UNIVERSITY OF SÃO PAULO

INSTITUTE OF CHEMISTRY

Postgraduate Program in Biological Sciences

(Biochemistry)

Angy Liseth Davalos Macías

Deciphering the mechanism of interaction between two key subunits from the Type IV Secretion System core complex: VirB9 and VirB7

Corrected version of the defended thesis

São Paulo

Date of deposit in the SPG:

18/04/2022

UNIVERSIDADE DE SÃO PAULO

INSTITUTO DE QUÍMICA

Programa de Pós-Graduação em Ciências Biológicas

(Bioquímica)

Angy Liseth Davalos Macías

Decifrando o mecanismo de interação entre duas subunidades-chave do core complex do Sistema de Secreção Tipo IV: VirB9 e VirB7

Versão corrigida da Tese defendida

São Paulo

Data de Depósito na SPG:

18/04/2022

Angy Liseth Davalos Macías

Deciphering the mechanism of interaction between two key subunits from the Type IV Secretion System core complex: VirB9 and VirB7

Tese apresentada ao Instituto de Química da Universidade de São Paulo para obtenção do Título de Doutora em Ciências (Bioquímica)

Orientador: Prof. Dr. Roberto Kopke Salinas

São Paulo

2022

Ficha Catalográfica elaborada eletronicamente pelo autor, utilizando o programa desenvolvido pela Seção Técnica de Informática do ICMC/USP e adaptado para a Divisão de Biblioteca e Documentação do Conjunto das Químicas da USP

Bibliotecária responsável pela orientação de catalogação da publicação:
Marlene Aparecida Vieira - CRB - 8/5562

Dávalos Macias, Angy Liseth
D245d Deciphering the mechanism of interaction between two key subunits from the Type IV Secretion System core complex: VirB9 and VirB7 / Angy Liseth Dávalos Macias. - São Paulo, 2022.
162 p.

Tese (doutorado) - Instituto de Química da Universidade de São Paulo. Departamento de Bioquímica.
Orientador: Salinas, Roberto Kopke Salinas

1. Cinética . 2. Proteínas. 3. Bioquímica. 4. Resonancia magnética nuclear. 5. Biofísica de proteínas. I. T. II. Salinas, Roberto Kopke Salinas, orientador.

Universidade de São Paulo
Instituto de Química

"Decifrando o mecanismo de interação entre duas sub-unidades chave do core complex do sistema de secreção tipo IV: VirB9 e VirB7"

ANGY LISETH DAVALOS MACIAS

Tese de Doutorado submetida ao Instituto de Química da Universidade de São Paulo como parte dos requisitos necessários à obtenção do grau de Doutora em Ciências obtido no Programa Ciências Biológicas (Bioquímica) - Área de Concentração: Bioquímica.

_____

Prof. Dr. Roberto Kopke Salinas
(Orientador e Presidente)

APROVADO(A) POR:

_____

Prof. Dr. Sandro Roberto Marana
IQ - USP

_____

Prof. Dr. Antônio José da Costa Filho
FFCLRP - USP

_____

Prof. Dr. Fabio Ceneviva Lacerda de Almeida
UFRJ

SÃO PAULO
14 de julho de 2022

Contents

List of Figures

List of Tables

Abstract

Dávalos, A. M. **Deciphering the mechanism of interaction between two key subunits from the Type IV Secretion System core complex: VirB9 and VirB7.** 2022. 161p. PhD Thesis - Postgraduate program in Biochemistry. Institute of Chemistry, University of São Paulo, São Paulo.

Intrinsically disordered proteins (IDPs) are implicated in the regulation of many important processes within the cell, a factor that explains their abundance in the proteome. Often, IDPs undergo local or global conformational rearrangements coupled to binding. However, the mechanisms by which IDPs interact with their partners have been a topic of debate and remain poorly understood. Here, we characterized the dynamics of the VirB9 C-terminal domain ($VirB9^{Ct}$), and its binding mechanism to the N-terminal tail of VirB7 ($VirB7^{Nt}$). The interaction between the two domains is essential for the assembly of a supramolecular complex, the Type IV Secretion System from the phytopathogen *Xanthomonas citri*, which is responsible for the secretion of toxins that lead to bacterial killing. $VirB7^{Nt}$ is completely disordered in the unbound state, while $VirB9^{Ct}$ has characteristics of a molten globule. The unbound state of $VirB9^{Ct}$ was characterized by a combination of Circular Dichroism spectroscopy, Isothermal Titration Calorimetry, Differen-

tial Scanning Calorimetry, Chemical Exchange Saturation Transfer (CEST) NMR experiments, and ANS fluorescence assays. We found that, in the unbound state, VirB9^{Ct} has similar secondary and tertiary structures as when bound to VirB7^{Nt}. Furthermore, in the unbound state VirB9^{Ct} samples the bound-conformation even in the absence of VirB7^{Nt}, which points to conformational selection as recognition pathway. The interaction mechanism between VirB9^{Ct} and VirB7^{Nt} was elucidated by quantitative analysis of CEST and fluorescence stopped flow experiments. The results support the view that their interaction occurs through conformational selection at 25°C, but becomes more complex at higher temperatures due to the intrinsic dynamics of VirB9^{Ct}, and, in this case a combined CS-IF mechanism predominates. Overall, these results highlight the need of combining different biophysical methods to elucidate protein-protein interaction mechanisms, and contribute to the understanding of molecular recognition pathways and their complexity.

Keywords: Nuclear magnetic resonance spectroscopy, Protein kinetics, Structural biology, Molecular biophysics, Type IV secretion system, Molecular recognition mechanisms.

Resumo

Dávalos, A. M. **Decifrando o mecanismo de interação entre duas subunidades chave do core complex do Sistema de Secreção Tipo IV: VirB9 e VirB7** 2022. 161p. Tesis de Doutorado - Programa de Pós-graduação em Bioquímica. Instituto de química, Universidade de São Paulo, São Paulo.

Proteínas intrinsicamente desordenadas, ou do inglês, IDPs, estão envolvidas na regulação de uma varieade de processos celulares, o que explica a sua abundância no proteoma. Frequentemente, IDPs sofrem rearranjos conformacionais locais ou globais, acoplados à ligação a outras moléculas. Os mecanismos pelos quais IDPs interagem com seus parceiros moleculares têm sido alvo de debates, e permanecem pouco compreendidos. Neste trabalho, nós caracterizamos a dinâmica conformacional do domínio C-terminal de VirB9 ($VirB9^{Ct}$), e o mecanismo de interação de $VirB9^{Ct}$ com a cauda N-terminal de VirB7 ($VirB7^{Nt}$). A interação entre os dois domínios é essencial para a montagem de um complexo supramolecular, o Sistema de Secreção do Tipo IV do fitopatógeno *Xanthomonas citri*, o qual é responsável pela secreção de toxinas que levam à morte baceriana. $VirB7^{Nt}$ é completamente desor-

denada na forma não ligada, enquanto que VirB9^{Ct} possui características de *molten globule*. VirB9^{Ct}, no estado não ligado, foi caracterizada por uma combinação de espectroscopia de Dicroísmo Circular, Calorimetria Isotérmica de Titulação, Calorimetria Diferencial de Varredura, experimentos de RMN de transferência de saturação (CEST), e medidas de fluorescência de ANS. Encontramos que VirB9^{Ct} no estado não ligado apresenta estruturas secundária e terciárias semelhantes à forma ligada. Além disso, observamos que VirB9^{Ct} visita a conformação ligada mesmo na ausência de VirB7^{Nt}, o que indica que o mecanismo de reconhecimento entre as duas proteínas ocorre por seleção de conformações. O mecanismo de reconhecimento entre VirB9^{Ct} e VirB7^{Nt} também foi estudado quantitativamente por CEST e medidas de fluorescência em *stopped-flow*. Os resultados obtidos suportam a visão de que a interação entre VirB9^{Ct} e VirB7^{Nt} ocorre através de seleção de conformações a 25C, mas torna-se mais complexa em maiores temperaturas devido à dinâmica intrínseca de VirB9^{Ct}. Neste caso um mecanismmo combinado CS-IF é predominante. De forma geral, estes resultados ressaltam a necessidade da combinação de diferentes métodos biofísicos para elucidar mecanismos de interação proteína-proteína, e contribuem para a compreensão dos mecanismos de reconhecimento molecular e sua complexidade.

Keywords: Resonancia magnética nuclear, Cinética de proteínas, Biologia estrutural, Biofísica, Sistema de secreção tipo IV, Mecanismos de reconhecimento molecular.

Acknowledgements

I must say, my PhD experience was frustrating and discouraging many times. But I am also thankful, for the great things it gave me and taught me, the good and the bad, because both are necessary to grow up and have made me the person I am now. I really enjoyed the process, I discovered my scientific curiosity and proved me in many areas of my life. I learned more about me, my strengths and weaknesses, and I also overcame my fears and prejudices. I feel more confident now than five years ago, but at the same time I feel humble because I know that there is still a long way to go in knowledge, personal experiences and in every aspect of life. A wonderful part of my postgraduate life was the possibility to get to know amazing people, as professionals and as humans. I am so grateful for having these people in my life, for all the help I received in my experiments, the scientific discussions, for the coffee afternoons, their support in difficult times and over all for their friendship. I really carry all of you in my thoughts and my heart. First, I want to thank God, the universe or whatever entity is out there that looks after me and my family. I want to thank my lovely family. For my happy childhood, for the opportunities I had in my life and for the values that have been instilled in my being. Mom, Luz Alba Macias, and Dad Taylor Dávalos, thank you so much for all your love, your effort, your example and support in my life. Little bro, Michael Dávalos, I am so proud of you and I love you so much. I'm so grateful for having such an intelligent, handsome and wonderful person as a brother. I want to make an special mention to my husband José David Rivera. I am the happiest woman by your side, you are the love of my life and everyday I reaffirm the best decision of having you as

my life partner. You have been not only an excellent support in this process, but also my colleague, my co-adviser, my friend and my psychologist lol. This work couldn't be done without your amazing contributions in all aspects, thank you so much once again my dear lizard. Now, I know for some of you the following may be weird but... I also want to thank my dogs. I have always had such a strong connection with animals, and my dogs have given me a peaceful and fun life. Whenever I was so stressed or sad, my dogs just made all the difference and turned my mood into a positive mindset, they have been a great component of my mental health, so thanks to Copito, Peggy, Lulú and Odin.

I want to specially thank my adviser Dr. Roberto Kopke Salinas. We have met each other way before I start my PhD, and I'm so grateful that I was lucky enough to meet you and work in your Lab. I really appreciate that you gave me the freedom to explore my ideas on my project, that you believed in them (sometimes rsrsrs) and I really wanted to thank you for your guidance in my project, for your contribution, great ideas, support, friendship and furthermore... for helping me so much in writing this thesis in a race against the clock (kkkkrying). I want to thank the group of prof. Roberto Salinas, my colleagues, for their support, the coffee breaks and their friendship and discussions in the lab life: Mario Rodriguez, Layara Abiko, Phelipe Vitale, Duílio Rodrigues, Edwin Neciosup, Made Castro, Andrés Casavilca, Luis Ricardo Pizzolato and Lucas Athayde. I also want to thank to prof. Chuck Farah for all the support and discussions and to his group for all the help I received, for the great talks, the churrascos and friendship. So, thankyou Edgar Llontop, Camila, Ilda, Bruno Matsuyama, Germán Sgro, Germán Dunger, Camila Jaime, Santiago Areválo, Tania Churascari, Rafael Teixera, Ethel Bayer, Gabriel Oka, Natalia Bueno, Diorge Souza, William Cenens and many others that I may forget right now.

I want to thank the people from the Institute of Chemistry (IQ), specially to Professor Sandro Marana. Thanks for the great discussions related to kinetics and for being always so kind to me. Thanks professor Frederico Gueiros for the support, opportunity and discussions. Professora Iolanda Cuccovia, thank you for allowing me to use your lab for my experiments

and for being the wonderful human being you are. A big thankyou and special acknowledge to my super collaborators in this project: Jose David Rivera, Gustavo Batessini, Carol Dutra, Denize Fávaro, Marcus and Ronaldo Junior. Your contributions made possible this work and I am so happy and grateful for that.

My amazing friends, I am so blessed to have you in my life. Tati and Layara, my partners in crime in the lab, I missed you so much and I will always remember our great days together, having fun and supporting each other. Mabel, my marriage Godmother, thank you for all these years of friendship. My Colombian collegue survivor group (that's how we used to call ourselves rsrs) thanks for those wonderful experiences at Universidad del Valle. To my dearest friends and support system in this journey, Juan Camilo Barona, Julian Cualciapud, Alejandro Herrera, Edgar Llontop, Felipe Akihiro, Tatiana Stabelini, Carlos Palechor and Julian Vargas. Thankyou for the laughs, for the drinking rsrs, for being always there for me, for supporting each other, I love you so much guys.

Last but not least, I want to acknowledge the Analytical Instrumentation Center of the University of São Paulo (Central Analítica) for the use of the 800 MHz NMR spectrometer and the Circular Dichroism instrument, and the Multiuser Center for Biomolecular Inovation (CMIB) from the Institute of Biosciences, Humanities and Exact Sciences (IBILCE) at the São Paulo State University "Júlio de Mesquita Filho" (UNESP) for the use of the 600 MHz NMR instrument. I also want to acknowledge the financial support of CNPq with my PhD fellowship (project: 870640/1998-4, process: 141398/2018-3). This work was also financed by a São Paulo Research Foundation research grant (FAPESP 2017/17303-7) and CAPES.

Introduction

Bacteria have developed supramolecular complexes known as secretion systems to transport macromolecules across the bacterial cell envelope. These secretion systems are essential for bacterial growth and survival [1, 2]. Among them, the Type IV Secretion System (T4SS) is used by gram-negative and gram-positive bacteria to transport a wide range of substrates, including proteins, and DNA-protein complexes, to the cell exterior, eventually injecting them into another cell [1, 3]. T4SSs mainly consist of two large subfamilies: (i) conjugation systems that mediate the transfer of DNA between bacteria, and (ii) effector translocators that deliver effector proteins into prokaryotic or eukaryotic cells [4, 5]. The conjugation and effector translocator systems typically require a direct donor-to-target contact. However, a few T4SSs import exogenous DNA or export DNA/proteins to the external milieu in a contact-independent manner [4, 5, 3].

Minimized T4SSs are formed by a set of 11 conserved VirB subunits, and VirD4, according to the nomenclature adopted for the prototypical T4SS from *Agrobacterium tumefaciens*. Other T4SSs, known as expanded systems, are formed by additional subunits that are system-specific. Examples of expanded systems are the Dot/Icm system from *Legionella pneumophila*, and the Cag effector system from *Helicobacter pylori*. Expanded and minimized systems share a minimum set of subunits homologous to VirB1-VirB11 and to VirD4, which are structurally arranged in three large subassemblies: the outer membrane core complex (OMCC), the inner membrane complex (IMC), and a set of ATPases located in the cytoplasm. In addition, those systems that inject effector molecules or DNA into host cells

also contain a pilus.

The VirB/VirD4 T4SS of the phytopathogen *Xanthomonas citri* ($T4SS_{XAC}$) is an effector translocator system that is specialized in the secretion of toxins that kill other gram-negative bacteria [6]. In this manner, the $T4SS_{XAC}$ is part of a warfare arsenal that enables *X. citri* to compete for space and resources with other bacteria [6]. Structural studies of the $T4SS_{XAC}$ by single-particle cryo-electron microscopy (Cryo-EM), showed that the OMCC subassembly is formed by 14 copies of the VirB7/VirB9/VirB10 heterotrimer, encompassing part of the periplasm and making a pore in the outer membrane (OM) [7]. The OMCC forms a cage that may be divided in two portions: the inner layer (I-layer) and outer layer (O-layer) [7]. The O-layer of the cage is composed by the lipoprotein VirB7 and the C-terminal domain of VirB9, while the C-terminal domain of VirB10 covers the interior of the cage and forms a pore in the OM. The OMCC I-layer is formed by the N-terminal domains of VirB9 and VirB10 [7]. The N-terminal domain of VirB10, encompassing the OMCC I-layer, is highly disordered and, hence, it was not observed in the Cryo-EM structure, except for a short α-helix that binds in a cleft between adjacent VirB9 N-terminal domains [7]. This architecture displays the same 14-fold symmetry that was observed for the OMCC from the conjugative plasmid pKM101 Tra-system ($T4SS_{pKM101}$), whose three-dimensional structure was solved by Cryo-EM at low resolution [8, 9]. The structure of the O-layer of the $T4SS_{pKM101}$ OMCC was additionally solved by X-ray crystallography at a resolution of 2.6 Å[10]. A significant difference between the $T4SS_{pKM101}$ and the $T4SS_{XAC}$ is that *X. citri*'s VirB7 subunit contains an additional globular domain, which forms an extra ring layer around the $T4SS_{XAC}$ OMCC [11]. Surprisingly, very few interactions between the O- and I- layers were observed in the $T4SS_{XAC}$ and $T4SS_{pKM101}$ OMCC structures [7, 8], which suggests that the O-layer and the I-layer may display independent motions, only limited by the long linkers connecting the N-terminal and C-terminal domains of VirB9 and VirB10.

The IMC subcomplex is formed by VirB10, which is anchored in the IM via a transmembrane helix [12, 5], the integral membrane proteins VirB3 and VirB6, and by VirB8 that is

also anchored in the IM via a transmembrane helix [4, 5]. IMC subassemblies of minimized or expanded T4SSs have not yet been isolated and structurally characterized. However, the low resolution density map of the conjugative plasmid R388 T4SS ($T4SS_{3-10}$), obtained by negative staining electron microscopy, showed a large asymmetric density corresponding to the IMC. This density was dominated by two hexameric barrels corresponding to the VirB4 ATPase at the cytoplasmic side of the complex [13]. In contrast, in situ visualization of the $T4SS_{pKM101}$ by cryo-electron tomography (Cryo-ET) in the native celular context indicated that the IMC displays 6-fold symmetry, and six VirB4 dimers form a collar at the cytoplasmic side [14]. Visualization of the $T4SS_{pKM101}$ by Cryo-ET, and of the $T4SS_{3-10}$ by negative staining Cryo-EM, showed that the IMC is connected to the OMCC by a thin stalk whose composition is still unclear [12, 5, 2, 13, 8]. These observations are in contrast with the visualization of the F-plasmid conjugative T4SS by Cryo-ET, which showed the presence of a channel $\sim$ 2 nm wide connecting the OMC to IMC [15]. Substrate translocation via the T4SS is powered by the ATPases (VirB4, VirB11 and VirD4) located at the cytoplasmic side of the IMC, that also coordinate the recruitment of substrates. The subunits VirB1, VirB2 and VirB5 are required for pilus-assembly [2, 3].

