded solutions or even highly concentrated proteins. It is important to recognize that this intrinsic dynamic may play an important role on the activity mechanism of the protein under study, esspecially when there is evidence of certain degree of structural flexibility, and it can easily be unperceived using only crystallography as a characterization method.

The very high quality of data obtained, judging from the linearity of Guinier plots, confirmed the monodisperse and monomeric state of WNV protease in solution. Interestingly, neither the open nor closed crystallographic structures (or a mixture thereof) of WNV NS2B-NS3pro adequately represented the conformational state of WNV protease in solution. These results are in agreement with the flexibility and/or elongated shape suggested by the pair-distance distribution plot and the frictional ratio observed by sedimentation velocity experiments. *Ab initio* reconstruction and rigid body modeling revealed an "average" conformation of WNV NS2B-NS3pro that suggested the protease does not adopt a stable compact structure in solution. Among the different degrees of flexibility tested in this work, models in which the C-terminus of NS2B was structured and the N-terminus adopted random coil conformations were not in agreement with the experimental data. This is interesting considering that disorder is often observed on extremes of a protein chain. Extending the disordered region of NS2B to the triad $D_{84}D_{85}D_{86}$ did not significantly improve the agreement between the ensemble model and the BioSAXS data. The ensemble model NS2B-CTf incorporated features suggested by previous NMR studies of WNV NS2B-NS3pro *(21, 22)*; unfortunately this model did not provide good agreement to our BioSAXS data. However, models that allowed the N-terminal region of NS2B to be fully mobile and included flexibility for the C-terminal portion of NS2B or of the linker between NS2B and NS3pro agreed very well with the BioSAXS data. The common feature in these models was the

mobility of the entire NS2B, with no segment of NS2B constrained to the positions shown in the crystallographic structures. Our FT-IR results suggested that the NS2B chain, although fully mobile, could retain the stable secondary structural elements observed in the open and closed crystallographic structures (e.g., the β-hairpin at NS2B residues 78-86). Thus, the NS2B-F model was in best agreement with our BioSAXS and FT-IR data; this was an ensemble of WNV NS2B-NS3pro conformations in which the entire NS2B cofactor was free to explore different conformations while preserving the NS2B secondary structure and NS3 tertiary structure observed by X-ray crystallography. Interestingly, this ensemble may contain two preferred sub-ensembles (Figure 3.11), the larger sub-ensemble (85 % of the population) being compact (Rg~22 Å) and the less populated sub-ensemble being more extended (Rg ~28 Å); the more populated sub-ensemble could be stabilized by transient interactions between NS2B and NS3pro.

With this work we provided a detailed structural characterization of the solution state of active WNV NS2B-NS3pro, free of mutations and in the absence of inhibitors. We established that the protease was best represented by an ensemble of structures in which the NS2B cofactor was fully dynamic and likely not bound to the stably-structured NS3pro domain. This insight must be considered when using structure-based approaches to guide inhibitor discovery and optimization.

CHAPTER 4

DETERMINATION OF THE SOLUTION STRUCTURE OF WNV PROTEASE IN PRESENCE OF SUBSTRATE ANALOGUE

4.1 Abstract

WNV protease structure has been solved by X-ray crystallography while being in presence of a substrate analogue, whereas, in its absence, critical mutations had to be included. Based on my previous results, the substrate free- WNV protease would exhibit a very flexible conformation especially focused on the NS2B cofactor, what makes me to propose the substrate analogue as a stabilizing and folding inductor of WNV protease. SAXS experiments revealed that a more compact and ordered conformation is adopted by the protease in presence of substrate, which is very well represented by the crystal structure in its closed state. Denaturing conditions were also tested exhibiting an elongated and totally unfolded state of the protease. Limited proteolysis experiments suggested a more susceptible WNV protease when it was in absence of substrate than in presence of it, confirming the finding of SAXS results. Chemical denaturation was assayed by intrinsic fluorescence experiments to determine the stabilizing effect of substrate analogue and crystallization conditions on the unfolding process of WNV protease. Whereas the presence of substrate analogue did not show any stabilizing effect under the tested conditions, crystallization buffer exhibited an evident increase of protein stability as its concentration was increased. These results reveal the stabilizing role of the substrate analogue on the WNV protease's fold and, interestingly, highlight the potential effects of buffers used for crystallization purposes on the intrinsic flexibility of proteins in solution.

4.2 Results

4.2.1 Determination of inhibition constant of peptide analogue by Kinetic assays. WNV NS2B-NS3pro was over-expressed in *E. coli* cells and purified by Nickel affinity followed by size exclusion chromatography according to standard protocols *(36)*. I tested the peptide-analogue Ac-Lys-Lys-Arg-NH_2 for inhibition of WNV NS2B-NS3pro protease activity. As previously established *(26)*, I used Boc-GKR-AMC as substrate of the reaction following conditions

described on Material and Methods section. WNV NS2B-NS3pro exhibited catalytic activity following a substrate inhibition mechanism, according to the analysis given by Dinafit *(27)*, which was evident at substrate concentrations > 600 μM (Figure 4.1). I tested the inhibition activity of Ac-Lys-Lys-Arg-NH_2 at 0, 50 and 100 μM as final concentrations. The best kinetic model to fit the experimental data predicted a competitive inhibition mechanism with parameters corresponding to K_d= 190 ± 21 μM, K_{d2} = 2218 ± 400 μM and k_{cat}= 0.4 ± 0.023 s^{-1}, whereas the inhibition constant for the Ac-Lys-Lys-Arg-NH_2 was determined to be K_i ~ 50 ± 3.8 μM.

Knowing the concentration of Ac-Lys-Lys-Arg-NH_2 required to inhibit WNV NS2B-NS3pro, it was possible for me to design different experiments to study its effect as substrate analogue on the folding and structural stability of the WNV protease.

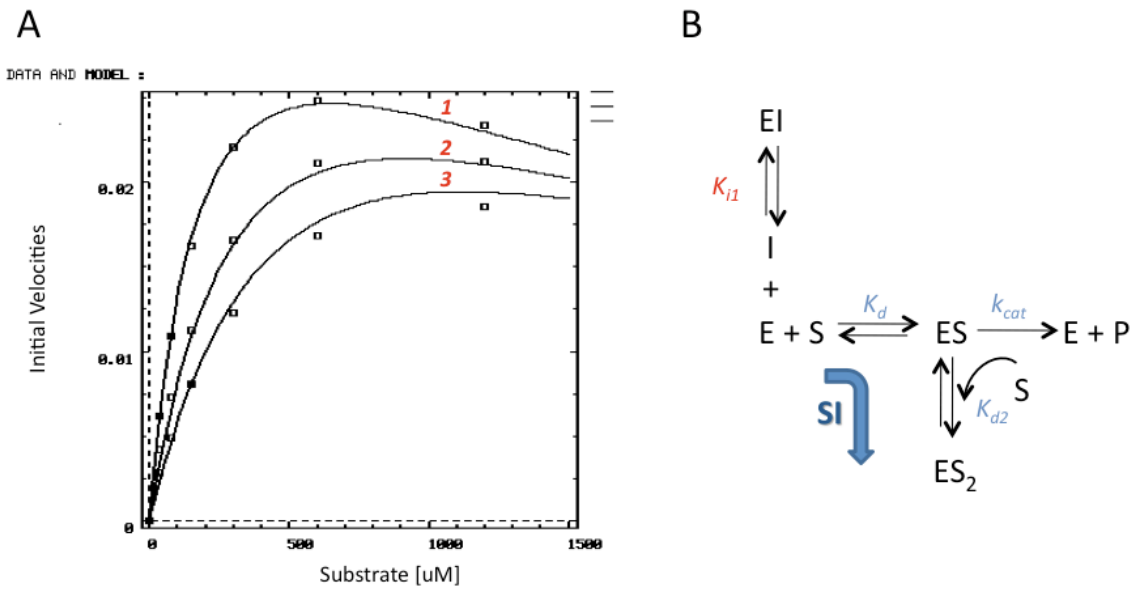

Figure 4.1. Test for inhibition of WNV protease activity by Ac-KKR-NH_2 as peptide analogue. A. Saturation curve showing WNV NS2B-NS3pro activity with a Boc-GKR-AMC peptide substrate. Final concentration of inhibitor *Ac-KKR-NH_2* was 0 μM (1), 50 μM (2) and 100 μM (3). A substrate inhibition model (solid curve) and best fit the experimental data points (open squares). B. Schematic representation of substrate inhibition model and competitive inhibition by Ac-Lys-Lys-Arg-NH_2. Kinetic parameters were determined to be K_d= 190 ± 21 μM, K_{d2} = 2218 ± 400 μM and k_{cat}= 0.4 ± 0.023 s^{-1}, whereas the inhibition constant for the Ac-Lys-Lys-Arg-NH_2 was K_i ~ 50 ± 3.8 μM.