The structural characterization of the isolated $T4SS_{XAC}$ subunits, VirB7 and VirB9, by X-ray crystallography or NMR spectroscopy, anticipated the main features of the atomic model of the OMCC that was obtained by Cryo-EM [11, 16]. Souza et al. (2011) used NMR spectroscopy to characterize the full length VirB7, and observed that the C-terminal globular domain, whose topology is similar to the N0 domains found in proteins involved in macromolecules transport across the outer membrane of bacteria, is involved in VirB7 self-association [11]. Subsequently, Oliveira et al. (2016) solved the solution NMR structure of the complex formed between a peptide corresponding to the N-terminal motif of VirB7 ($VirB7^{Nt}$) and the C-terminal domain of VirB9 ($VirB9^{Ct}$) [16]. In the bound state, $VirB9^{Ct}$ adopts a well folded conformation that consists of a β sandwich formed by two β sheets, each containing four antiparallel β strands. This work also showed that *X. citri* strains knockout

for VirB7, but harboring a complementation plasmid coding for VirB7 mutants that do not self-associate, or that display the C-terminal domain deleted, do not show antibacterial activity despite the fact that $T4SS_{XAC}$ assembly was not affected. However, VirB7 mutants that do not bind to $VirB9^{Ct}$ impair the assembly and functionality of $T4SS_{XAC}$ [16].

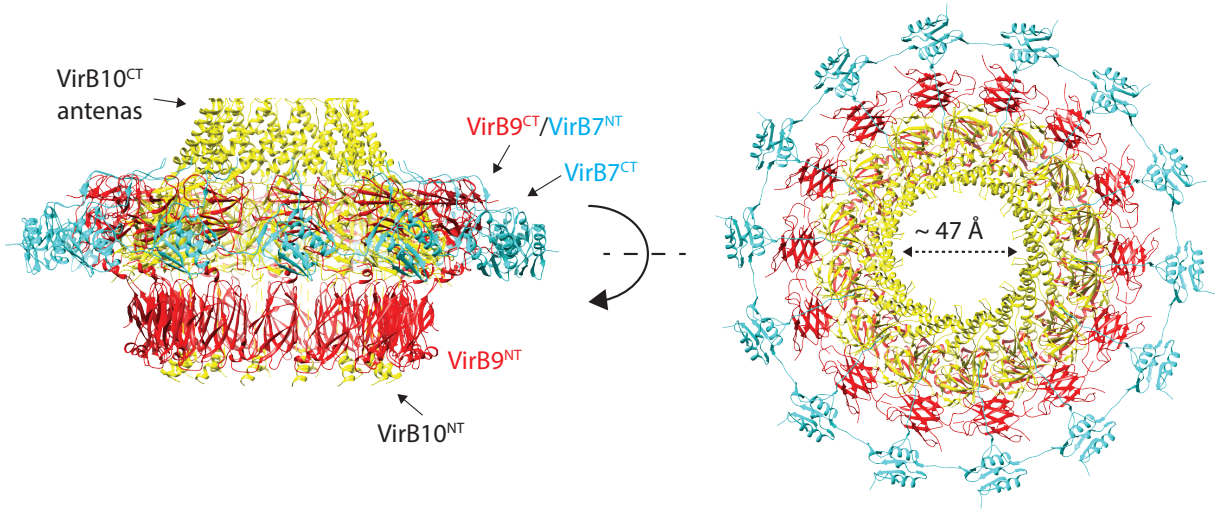


Figure 1: Cryo-EM structure of the XAC T4SS core complex, showing VirB10 in yellow, VirB9 in red and $VirB7^{Nt}$ in cyan.

Souza et al. (2011) observed early on that the ^{1}H-^{15}N HSQC spectrum of $VirB9^{Ct}$ in the free state lacks expected peaks and displays a broad distribution of line widths, which is suggestive of a molten globule [11, 16]. In addition, Souza et al. (2011) reported that $VirB7^{Nt}$ is intrinsically disordered in the free state [11]. The two proteins fold and assume rigid conformations upon binding to each other. $VirB7^{Nt}$ folds into a short β-strand that complements one antiparallel β-sheet of $VirB9^{Ct}$ [16]. These observations indicate that the isolated $VirB9^{Ct}$ and $VirB7^{Nt}$ subunits belong to a particular group of proteins known as Intrinsically Disordered Proteins.

Partially or Intrinsically disordered proteins (IDPs) are characterized by the absence of a defined folded structure. On the contrary to the classical protein structure-function paradigm, IDPs are fully functional despite the absence of a folded conformation. IDPs

may display various degrees of residual secondary structure, or behave as a random coil. When combined with folded domains, they are called intrinsically disordered regions or IDRs [17, 18, 19, 20, 21, 22, 23]. IDPs and IDRs exist as dynamic ensembles of rapid inter-converting conformations in equilibrium. These conformations may be compact (molten-globule like) or extended (random-coil like or pre-molten globule like) [20, 18]. Bioinformatic studies indicated that IDPs and IDRs are frequently found in the proteome of all species, and more abundant in eukaryotes than in prokaryotes [24, 25, 26, 27]. Furthermore, it has been shown that IDPs are involved in crucial processes within the cell, such as signaling and regulatory pathways, post-translational modification, cell cycle regulation, transcription regulation, mRNA processing, apoptosis, among others [28, 29, 30, 31, 20, 32]. It was even proposed that there is a correlation between the amount of intrinsic disorder and the organism complexity, and that perhaps IDPs are used by nature to deal with the increased cell complexity that comes with the appearance of several cellular compartments [20, 26]. However, bacteria also use IDPs/IDRs in a wide array of biological processes including sporulation, translation, catabolic and metabolic processes, pathogenesis (such as the role of intrinsic disorder in effector proteins produced by plant pathogens), and chromosome condensation [27, 33].

In contrast to folded proteins, the majority of IDPs present low sequence complexity, which is related to the over-representation of a few residues that differ from the composition and complexity normally associated to globular proteins [34, 35]. IDPs are often enriched in charged and polar amino acids and depleted in bulky hydrophobic groups [36, 21, 19]. The distribution of charges maintain their flexibility and leads to charge-charge repulsion, while the low hydrophobicity content impairs the driving force for a compact structure [37, 38]. Short tandem repeats of proline, glycine, glutamine, lysine, asparate, arginine, histidine and threonine are commonly found in IDPs and IDRs. Repeats of specific composition have been linked to specific functions, for example transcriptional activation in the case of proline-rich, glutamine-rich and acidic activation domains [21, 36].

The lack of a unique folded conformation allows IDPs to bind to different partners, adopting different conformations upon binding. IDPs often display extended binding interfaces, which may lead to an increase in their association rate constants. This phenomenon is known as "fly-casting" effect, and results advantageous in protein-protein interactions [21, 19, 39]. Although "fuzzy complexes", which retain a high degree of disorder in the bound state, have been described [40, 39], IDPs usually fold into rigid conformations as a consequence of the binding event, a phenomenon known as **folding coupled to binding**. How these interactions are accomplished has been object of discussion. The induced-fit and conformational selection mechanisms have been considered as limiting cases to explain coupled folding and binding processes by IDPs/IDRs. In the induced fit mechanism, the IDP unfolded conformations interact weakly with the binding partner to form an encounter complex which, subsequently, folds into the bound conformation. On the contrary, in the conformational-selection mechanism the IDP samples unfolded and pre-folded conformations, and only the pre-folded conformation is selected for binding. Furthermore, complex mixtures of these two extreme cases are also possible [22, 23, 20, 39, 21, 41].

The development of new spectroscopy and computational techniques to characterize low populated states and ensembles of conformations have allowed a better understanding of IDPs conformational properties and behaviour [42, 39, 43]. In particular, NMR spectroscopy has been extremely useful for the study of IDPs, since it can be used to characterize the conformations and dynamics of the free and bound states, as well as the binding process itself [42, 44, 45, 46, 47]. Molecular dynamics simulations yield atomic-level description of the structure and dynamic properties of IDPs. The development of better force fields and of advanced sampling techniques have allowed the investigation of the dynamics and molecular mechanisms of IDP interactions from computer simulations [48, 49, 43]. On the other hand, kinetic measurements using stopped-flow techniques have proved to be useful to experimentally determine the mechanisms of protein interactions. Despite these advances, the study of IDPs, and particularly of the mechanism of folding coupled to binding, remains

challenging due to the difficulties associated with the characterization of low populated species [39, 20].

Chapter 1

Background theory

1.1 Kinetics of protein-protein interactions

The study of a chemical reaction kinetics is based on the observation of the change in concentration of a particular chemical species by means of a given property, i.e. spectroscopy signal (NMR, fluorescence, absorbance, etc), as a function of time. The reaction kinetic properties are the rate constants, which are key to understand the mechanism of interaction between molecules [50]. In general, kinetic experiments consist of manually mixing reactants solutions, followed by measuring a signal or property change as a consequence of reagent consumption or product formation. This procedure is suitable for slow reactions. In the case of protein-protein interactions rapid methods are necessary to access important kinetic information, which are completed within the first milliseconds of the binding reaction. Rapid mixing methods, such as stopped-flow, were developed to monitor reactions that are fast and completed in few seconds. If the reaction main steps are experimentally accessible to the stopped-flow measurement time, then we will obtain mechanistic information [51]. For instance, we will consider a reaction between a protein molecule and a ligand, forming a stable complex. The change in the concentration of each observed species as a function of

time, $x(t)$, follows the general form:

$$x(t) = A_0 + \sum_{i=1}^{n} A_i \exp(\lambda_1 t), \tag{1.1}$$

where λ corresponds to the n non zero eigenvalues arising from the matrix composed by the differential equations determined by the kinetic model, and $\{A_i\}$ are related to the initial conditions and the eigenvectors. The eigenvalues of the "kinetic matrix", λ_i, are commonly known as "observed rate constants" or k_{obs} [51]:

$$x(t) = A_0 + \sum_{i=1}^{n} A_i \exp(-k_{\text{obs},i}\, t), \tag{1.2}$$

The term "rate constant" makes reference to a proportionality constant that relates the rate to concentration(s), and has a unique value for every particular condition. In many cases k_{obs} is the result of the combination of two or more fundamental rate constants such as k_{on} and k_{off}. But what does the value of k_{obs} really mean? Well, the easiest way to understand this term is to consider its reciprocal $\tau \equiv 1/k$. Here τ is the time constant and provides insight on the timescale of the reaction [50].

1.1.1 One step reaction mechanism

We will now explore the simplest model to describe protein-protein interactions: the lock and key mechanism, first proposed by Emil Fischer in 1894 [52]. In this model, protein-ligand association is achieved by a nearly perfect shape complementarity between the macromolecule and its ligand, such as a lock and its key. Thus, the "lock and key" model implies the assumption that biological macromolecules behave as rigid bodies, rationalizing the association as a consequence of the binding event only (Eqn. 1.3):

$$\mathrm{P} + \mathrm{L} \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} \mathrm{PL}. \tag{1.3}$$

The lock and key model is characterized by a second order rate constant, k_{on}, measured in $\text{s}^{-1}\text{M}^{-1}$, and a first order rate constant, k_{off}, measured in s^{-1}, characterizing the velocity of the forward and reverse chemical reactions in the formation of the complex PL. The equilibrium is reached at the limit $\text{t} \to \infty$ when the ratio:

$$K_{\text{a}} = \frac{k_{\text{on}}}{k_{\text{off}}} = \frac{[\text{PL}]}{[\text{P}][\text{L}]}, \tag{1.4}$$

becomes constant. K_{a} defines the thermodynamic equilibrium association constant under a given set of conditions. Information on K_{a} can be obtained from experiments performed at equilibrium, where it gives information on the extent of the reaction completeness, and may be calculated from the concentrations of reactants and products. Alternatively, K_{a} may be obtained from kinetic experiments as the ratio of the forward and reverse rate constants. The advantage of the kinetic method is that the kinetic constants give information on the reaction mechanism [53].

In order to simplify the math and experimentation, we will assume that the ligand (L) is in large excess over the macromolecule (P) and that its concentration does not change significantly over the course of the reaction. This is the so called "pseudo-first order condition", an approximation used extensively in practice [54]. Under this condition, we can re-write Eq. 1.3 as follows:

$$\text{P} \underset{k_{\text{off}}}{\overset{k_{\text{on}}[\text{L}]}{\rightleftharpoons}} \text{PL}, \tag{1.5}$$

where k_{on} ($\text{M}^{-1}\ \text{s}^{-1}$) is the second-order rate constant for the association of P with L, and k_{off} (s^{-1}) is the first-order rate constant for the dissociation of PL. The set of differential equations associated to this model is [55]:

$$\frac{\text{d}}{\text{d}t}\begin{bmatrix} [\text{P}] \\ [\text{PL}] \end{bmatrix} = \begin{bmatrix} -k_{\text{on}}[\text{L}] & k_{\text{off}} \\ k_{\text{on}}[\text{L}] & -k_{\text{off}} \end{bmatrix} \begin{bmatrix} [\text{P}] \\ [\text{PL}] \end{bmatrix}. \tag{1.6}$$

We can obtain an expression for the experimentally measured k_{obs} by calculating the non-zero eigenvalue from the kinetic matrix (Eq. 1.6) as described in Appendix A. Therefore,

$$-\lambda = \mathrm{k_{obs}} = k_{\text{on}}[L] + k_{\text{off}}. \tag{1.7}$$

As observed in Eq. 1.7, k_{obs} increases linearly with the concentration of L. This is the overall behaviour of the observed rate constant as the reaction approaches the equilibrium. Thus, in a plot of k_{obs} as a function of [L] the slope will be equal to k_{on}, while the intercept will determine k_{off} [55]. Hereafter, we will refer to $\mathrm{k_{obs}}$ of the lock and key model as the "binding exchange rate" or k^b_{ex}.

1.1.2 Binding associated to conformational transitions: Two step reaction mechanism

The Lock and key mechanism assumes that the ligand and its target molecule interact by the complementarity of their shapes, and does not consider any conformational change linked to binding. With the advances of the protein field, it was realized that biological macromolecules are dynamic entities and, hence, dynamics must play an important role in protein-ligand interactions. Conformational transitions linked to binding have an important role in the kinetic mechanism to form the active complex, and they may either follow or precede the binding step [54, 55]. These two limiting cases are discussed below.

In the first case, the conformational change follows the binding step and defines the induced-fit mechanism, first suggested by Koshland in 1958. Koshland postulated that the substrate induces a change in the protein structure upon binding [56]. Later on, the induced-fit theory was developed by Koshland, Némethy and Filmer leading to the sequential KNF model [57]. The reaction scheme for the IF mechanism is:

$$\mathrm{P^*} \underset{k_{\text{off}}}{\overset{k_{\text{on}}[\mathrm{L}]}{\rightleftharpoons}} \mathrm{P^*L} \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} \mathrm{PL} \tag{1.8}$$

Note that the first step in the reaction scheme is identical to the lock and key mechanism, while the second step describes a conformational rearrangement of the P*L complex to a new conformational state PL.

Investigations made on hemoglobin and other metabolic enzymes led to the development of the allostery concept early on in the 1960s [58, 59, 60]. A structural interpretation of the allostery phenomenon was proposed by Wyman, Monod and Changeux in 1965 as the simmetry MWC model. The MWC model explains the homotropic allosteric effect of oxygen binding to hemoglobin, as being a consequence of the existence of discrete states in equilibrium in the absence of the ligand [61]. In this way, the MWC model assumes that the conformational change precedes the binding step, originating the conformational selection (CS) mechanism. The reaction scheme for the CS mechanism is:

$$\mathrm{P}^* \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} \mathrm{P} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}[\mathrm{L}]}{\rightleftharpoons}} \mathrm{PL}. \tag{1.9}$$

In the CS scheme, the first step is a conformational rearrangement between states P* and P that preexist before ligand binding, followed by the lock and key binding step.

Both limiting cases have been demonstrated in protein-ligand interaction studies [62, 63, 64, 65, 66, 67, 68, 69]. Boehr, Nussinov and Wright (2009) used a thermodynamic cycle and the energy landscape theory to explain molecular recognition. Depending on the values of the kinetic rate constants, either the induced fit mechanism, or the conformational selection mechanism may dominate as the kinetic pathway. However, they emphasize that it is likely that both mechanisms play important roles in molecular recognition [70].

Because the IF and CS mechanisms involve two events (reaction schemes 1.8 and 1.9), there are two independent non-zero kinetic "relaxation" processes towards equilibrium, or, in other words, there are two observed rate constants:

$$x(t) = A_0 + A_1 \exp\big(-k^1_{\mathrm{obs}}\, t\big) + A_2 \exp\big(-k^2_{\mathrm{obs}}\, t\big), \tag{1.10}$$

where one may reflect binding and eventually increase linearly with the ligand concentration, while the other may become dominated by conformational transitions and eventually reach saturation at high values of ligand concentration [54]. As note of caution, here the term "relaxation" does not refer to relaxation of spin quantum states, but rather to the decay of chemical species populations towards equilibrium.

In both the CS and the IF schemes, the observed rate associated with binding is defined by Eq. 1.7, while the observed rate constant due to conformational transitions is given by:

$$k_{\mathrm{ex}} = k_{\mathrm{r}} + k_{-\mathrm{r}}, \tag{1.11}$$

where k_{r} and $k_{-\mathrm{r}}$ define the rates of protein conformational exchange between states P* and P or P*L and PL. Note that the rate constant associated with conformational transitions, $\mathrm{k_{ex}}$, is different from the rate constant of the binding equilibrium: $\mathrm{k_{ex}^{b}}$.

Expressions for k_{obs}^{1} and k_{obs}^{2} based on the calculation of the eigenvalues of a three-site chemical exchange kinetic matrix were derived in the appendix B. As it can be seen, k_{obs}^{1} and k_{obs}^{2} contain contributions of both binding and conformational transition events, depending on the four fundamental rate constants. Therefore, it is generally not possible to associate each of them with one or the other event exclusively. However, in the so called *rapid equilibrium approximation*, the binding step may be separated from the conformational transition step.

Given that CS and IF mechanisms feature similar steps but in reverse order, they may display very different kinetics [54]. This becomes clear when one analyzes the limiting values of the observed rate constants. Consider the limit when the ligand concentration decreases ($[\mathrm{L}] \to 0$). In this scenario IF collapses towards a single conformation P*, whereas CS reflects the pre-equilibrium $\mathrm{P}^* \rightleftarrows \mathrm{P}$. In contrast, when $[\mathrm{L}] \to \infty$, IF defines the $\mathrm{P^*L} \rightleftarrows \mathrm{PL}$ equilibrium between the two conformations of the complex, while CS collapses toward the single conformation PL (Fig. 1.1). This implies that IF generates new conformations as a

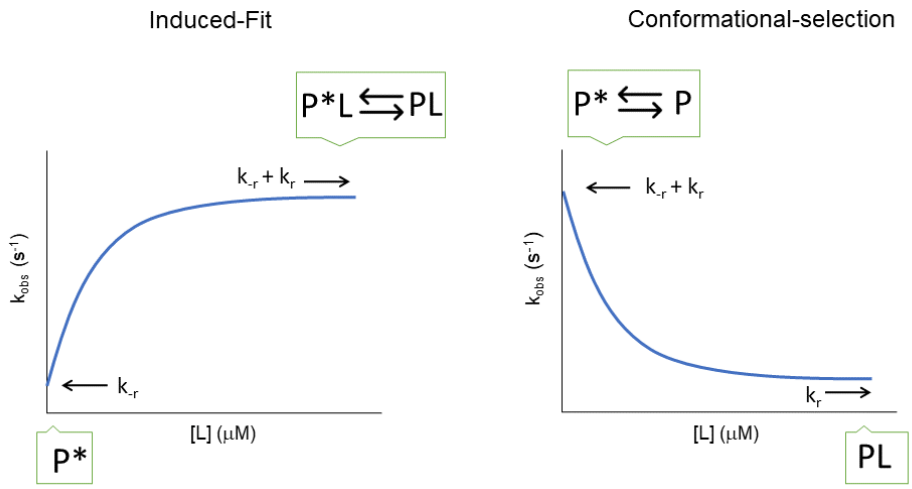


Figure 1.1: Observed rate constants for the Induced-fit and conformational-selection mechanisms under the rapid equilibrium approximation. Inside the green squares are indicated the protein species that will predominantly exist in the limits when [L]→0 and when [L]→ ∞, and the arrows indicate the approaching value of k_{obs} in those limits.

result of the binding event, while CS reduces the number of conformations as a consequence of the binding event[54] (Fig. 1.1). The above is an intuitive analysis about the expected kinetics for schemes 1.8 and 1.9 from simple inspection of the chemical steps involved. Therefore, IF and CS are mutually exclusive mechanisms of recognition and offer distinct interpretations of the linkage between binding and conformational transitions [54].

The rapid equilibrium approximation

The *rapid equilibrium approximation* assumes that the binding event takes place at a significantly faster timescale than the conformational transitions [55]. This condition can be understood by comparing the rates of kinetic relaxation to equilibrium for conformational transitions (Eq. 1.11) and binding (Eq. 1.7). While the observed rate constant remains constant in the former case (Eq. 1.11), it increases linearly with [L] in the latter (Eq. 1.7) [54]. When [L] reaches high values, binding will always be faster than any linked conformational

change. When the dissociation rate k_{off} is faster than the rates associated with conformational transitions, binding will reach equilibrium faster that the linked conformational change and the two relaxations (associated with IF and CS) become separated by widely different time scales. If binding is too fast then it is probably not detected by the stopped flow technique, otherwise if detected it will follow a straight line according to Eq. 1.7. Conformational changes on the contrary will occur over a slower timescale detectable by the stopped flow [54]. Therefore, under the "rapid equilibrium approximation" the kinetics of the system is governed by a single observed rate in most of the cases. This slow kinetic relaxation is saturable as a function of the free ligand concentration.

The rate of the kinetic relaxation to equilibrium for the IF model under the rapid equilibrium approximation is defined by the following expression:

$$\lambda = k_{\text{obs}} = k_{-\text{r}} + k_{\text{r}} \frac{[L]}{K_{\text{d}} + [L]}, \tag{1.12}$$

where the conformational transition on-rate, k_{r}, is weighted by the population of the PL bound species. In contrast, for the CS mechanism, the expression for the rate of kinetic relaxation to equilibrium under the rapid equilibrium approximation is:

$$\lambda = k_{\text{obs}} = k_{\text{r}} + k_{-\text{r}} \frac{K_{\text{d}}}{K_{\text{d}} + [L]}, \tag{1.13}$$

where the rate constant for the reverse conformational transition, $k_{-\text{r}}$, appears multiplied by the population of P^* in the free state.

Therefore, under the rapid equilibrium approximation one may determine k_{r}, $k_{-\text{r}}$ and K_{d} within a single experiment. Note here that k_{on} and k_{off} are not defined in this approximation, but the dissociation equilibrium constant K_{d} is. Another advantage of this simplification is that it is easy to experimentally distinguish the IF and CS mechanisms because the slow kinetic phase displays different saturation properties with L as shown in Fig. 1.1. It follows from Eq. 1.12 that in the IF scheme k_{obs} will increase hyperbolically from $k_{\text{obs}} \approx k_{\text{r}}$ at low

concentrations to a plateau of $k_r + k_{-r}$ at high concentrations of L. In contrast, in the CS scheme, it follows from Eq. 1.13 that k_{obs} will decrease hyperbolically from $k_{obs} \approx k_r + k_{-r}$ at low values of [L] to a plateau of k_r at high [L]. Figure 1.1 shows the expected kinetic signatures of the IF and CS mechanisms under the rapid equilibrium approximation.

Limitations of the rapid equilibrium approximation

Although the rapid equilibrium approximation allows an easy interpretation of the kinetic curves and provides a simplified math for the fit of experimental data, it has limitations and may bias the conclusions about the binding mechanism [55, 54]. In fact, CS can also give rise to relaxations that increase hyperbolically with [L] in conditions where the approximation is no longer valid as will be explained later in this chapter [55]. Moreover, the observation of a single relaxation that saturates with [L] can not be taken as a proof for fast binding and neither as a reason for using the rapid equilibrium approximation to treat experimental data. Furthermore, the absence of a kinetic relaxation can be explained by a rate constant that is too fast to be measured experimentally or by a low amplitude of the spectroscopic signal to be reliably observed [22, 54].

In the case when the rapid equilibrium approximation may not be considered, the observed rate constants for IF and CS will depend on the four fundamental rate constants (k_r, k_{-r}, k_{on} and k_{off}) as described in the Appendix B. When no approximations are made and under pseudo first-order conditions, both mechanism CS and IF are predicted to produce a kinetics that depends on two exponentials functions, each one associated to an observed rate constant (Eq. 1.10). The fast observed rate constant known as k_{fast} ($-\lambda_1$ or k^1_{obs}) eventually increases linearly with [L], and the slow observed rate constant known as k_{slow} ($-\lambda_2$ or k^2_{obs}) that saturates at high values of [L] [55]. Determination of all four independent rate constants is only possible when both the faster and slower relaxations to equilibrium (observed rate constants) are accessed experimentally.

Induced-fit in the general case

When the rapid equilibrium approximation is not considered, the experimental kinetic data is modeled using the analytical solutions for the IF and CS mechanisms. Here I will explain the theory and the implications of these solutions in the general case. Let's start with the IF case represented by scheme 1.8. Under the assumption of pseudo first order conditions, the set of differential equations for the IF model (Eq. 1.14) associated to the scheme 1.8 is:

$$\begin{bmatrix} \mathrm{d[P^*]}/\mathrm{d}t \\ \mathrm{d[P^*L]}/\mathrm{d}t \\ \mathrm{d[PL]}/\mathrm{d}t \end{bmatrix} = \begin{bmatrix} -k_{\mathrm{on}}[\mathrm{L}] & k_{\mathrm{off}} & 0 \\ k_{\mathrm{on}}[\mathrm{L}] & -k_r - k_{\mathrm{off}} & k_{-r} \\ 0 & k_r & -k_{-r} \end{bmatrix} \begin{bmatrix} [\mathrm{P}] \\ [\mathrm{PL}] \\ [\mathrm{PL^*}] \end{bmatrix}. \tag{1.14}$$

The non-zero eigenvalues of the kinetic matrix (Eq.1.14), which corresponds to the observed rate constants, are given by Eq. 1.15 (for the derivation see the Appendix B):

$$-\lambda_{1,2}^{\mathrm{IF}} = \frac{1}{2}\left(k_{-r} + k_r + k_{\mathrm{off}} + k_{\mathrm{on}}[\mathrm{L}] \pm \sqrt{(k_{\mathrm{off}} + k_{\mathrm{on}}[\mathrm{L}] - k_{-r} - k_r)^2 + 4k_r k_{\mathrm{off}}}\right). \tag{1.15}$$

The expression above shows the general form of both rates of relaxation to equilibrium (λ_1 and λ_2) for the IF mechanism. λ_1 is the fast relaxation (the largest value) also known as k_{fast} or $\mathrm{k}^1_{\mathrm{obs}}$, and λ_2 is the slow relaxation (the smallest value) also known as k_{slow} or $\mathrm{k}^2_{\mathrm{obs}}$. These equations imply that in the limit, when $[\mathrm{L}] \to \infty$, $-\lambda_1$ behaves as $k_{\mathrm{on}}[L]$, while $-\lambda_2$ approaches the value of $k_{-\mathrm{r}} + k_{\mathrm{r}}$ as shown in the simulation in Fig. 1.2, [55].

Conformational-selection in the general case

The set of differential equations describing the CS model (scheme 1.9) is:

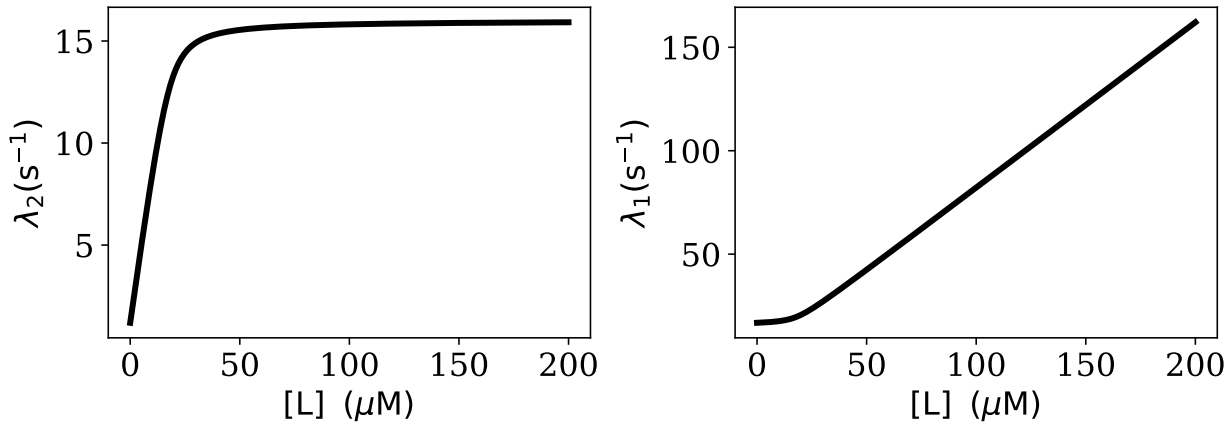


Figure 1.2: Simulation of the observed rate constants ($-\lambda_1$ and $-\lambda_2$) using Eq.1.15 for the IF mechanism. The values for the fundamental rate constants used in this simulation were: $k_{\mathrm{r}} = 6,0\ s^{-1}$, $k_{-\mathrm{r}} = 10,0\ s^{-1}$, $k_{\mathrm{on}} = 0,8\ s^{-1}$ and $k_{\mathrm{off}} = 2\ s^{-1}$.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}[\mathrm{P}^*]\\ [\mathrm{P}]\\ [\mathrm{PL}]\end{bmatrix} = \begin{bmatrix}-k_r & k_{-r} & 0\\ k_r & -k_{-r}-k_{\mathrm{on}}[\mathrm{L}] & k_{\mathrm{off}}\\ 0 & k_{\mathrm{on}}[\mathrm{L}] & -k_{\mathrm{off}}\end{bmatrix}\begin{bmatrix}[\mathrm{P}^*]\\ [\mathrm{P}]\\ [\mathrm{PL}]\end{bmatrix}. \tag{1.16}$$

The observed rate constants, $\lambda_{1,2}$, correspond to the non-zero eigenvalues of Eq. 1.16 as calculated in the Appendix:

$$-\lambda_{1,2}^{\mathrm{CS}} = \frac{1}{2}\left(k_{-r}+k_r+k_{\mathrm{off}}+k_{\mathrm{on}}[\mathrm{L}] \pm \sqrt{(k_{\mathrm{off}}+k_{\mathrm{on}}[\mathrm{L}]-k_{-r}-k_r)^2+4k_{-r}k_{\mathrm{on}}[\mathrm{L}]}\right), \tag{1.17}$$

where $-\lambda_1$ is the largest eigenvalue or k_{fast} ($\mathrm{k}^1_{\mathrm{obs}}$), and $-\lambda_2$ the smallest eigenvalue or k_{slow} ($\mathrm{k}^2_{\mathrm{obs}}$). Note that, in contrast with the IF case (Eq. 1.15), now Eq. 1.17 displays an additional dependence on [L] within the square root (shown in red). This additional dependence on [L] will generate a complex repertoire of kinetic behaviors. When $[\mathrm{L}] \longrightarrow \infty$, $-\lambda_1$ will tend to

$k_{on}[L]$, while $-\lambda_2$ tends to k_r as shown in Fig. 1.3. On the other hand, when [L] $\longrightarrow 0$, the expression in Eq. 1.17 becomes:

$$-\lambda_1^{CS} = \frac{1}{2}\left(k_{-r} + k_r + k_{off} + \mid k_{off} - k_{-r} - k_r \mid\right), \tag{1.18}$$

$$-\lambda_2^{CS} = \frac{1}{2}\left(k_{-r} + k_r + k_{off} - \mid k_{off} - k_{-r} - k_r \mid\right). \tag{1.19}$$

Here there are two alternatives related to the absolute value: if $\mathbf{k_{off} > k_{-r} + k_r}$ ($k_{off} > k_{ex}$) $-\lambda_2$ (k^2_{obs}) will display a hyperbolic decay as shown in Fig. 1.3. In contrast, if $\mathbf{k_{off} < k_{-r} + k_r}$ ($k_{off} < k_{ex}$), $-\lambda_2$ (k^2_{obs}) may exhibit different patterns according to the relative values of k_{off}, k_r and k_{-r} as observed in Fig. 1.4.

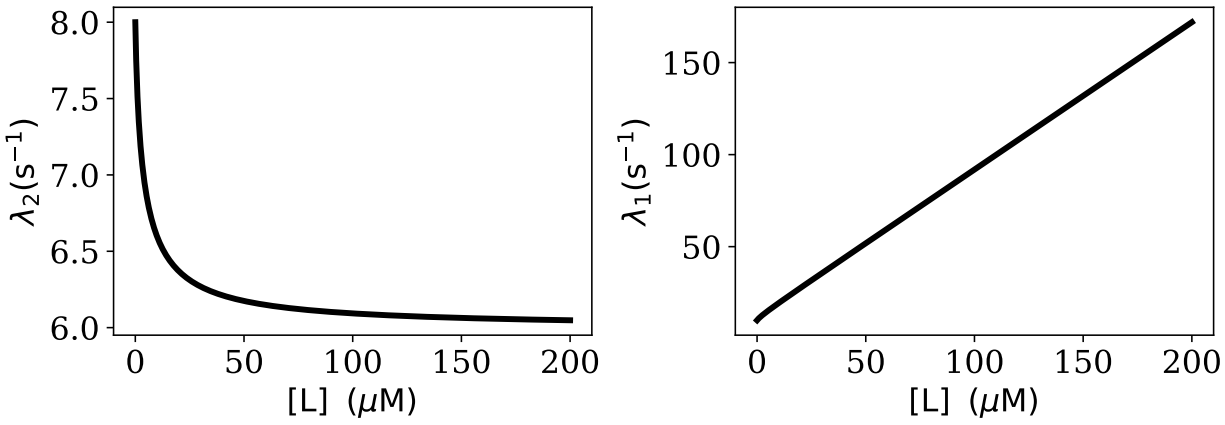


Figure 1.3: Simulation of the observed rate constants ($-\lambda_1$ (k^1_{obs}) and $-\lambda_2$ (k^2_{obs})) using Eq.1.17 for the CS mechanism, considering the case in which $k_{off} > k_r + k_{-r}$. The values for the fundamental rate constants used in this simulation were: $k_r = 6,0\ s^{-1}$, $k_{-r} = 2,0\ s^{-1}$, $k_{on} = 0,8\ s^{-1}$ and $k_{off} = 10,0\ s^{-1}$.