4.2.2 Characterization of substrate free-, substrate bound- and denatured states of WNV NS2B-NS3pro by BioSAXS.

4.2.2.a Effect of substrate analogue Ac-Lys-Lys-Arg-NH_2 on the WNV NS2B-NS3pro folding state. I performed BioSAXS experiments to characterize the solution-state conformation of WNV NS2B-NS3pro and the effect of Ac-Lys-Lys-Arg-NH_2 as a folding inductor. Data collection for 60 minutes from solutions of 2.5 mg/ml WNV NS2B-NS3pro produced reproducible scattering curves for q values > 0.2 (Figure 4.2 A). Linearity of the Guinier plot at qRg < 1.3 (r=0.98) (Figure 4.2 B) was an indication of good data quality with no evidence of aggregation or inter-particle repulsion *(28)*. The Kratky plot of WNV NS2B-NS3pro in presence of substrate analogue showed a local maximum at q ~ 0.1 Å, smoothly decaying until q ~ 0.2 Å (Figure 4.2 C). A pair distance distribution [P(r)] plot exhibited a symmetrical shape with a D_{max} of ~ 60 Å (Figure 4.2 D) suggesting that WNV NS2B-NS3pro in presence of substrate adopts a compact and symmetric conformation.

4.2.2.b Effect of Urea as a denaturing agent. In order to characterize a completely unfolded WNV protease to then compare the resulting curves with the ones obtained with the protein in substrate free- and substrate bound- conformations, I performed BioSAXS experiments with WNV NS2B-NS3pro in presence of 6 M Urea as a denaturing agent. After overnight incubation at 4 ˚C, I collected SAXS data for 60 minutes from solutions of 2.5 mg/ml WNV NS2B-NS3pro (Figure 4.2 A). Linearity of the Guinier plot at qRg < 1.3 (r=0.99) (Figure 4.2 B) was an indication of good data quality with no evidence of aggregation or inter-particle repulsion *(28)*. The Kratky plot of WNV NS2B-NS3pro under denaturing conditions showed absence of a local maximum (Figure 4.2 C), which is typically observed in completely unfolded proteins. The corresponding pair distance distribution [P(r)] plot exhibited an evident more elongated tail with a D_{max} of ~ 160 Å (Figure 4.2 D) suggesting that WNV NS2B-NS3pro in presence of 6 M Urea is denatured.

The BioSAXS results of WNV NS2B-NS3pro under the three previously tested conditions were compared and analyzed. From the experimental data (Figure 4.2 A) it is possible to observe differences on the scattering curve decay for the samples in presence of the substrate analogue, in absence of it and under denaturing conditions, in which the protein sample in presence of Urea showed the fastest decay at very low q values compared with the protein

samples in native conditions. This differences are more evident when the Guinier plots are analyzed (Figure 4.2 B), in which the slope observed for the protein sample in presence and absence of substrate analogue are slightly different, whereas the sample in denaturing conditions showed a higher slope. In all the cases, however, I observed excellent linearity indicating monodispersity, with no aggregation or electrostatic repulsion.

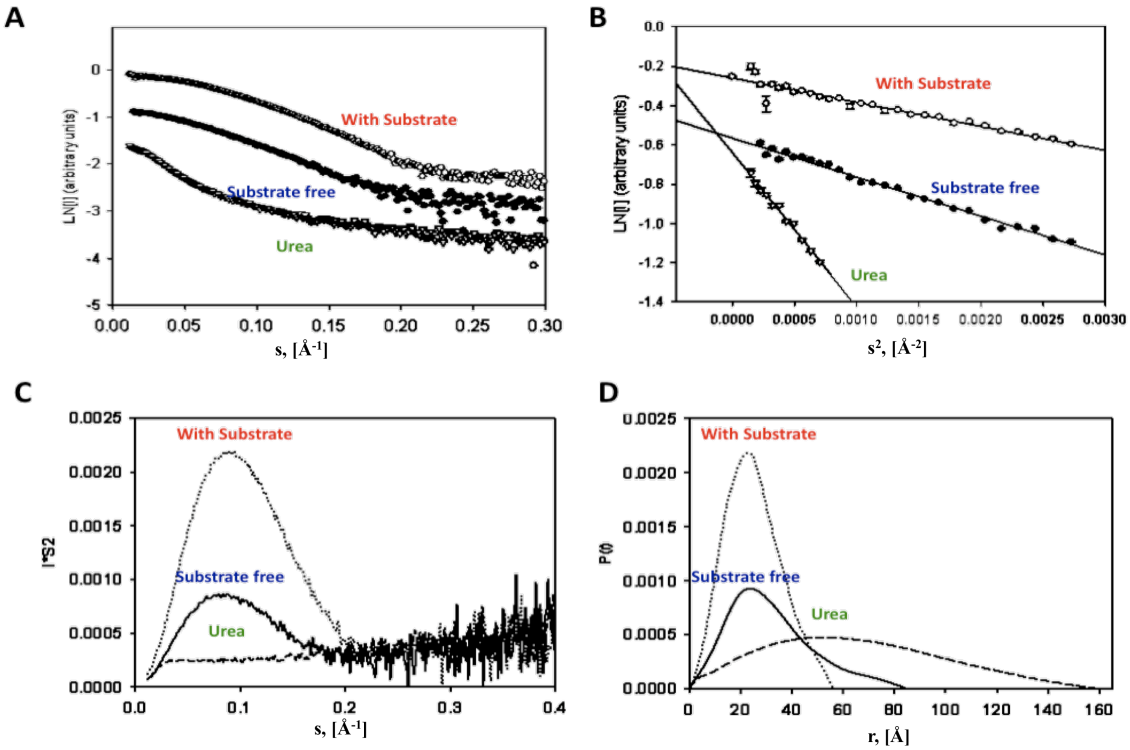

Figure 4.2. BioSAXS solution characterization of WNV NS2B-NS3pro in absence, presence of substrate and denatured states. A. Experimental SAXS data collected from 2.5 mg/ml WNV protease after overnight dialysis against BioSAXS buffer, 1.3 mM substrate analogue and 6 M Urea, as indicated in each case. Average of 3 scans. *B.* Guinier plots. The observed linearity at $sR_g<1.3$ indicated no aggregation or electrostatic repulsion on the substrate free-, substrate bound- and denatured states. *C.* Kratky plots. Scattering patterns as s^2I *vs* q. *D.* Pair distance distribution plot from experimental data obtained using GNOM software *(28)*.

The effect of the substrate analogue on the folding state of WNV NS2B-NS3pro in solution was analyzed through Kratky plots (Iq^2 *vs* q) *(28)*. In general, a compact globular protein will exhibit a defined peak, while flexible or elongated proteins will show a tendency for increased Iq^2 values at high q values *(29)*. Figure 4.2 C indicates clearly a more structurally

ordered conformation for the substrate bound- conformation of WNV NS2B-NS3pro, a partially folded or more flexible state characterizes the substrate free- state of the protease, while a fully unfolded state is observed when the protease is in presence of Urea. In terms of the shape of the particle in solution predicted by the pair distance distribution [P(r)] plot (Figure 4.2 D), I observed that the maximum distance between all electrons within the macromolecule was going from ~ 80 Å to ~ 60 Å when the protein was incubated with the substrate analogue, whereas in denaturing conditions a highly extended conformation was described with a Dmax of ~ 160 Å. Therefore I conclude that the substrate bound- state of WNV NS2B-NS3pro exhibits a more compact and ordered structural conformation than the protein in absence of substrate analogue, while a fully denatured conformation is observed in presence of Urea.

To quantify the structural differences of substrate free- substrate bound- and denatured state of WNV NS2B-NS3pro I determined the SAXS parameters from the experimental scattering curves (Table 4.1). In addition to the previously mentioned Dmax value, I determined the radius of gyration (Rg) in each case using the Guinier approximation. Thus the substrate bound-state of WNV NS2B-NS3pro exhibited a more constrained Rg of 19.6 ± 0.1 Å compared with its substrate free- state (22.7 ± 0.8 Å), while the fully denatured state exhibited a Rg value of 52.4 ± 0.4 Å, confirming its disordered state.