From the above discussion, it is clear that distinguishing between the CS and IF model by observing the kinetic signatures of $-\lambda_2$ (k^2_{obs}) can only be done in certain cases for the CS mechanism. Values of $-\lambda_2$ (k^2_{obs}) that decrease with or are independent from [L] are sufficient evidence to assign CS as the interaction mechanism. In contrast, relaxations that

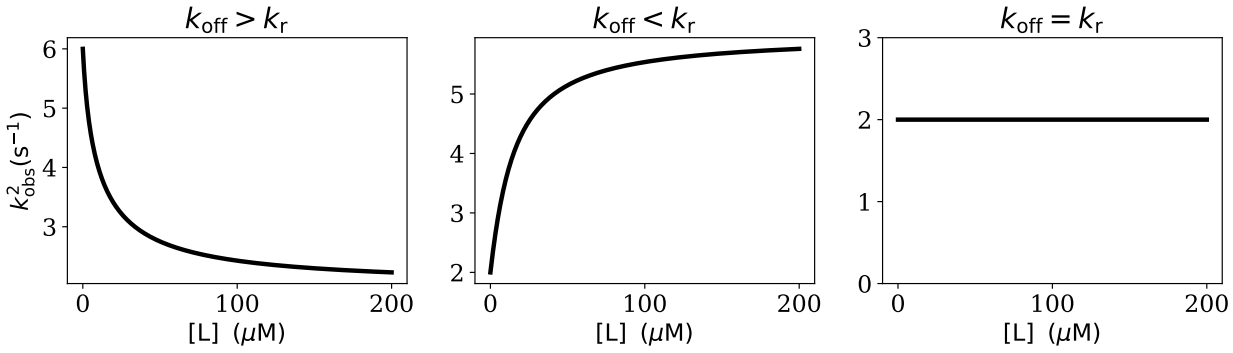


Figure 1.4: Simulations of the observed rate constant $-\lambda_2$ (k^2_{obs}) using Eq.1.17 for the CS mechanism considering the case in which $k_{off} < k_r + k_{-r}$. From left to right, the values of the fundamental rate constants used in these simulations: (left) $k_r = 2,0\ s^{-1}$, $k_{-r} = 10,0\ s^{-1}$, $k_{on} = 0,8\ s^{-1}$ and $k_{off} = 6,0\ s^{-1}$; (middle) $k_r = 6,0\ s^{-1}$, $k_{-r} = 10,0\ s^{-1}$, $k_{on} = 0,8\ s^{-1}$ and $k_{off} = 2,0\ s^{-1}$; (right) $k_r = 2,0\ s^{-1}$, $k_{-r} = 10,0\ s^{-1}$, $k_{on} = 0,8\ s^{-1}$ and $k_{off} = 2,0\ s^{-1}$.

hyperbolically increase with [L] can not distinguish between the CS and IF model, because both models can give rise to values of $-\lambda_2$ (k^2_{obs}) that increase with [L]. Notice that Eq. 1.17 and 1.15 are the same except for the expression colored in red, which shows a dependency on [L] in the CS but not for IF. This difference results in two eigenvalues that always increases with [L] in the IF scenario, in contrast to CS [55]. Table 1.1 shows a concise recapitulation on how $-\lambda_1$ and $-\lambda_2$ varies in its extremes cases when [L]$\rightarrow \infty$ and [L]$\rightarrow$0 in both CS and IF using the general case equations.

How to distinguish between CS and IF

The premise that a slow kinetic relaxation (λ_2, k^2_{obs}) increasing hyperbolically with the ligand concentration is consistent with both CS and IF mechanisms, leads to the need of finding strategies to discriminate between both mechanisms through experimental data. Structural information on the free and bound forms of the macromolecule is of great importance in this matter, since the existence of multiple conformations of the free form supports CS, while

Table 1.1: Kinetic behaviour of the rate constants $-\lambda_1$ (k^1_{obs}) and $-\lambda_2$ (k^2_{obs}), in the general case for the conformational selection and induced fit models. Table taken from [55].

Conformational Selection

$$\frac{1}{2}\left(k_{-\text{r}} + k_{\text{r}} + k_{\text{off}} + k_{\text{on}}[\text{L}] \pm \sqrt{\left(k_{\text{off}} + k_{\text{on}}[\text{L}] - k_{-\text{r}} - k_{\text{r}}\right)^2 + 4k_{-\text{r}}k_{\text{on}}[\text{L}]}\right)$$

$-\lambda_{1,2}$	$\mathbf{k}_{\text{off}} > \mathbf{k}_{-\text{r}} + \mathbf{k}_{\text{r}}$	$\mathbf{k}_{\text{off}} < \mathbf{k}_{-\text{r}} + \mathbf{k}_{\text{r}}$
$-\lambda_1([\text{L}] = 0)$	k_{off}	$k_{-\text{r}} + k_{\text{r}}$
$-\lambda_1([\text{L}] = \infty)$	$k_{\text{on}}[\text{L}]$	$k_{\text{on}}[\text{L}]$
$-\lambda_2([\text{L}] = 0)$	$k_{-\text{r}} + k_{\text{r}}$	k_{off}
$-\lambda_2([\text{L}] = \infty)$	k_{r}	k_{r}
$k^2_{\text{obs}} = \lambda_2$	Always decrease with [L]	Decreases, increases or is independent from [L]

Induce Fit

$$\frac{1}{2}\left(k_{-\text{r}} + k_{\text{r}} + k_{\text{off}} + k_{\text{on}}[\text{L}] \pm \sqrt{\left(k_{\text{off}} + k_{\text{on}}[\text{L}] - k_{-\text{r}} - k_{\text{r}}\right)^2 + 4k_{\text{r}}k_{\text{off}}}\right)$$

$-\lambda_1([\text{L}] = 0)$	$\frac{1}{2}\left(k_{-\text{r}} + k_{\text{r}} + k_{\text{off}} + \sqrt{\left(k_{\text{off}} - k_{-\text{r}} - k_{\text{r}}\right)^2 + 4k_{\text{r}}k_{\text{off}}}\right)$
$-\lambda_1([\text{L}] = \infty)$	$k_{\text{on}}[\text{L}]$
$-\lambda_2([\text{L}] = 0)$	$\frac{1}{2}\left(k_{-\text{r}} + k_{\text{r}} + k_{\text{off}} - \sqrt{\left(k_{\text{off}} - k_{-\text{r}} - k_{\text{r}}\right)^2 + 4k_{\text{r}}k_{\text{off}}}\right) < (k_{-\text{r}} + k_{\text{r}})$
$-\lambda_2([\text{L}] = \infty)$	$k_{-\text{r}} + k_{\text{r}}$
$k^2_{\text{obs}} = \lambda_2$	Always increases with [L]

the existence of multiple conformations of the bound form supports IF. When structural information is not available or inconclusive, the distinction between both mechanisms must be made by proper analysis of the observed rate constants [54]. So far, we focused on describing the kinetic signatures and mathematical expressions for the expected k_{obs} ($-\lambda_{1,2}$) under pseudo-first order conditions with respect to ligand concentration, i.e, under a large excess of ligand over the macromolecule. An interesting alternative approach [71, 72, 54, 63], consists of performing measurements in reverse pseudo first order conditions, in which the macromolecule is in large excess over the ligand. For instance, under pseudo-first order conditions for the ligand ([ligand] $\gg$ [protein]), the IF scheme simplifies to that previously shown 1.8:

$$\mathrm{P}^* \underset{k_{\text{off}}}{\overset{k_{\text{on}}[\mathrm{L}]}{\rightleftharpoons}} \mathrm{P}^*\mathrm{L} \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} \mathrm{PL}, \tag{1.8}$$

while under pseudo-first order conditions for the macromolecule ([protein] $\gg$ [ligand]) we have that:

$$\mathrm{L} \underset{k_{\text{off}}}{\overset{k_{\text{on}}[\mathrm{P}^*]}{\rightleftharpoons}} \mathrm{P}^*\mathrm{L} \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} \mathrm{PL}, \tag{1.20}$$

Therefore, in excess of either the macromolecule or the ligand, the topology of the IF scheme is the same, hence, the analytical solution corresponds to the expression shown in Eq. 1.15, and the behavior of the slow relaxation (λ_2) remains unaltered [63].

Under pseudo-first order conditions with respect to the ligand, the CS scheme simplifies to that previously shown in Eq. 1.9

$$\mathrm{P}^* \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} \mathrm{P} \underset{k_{\text{off}}}{\overset{k_{\text{on}}[\mathrm{L}]}{\rightleftharpoons}} \mathrm{PL}, \tag{1.9}$$

with the corresponding eigenvalues given by Eq. 1.17. On the other hand, if [protein] $\gg$ [ligand], all species regarding the free macromolecule will be in excess, particularly the free protein species P^* and P. Because the protein is now in large excess, the fraction of it in the bound state is very small even at the initial protein concentration [71]. Thus, under a large excess of the macromolecule, binding of the ligand is insufficient to significantly perturb the

protein pre-existing equilibrium [63]. Consequently, the kinetic relaxation associated with the conformational transition is undetectable and the $P^* \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} P$ equilibrium is not observed. As a result, only the fast relaxation λ_1 (k^1_{obs}) associated to ligand binding ($L \underset{k_{off}}{\overset{k_{on}[P]}{\rightleftharpoons}} PL$, in excess of macromolecule) is observed experimentally and the scheme becomes identical to that of the lock and key mechanism (Scheme 1.5) [71, 63, 54]. In summary, a similar kinetics will be observed in excess of ligand and in excess of macromolecule in the IF mechanism. In contrast, the hyperbolic λ_2 (k^2_{obs}) increase in excess of ligand will become a single straight line under pseudo-first order condition with respec to the macromolecule in the CS mechanism [54].

1.1.3 The effect of conformational changes in binding affinities

The interaction between a ligand and its target molecule is of significant importance in biological processes. The success in molecular recognition is mostly defined by the specificity and affinity of the encounter complex [73]. The strength of this interaction is commonly expressed in terms of the intrinsic equilibrium dissociation constant K_d, defined as:

$$K_d \equiv \frac{k_{off}}{k_{on}} = \frac{[P][L]}{[PL]}, \tag{1.21}$$

where [PL] denotes the concentration of the formed complex, [P] and [L] the concentration of the free protein and ligand, respectively. Note that the inverse of Eq. 1.21 is the association equilibrium constant K_a.

The fractional saturation, ν, is the fraction of macromolecules bound to the ligand and is defined by Eq. 1.22.

$$\nu = \frac{[PL]}{[P] + [PL]}. \tag{1.22}$$

Replacing K_a into Eq. 1.22 we obtain:

$$\nu = \frac{K_a[L]}{1 + K_a[L]}, \tag{1.23}$$

which represents the fraction of bound macromolecules as a function of [L] in a system where just binding is taking place. When binding is coupled to conformational transitions the existence of alternative conformations can considerably affect K_{a}, giving rise to an apparent equilibrium association constant $K_{\mathrm{a}}^{\mathrm{app}}$. Therefore, a detailed explanation about $K_{\mathrm{a}}^{\mathrm{app}}$ is developed here for the IF and CS models, and for a combination of these two models.

First consider the IF mechanism (scheme 1.8). The association and conformational exchange equilibrium constants may be defined as:

$$K_{\mathrm{a}} \equiv \frac{k_{\mathrm{on}}}{k_{\mathrm{off}}} = \frac{[\mathrm{P^*L}]}{[\mathrm{P^*}][\mathrm{L}]}, \quad K_{\mathrm{ex}} \equiv \frac{[\mathrm{PL}]}{[\mathrm{P^*L}]} = \frac{k_{\mathrm{r}}}{k_{-\mathrm{r}}}, \tag{1.24}$$

where K_{a} is the association equilibrium constant and K_{ex} is the conformational exchange equilibrium constant for the IF model (do not confuse with k_{ex}, which is the conformational exchange rate constant defined as the sum of k_{r} and $k_{-\mathrm{r}}$). The fractional saturation of the complex ν_{IF} is given by:

$$\nu_{\mathrm{IF}} = \frac{[\mathrm{P^*L}] + [\mathrm{PL}]}{[\mathrm{P}]_{\mathrm{tot}}}, \tag{1.25}$$

where $[\mathrm{P}]_{\mathrm{tot}} = [\mathrm{P}] + [\mathrm{P^*L}] + [\mathrm{PL}]$ is the total protein concentration. Using the expressions from Eq. 1.24 and substituting in 1.25 we get:

$$\nu_{IF} = \frac{K_{\mathrm{a}}[\mathrm{L}] + K_{\mathrm{ex}}K_{\mathrm{a}}[\mathrm{L}]}{1 + K_{\mathrm{a}}[\mathrm{L}] + K_{\mathrm{ex}}K_{\mathrm{a}}[\mathrm{L}]}.$$

Rearranging the expression we arrive at:

$$\nu_{\mathrm{IF}} = \frac{K_{\mathrm{a}}(1 + K_{\mathrm{ex}})[\mathrm{L}]}{1 + K_{\mathrm{a}}(1 + K_{\mathrm{ex}})[\mathrm{L}]},$$

from which we may define the apparent equilibrium association constant for the induced-fit mechanism as:

$$K_{\mathrm{a}}^{\mathrm{app,IF}} \equiv K_{\mathrm{a}}(1 + K_{\mathrm{ex}}). \tag{1.26}$$

Finally, the fractional saturation for the IF mechanism becomes:

$$\nu_{\mathrm{IF}} = \frac{K_{\mathrm{a}}^{\mathrm{app,\,IF}}[\mathrm{L}]}{1 + K_{\mathrm{a}}^{\mathrm{app,\,IF}}[\mathrm{L}]}.$$

Similarly, the $K_{\mathrm{d}}^{\mathrm{app}}$ for the IF mechanism is defined as:

$$K_d^{\mathrm{app,IF}} \equiv \frac{1}{K_a^{\mathrm{app,IF}}} \equiv \frac{K_{\mathrm{d}}}{(1 + K_{\mathrm{ex}})}. \tag{1.27}$$

Note that IF causes an apparent strengthening of the initial interaction with a resulting apparent equilibrium dissociation constant that is lower than the intrinsic K_{d} measured as the ratio $k_{\mathrm{off}}/k_{\mathrm{on}}$. When the interaction occurs via IF, the affinity measured at equilibrium ($K_{\mathrm{d}}^{\mathrm{app}}$) is the result of an optimized fit and always overestimates the affinity of the initial encounter complex [54].

Similarly, using the chemical reactions defined in Eq. 1.9 and considering the equilibrium conditions, we get the following relationships for the CS mechanism:

$$K_{\mathrm{a}} \equiv \frac{k_{\mathrm{on}}}{k_{\mathrm{off}}} = \frac{[\mathrm{PL}]}{[\mathrm{P}][\mathrm{L}]}, \quad K_{\mathrm{ex}} \equiv \frac{[\mathrm{P}]}{[\mathrm{P}^*]} = \frac{k_{\mathrm{r}}}{k_{-\mathrm{r}}}, \tag{1.28}$$

where K_{a} is the association equilibrium constant and K_{ex} is the conformational exchange equilibrium constant for the CS model. Based on the same approach used for IF, we may calculate the fractional saturation ν_{CS}:

$$\nu_{\mathrm{CS}} = \frac{[\mathrm{PL}]}{[\mathrm{P}]_{\mathrm{tot}}}, \tag{1.29}$$

where $[\mathrm{P}]_{\mathrm{tot}} = [\mathrm{P}^*] + [\mathrm{P}] + [\mathrm{PL}]$ is the total protein concentration. Using the expressions from Eq. 1.28, the fractional saturation may be rewritten as:

$$\nu_{\mathrm{CS}} = \frac{K_{\mathrm{a}}[\mathrm{L}]}{1 + K_{\mathrm{ex}}^{-1} + K_{\mathrm{a}}[\mathrm{L}]}.$$

The above equation may be rearranged to define an apparent association equilibrium constant for the CS mechanism:

$$K_{\mathrm{a}}^{\mathrm{app,CS}} \equiv \frac{K_{\mathrm{a}}}{K_{\mathrm{ex}}^{-1}+1}, \tag{1.30}$$

from which the fractional saturation may be formulated:

$$\nu_{\mathrm{CS}} = \frac{K_{\mathrm{a}}^{\mathrm{app,CS}}[\mathrm{L}]}{1+K_{\mathrm{a}}^{\mathrm{app,CS}}[\mathrm{L}]}.$$

Therefore, the apparent equilibrium dissociation constant, $K_{\mathrm{a}}^{\mathrm{app}}$, for CS is defined as:

$$K_{\mathrm{d}}^{\mathrm{app,CS}} \equiv \frac{1}{K_{\mathrm{a}}^{\mathrm{app,CS}}} \equiv (K_{\mathrm{ex}}^{-1}+1)K_{\mathrm{d}}. \tag{1.31}$$

As shown above, the CS mechanism considers that the free protein exists in equilibrium between different conformations, hence, when binding occurs, the affinity measured at equilibrium is the result of ligand binding to a fraction of the total protein population, which always leads to an underestimation of the intrinsic affinity [54].

Finally, we will consider a more complex model that is a combination of CS and IF. This binding mechanism is characterized by a conformational pre-equilibrium of the free protein ($\mathrm{P} \Longleftrightarrow \mathrm{P}^*$) followed by ligand binding with a final conformational rearrangement of the complex:

$$\mathrm{P} \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} \mathrm{P}^* \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}[\mathrm{L}]}{\rightleftharpoons}} \mathrm{C}^* \underset{k_{\mathrm{u}}}{\overset{k_{\mathrm{f}}}{\rightleftharpoons}} \mathrm{C}, \tag{1.32}$$

where, in analogy with a folding/unfolding reaction, k_{f} and k_{u} represent the kinetic rate constants for the IF step after ligand binding. Using the chemical equilibrium equations for the CS-IF model we arrived to the following relationships:

$$[\mathrm{P}] = \frac{\mathrm{k_{-r}}}{\mathrm{k_r}}[\mathrm{P}^*], \quad [\mathrm{C}^*] = \frac{\mathrm{k_{on}}[\mathrm{P}^*][\mathrm{L}]}{\mathrm{k_{off}}}, \quad [\mathrm{C}] = \left(\frac{\mathrm{k_{on}}}{\mathrm{k_{off}}}\right)\left(\frac{\mathrm{k_f}}{\mathrm{k_u}}\right)[\mathrm{P}^*][\mathrm{L}], \tag{1.33}$$

Based on the same approach used before, we may formulate the protein fractional saturation

$\nu_{\mathrm{CS-IF}}$ for this model as:

$$\nu_{\mathrm{CS}} = \frac{[\mathrm{C}^*] + [\mathrm{C}]}{[\mathrm{P}]_{\mathrm{tot}}}, \tag{1.34}$$

where $[\mathrm{P}]_{\mathrm{tot}} = [\mathrm{P}] + [\mathrm{P}^*] + [\mathrm{C}^*] + [\mathrm{C}]$ is the total protein concentration. Using the expressions from Eq. 1.33, the protein fractional saturation may be rewritten as:

$$\nu_{\mathrm{CS-IF}} = \frac{\frac{k_{\mathrm{on}}}{k_{\mathrm{off}}}(1 + \frac{k_{\mathrm{f}}}{k_{\mathrm{u}}})[\mathrm{L}]}{(1 + \frac{k_{-\mathrm{r}}}{k_{\mathrm{r}}}) + \frac{k_{\mathrm{on}}}{k_{\mathrm{off}}}(1 + \frac{k_{\mathrm{f}}}{k_{\mathrm{u}}})[\mathrm{L}]}, \tag{1.35}$$

which in a more compact form becomes:

$$\nu_{\mathrm{CS-IF}} = \frac{K_{\mathrm{a}}^{\mathrm{app,CS-IF}}[\mathrm{L}]}{1 + K_{\mathrm{a}}^{\mathrm{app,CS-IF}}[\mathrm{L}]},$$

where:

$$K_{\mathrm{a}}^{\mathrm{app,CS-IF}} \equiv \frac{\frac{k_{\mathrm{on}}}{k_{\mathrm{off}}}\left(1 + \frac{k_{\mathrm{f}}}{k_{\mathrm{u}}}\right)}{\left(1 + \frac{k_{-\mathrm{r}}}{k_{\mathrm{r}}}\right)}. \tag{1.36}$$

The apparent equilibrium dissociation constant $K_{\mathrm{d}}^{\mathrm{app}}$ for the CS-IF model is given by the inverse of $K_{\mathrm{a}}^{\mathrm{app,CS-IF}}$:

$$K_{\mathrm{d}}^{\mathrm{app,CS-IF}} \equiv \frac{\left(1 + \frac{k_{-\mathrm{r}}}{k_{\mathrm{r}}}\right) K_{\mathrm{d}}}{\left(1 + \frac{k_{\mathrm{f}}}{k_{\mathrm{u}}}\right)}, \tag{1.37}$$

where the ratios of the conformational exchange rate constants $\frac{k_{\mathrm{r}}}{k_{-\mathrm{r}}}$ and $\frac{k_{\mathrm{f}}}{k_{\mathrm{u}}}$ are related to the CS and the IF steps, respectively. Note that when $k_{-\mathrm{r}} \gg k_{\mathrm{r}}$, the CS step will significantly contribute to the value of $K_{\mathrm{d}}^{\mathrm{app,CS-IF}}$, and the apparent dissociation equilibrium constant will be higher than the intrinsic K_{d}. On the other hand, when $k_{\mathrm{f}} \gg k_{\mathrm{u}}$ the ratio $k_{\mathrm{f}}/k_{\mathrm{u}}$ ($K_{\mathrm{ex}}^{\mathrm{IF}}$) will significantly affect the value of $K_{\mathrm{d}}^{\mathrm{app,CS-IF}}$. Therefore, when $K_{\mathrm{ex}}^{\mathrm{IF}} > 0$ then the denominator will be always bigger than 1, leading to an apparent dissociation equilibrium constant $K_{\mathrm{d}}^{\mathrm{app,CS-IF}}$ that is always lower than the intrinsic K_{d}. Therefore, the CS and the IF conformational exchange steps will contribute in opposite ways to the value of the $K_{\mathrm{d}}^{\mathrm{app,CS-IF}}$ with respect to the intrinsic K_{d}.

1.2 NMR spectroscopy

NMR spectroscopy is an attractive technique to study the structure of biomolecules in solution at physiological temperatures, in a native-like environment [74]. NMR spectroscopy is also powerful to investigate protein-protein and protein-ligand interactions. It yields residue-specific information on binding sites, affinities, and exchange constants [39]. Typically, binding is monitored by observing the NMR resonances of an isotopically labeled protein in the presence of unlabeled ligand. The binding reaction induces changes in the chemical environment of spins from residues located at the binding site or undergoing structural rearrangements due to the binding process. These changes in the chemical environment will be reflected in the chemical shifts of NMR resonances and/or their line shapes [39, 75]. In addition to provide structural information, NMR is unique in detecting molecular motions, or dynamics, over a wide range of timescales [74, 75, 47]. Furthermore, NMR has become extremely useful to study intrinsically disordered proteins, as well as sparsely populated and transiently formed states that are otherwise difficult to study by other biophysical techniques [42, 45, 74, 46, 76, 47]. Since NMR spectroscopy was of fundamental importance to the realization of this thesis, I dedicated a specific chapter to the theory of NMR, starting with a description of the technique from a semi classical point of view, and ending with a simple explanation of the most important NMR experiments used in this work.

1.2.1 NMR spectroscopy: A classical view

NMR spectroscopy is similar to other types of spectroscopy, in that light absorption is necessary to promote transitions between different quantum states. In the case of NMR, it is the nuclear spin which is perturbed through radiation absorption that promote transitions from the ground to the excited state. Particularly, NMR exploits the property of nuclear spin angular momentum of atomic nuclei. The nuclear spin angular momentum $\vec{\mathcal{S}}$ is an intrinsic nuclear property, just as mass or charge. The presence of spin angular momentum

and charge creates a nuclear magnetic moment $\vec{\mu}$ described by Eq. 1.38:

$$\vec{\mu} = \gamma \vec{\mathcal{S}}, \tag{1.38}$$

where γ is the gyro-magnetic ratio, a quantity that is characteristic of each nuclei. Thus, we can imagine the nuclear magnetic moment as a magnetic bar.

In general, we will be interested in the z-component of the spin angular momentum, $\mathcal{S}_z$, and in particular in its eigenfunctions and eigenvalues applied to spin one-half nuclei.
In the presence of a static magnetic field applied along the z-direction of the laboratory reference frame, the z-component of the spin angular momentum interacts with the applied external magnetic field. For spin one-half nuclei, the operator $\mathcal{S}_z$ has just two eigenfunctions, $\psi_{+1/2}$ and $\psi_{-1/2}$, that obey the corresponding eigenvalue equations:

$$\mathcal{S}_z\psi_{+1/2} = +\frac{\hbar}{2}\psi_{+1/2}, \quad \mathcal{S}_z\psi_{-1/2} = -\frac{\hbar}{2}\psi_{-1/2}, \tag{1.39}$$

where $+\frac{\hbar}{2}$ and $-\frac{\hbar}{2}$ are the eigenvalues of $\psi_{+1/2}$ and $\psi_{-1/2}$, respectively. Please note that spin states $\psi_{+1/2}$ and $\psi_{-1/2}$ are also called ψ_α and ψ_β, respectively. This new nomenclature will be adopted hereafter.

In the absence of a magnetic field, the nuclear magnetic moment assumes all possible orientations. When an external field B_0 is applied (along the z-axis) the magnetic moments of one-half spins will tend to align with respect to the field and may assume two possible orientations along the z-direction: the magnetic moments oriented in the same direction as the magnetic field will be associated with the $\psi_{+1/2}$ state, while those oriented in the opposite direction to the magnetic field will be associated with the $\psi_{-1/2}$ state. The energy of the $\psi_{+1/2}$ or α state, and of the $\psi_{-1/2}$ or β state, depends on the strength of the interaction between the magnetic field and the nuclear magnetic moments, as given by the scalar product:

$$E = -\vec{\mu} \cdot \vec{B}, \tag{1.40}$$

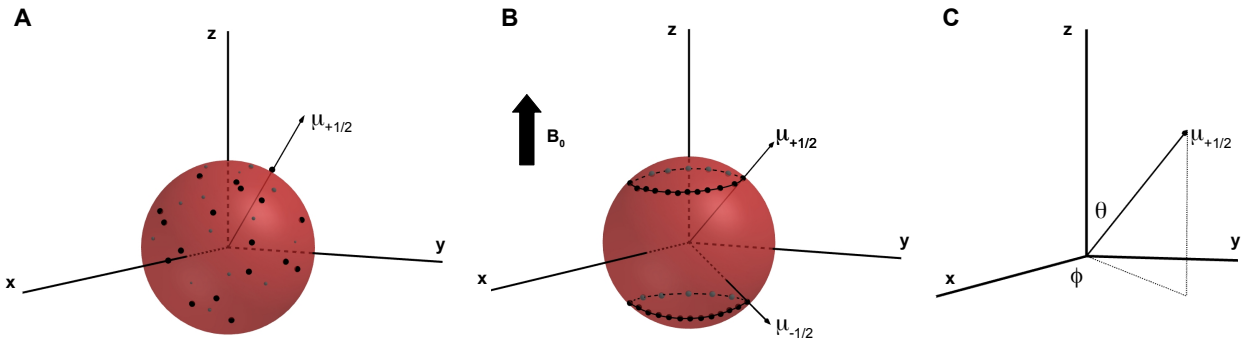


Figure 1.5: Orientation of the spin magnetic moments in the absence (A) and presence (B) of a magnetic field B_0 oriented along z. Orientation of a magnetic moment in spherical coordinates (C).

in which $\vec{B}$ is the magnetic field vector. Since the magnetic field is oriented along the z-axis, and knowing the eigenvalues and eigenvectors of the $\mathcal{S}_z$ operator, we can use use Eqs. (1.38) to (1.40) to obtain the expression:

$$E = -\gamma \hbar m_s B_0, \tag{1.41}$$

where B_0 refers to the static magnetic field along z, γ is the gyromagnetic ratio, an intrinsic property of each nuclei, and m_s is the magnetic quantum number that will be equal to $+\frac{1}{2}$ for the ground state ($\psi_{+1/2}$), and $-\frac{1}{2}$ for the excited state ($\psi_{-1/2}$) of the spin.

The energy difference between these two states can be easily obtained with Eq. (1.41):

$$\Delta E = E_{\psi_{-1/2}} - E_{\psi_{+1/2}} = \gamma \hbar B_0. \tag{1.42}$$

Using the relationship, $E = \hbar\omega$, we may express ΔE in frequency units, the so called Larmor equation:

$$\omega_0 = -\gamma B_0, \tag{1.43}$$

where ω_0 refers to the Larmor frequency of the nucleus. The Larmor equation states that the absorption frequency of a transition is given by the value of γ times the strength of the static magnetic field (1.6). The energy required to promote an NMR transition is low, thus radio-waves are used to excite the spins. At equilibrium, and in the presence of the

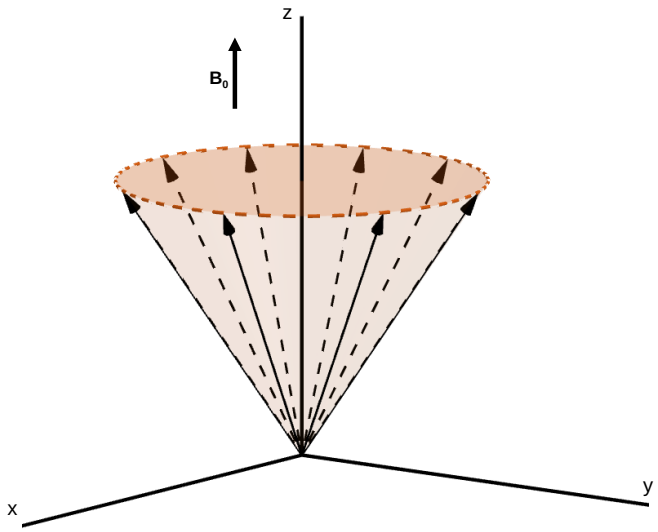


Figure 1.6: Precession of a magnetic moment with larmor frequency ω_0 as result of the presence of the $\vec{B}_0$ static field oriented in the z direction.

static B_0 field, the population in the energy states is not equal. There will always be a little excess of population in the state with the lowest energy since it is the most probable state. This population distribution between states is given by the Boltzmann's distribution law, according to which the ratio of spin populations in the α and β states will be given by:

$$\frac{N_\beta}{N_\alpha} = \exp\left\{\frac{-\gamma\hbar B}{kT}\right\} \approx 1 - \frac{\gamma\hbar B}{kT}. \tag{1.44}$$

Here we assumed that $kT \gg -\gamma\hbar B$ in the above equation.

1.2.2 Before the pulse: Magnetization at equilibrium

When an external magnetic field is applied, there will be a slight excess of magnetic moments oriented towards the direction of the external field (i.e the lowest energy state). This slight excess will lead to a bulk magnetization. This net equilibrium magnetization is oriented along z, and it is often called **longitudinal equilibrium magnetization (Fig. 1.7)**:

$$M_z = M_0. \tag{1.45}$$

Since the static magnetic field was applied along the z-direction, the x and y components of the magnetization **(transverse magnetization)** are randomly distributed in the xy plane, thus the bulk transverse magnetization at equilibrium is zero **(Fig. 1.7)**:

$$M_x = M_y = 0, \tag{1.46}$$

Therefore, at equilibrium the bulk magnetization is represented by a static vector, $\vec{\mathcal{M}}$, which is aligned along the z axis, and is fixed in size and direction, it does nor vary over time [77].

1.2.3 Out of equilibrium

When the equilibrium is perturbed, the bulk magnetization moves away from its original position and, hence, it feels a torque that causes it to precess around the external field at the Larmor angular frequency ω_0 (Fig. 1.8). The torque is a consequence of the interaction between the bulk spin angular momentum (the magnetization) and the external field. It produces a change in the angular momentum, an effect that is analogous to that felt by a spinning top due to the earth gravitational field:

$$\Gamma = \frac{d\mathcal{L}}{dt} = \vec{\mathcal{M}} \times \vec{B}. \tag{1.47}$$

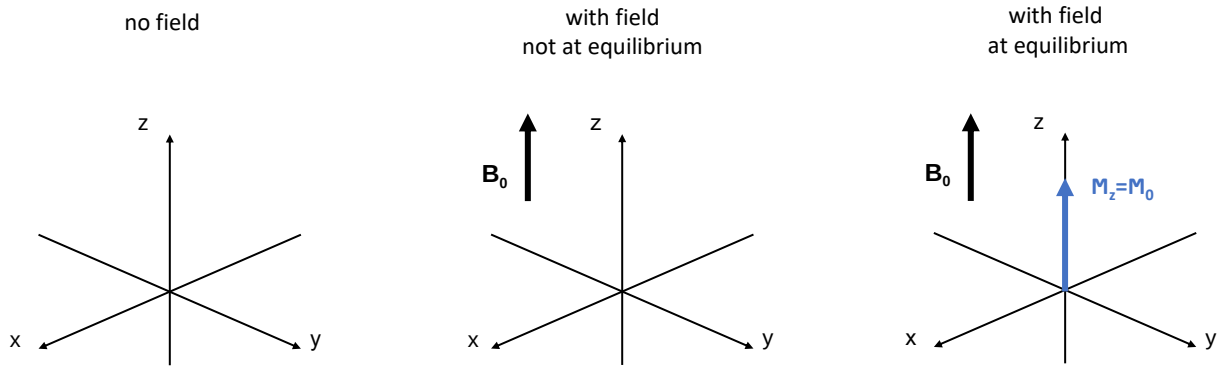


Figure 1.7: In the absence of a magnetic field, the magnetic moments are randomly distributed (left). In the presence of a magnetic field, the sample magnetic moments can adopt two orientations, α and β (center). At equilibrium, the sample has a net magnetization along the magnetic field direction, represented by the vector $\mathcal{M}_0$ (right)

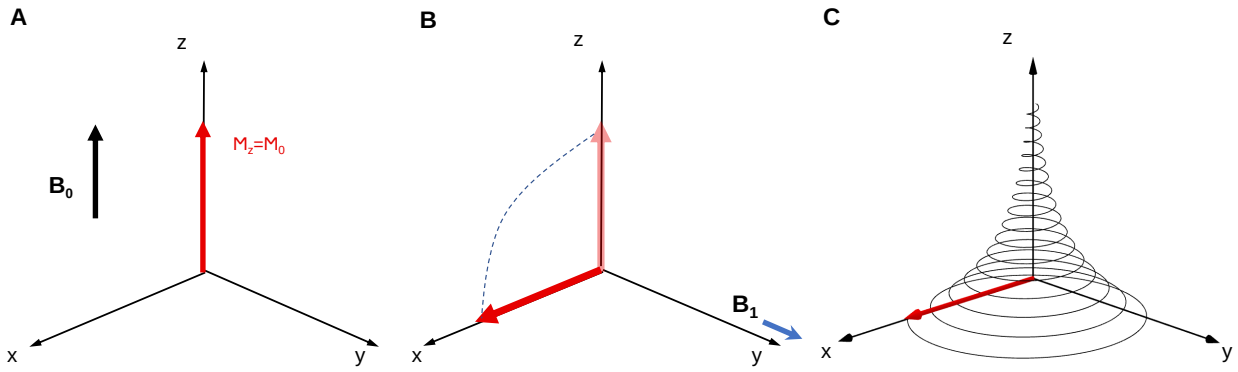


Figure 1.8: The bulk magnetization in presence of the static field B_0 (A) at equilibrium, (B) coming out of equilibrium toward the x-axis due to a pulse on the y-axis, and (C) returning to equilibrium after the pulse is turned off.

Multiplying by γ at both sides of the expression and using, $\vec{\mathcal{M}} = \gamma\vec{\mathcal{L}}$, we obtain:

$$\frac{d\vec{M}}{dt} = \vec{\mathcal{M}} \times \gamma\vec{B}. \tag{1.48}$$

The above equation describes the motion of the macroscopic magnetization in the presence of an external field in the special case when relaxation is ignored. It is a particular case of the so called Larmor equations.