4.2.3 Comparison of in solution conformation of WNV NS2B-NS3pro in presence of substrate with crystallographic structures. To determine if the substrate bound- conformation of WNV NS2B-NS3pro in solution was represented by the X-ray crystal structures in open state (2GGV.pdb) or inhibitor-bound conformation or closed state (3E90.pdb) *(19, 20)*, I generated theoretical scattering curves from each crystallographic structure and I compared them to the experimental solution scattering data. So, similar to the methodology used in Chapter 3, the level of agreement was quantitatively evaluated using χ^2 values, where and χ^2=1 indicates an exact fit between model and experimental data and higher values would represent a poor fit between model and experimental data *(28)*. Figure 4.3 A shows the theoretical scattering curves for the open conformation and the fitting with the experimental BioSAXS data. A χ^2 value of 2.1 suggested that the X-ray crystal structure does not exactly represent the folding in solution of WNV NS2B-NS3pro in its substrate bound- state. Figure 4.3 C shows a similar comparison but with the X-ray crystal structure in closed conformation. The χ^2 value of 1.57 suggested that the

X-ray crystal structure (3E90.pdb) represent well the folding state in solution of WNV NS2B-NS3pro. In both cases, the crystal structures presented gaps in the 3D coordinate files that were modeled to mimic the whole conformation using CORAL *(28)*. In the case of the open conformation (2GGV.pdb), a segment corresponding to the linker region and the C-terminal of NS3pro was modeled, while the closed conformation (3E90.pdb) required just the modeling of the C-terminal of NS3pro.

Table 4.1. BioSAXS parameters of substrate free-, bound to substrate- and denatured states of WNV Protease

Parameter	Experimental Data			Theoretical data	
	WNpro[a]	**WNVpro + Substrate**	**WNVpro + 6M Urea**	**Closed state (3e90)**	**Open state (2ggv)**
Dmax (Å)	80 ± 1	60 ± 1	160 ± 1	57 ± 1	63 ± 1
Rg (Å)	22.7 ± 0.8	19.6 ± 0.1	52.4 ± 0.4	19.8	20.0
Porod Volume estimate (A^3)	45,958 ± 4,600	36,342 ± 3,600	nd	32,980	32,430
MW (kDa)[b]	27 ± 2.7	21.4 ± 2.1	nd	32.9 ± 3.3	32.4 ± 3.2
Expected MW (kDa)	26.6	26.6	26.6	23.9	23.3

[a] From average of 3 data sets. Error: Standard deviation

[b] MW was determined from Porod volume. Error: 10%

nd = not determined

In addition, I analyzed the structural parameters extrapolated from the theoretical scattering curves and experimental data to determine if a crystallographic structure might be consistent with the experimental data (Table 4.1). The open and closed conformations from crystallographic structures had similar Dmax values, ~ 63 and ~ 57 Å, respectively, as expected from the similarities of both structures. A comparison with the experimental results revealed that

these conformations are closer to the Dmax observed on the substrate bound- state of WNV NS2B-NS3pro determined experimentally as ~ 60 Å than the substrate free- (~80 Å) and denatured (~160 Å) conformations. A similar tendency is observed for Rg values, where closed and open crystal structures showed 19.8 and 20.0 Å, respectively, whereas the substrate bound-state for WNV NS2B-NS3pro was determined to be 19.6 Å, closer that the Rg value observed in absence of substrate. Similarly, Porod volumes calculated for the open and closed crystallographic structures suggested they had more similar compact folding states to the WNV NS2B-NS3pro in presence of substrate than to the WNV protease in absence of it (Table 4.1).

I then determined the approximate shape and compactness of WNV NS2B-NS3pro in presence of the substrate analogue through *ab initio* models. Following the methodology explained in detail on the Material and Methods section, twenty independent models were produced with no symmetry restrictions. The generated averaged envelope exhibited a roughly globular shape with a clear bump. The averaged envelope was compared with the WNV NS2B-NS3pro crystallographic structures. This produced NSD values of 1.02 and 0.98 for the open and closed conformations, respectively. These results suggest that the crystal structure of WNV NS2B-NS3pro in presence of substrate analogue may represent very accurately what I observed in solution by SAXS experiments. On the other hand, the open conformation as observed on the crystal structure seems to be just a variant of the closed conformation that was stabilized by the double mutation or even the crystallization conditions, thus resulting in a modest fitting when compared with the experimental SAXS data.

4.2.4 Structural stability of WNV NS2B-NS3pro followed by Limited Proteolysis.

I used an alternative approach to establish differences on structural stability of WNV NS2B-NS3pro in absence and presence of Ac-KKR-NH_2 as a substrate analogue. Proteinase K has been used to develop limited proteolysis assays, in which random coil segments of any protein that are also accessible enough to the solvent, are cleaved by Proteinase K (PK) without any kind of sequence specificity. The controlled protein degradation can be achieved through temperature, incubation time and PK concentration. Figure 4.4 shows an experiment in which different incubation times were tested (10 and 60 min) at 37 ºC. Increased PK concentrations are indicated on each gel with a triangle. The limited proteolysis experiment seems to work in both cases, where a degradation pattern is easy to observe. I choose a 10 minutes incubation time since it

allowed me to explore new intermediary PK concentrations for the next experiments. It is important to notice the presence of the PK band on lines 6 and 7 in both gels indicating that both proteins, WNV NS2B-NS3pro and PK are migrating in a similar fashion, so extra control samples were considered where just PK is loaded into the gel.

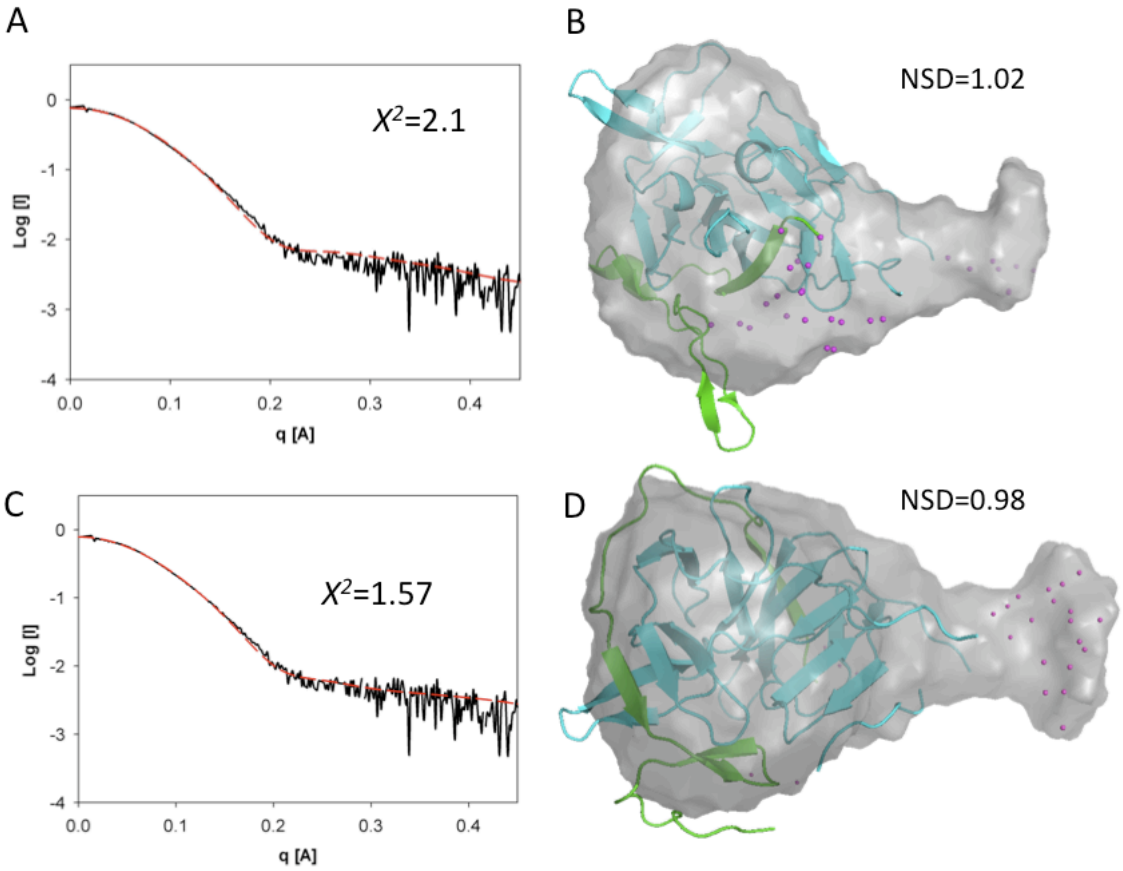

Figure 4.3. Comparison of SAXS data collected from WNV NS2B-NS3pro in presence of substrate analogue and from NS2B-NS3pro crystallographic structures. (A) Experimental scattering curve from WNV protease in presence of substrate (black line) and theoretical scattering curve calculated from the crystal structure in open conformation (red line). (B) WNV crystal structure in open conformation superimposed into the envelope produced from averaged *ab initio* models that best fit experimental BioSAXS data of WNV protease. Modeled amino acids are depicted with pink dots. (C) Experimental scattering curve from WNV protease in presence of substrate (black line) and theoretical scattering curve calculated from crystal structure in closed conformation (red lines). (D) WNV crystal structure in closed conformation superimposed into the envelope produced from averaged *ab initio* models that best fit experimental BioSAXS data of WNV protease. Crystallographic structures are depicted as ribbons with the NS2B and NS3pro chains colored green and cyan, respectively. Modeled C-terminal NS3pro is depicted with pink dots.