1.2.4 The B_1 pulse: Excitation of nuclear spins

The ultimate goal of an NMR experiment is to measure the precession of the macroscopic magnetization to determine the Larmor angular frequency, for which we need to take the magnetization out of equilibrium. The magnetization is rotated away from its equilibrium position by the application of a weak oscillating magnetic field along the x or y axes. The frequency of oscillation of this transverse magnetic field, ω_{rf}, needs to match the Larmor precession frequency. To facilitate the description and calculations that come with the appearance of this time-dependent RF field, we need to move to a rotating frame of reference. This new reference frame is chosen to rotate around the z axis at angular frequency ω_{rf}. Hence, in the rotating frame, the magnetization angular frequency will be called **offset** and will be represented by Ω, which is equivalent to the difference between ω_0 (Eq. 1.43) (Larmor frequency in the laboratory frame) and ω_{rf} (frequency of the RF field modulation):

$$\Omega = \omega_0 - \omega_{\mathrm{rf}} = -\gamma B_0 - \omega_{\mathrm{rf}}. \tag{1.49}$$

Considering that, in the rotating frame of reference, the magnetization precess around B_0 with the offset angular frequency, then the Larmor equation may be re-written as:

$$\Omega = -\gamma\Delta B. \tag{1.50}$$

Combining Eqs. 1.49 and 1.50 brings us to an expression for the static field in the rotating frame:

$$\Delta B = -\frac{(\omega_0 - \omega_{rf})}{\gamma}. \tag{1.51}$$

It is noteworthy that the "reduced static field" ΔB in the rotating frame will approach zero when the RF frequency matches the Larmor frequency, which is known as **"resonance condition"**. Under this condition, the static magnetic field disappears and the magnetization will precess only around the transverse RF field. The form of the RF field (along x) in the Laboratory frame $\vec{B}_{\text{rf}}$ is:

$$\vec{B}_{\text{rf}} = 2B_1 \cos(\omega_{rf} t + \phi)\hat{\imath}. \tag{1.52}$$

The RF field is commonly named the B_1 field. With some basic math we can re-write the expression for the B_1 field as:

$$\vec{B}_1 = B_1[\cos(\omega_{rf} t + \phi)\hat{\imath} + \sin(\omega_{rf} t + \phi)\hat{\jmath}] + B_1[\cos(\omega_{rf} t + \phi)\hat{\imath} - \sin(\omega_{rf} t + \phi)\hat{\jmath}]. \tag{1.53}$$

The perturbing field B_1 is time independent in the rotating frame. As long as the RF pulse is switched on, there will be two magnetic fields: the reduced field ΔB given by $-\Omega/\gamma$, and the B_1 field. The reduced field ΔB and the B_1 field add vectorially to give an effective field B_{eff} defined as (Fig. 1.9):

$$\vec{B}_{\text{eff}} = B_1 \cos\phi\hat{\imath} + B_1 \sin\phi\hat{\jmath} + \Delta B\hat{k}. \tag{1.54}$$

In the case when the RF field phase angle is $B_x = B_1$, such that it is fixed and oriented along the x axis, and $B_z = B_0$, the magnitude of the effective field is given by

$$B_{\text{eff}} = \sqrt{(B_1)^2 + (\Delta B)^2}. \tag{1.55}$$

About the Effective field and its construction For each field, the precession frequency is proportional to the magnetic field with the gyromagnetic ratio γ as the proportionality constant.

$$\Omega = -\gamma \Delta B \qquad \textbf{Larmor frequency in the rotating frame} \tag{1.56}$$

$$\omega_1 = -\gamma B_1 \qquad \textbf{Precessional frequency about the } B_1 \textbf{ field} \tag{1.57}$$

$$\omega_{\text{eff}} = -\gamma B_{\text{eff}} \qquad \textbf{Precessional frequency about the effective field} \tag{1.58}$$

We can re-write the expression of Eq. (1.55) **in terms of frequencies:**

$$\omega_{\text{eff}} = \sqrt{(\omega_1)^2 + (\Omega)^2}. \tag{1.59}$$

The important idea here is that the magnetization precesses around the effective field in the rotating frame, and that the precessional frequency ω_{eff} is proportional to the effective field strength:

$$\omega_{\text{eff}} = -\gamma B_{\text{eff}}, \tag{1.60}$$

As the frequency of the B_1 field, ω_{rf}, approaches the Larmor frequency ω_0, the value of Ω gets closer to zero (resonance condition). Thus, an on-resonance pulse is all about making $\omega_{rf} = \omega_0$ and consequently $\Omega = 0$. In this case the effective field lies along the x axis. During the on-resonance pulse ($\Omega = 0$) the equilibrium magnetization is rotated from the z axis to the xy plane. The precession frequency is ω_1 and so if the RF field is applied for a time τ_p, the angle β that has been covered while rotating the magnetization towards the xy plane is given by:

$$\beta = \omega_1 \tau_p, \tag{1.61}$$

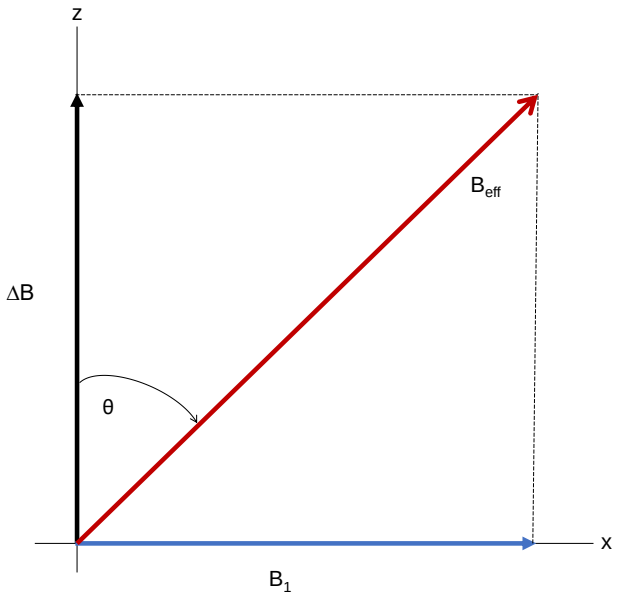


Figure 1.9: The effective field $\vec{B}_{\text{eff}}$ is the vector sum of the reduced field $\Delta\vec{B}$ and the $\vec{B}_1$ field in the rotating frame. The angle between $\Delta\vec{B}$ and $\vec{B}_{\text{eff}}$ is defined as θ.

where β is called the flip angle of the pulse. By altering the duration of the RF pulse, we control the angle through which the magnetization is rotated. The most common flip angles are $\pi/2$ and π. The magnetization rotates around the B_1 field according to the right-hand rule. Therefore, given that the RF (B_1) field is aligned along the x axis, the 90° pulse rotates the equilibrium magnetization from the z to the $-y$ axis. The magnetization ends up along $-y$ because the rotation is in a positive sense about the x-axis. In general, for a flip angle β, the z and y components of the magnetization are:

$$M_z = M_0 \cos\beta, \quad M_y = -M_0 \sin\beta. \tag{1.62}$$

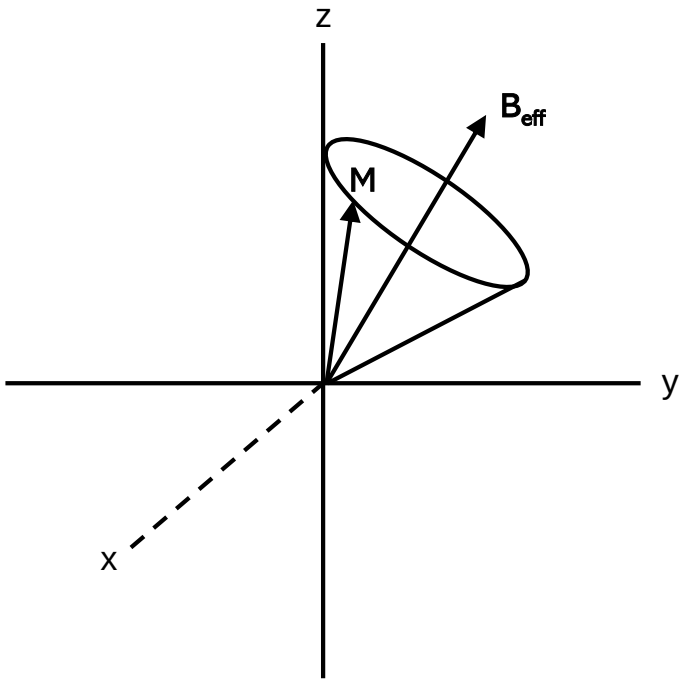


Figure 1.10: Effect of the applied rf field with x-phase. The effective field, $\vec{B}_{\text{eff}}$ is in the y–z plane in the rotating reference frame, and the magnetization vector, $\vec{M}$, precesses around $\vec{B}_{\text{eff}}$.

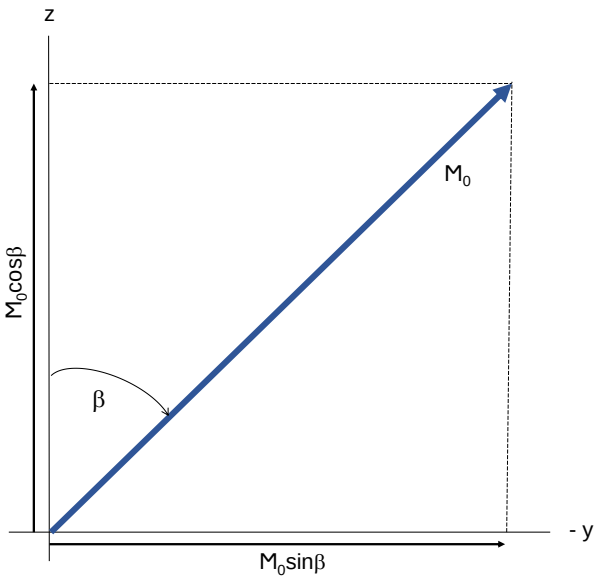


Figure 1.11: During a pulse, the angle β, through which the magnetization has been rotated is defined by Eq. 1.61. With simple geometry we can determine the y- and z-components of the magnetization.

1.2.5 After the pulse: Detection of the NMR signal

After the B_1 pulse is turned off, the magnetization precesses around the static field B_0 at the Larmor angular frequency ω_0 as it relaxes back to its thermal equilibrium position along z. The only difference is that at this time the transverse magnetization is coherent and give rise to a non-zero magnetic moment in the xy plane. This precession of the magnetization vector is what we actually detect in a pulsed NMR experiment. The precession of the coherent magnetization in the xy plane induces a time dependent current in the receiver coil. This signal is called the **free induction decay** (FID), and represents bulk magnetization that exists in the xy plane. Since the magnetization precesses at frequency ω_0, the frequency of the induced signal is exactly the same resonance frequency of the nuclear spin transition. However, for reasons that I will not explain at this moment, the NMR instrument actually detects the offset angular frequency: Ω.

We can visualize what happens to the magnetization while it precess around B_0 by taking

the projections of the magnetization vector on the xy plane. As shown in Fig 1.12, at time zero the x component of this vector is $M_0 \sin\beta$. Then, as the vector precess with angular frequency Ω around the z-axis, this initial x-component rotates at the same frequency in the xy plane. At time t the angle through which the vector has rotated is Ωt. The x and y components are thus proportional to $r\cos(\Omega_0)t$ and $r\sin(\Omega_0)t$, respectively, where $r = M_0 \sin\beta$ is the length of the vector. Thus, neglecting relaxation, M_x and M_y at a given

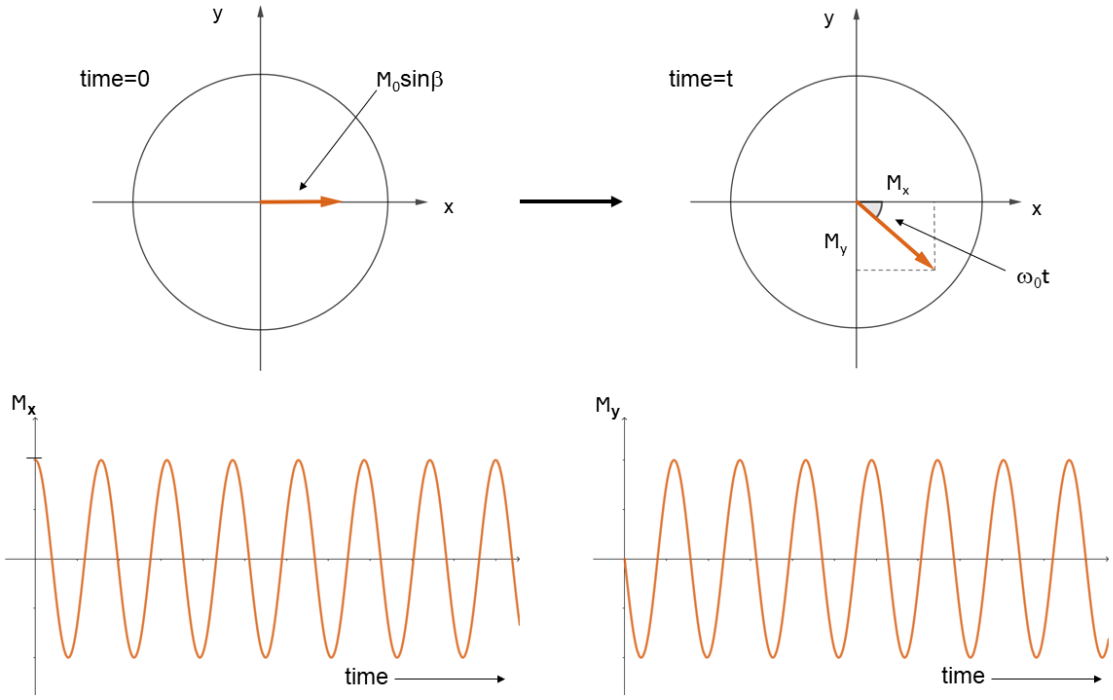


Figure 1.12: Precession of the magnetization vector in the xy-plane. The angle through which the vector has precessed is given by $\omega_0 t$ (Top). The x and y components of the magnetization as a function of time (Bottom). The Fourier transformation of these signals will give rise to the usual spectrum. Taken from [77]

instant t are given by:

$$M_x = M_0 \sin\beta \cos\omega_0 t, \quad M_y = -M_0 \sin\beta \sin\omega_0 t. \tag{1.63}$$

There are two processes that account for the relaxation of the magnetization back to equilibrium. The first relaxation process arises from a re-alignment of the bulk magnetic moment along the static field to regenerate the original Boltzman equilibrium distribution associated with M_z. This relaxation process is known as **spin-lattice** or **longitudinal relaxation**. It involves the transfer of energy from the excited state to the surroundings, and is characterized by a relaxation time, T_1, or by a relaxation rate constant, R_1 ($R_1 = 1/T_1$). The second relaxation process is associated with the dephasing of the coherent transverse magnetization as it precesses around B_0. The time constant for this process is T_2 (or $R_2 = 1/T_2$), and it is known as **spin-spin** or **transverse relaxation**. R_1 and R_2 relaxation are caused by stocastic fluctuations of dipolar coupling interactions between magnetic dipoles ($R_{1,2}^{\mathrm{DD}}$) and the anisotropy in the electron density surrounding the nucleus, in other words, the chemical shift anysotropy (CSA) ($R_{1,2}^{\mathrm{CSA}}$), as molecules move in brownian motion. The loss of coherence of the transverse magnetization will be additionally caused by inhomogenity in the magnetic field: R_2^*.

1.2.6 The Bloch Equation

In this thesis we will use the matrix form of the Bloch-McConnell equation to describe chemical exchange effects in NMR spectroscopy. We will first briefly describe the theory behind the Bloch equation below.

In order to monitor the motion of the bulk magnetic moment in the presence of the static magnetic field B_0, we will use the classical description developed by Bloch [78]. Therefore, Eqn. 1.48 may be expanded as follows:

$$\frac{dM_x}{dt} = \gamma[M_y B_z - M_z B_y], \tag{1.64}$$

$$\frac{dM_y}{dt} = \gamma[M_z B_x - M_x B_z], \tag{1.65}$$

$$\frac{dM_z}{dt} = \gamma[M_x B_y - M_y B_x]. \tag{1.66}$$

In the absence of the applied RF field, the effective field B_{eff} corresponds only to the static magnetic field along z, i.e. $\vec{B}_{\text{eff}} = \Delta B\,\hat{\mathrm{k}}$, therefore, we may write:

$$\frac{dM_x}{dt} = M_y \gamma \Delta B, \tag{1.67}$$
$$\frac{dM_y}{dt} = -M_x \gamma \Delta B, \tag{1.68}$$
$$\frac{dM_z}{dt} = 0. \tag{1.69}$$

Adding the R_1 and R_2 relaxation contributions, and using the definition of the Larmor frequency in the rotating frame (Eq. 1.50), we get:

$$\frac{dM_x}{dt} = -\Omega M_y(t) - R_2 M_x(t), \tag{1.70}$$
$$\frac{dM_y}{dt} = \Omega M_x(t) - R_2 M_y(t), \tag{1.71}$$
$$\frac{dM_z}{dt} = -R_1(M_z(t) - M_0). \tag{1.72}$$

These equations are the famous Bloch equations. They may be written in the more compact matrix form:

$$\frac{d\vec{M}}{dt} = \begin{bmatrix} -R_2 & -\Omega & 0 \\ \Omega & -R_2 & 0 \\ 0 & 0 & -R_1 \end{bmatrix} \vec{M}(t) + R_1 \vec{M}_0, \tag{1.73}$$

in which $\vec{M}(t)$ is:

$$\vec{M}(t) = [M_x(t),\, M_y(t),\, M_z(t)], \tag{1.74}$$

and $\vec{M}_0$ is the equilibrium magnetization. In the presence of an RF field with phase ϕ, the effective field will assume the form given by Eqn. 1.54, and, hence, the Bloch equations

become:

$$\frac{dM_x}{dt} = -\Omega M_y(t) + \omega_1 \sin\phi M_z(t) - R_2 M_x(t), \tag{1.75}$$

$$\frac{dM_x}{dt} = \Omega M_x(t) - \omega_1 \cos\phi M_z(t) - R_2 M_y(t), \tag{1.76}$$

$$\frac{dM_x}{dt} = -\omega_1 \sin\phi M_x(t) + \omega_1 \cos\phi M_y(t) - R_1[M_z(t) - M_0]. \tag{1.77}$$

These equations may also be written in the compact matrix form:

$$\frac{d\vec{M}}{dt} = \begin{bmatrix} -R_2 & -\Omega & \omega_1 \sin\phi \\ \Omega & -R_2 & -\omega_1 \cos\phi \\ -\omega_1 \sin\phi & \omega_1 \cos\phi & -R_1 \end{bmatrix} \vec{M}(t) + \begin{bmatrix} 0 \\ 0 \\ R_1 M_0; \end{bmatrix}, \tag{1.78}$$

1.2.7 The Bloch-McConnell equation

Up to now we have been discussing the behaviour of the magnetization originated from an isolated spin in the presence of the effective field and how we can follow its evolution in time using the Bloch equations. Now we will see a new scenario in which chemical exchange is present. Let's consider a molecule that exists in two different conformations, called states A and B, that are associated to two different chemical environments resulting in different NMR chemical shifts for the same spin:

$$\mathrm{A} \underset{k_\mathrm{B}}{\overset{k_\mathrm{A}}{\rightleftharpoons}} \mathrm{B}, \tag{1.79}$$

where k_A and k_B account for the rate constants of the forward and reverse reactions. This situation is analogous to the lock and key model described in section 2.1.1. Therefore, we will apply the kinetic matrix of the lock and key mechanism (Eq. 1.6), except that the free

and bound protein states will be from now on called "A" and "B", respectively:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} p\mathrm{A} \\ \\ p\mathrm{B} \end{bmatrix} = \begin{bmatrix} -k_\mathrm{A} & k_\mathrm{B} \\ \\ k_\mathrm{A} & -k_\mathrm{B} \end{bmatrix} \begin{bmatrix} p\mathrm{A} \\ \\ p\mathrm{B} \end{bmatrix}, \tag{1.80}$$

where p_A and p_B are the populations of states A and B, respectively, and they satisfy the condition $p_A + p_B = 1$. Assuming that the magnetization of each species is proportional to its population, we may re-write Eq. 1.80 as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \vec{M}^A \\ \\ \vec{M}^B \end{bmatrix} = \begin{bmatrix} -k_\mathrm{A} & k_\mathrm{B} \\ \\ k_\mathrm{A} & -k_\mathrm{B} \end{bmatrix} \begin{bmatrix} \vec{M}^A \\ \\ \vec{M}^B \end{bmatrix}, \tag{1.81}$$

where $\vec{M}^A$ and $\vec{M}^B$ stands for the magnetization of states A and B, respectively. Since $\vec{M}$ is a three-component column vector as described by Eq. 1.74, the above equation is a 6×6 matrix. In a compact form, this matrix may be written as:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \vec{M}^A \\ \\ \vec{M}^B \end{bmatrix} = \begin{bmatrix} -k_\mathrm{A}\,\mathcal{I}_3 & k_\mathrm{B}\,\mathcal{I}_3 \\ \\ k_\mathrm{A}\,\mathcal{I}_3 & -k_\mathrm{B}\,\mathcal{I}_3 \end{bmatrix} \begin{bmatrix} \vec{M}^A \\ \\ \vec{M}^B \end{bmatrix}, \tag{1.82}$$

where I_3 is the 3-dimensional identity matrix. The above matrix represents the two-spins magnetization evolution due to exchange. Summing up the exchange matrix with the Bloch equation (Eq. 1.78) gives rise to the so called Bloch-McConnell equation that describes the time evolution of the magnetization of a system of two spins under chemical exchange in the

presence of an effective magnetic field:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\vec{M}^A\\ \\ \vec{M}^B\end{bmatrix} = \begin{bmatrix}\vec{M}_0^A\\ \\ \vec{M}_0^B\end{bmatrix} + \begin{bmatrix}\mathcal{M}^A & \mathcal{I}_3\, k_B\\ \\ \mathcal{I}_3\, k_A & \mathcal{M}^B\end{bmatrix}\begin{bmatrix}\vec{M}^A\\ \\ \vec{M}^B\end{bmatrix}, \tag{1.83}$$

where

$$\vec{M}_0^i = \begin{bmatrix}0\\ \\ 0\\ \\ R_1^i p_i\end{bmatrix}, \quad \mathcal{M}^i = \begin{bmatrix}-(R_2^i + k_i) & -\Omega_i & \omega_1 \sin(\phi)\\ \\ \Omega_i & -(R_2^i + k_i) & -\omega_1 \cos(\phi)\\ \\ -\omega_1 \sin(\phi) & \omega_1 \cos(\phi) & -(R_1^i + k_i)\end{bmatrix}, \tag{1.84}$$

for $i = \{A, B\}$. Here $\vec{M}^i$ is the magnetization of the i-state, R_j^i is the longitudinal ($j = 1$) or transverse ($j = 2$) relaxation rate of the i-state, Ω_A and Ω_B are the offsets.

We will use the Bloch-McConnell equation (Eq. 1.83) to analyse the CEST (Chemical Exchange Saturation Transfer) experiments to obtain information on exchange parameters, which will be discussed next.

1.2.8 The Chemical Exchange Saturation Transfer (CEST) experiment

NMR has proved to be a powerful technique to study biological macromolecules in solution. Biological macromolecules are intrinsically dynamic, exchanging between different conformational states over a wide range of timescales. Often, these states correspond to different conformations, and, consequently, different chemical environments. Because chemical shifts depend on the unique chemical environment of the nuclei, NMR can be used to study

these conformers. Furthermore, NMR becomes fundamental in systems involving sparsely or transiently populated states, which are otherwise difficult to address by other biophysical techniques [76].

One of the current NMR methodologies to study chemical exchange phenomena in the slow exchange time regime is the Chemical Exchange Saturation Transfer (CEST) experiment [76, 79]. By using the CEST tecnnique, it is possible to amplify the signal of a low populated 'invisible' excited state in exchange with a major ground state of the molecule. Therefore, one may detect a low populated species if it exists in chemical exchange with a 'visible' highly populated state. In the CEST experiment, a weak field B_1 is applied on-resonance to the excited or major state. This perturbation is then transferred to the interconverting states via chemical exchange [76]. The frequency offset of the applied B_1 field is incremented stepwisely. When it reaches the frequency offset of one or the other exchanging partners, it will induce a perturbation leading to a signal saturation if the applied frequency is close (on-resonance) to the offset of one of the exchanging partners, otherwise it will have no effect [79]. In summary, the resonances of the minor invisible excited-state of a molecule can then be visualized through the corresponding signal of the visible and highly populated ground-state due to saturation transfer from one state to the other [76].
Considering a protein that exchanges between two conformations, A and B, where A is the highly populated ground-state conformer and B is the sparsely populated excited-state:

$$\mathrm{A} \underset{k_\mathrm{B}}{\overset{k_\mathrm{A}}{\rightleftharpoons}} \mathrm{B}, \tag{1.85}$$

the population of each species are given by $p_\mathrm{A} = k_\mathrm{BA}/k_\mathrm{ex}$ for state A and $p_\mathrm{B} = k_\mathrm{AB}/k_\mathrm{ex}$ for state B, where $p_\mathrm{A} + p_\mathrm{B} = 1$, and $k_\mathrm{ex} = k_\mathrm{A} + k_\mathrm{B}$ is the exchange rate (Eq. 1.7). As the state A is the most populated state, then $p_\mathrm{A} \gg p_\mathrm{B}$. The resonance frequencies of states A and B are ω_A and ω_B, respectively. The result of this experiment is visualized as a profile of I/I_0 as a function of the saturation frequency, where I and I_0 correspond to the signal intensity in

the presence and absence of saturation, respectively. As $p_{\mathrm{A}} \gg p_{\mathrm{B}}$, a major and intense dip is observed at the chemical shift of state A. If chemical exchange occurs between states A and B, and ω_{A} is sufficiently different from ω_{B}, then we will also observe a minor dip at the chemical shift corresponding to the resonance frequency of state B. The Bloch-McConnell equation is used to fit the CEST profiles to extract important information, such as the value of k_{ex}, the rate of exchange between species A and B, the population and the resonance of the 'invisible' excited-state, p_{B} and ω_{B}, respectively.

In our particular case, we will use the Bloch-McConnell equation to decipher $k_{\mathrm{ex}}^{\mathrm{b}}$, which is the rate of exchange between the VirB7$^{\mathrm{Nt}}$-bound and VirB7$^{\mathrm{Nt}}$-unbound states of VirB9$^{\mathrm{Ct}}$. Here, $k_{\mathrm{ex}}^{\mathrm{b}}$ will be equal to $k_{\mathrm{on}}[\mathrm{L}] + k_{\mathrm{off}}$.

1.3 Protein thermal stability analysis

The thermal stability of a protein makes reference to the protein's ability to resist to thermal denaturation. During the thermal denaturation process the population of the denatured state (S_D) increases, while the population of the native state (S_N) decreases as a function of temperature. At each temperature the sample should be allowed to equilibrate for a certain time period before the measurement is performed. Considering a protein that unfolds without the forming intermediates, the equilibrium between the native and denatured states is written as:

$$N \rightleftharpoons D, \tag{1.86}$$

where N denotes the native state and D the denatured state. The denaturation equilibrium constant is defined by:

$$K = \frac{[D]}{[N]} = \frac{\chi_D}{\chi_N}, \tag{1.87}$$

where [D] and [N] are the concentrations of the protein in the denatured and native states, respectively. Similarly, χ_D and χ_N denotes the mole fractions of the denatured and native

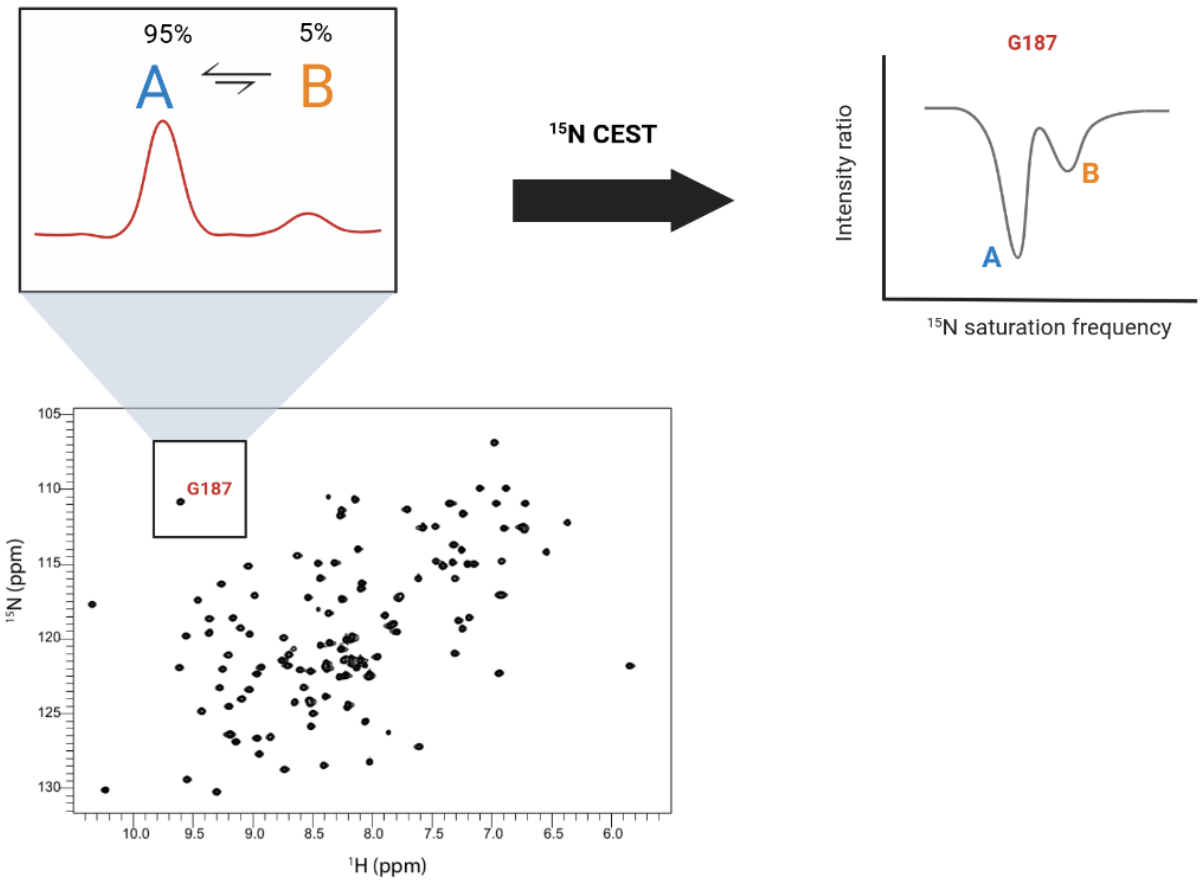


Figure 1.13: Graphic representation of the CEST experiment. In a system where states A and B are under exchange at equilibrium, we can observe the signal of the invisible state B through the saturation transfer to the visible state A, during the CEST experiment. The resulting CEST profile is a plot of the intensity ratio as a function of the saturation frequency. The observed dips in the CEST plot correspond to the most populated state (state A, larger dip) and the least populated state (state B, smaller dip).

states, respectively.

The stability of a protein is quantified in terms of the Gibbs free energy difference between the native and denatured states (ΔG). Under standard conditions, ΔG^o is:

$$\Delta G^o = \Delta H^o - T\Delta S^o, \tag{1.88}$$

where ΔH^o is the enthalpy change of unfolding, T denotes the temperature in Kelvin, and ΔS^o is the entropy change. At equilibrium:

$$\Delta G^o = -RT \ln K, \tag{1.89}$$

where R is the gas constant, and K is the equilibrium constant.

The midpoint of a thermal denaturation transition, T_{m}, is defined as the temperature where the concentration of molecules in the denatured and native states are equal. At this point, $K = 1$ and $\Delta G^o = 0$ according to Eq. 1.87 and Eq. 1.89. Therefore, Eq. 1.88 can be re-written as:

$$\Delta S^0 = \frac{\Delta H^o}{T_{\mathrm{m}}}. \tag{1.90}$$

If we combine Eq. 1.88 and Eq. 1.89 gives the formula:

$$K = \exp\left\{\frac{T\Delta S^0 - \Delta H^o}{RT}\right\}. \tag{1.91}$$

Combining Eq. 1.90 and Eq. 1.91 gives:

$$K = \exp\left\{\frac{\Delta H^o}{R}\left(\frac{1}{T_{\mathrm{m}}} - \frac{1}{T}\right)\right\} = \exp\left\{\frac{-\Delta H^o}{RT}\left(1 - \frac{T}{T_{\mathrm{m}}}\right)\right\}. \tag{1.92}$$

In a protein denaturation experiment that is being monitored by circular dichroism (CD) spectroscopy, the measured CD signal at each temperature is the weighted sum of the spectrum of the native and denatured states. Assuming that there's a linear relationship between

the CD signal for each state, and the temperature, the CD spectrum may be expressed as:

$$\Theta_\lambda(T) = (S_N \times \chi_N) + (S_D \times \chi_D), \tag{1.93}$$

where $\Theta_\lambda(T)$ denotes the total measured signal at a given wavelength as a function of temperature, while S_N and χ_N, S_D and χ_D, denote the CD signal and the protein population in the native and denatured states, respectively. Using Eq. 1.87 and the law of conservation of mass ($\chi_D + \chi_N = 1$) we can obtain the following expression:

$$\chi_D = \frac{K}{K+1}; \;\; \chi_N = \frac{1}{K+1}. \tag{1.94}$$

Replacing Eq. 1.94 and in Eq. 1.93:

$$\Theta_\lambda(T) = S_N \left(\frac{1}{K+1}\right) + S_D \left(\frac{K}{K+1}\right). \tag{1.95}$$

Substituting K using Eq. 1.92 gives the following expression:

$$\Theta_\lambda(T) = S_N \left(\frac{1}{1+\exp\left\{\frac{-\Delta H}{RT}\left(1-\frac{T}{T_\mathrm{m}}\right)\right\}}\right) + S_D \left(\frac{\exp\left\{\frac{-\Delta H}{RT}\left(1-\frac{T}{T_\mathrm{m}}\right)\right\}}{1+\exp\left\{\frac{-\Delta H}{RT}\left(1-\frac{T}{T_\mathrm{m}}\right)\right\}}\right). \tag{1.96}$$

Rearranging the above expression we obtain:

$$\Theta_\lambda(T) = S_N \left(\frac{1}{1+\exp\left\{\frac{-\Delta H}{RT}\left(1-\frac{T}{T_m}\right)\right\}}\right) + S_D \left(\frac{1}{1+\exp\left\{\frac{\Delta H}{RT}\left(1-\frac{T}{T_m}\right)\right\}}\right). \tag{1.97}$$

Eq. 1.97 describes denaturation curves obtained from CD experimental data having as free parameters ΔH, T, S_N and S_D. Thus, the value of the melting temperature, T_m, and the denaturation enthalpy, ΔH, may be obtained from a non-linear fitting of the experimental data to Eq 1.97. Often, the CD denaturation curves display sloping CD baselines at the beginning and/or at the end of the unfolding transitions, where the ellipticity shows a change

that is almost linear as a function of temperature. This effect could be the result of the cooperativity of the system, and may be modeled replacing the S_N and/or S_D signal with a straight line equation as function of temperature, as observed in Eq 1.98 [80, 81, 82].

$$S_{N/D}(T) = mT + b, \tag{1.98}$$

where T is the temperature, m is the slope and b is the intercept in the y axis.