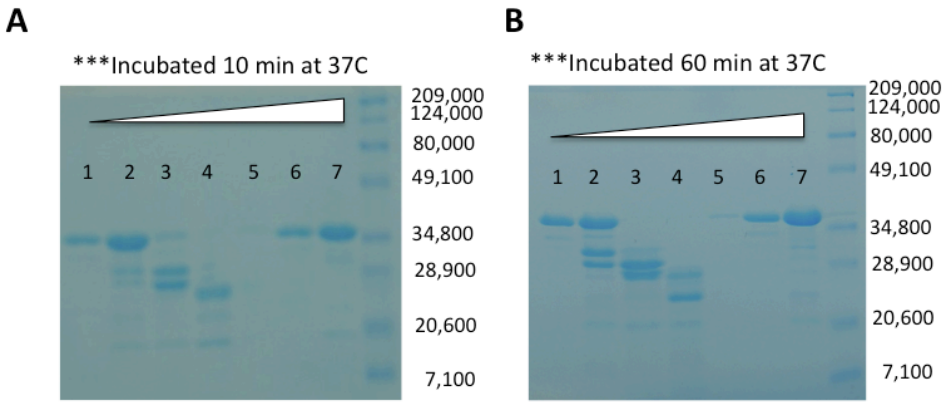

Figure 4.4. Limited proteolysis on WNV NS2B-NS3pro by proteinase K at different incubation times. A. 10 min incubation at 37 ºC. B. 60 min incubation at 37 ºC. In both gels: *lane 1*: Buffer:WNpro (1:1), *lane 2*: PK:WNpro (1:500), *lane 3*: PK:WNpro (1:100), *lane 4*: PK:WNpro (1:50), *lane 5*: PK:WNpro (1:10), *lane 6*: PK:WNpro (10:1), *lane 7:* PK:WNpro (100:1). Standard molecular weight is shown on the right side of each gel.

Thus, I performed a limited proteolysis experiment to study the potential differences on the degradation pattern of WNV NS2B-NS3pro in absence and presence of the substrate analogue Ac-KKR-NH2. In order to make sure that the higher band migrating around ~36 kDa corresponds to the full length WNV NS2B-NS3pro construct that I am working with I analyzed several samples by Mass spectrometry, and confirmed their identity (data not shown). Figure 4.5 A shows the degradation pattern of WNV NS2B-NS3pro in absence of substrate after 10 min incubation at 37 ºC with increased PK concentrations. As observed in lanes 1, 2, 4 and 6, the full length WNV NS2B-NS3pro is present without any evident degradation due to the PK activity, nonetheless in lane 8 it is possible to observe that smaller fragments start to populate the gel, which are more evident in lane 10, where the full length WNV NS2B-NS3pro starts to decay. Lanes 14 and 15 correspond to the next PK concentration instead of lanes 12 and 13 (error at loading samples to gel). Lane 14 shows the total degradation of the full length WNV NS2-NS3pro, whereas just smaller fragments populate the reaction. Finally, lane 12 exhibits even smaller bands which at the highest PK concentration (lane 16) are totally degraded. The absence of band on the lanes where just PK is present indicates that PK is not contributing on the observed protein bands. When comparing the degradation pattern of WNV NS2B-NS3pro in the absence (Figure 4.5 A) or the presence (Figure 4.5 B) of substrate analogue Ac-KKR-NH_2 it is possible to

appreciate small differences on the presence of small fragments, which are slightly more intense on lane 10 of gel A than lane 10 of gel B, indicating that a protection against PK may be occurring due to the presence of the substrate. This effect is better appreciated at observing lane 14 in both gels, where the full length WNV NS2B-NS3pro is still present in gel B and it is totally absent on gel A. This suggests that PK was not able to access the cleavage sites when the WNV NS2B-NS3pro was in the presence of substrate, whereas in the absence of substrate the degradation took place with no obstruction. Lanes 12 where small fragments seem to be more populated in gel B than in gel A, also show differences, in which band are weaker in intensity. Lanes 16 and 17 in both gels exhibit total protein degradation.

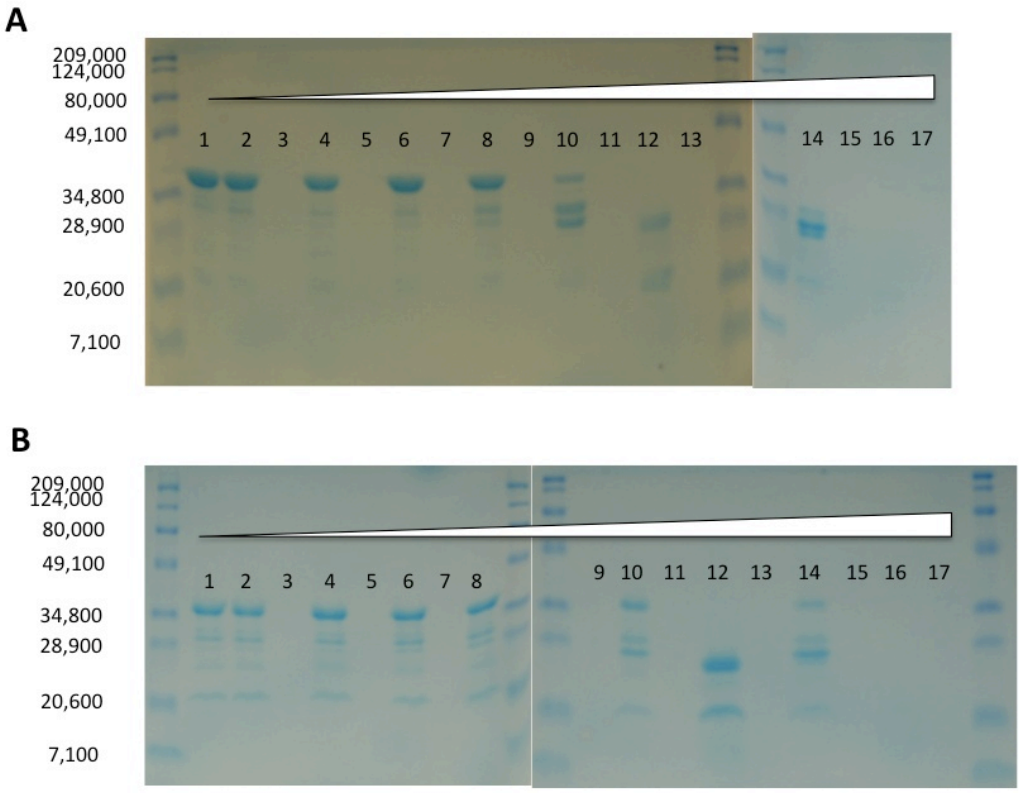

Figure 4.5. Limited proteolysis on WNV NS2B-NS3pro by proteinase K. A. Absence of substrate. B. Presence of 41 nM substrate. In both gels: *lane 1*: Buffer:WNpro (1:1), *lane 2*: PK:WNpro (1:10,000), *lane 3*: PK:Buffer (1:10,000), *lane 4*: PK:WNpro (1:5,000), *lane 5*: PK:Buffer (1:5,000), *lane 6*: PK:WNpro (1:1,000), *lane 7*: PK:Bufer (1:1,000), *lane 8*: PK:WNpro (1:500), *lane 9*: PK:Buffer (1:500), *lane 10*: PK:WNpro (1:100), *lane 11*: PK:Buffer (1:100), *lane 12*: PK:WNpro (1:50), *lane 13*: PK:Buffer (1:50), lane 14: PK:WNpro (1:80), *lane 15*: PK:Buffer (1:80), *lane 16*: PK:WNpro (1:10), *lane 17*: PK:Buffer (1:10).

4.2.5 Chemical denaturation of WNV NS2B-NS3pro followed by Intrinsic Fluorescence. The presence of natural fluorescent probes such as Tryptophan residues on the WNV protease sequence provides a very useful tool to study its structural stability and unfolding process. The emission signal coming from the Tryptophan residues after the appropriate excitation works as a signature of its degree of exposition to the solvent, highlighting a potential unfolding pathway. Specifically, the WNV protease sequence used in this work has seven Tryptophan residues. After analyzing the crystal structure in its bound- state (3E90.pdb), I determined that three of them have a percentage of accessible solvent area (% ASA) higher that 20%, so they were classified as exposed to the solvent, while the rest of them were considered buried into the structure (Figure 4.6). Thus, those Tryptophan residues that appear to be buried into the core of the protein may work as fluorescence probes that can report an average signal of the unfolding process of WNV NS2B-NS3pro. Then, a study of chemical denaturation of WNV NS2B-NS3pro can be done without perturbing the protein sequence with mutations or adding extrinsic fluorescent probes.

Figure 4.6. Trypthophan residues observed on the crystal structure of WNV NS2B-NS3pro. WNV protease NS3pro is depicted in cyan while NS2B is shown in yellow. Peptidic inhibitor present on the crystal structure is shown in pink sticks. Tryptophan residues that show higher percentage of ASA (>20%) are depicted in blue and are considered exposed to the solvent. Tryptophan residues that show less percentage of ASA (<20%) are depicted in red and are considered buried to the solvent.

WNV NS2B-NS3pro was incubated with increasing Urea concentrations as described in Material and Methods section, and the intrinsic fluorescence signal was measured to determine the unfolding state of the protein under each condition. Figure 4.7 A shows the emission signal corresponding to excited Tryptophan residues of the WNV NS2B-NS3pro under different denaturation conditions. Increasing concentrations of Urea (0-8 M) were related to changes on fluorescence curve that included a shift of the maximum wavelength to the low energy region, from ~ 336 nm to ~ 350 nm, and a decreased maximum in fluorescence intensity.

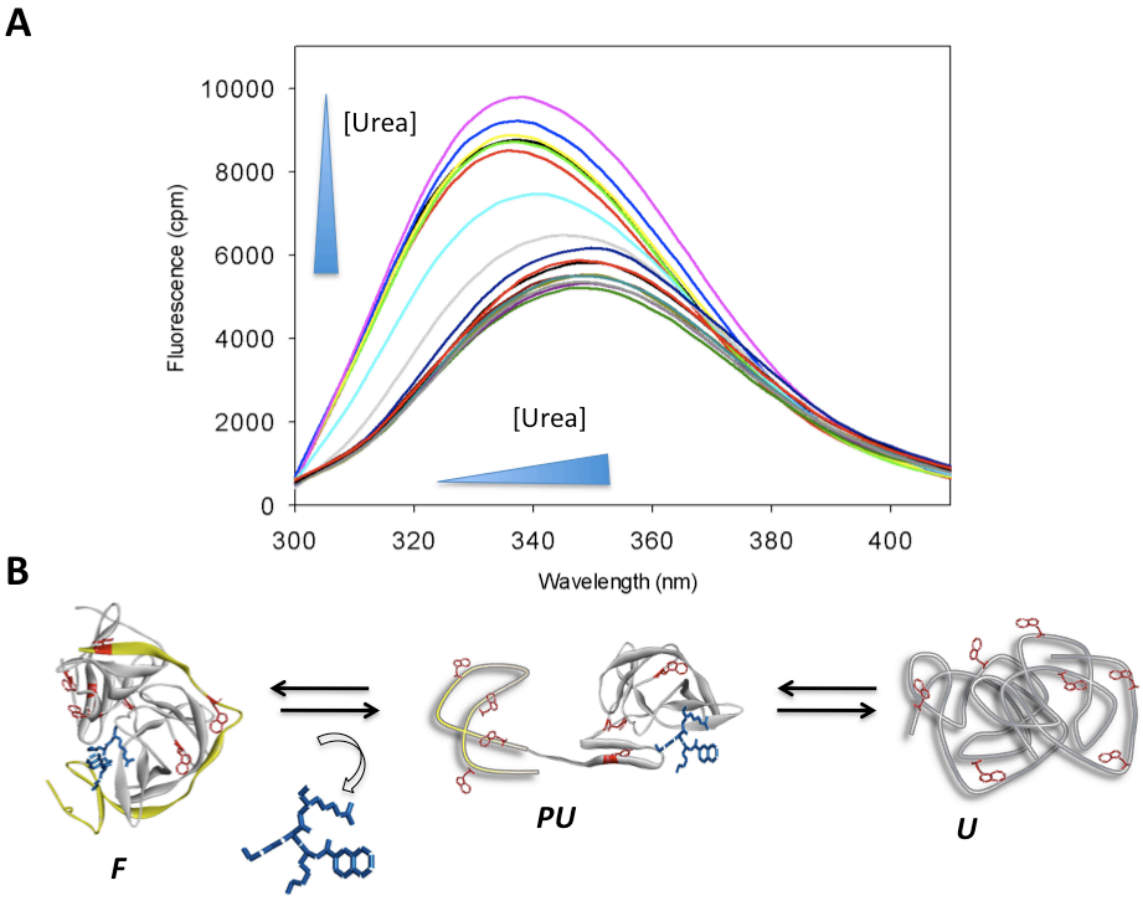

Figure 4.7. Chemical denaturation of WNV NS2B-NS3pro followed by intrinsic fluorescence. A. Intrinsic fluorescence curves of WNV NS2B-NS3pro under increasing Urea concentrations. Excitation wavelength was set at 280 nm, while emission was collected at 300-450 nm. *B*. Theoretical equilibrium pathway of unfolding process of WNV NS2B-NS3pro followed by Tryptophan fluorescence. F-state represents a native condition where a totally folded protein is described as in the crystal structure. NS3pro is colored in gray, NS2B cofactor in yellow and inhibitor is depicted in blue. Tryptophan residues are shown in red. A

theoretical PU-state describes a partially unfolded state with some of the Tryptophan residues being exposed to the solvent emitting the corresponding fluorescence signal. At this point, the inhibitor may follow two alternative pathways, either stay interacting with the partially unfolded protein or be released at low concentrations of denaturing agent due to low affinity to the protease. U-state describes a totally unfolded state, in which all the tryptophan residues are exposed to the solvent.

Considering the maximum wavelength given by the fluorescence intensity at each Urea concentration, it was possible to observe a two-state unfolding curve for the WNV NS2B-NS3pro (black circles), with a midpoint close to 3 M Urea suggesting two main conformations that predominate the equilibrium reaction (F- and U- states). In order to analyze any change on the protein stability due to the presence of the substrate analogue Ac-KKR-NH_2 or due to the experimental condition used when WNV NS2B-NS3pro was crystallized, I performed similar experiments with special emphasis on any shift around the midpoint of the unfolding curve. Thus, a shift to a higher Urea concentration would indicate a more stable protein structure under those conditions, whereas a shift to the lower Urea concentrations would indicate a less stable protein conformation.

4.2.5.a Effect of the substrate analogue on the protein stability of WNV NS2B-NS3pro followed by intrinsic fluorescence. WNV NS2B-NS3pro was incubated with Ac-KKR-NH_2 as a substrate analogue at 1.3 mM as final concentration as described on the Materials and Method section. After 30 min incubation at 37 ºC, fluorescence signal was measured previous excitation at 280 nm. Figure 4.8 A shows the maximum fluorescence intensity at each Urea concentration, where the two-state unfolding process observed with WNV NS2B-NS3pro in presence of the substrate analogue (red triangles) shows a midpoint at 2.5 M Urea similar to that observed when the protein is in absence of substrate. This suggests that at these experimental conditions no significant changes on protein stability were evident. This may have two explanations; first, that the substrate analogue is actually not stabilizing the protein structure of WNV NS2B-NS3pro upon binding or second, that the affinity of the substrate to the protease is not strong enough to keep the ligand interacting with the protease even at very low Urea concentrations, so that the observed unfolding process would correspond to the WNV NS2B-NS3pro itself as it was observed in absence of substrate. These possible explanations may be proved using a substrate analogue with higher affinity by the WNV NS2B-NS3pro with a Ki in the order of nanomolar instead of the 50 μM determined by the Ac-KKR-NH_2 used in this experiment.

4.2.5.b Effect of Crystallization buffer on the protein stability of WNV NS2B-NS3pro followed by intrinsic fluorescence. It is well known that most of the buffers used for protein crystallization are often saturated with high salt concentrations, crowding agents, PEG, etc. that can lead to either stabilization of the protein structure or in the worst scenario to protein aggregation. However, when there is some evidence indicating that the protein under study has an intrinsic flexibility associated to its folding state, it is difficult to assess whether or not these extreme conditions will stabilize a conformation that is representative of the real folding state that the protein adopt at physiologically relevant solution conditions. To analyze if the experimental environment has an effect on the stabilization of the WNV NS2B-NS3pro structure as it was determined by X-ray crystallography, I performed a denaturation curve in presence of increasing concentrations of Sodium Formate buffer pH 7.2 (NaFo), which was used to obtain the crystal structure in its bound conformation *(20)*. Figure 4.8 B shows the final denaturation curves of the maximum wavelength given by the fluorescence intensity versus the Urea concentration. I performed the experiment at 0, 1 and 3.5 M NaFo, which are represented with black, green and yellow symbols. As observed in the Figure 4.8 B there is an evident shift of the midpoint of each curve to higher Urea concentration as the NaFo concentration is increased. The midpoint at 0 M NaFo corresponded to ~ 2.5 M Urea, at 1 M NaFo was shifted to ~ 3.5 M, whereas at 3.5 M NaFo (actual concentration used to get a stable enough structure to obtain a crystal protein) the midpoint was shifted to 7 M Urea. This finding suggests that the buffer used for crystallization purposes has a strong effect on the structural stability of the protein, and this effect may alter the folding state of the protein in an important manner if the protein exhibits certain degree of flexibility. Therefore, the conformation revealed by the crystallographic structure will not necessarily represent the folding state of the protein in solution.

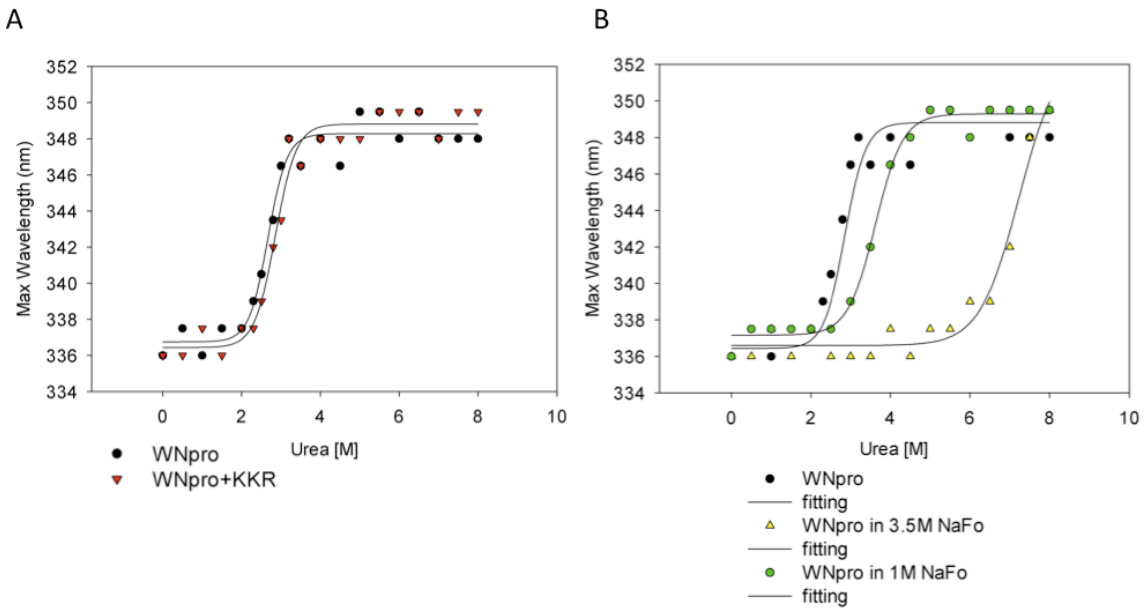

Figure 4.8. Unfolding curves of WNV NS2B-NS3pro followed by intrinsic fluorescence. A. Effect of the substrate analogue Ac-KKR-NH_2 on the protein stability. WNV NS2B-NS3pro in absence of substrate (WNpro) is indicated with black circles whereas the presence of substrate (WNpro+KKR) is shown with red triangles. B. Effect of Crystallization buffer on the protein stability. 0 M NaFo is represented with black circle, 1 M NaFo with green circles, and 3.5 M NaFo is indicated with yellow triangles. Fitting curves are depicted in solid black lines in both plots.

4.3 Discussion

WNV protease structure has been solved by X-ray crystallography in presence of a peptidic inhibitor *(20)* and aprotinin (BPTI) *(19)*. In both cases, the folding state of the WNV protease adopts a compact and globular conformation, with NS2B cofactor wrapped around NS3pro and forming part of the active site. This conformation is called closed state. These high resolution structures are in good agreement with NMR experiments that characterize the protease in presence of small compounds or non-peptidic inhibitors *(21-22)*, suggesting that, in presence of substrate analogue, WNV protease exhibits a compact conformation in both, solution and crystal state.

Since my previous results (Chapter 3) indicate that WNV protease in absence of substrate would be highly flexible in solution and therefore existing as a conformational ensemble, I decided to examine a substrate analogue as a potential folding inductor. To approach this hypothesis I used Ac-KKR-NH_2, a similar peptide to that used for crystallography (2-naphthoyl-KKR-H) *(20)*. Through a kinetic assay, I determined that Ac-KKR-NH_2 follows a competitive

inhibition mechanism with an inhibition constant of 50 μM, which is weaker compared to the 41 ηM indicated for 2-naphthoyl-KKR-H *(20)*. Once the required substrate analogue concentration to populate the protease-substrate complex was determined, I performed SAXS experiments to characterize the folding state of WNV NS2B-NS3pro in presence and absence of a substrate analogue, as well as under denaturing conditions. These results clearly indicate that a significant conformational change of the protease is induced upon interaction with Ac-KKR-NH_2 leading to the conclusion that a more compact folding state is better in representing the WNV NS2B-NS3pro in solution. In addition, an expected more elongated conformation was characteristic of the protein exposed to the denaturing environment. Interestingly, the "closed" crystallographic structure of WNV NS2B-NS3pro adequately represented the conformational state of WNV protease in solution after modeling onto the structure the C-terminal segment of NS3pro that was absent on the crystal structure. When the experimental scattering data was compared to the theoretical data from the "open" crystal structure *(19)*, the fitting curves revealed a poorer degree of agreement. This finding may be due to the different NS2B conformation, which could be induced either by the double mutation K104R/H51A that had to be introduced to obtain the crystal structure in absence of a substrate analogue or even by the crystallization conditions.

In order to test the possibility of the substrate analogue being able to increase the structural stability of WNV NS2B-NS3pro I used an alternative approach, which involved the presence of Proteinase K (PK) to perform limited proteolysis on WNV NS2B-NS3pro. The results indicated a more inaccessible conformation to PK when WNV protease was in presence of the substrate analogue, than when it was in absence of it, suggesting a flexible conformation of the substrate free-state protease than become more structured and compact in presence of the substrate analogue. This result is in agreement with the outcome observed by SAXS experiments.

The structural stability of WNV NS2B-NS3pro as well as the effect of the substrate analogue on its folding state was analyzed by a parallel approach, which takes advantage of the intrinsic fluorescence of WNV protease. Thus, WNV protease was exposed to a range of denaturant agent concentrations to establish the level of exposure of tryptophan residues naturally occurring in its sequence. I determined that WNV NS2B-NSpro undergoes a two state unfolding process, which seems to be unaltered due to the presence of substrate analogue, with no observed shift on the midpoint curve. This finding may be explained by simply suggesting that the

substrate used was not able to increase the protein stability when studied by this technique, or, a second and more likely explanation, is that even low concentrations of denaturant agent and weak interaction between the substrate analogue and the protease (determined previously by kinetic assay) facilitated the release of the substrate analogue from the protease at early states of the unfolding process characterizing WNV protease in its substrate free-conformation. The high flexibility of WNV NS2B-NS3pro observed in solution by SAXS experiments and the compact and well-structured conformation observed by X-ray crystallography led me think about the possibility that crystallization buffer (Sodium Formate) at high concentrations may play a role on the stabilization of the protein fold. Therefore, when I tested this idea using the same methodology, I observed an evident shift of the midpoint of the unfolding curve towards higher Urea concentrations as the Sodium formate concentration was closer to the one used for crystallization. Therefore, there is a clear role of the crystallization buffer as a stabilizing factor of the WNV protease fold. The potential effect of this salt on the intrinsic flexibility and dynamic of WNV protease may raise concerns about the real nature of the structure of the protease in simple aqueous solution, especially in the case of the crystal structure in absence of substrate.

CHAPTER 5

GENERAL CONCLUSIONS

The dissertation work described on this thesis intended to contribute to the drug design and development against West Nile virus protease from the perspective of the target protein used in structure-based computational screens. Thus, a detailed characterization in solution of the WNV protease used for inhibition assays and crystallization experiments was made to provide a better understanding of the conformational state of the target protein. Previous studies have successfully characterized the WNV protease's fold in the presence of peptidic inhibitors or small compounds. However in their absence, the protein seems to exhibit an excessive flexibility that makes difficult an accurate determination of its folding state. Therefore, introducing mutations to obtain the crystal structure combined with the non-straightforward assignment of broad NMR chemical shifts may underline the so far fruitless efforts to generate an effective drug against the WNV protease.

The results from this research indicate that first, active and non-mutated wild type WNV protease is partially folded and exhibits intrinsic flexibility that should be taken into consideration when performing structure-based drug design. Incorporation of this knowledge may improve the development of more efficient inhibitors against WNV protease. Second, the ordered and compact folding state of WNV protease as observed in X-ray crystallography seems to be partially due to the presence of inhibitors or substrate analogues and/or the crystallization conditions.

In order to accurately determine the conformational state of WNV protease utilized for drug discovery, this research was focused on the same construct used by the drug design field for inhibition assays and crystallization experiments. The artificial construct utilized in this work included an artificial linker connecting NS2B-NS3pro and removal of the transmembrane segments of NS2B cofactor. The specific impact of these modifications to the overall folding state of the protein remains unknown as well as their in vivo relevance. However, the complexity

of the polyprotein assembly process in the membrane plus the high insolubility of a more accurate protease construct makes detailed structural analyses virtually intractable. Several prospective rationally-designed drugs against the protease, though exhibiting low affinity, have shown acceptable specificity which provides evidence that the currently used constructs can mimic the in vivo target relatively well. We think this work should greatly contribute toward designing promising new drugs.

BIBLIOGRAPHY

1. **Mueller NH, Yon C, Ganesh VK, Padmanabhan R.** 2007. Characterization of the West Nile virus protease substrate specificity and inhibitors. Int J Biochem Cell Biol. 39:606-614.

2. **Suthar MS, Diamond MS, Gale M Jr.** 2013. West Nile virus infection and immunity. Nat Rev Microbiol. 11:115-128.

3. **Chappell KJ, Stoermer MJ, Fairlie DP, Young PR.** 2008. West Nile Virus NS2B/NS3 protease as an antiviral target. Curr Med Chem. 15:2771-2784.

4. **Colpitts TM, Conway MJ, Montgomery RR, Fikrig E.** 2012. West Nile Virus: biology, transmission, and human infection. Clin Microbiol Rev. 25:635-648.

5. **Cho H, Diamond MS.** 2012. Immune responses to West Nile virus infection in the central nervous system. Viruses. 4:3812-30.

6. **Shiryaev SA, Kozlov IA, Ratnikov BI, Smith, JW, Lebl M, Strongin, AY.** 2007. Cleavage preference distinguishes the two-component NS2B-NS3 serine proteinases of Dengue and West Nile viruses. Biochem J. 401:743-752.

7. **Bera AK, Kuhn RJ, Smith JL.** 2007. Functional characterization of cis and trans activity of the Flavivirus NS2B-NS3 protease. J Biol Chem. 282:12883-12892.

7b. **Assenberg R, Mastrangelo E, Walter TS, Verma A, Milani M, Owens RJ, Stuart DI, Grimes JM, Mancini EJ.** 2009. Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication. J Virol. 83:12895-12906.

8. **Falgout B, Pethel M, Zhang YM, Lai CJ.** 1991. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. J Virol. 65:2467-2475.

9. **Nall TA, Chappell KJ, Stoermer MJ, Fang NX, Tyndall JD, Young PR, Fairlie DP.** 2004. Enzymatic characterization and homology model of a catalytically active recombinant West Nile virus NS3 protease. J Biol Chem. 279:48535-48542.

10. **Chernov AV, Shiryaev SA, Aleshin AE, Ratnikov BI, Smith JW, Liddington RC, Strongin AY.** 2008. The two-component NS2B-NS3 proteinase represses DNA unwinding activity of the West Nile virus NS3 helicase. J Biol Chem. 283:17270-17278.

11. **Luo D, Wei N, Doan DN, Paradkar PN, Chong Y, Davidson AD, Kotaka M, Lescar J, Vasudevan SG**. 2010. Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. J Biol Chem. 285:18817-18827.

12. **Shiryaev SA, Ratnikov BI, Chekanov AV, Sikora S, Rozanov DV, Godzik A, Wang J, Smith JW, Huang Z, Lindberg I, Samuel MA, Diamond MS, Strongin AY.** 2006. Cleavage targets and the D-arginine-based inhibitors of the West Nile virus NS3 processing proteinase. Biochem J. 393:503-511.

13. **Knox JE, Ma NL, Yin Z, Patel SJ, Wang WL, Chan WL, Ranga Rao KR, Wang G, Ngew X, Patel V, Beer D, Lim SP, Vasudevan SG, Keller TH.** 2006. Peptide inhibitors of West Nile NS3 protease: SAR study of tetrapeptide aldehyde inhibitors. J Med Chem. 49:6585-6590.

14. **Schuller A, Yin Z, Brian Chia CS, Doan DN, Kim HK, Shang L, Loh TP, Hill J, Vasudevan SG**. 2011. Tripeptide inhibitors of dengue and West Nile virus NS2B-NS3 protease. Antiviral Res. 92:96-101.

15. **Ganesh VK, Muller N, Judge K, Luan CH, Padmanabhan R, Murthy KH.** 2005. Identification and characterization of nonsubstrate based inhibitors of the essential dengue and West Nile virus proteases. Bioorg Med Chem. 13:257-264.

16. **Johnston PA, Phillips J, Shun TY, Shinde S, Lazo JS, Huryn DM, Myers MC, Ratnikov BI, Smith JW, Su Y, Dahl R, Cosford ND, Shiryaev SA, Strongin AY.** 2007. HTS identifies novel and specific uncompetitive inhibitors of the two-component NS2B-NS3 proteinase of West Nile virus. Assay Drug Dev Technol. 5:737-750.

17. **Mueller NH, Pattabiraman N, Ansarah-Sobrinho C, Viswanathan P, Pierson, TC, Padmanabhan R.** 2008. Identification and biochemical characterization of small-molecule inhibitors of west nile virus serine protease by a high-throughput screen. Antimicrob Agents Chemother, 52: 3385-3393.

18. **Ekonomiuk D, Su XC, Ozawa K, Bodenreider C, Lim SP, Yin Z, Keller TH, Beer D, Patel V, Otting G, Caflisch A, and Huang D.** 2009. Discovery of a non-peptidic inhibitor of west nile virus NS3 protease by high-throughput docking. PLoS Negl Trop Dis. 3:e356.

19. **Aleshin AE, Shiryaev SA, Strongin AY, Liddington RC.** 2007. Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. Protein Sci.16:795-806.

20. **Robin G, Chappell K, Stoermer MJ, Hu SH, Young PR, Fairlie DP, Martin JL.** 2009. Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation. J Mol Biol. 385:1568-1577.

21. **Su XC, Ozawa K, Qi R, Vasudevan SG, Lim SP, Otting G.** 2009 NMR analysis of the dynamic exchange of the NS2B cofactor between open and closed conformations of the West Nile virus NS2B-NS3 protease. PLoS Negl Trop Dis. 3:e561.

22. **Su XC, Ozawa K, Yagi H, Lim SP, Wen D, Ekonomiuk D, Huang D, Keller TH, Sonntag S, Caflisch A, Vasudevan SG, Otting G.** 2009. NMR study of complexes between low molecular mass inhibitors and the West Nile virus NS2B-NS3 protease. FEBS J. 276:4244-4255.

23. **Gayen S, Chen AS, Huang Q, Kang C.** West Nile Virus (WNV) protease and membrane interactions revealed by NMR spectroscopy. Biochem Biophys Res Commun. 2012 Jul 13;423(4):799-804.

24. **Schuck P.** 2000. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophys J. 78:1606-1619.

25. **Ortega A, Amorós D, García de la Torre J.** 2011. Prediction of hydrodynamic and other

solution properties of rigid proteins from atomic- and residue-level models. Biophys J. 17:892-898.

26. **Tomlinson SM, Watowich SJ.** 2008. Substrate inhibition kinetic model for West Nile virus NS2B-NS3 protease. Biochemistry. 47:11763-11770.

27. **Kuzmic P.** 1996. Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. Anal Biochem. 237:260-273.

28. **Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI.** 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Cryst. 36:1277-1282.

29. **Mertens HD, Svergun DI.** 2010. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. J Struct Biol. 172:128-141.

30. **Svergun DI.** 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Cryst. 25:495-503.

31. **Svergun DI, Barberato C, Koch MHJ.** 1995. CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J. Appl. Cryst. 28:768-773.

32. **Franke D, Svergun DI.** 2009. DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. J. Appl. Cryst. 42:342-346.

33. **Volkov VV, Svergun DI.** 2003. Uniqueness of ab-initio shape determination in small-angle scattering. J. Appl. Cryst. 36:860-864.

34. **Bernado P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI.** 2007. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. J. Am. Chem. Soc. 129:5656-5664.

35. **KEN'ICHI A.** 2002. Development and Application of Secondary Structural Estimation Program of Protein for FTIR (IR-SSE). Jasco Rep. 44:54-57.

36. **Chappell KJ, Stoermer MJ, Fairlie DP, Young PR.** 2007. Generation and characterization of proteolytically active and highly stable truncated and full-length recombinant West Nile virus NS3. Protein Expr Purif. 53:87-96.

37. **Dam J, Schuck P.** 2004. Calculating sedimentation coefficient distributions by direct modeling of sedimentation velocity concentration profiles. Methods Enzymol. 384:185-212.

38. **Bulheller BM, Hirst JD.** 2009. DichroCalc--circular and linear dichroism online. Bioinformatics. 25:539-40.

39. **Uversky VN.** 2009. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. Protein J. 28:305-325.

40. **Das BK, Bhattacharyya T, Roy S.** 1995. Characterization of a urea induced molten globule intermediate state of glutaminyl-tRNA synthetase from Escherichia coli. Biochemistry. 34:5242-5247.

41. **Ansari MA, Zubair S, Atif SM, Kashif M, Khan N, Rehan M, Anwar T, Iqbal A, Owais M.** 2010. Identification and characterization of molten globule-like state of hen egg-white lysozyme in

presence of salts under alkaline conditions. Protein Pept Lett. 17:11-17.

42. **Celej MS, Montich GG, Fidelio GD.** 2003 Protein stability induced by ligand binding correlates with changes in protein flexibility. Protein Sci. 12:1496-1506.

43. **Manning MC.** 2005. Use of infrared spectroscopy to monitor protein structure and stability. Expert Rev Proteomics. 2:731-743.

44. **Byler DM, Susi H.** 1986. Examination of the secondary structure of proteins by deconvolved FTIR spectra. Biopolymers. 25:469-487.

45. **Haris PI, Lee DC, Chapman D.** 1986. A Fourier transform infrared investigation of the structural differences between ribonuclease A and ribonuclease S. Biochim Biophys Acta. 874:255-265.

46. **Kong J, Yu S.** 2007. Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta Biochim Biophys Sin (Shanghai). 39:549-559.

47. **Sreerama N, Manning MC, Powers ME, Zhang JX, Goldenberg DP, Woody RW.** 1999. Tyrosine, phenylalanine, and disulfide contributions to the circular dichroism of proteins: circular dichroism spectra of wild-type and mutant bovine pancreatic trypsin inhibitor. Biochemistry. 38:10814-22.

VITA

Andrea del Pilar Garcés Fernández was born on March 27th, 1979, Valparaíso, Chile. Her parents are Teresa Fernández Paganini and Carlos Garcés Astudillo. Her youngest sister is Katherine. All his primary and secondary education was carried in Liceo Juana Ross de Edwards, Valparaíso. In 1998, she was accepted to start college studies in the Universidad de Chile, pursuing a major in Molecular Biotechnology Engineering, where she met Rodrigo Díaz-Espinoza, who today is the father of her soon-to-be-born daughter. In 2005, she obtained the Bachelor of Sciences degree with honors. After conducting her undergraduate thesis work under Dr. Octavio Monasterio mentoring, which was published later in 2007, Andrea obtained the Engineer degree in 2007, with *summa cum laude* honor. During his undergraduate studies, she worked as a teaching assistant in several courses and presented her work in several national meetings.

In 2007, she joined the laboratory of Dr. James C. Lee at The University of Texas Medical Branch, USA, as visiting scientist. She presented her research in meetings, such as The Gibbs Conference (Illinois, USA, 2007, 2008 and 2009). Her work in this laboratory was published later in 2012.

In 2008 she was accepted by the Graduate School of Basic Science in the same university to pursue graduate studies in the Structural and Molecular Biophysics PhD Program. There, she rotated in three different labs conducted by: Dr. Andres Oberhauser, Dr. Vincent Hilser and Dr. Junji Iwajara. In 2010, she joined Dr. Stanley Watowich's lab to continue her PhD thesis work. Her graduate research was presented in the 18th Annual Structural Biology Symposium, UTMB, Galveston (2013). She finished her graduate coursework with a perfect GPA. Andrea participated actively during 2012-2013 as mentor in the Scientific Research and Design: Bench Tutorials Program for high school students.

Following the completion of her dissertation, Andrea will dedicate her short-term future to the most challenging task she ever imagined, being mom.

Education

2005. Bachelor degree in Molecular Biotechnology Engineering, University of Chile. Chile.
2007. Engineer in Molecular Biotechnology degree, University of Chile. Chile.

Publications

Garces, A. P.; Watowich S. J. "Intrinsic flexibility of West Nile virus protease in solution characterized using small angle X-ray scattering". (Submitted to Biochemistry)

Garces, A. P.; Watowich S. J. "Insights to the mechanism of cleavage of West Nile virus protease and drug design". (in preparation)

Lee-Chuan C. Yeh, Wilfredo E. Falcon, Andrea Garces, J. Ching Lee, and John C. Lee. A Host-Guest Relationship in Bone Morphogenetic Protein Receptor-II Defines Specificity in Ligand-Receptor Recognition. *Biochemistry*. 2012 Sep 4;51(35):6968-80

Diaz-Espinoza, R., Garces, A.P., Arbildua, J.J., Montecinos, F., Brunet, J.E., Lagos, R., Monasterio, O. Domain folding and flexibility of Escherichia coli FtsZ determined by tryptophan site-directed mutagenesis. *Protein Sci*. 16: 1543-1556; 2007.

Tomlinson, S. Garces, A., Viswanathan U. and Watowich, S.J, "Computational and structural approaches to develop West Nile Virus neurotropic protease inhibitors" *Computational and structural Biotechnology Journal* (in preparation).

Abstracts

- Garces, A; Watowich, SJ. "Intrinsic flexibility of West Nile virus protease in solution characterized using small angle X-ray scattering". 18th Annual Structural Biology Symposium, UTMB, Galveston. May, 2013.

- Garces, A.P., Hilser,V.J. and Lee, J.C. "Towards an Understanding of the Mechanism of Resistance to Antibody Neutralization: Involvement of Two Apparently Different Networks of Residues in Domain III of Envelope Protein in West Nile and Dengue 2 flaviviruses". *23st Annual Gibbs Conference on Biothermodynamics*, Carbondale, Illinois. October, 2009.

- Diaz-Espinoza, R., Garces, A.P., Whitten, S., Soto, C., Hilser, V. "Denatured-Ensemble-Based Thermodynamic Analysis of Prion and Amyloid Sequences". *23st Annual Gibbs Conference on Biothermodynamics*, Carbondale, Illinois. October, 2009.

- Garces, A.P.; Whitten, S. and Lee, J.C. "Towards an Understanding of the Sequence-Structure-Stability-Function Relationship". *22st Annual Gibbs Conference on Biothermodynamics*, Carbondale, Illinois. October, 2008.

- Garces, A.P., Maillard, R.A., Hilser, V.J. Lee, J.C. "Viral strategy to modulate the resistance of West Nile virus to antibody-mediated neutralization: All the residues in the envelope protein domain III are potentially involved. *21st Annual Gibbs Conference on Biothermodynamics*, Carbondale, Illinois. September, 2007.