Chapter 2

Methodology

2.1 Expression and purification of VirB9^{Ct}

BL21(DE3) *Escherichia coli* cells harboring the plasmid pET28a containing the gene encoding VirB9^{Ct} (residues 154-255) [16] were grown in either M9 or LB medium supplemented with 50 μg/ml of kanamycin at 37°C. Protein expression was induced by the addition of 0.5 mM of isopropyl-1-thio-β-D-galactopyranoside (IPTG) when the optical density at 600 nm ($O.D_{600}$) reached 0.8, and was carried out overnight at 18°C. Cells were harvested by 1 hour centrifugation at 5000 rpm and 4°C, and lysed by sonication in lysis buffer (20 mM Tris/HCL pH 7.5, 1 mM PMSF, 0.5 mg/mL lysozyme). The lysate was clarified by centrifugation at 13000 rpm on a Beckman centrifuge equipped with rotor SS34. The supernatant was loaded into a Hiprep FF 16/10 SP Sepharose column (GE Healthcare), pre-equilibrated with buffer A (20 mM Tris/HCL pH 7.5), for separation by cation exchange chromatography. VirB9^{Ct} eluted during a linear gradient of NaCl with buffer B (20 mM Tris/HCl pH 7.5 and 700 mM NaCl) over 14 column volumes (CV). Eluted fractions were pooled and loaded into a Ni^{2+}-HiTrap column (GE Healthcare), pre-equilibrated with buffer A (20 mM Tris/HCl pH 7.5, 200 mM NaCl and 20 mM imidazole), for purification by affinity chromatography. VirB9^{Ct} eluted during a linear concentration gradient of imidazole with buffer B (20 mM Tris/HCl pH

7.5, 200 mM NaCl and 500 mM imidazole). The imidazole was removed by buffer exchange to 20 mM Tris/HCl pH 7.5, and VirB9^{Ct} was concentrated to a final volume of 1 ml using an Amicon Ultra Device with cutoff of 3 kDa. Cleavage of the histidines tag was carried out for three hours by incubation with 100-200 μl of thrombin-agarose resin (Sigma-Aldrich) in 50 mM Tris-HCl pH 8.0 containing 10 mM $CaCl_2$. Subsequently, this sample was loaded into the Ni^{2+}-HiTrap column equilibrated with buffer A and cut VirB9^{Ct} was found in the flow-through while the histidines-tag fragment remained tightly bound to the column. The isolated protein buffer was exchanged to 20 mM sodium acetate pH 5.5 containing 50 mM NaCl. Concentrated protein samples were stored at 4°C until further use. Protein concentration was calculated from the absorbance at 280 nm using the molar extinction coefficient $\epsilon = 17420 M^{-1}cm^{-1}$. Isotopically enriched protein samples were prepared as described above, except that the M9 medium contained ^{13}C-glucose and ^{15}N ammonium sulfate, or only ^{15}N ammonium sulfate, as the only sources of carbon and nitrogen, respectively. Stable isotopes were used at concentrations of 0.5 g/l and 2 g/l for ^{15}N and ^{13}C, respectively. Isotopically labeled VirB9^{Ct} samples were used exclusively for NMR experiments as shown in Fig 2.1.

2.2 Expression and purification of VirB9^{Ct}-5OHW

Incorporation of 5-hydroxy L-tryptophan (5OHW) into VirB9^{Ct} was carried out by protein expression in the *E. coli* Trp-auxotropic strain CY(DE3) following a protocol based on the suggestions of Dr.Edgar Llontop and published work by Oliveira-Souza and co-workers [83]. Briefly, *E. coli* CY(DE3) cells harboring the plasmid pET28a-VirB9^{Ct} (residues 154-255) were grown in M9 medium supplemented with 50 μg/ml kanamycin, 34 μg/ml chloramphenicol, L-Trp (4 g/l), and thiamine (10μg/ml), under constant agitation at 37°C. Protein expression was induced by the addition of 0.5 mM of IPTG when the OD_{600} reached 0.6-0.8, followed by incubation at 37°C for 30 minutes to allow the build up of a sufficient amount of T7 RNA polymerase. After 30 minutes of induction, cells were harvested and washed

with M9 medium without any Trp, and centrifugated. The cell pellet was resuspended in M9 medium supplemented with 50 μg/ml kanamycin, 34 μg/ml chloramphenicol, thiamine (10μg/ml), and 8 g/L of 5-hydroxy-tryptophan (5OHW), followed by the addition of 0.5 mM of IPTG, and incubated overnight at 28°C. Cells were harvested by centrifugation at 5000 rpm and lysated by sonication in lysis buffer (40 mM Tris/HCl pH 7.5, 1% Triton X-100, 0.1% SDS, 50 mM Arginine and 50 mM Glutamine) with the previous addition of 1 mM PMSF and lysozyme (0.5 mg/mL). The supernatant from the clarified lysate was loaded into a Ni^{2+}-HiTrap column (GE Healthcare) previously equilibrated with buffer A (50 mM Tris/HCl pH 7.5, 300 mM NaCl and 20 mM imidazole) for separation by affinity cromatography. VirB9^{Ct} eluted during a linear concentration gradient of imidazole using buffer B (50 mM Tris/HCl, 300 mM NaCl and 500 mM imidazole). Fractions containing VirB9^{Ct} were pooled and two-times diluted in 50 mM Tris/HCl pH 7.5 (some precipitate was observed during dilution and the solution became a little blurry). The diluted sample was loaded into a Hiprep FF 16/10 SP Sepharose column (GE Healthcare), pre-equilibrated with 50 mM Tris/HCl pH 7.5 for cation exchange chromatography. The protein was eluted during a linear concentration gradient of 50 mM Tris/HCl pH 7.5 containing 700 mM NaCl. Fractions containing VirB9^{Ct}-5OHW were pooled and concentrated up to 2 ml using an Amicon Ultra device with a cutoff of 3 kDa. Removal of the His-tag, and protein concentration for storage at 4°C was carried out as described above.

2.3 Preparation of the VirB7^{Nt} peptide

A synthetic peptide with the amino acid sequence TKPAPDFGGR WKHVNHFDEAPTE, corresponding to VirB7 residues 24 to 46 (VirB7^{Nt}), was purchased from Biomatik. The peptide is acetylated at the N-terminal end, and amidated at the C-terminal end. Samples of VirB7^{Nt} were prepared by solubilization of the peptide in 20 mM sodium acetate pH 5.5 with 50 mM NaCl, or in 20 mM sodium phosphate pH 7.5 with 10 mM NaCl. Peptide con-

centration was calculated by the absorbance at 280 nm using its molar extinction coefficient (5500 $M^{-1}cm^{-1}$).

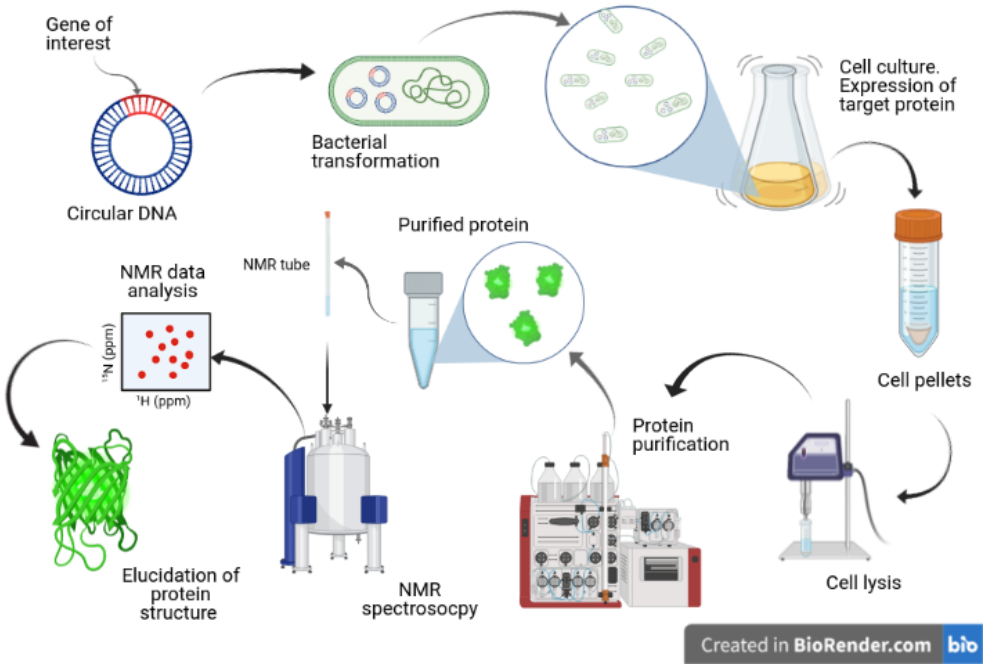


Figure 2.1: General procedure adopted for the expression and purification of the recombinant proteins used for NMR experiments

2.4 NMR measurements

2.4.1 Sample preparation

Samples for NMR measurements were prepared by buffer exchanging $VirB9^{Ct}$ to 20 mM sodium acetate pH 5.5, containing 50 mM NaCl, using an Amicon Ultra device with 3 kDa cutoff. Typical protein concentration was 500μM, and all samples contained 10% of D_2O. $VirB9^{Ct}$:$VirB7^{Nt}$ complexes used for the CEST experiments were prepared at the 1:0.7 and 1:0.8 molar ratios, corresponding to 350 μM and 400 μM nominal $VirB7^{Nt}$ concentrations, respectively.

NMR experiments

NMR measurements were performed on a Bruker AVANCE III spectrometer operating at 800 MHz (^{1}H frequency) and equipped with a TCI cryoprobe. The ^{1}HN-CEST experiment was recorded on a 600 MHz Bruker AVANCE III HD spectrometer, also equipped with a TCI cryoprobe. All NMR spectra were recorded using 5 mm NMR tubes, and at 35°C unless otherwise stated. All NMR data were processed with NMRPipe and analyzed using the CcpNmr Analysis software version 2.4 [84, 85]. Backbone resonance assignments for VirB9^{Ct} in complex with VirB7^{Nt} were obtained from the BMRB entry 25512 [16]. Assignments for VirB9^{Ct} in the unbound state were initially performed by Dr. Denize Favaro, who analysed a set of triple resonance NMR experiments recorded at 7°C . These assignments are found in Appendix C. The ^{15}N ZZ-exchange HSQC experiment [86] was recorded at 35 °C, on a sample of ^{15}N labeled VirB9^{Ct} and unlabeled VirB7^{Nt} at the substoichiometric 1:06 molar ratio (VirB9^{Ct}:VirB7^{Nt}), using mixing times of 400 ms. The experiment pulse sequence was based on the fast-hsqc pulse sequence [87], which was altered to allow the exchange of longitudinal N_z magnetization during the second INEPT transfer period (Fig. 2.2). ^{15}N-CEST experiments were recorded using a Bruker implementation of the experiment described by Vallurupalli and co-workers (2012) [79]. The ^{13}CO CEST experiment was recorded in similar fashion as described by Vallurupalli and Kay (2013) [88]. Briefly, the HNCO pulse sequence was altered to omit the ^{13}CO evolution period, and an exchange period of longitudinal two-spin order $2N_zCO_z$ magnetization was included just before ^{15}N chemical shift evolution (Fig 2.3). The pulse sequence used for the ^{1}HN-CEST experiment was based on the experiment described by Sekhar et al. (2016)[89]. Briefly, an exchange period for longitudinal two-spin order $2H_zN_z$ magnetization was included in the middle of the second INEPT transfer block of the HSQC experiment (Fig 2.4). All CEST experiments were recorded in pseudo-3D fashion: the carrier frequency of the weak $\gamma_i B_1/2\pi$ (where $i =$ ^{1}H, ^{13}C, or ^{15}N) saturation field was incremented stepwisely, followed by the incrementation of the ^{15}N dimension. Calibration of the $\gamma B_1/2\pi$ saturation fields was done automatically,

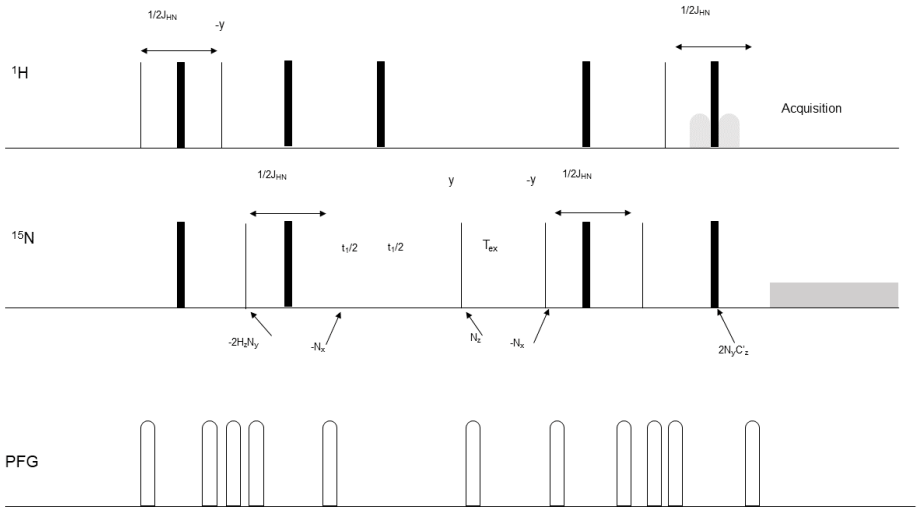


Figure 2.2: Pulse sequence used for the ^{15}N ZZ-exchange experiments. The most relevant elements are indicated. Power levels, delays, and pulse durations are not in scale.

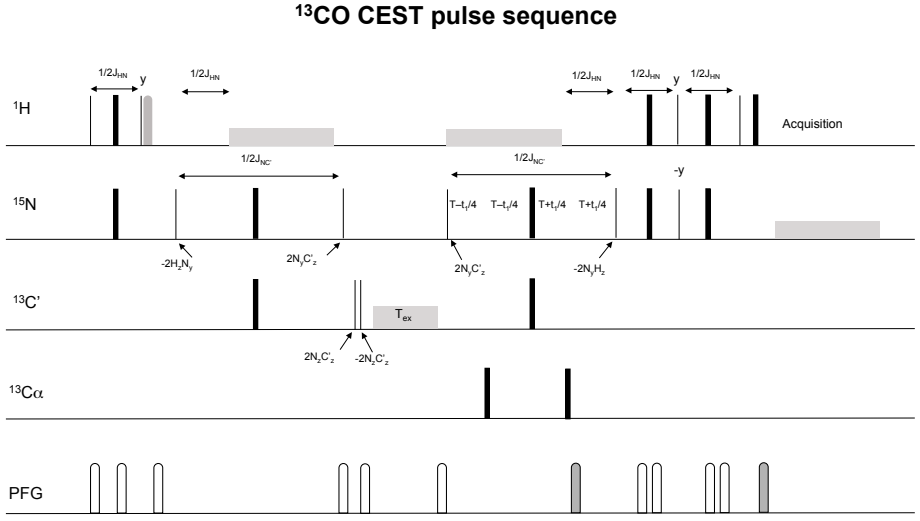


Figure 2.3: Pulse sequence used for the ^{13}CO-CEST experiment. The most relevant elements are indicated. All pulses applied on ^{13}C were selective pulses. Power levels, delays, and pulse durations are not in scale.

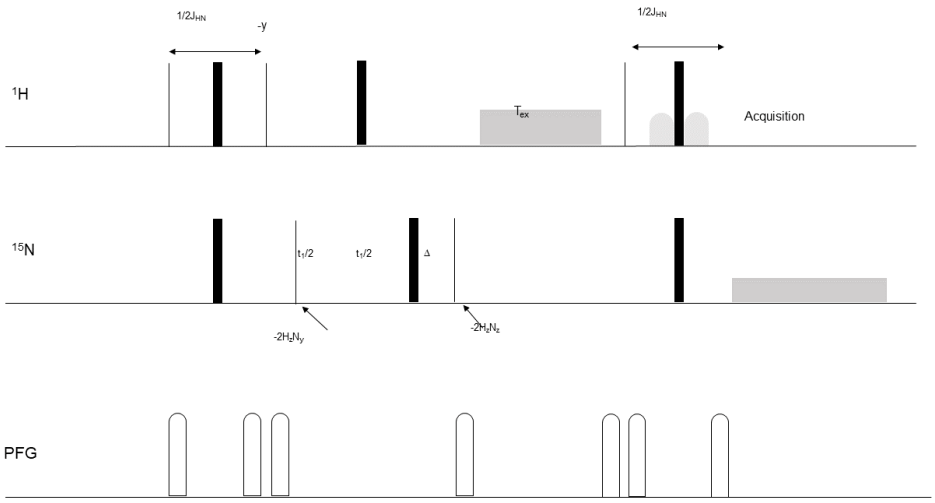


Figure 2.4: Pulse sequence used for the ^{1}HN-CEST experiment.The most relevant elements are indicated. Power levels, delays, and pulse durations are not in scale.

assuming linear amplifiers. Weak saturation fields of $\gamma_{\mathrm{N}}B_1 = 15$ and 30 Hz, $\gamma_C B_1 = 25$ Hz, and $\gamma_H B_1 = 50$ Hz, were used to record the CEST experiments on ^{15}N, ^{13}CO and ^{1}HN nuclei, respectively. The exchange period, T_{ex}, was set to 400 ms in all cases, except for the ^{1}HN-CEST experiment, in which an exchange period of 150 ms was used due to the faster $2\mathrm{H_zN_z}$ relaxation rate. CEST I/I_0 profiles were calculated and visualized using *in-house* scripts made for MATLAB R2015a. ^{15}N-CEST I/I_0 profiles, recorded with $\gamma_{\mathrm{N}}B_1/2\pi = 15$ and 30 Hz, were simultaneously fitted to the Bloch-McConnell equation assuming a two-state exchange model (see Eq. 1.83). The CEST fitting was carried out using Bayesian statistics, as implemented in an in-house python script described by Cardoso and Rivera at al. 2020 [90]. Uncertainties in cross peak intensities were taken from the noise level in each spectrum. The $\gamma_{\mathrm{N}}B_1/2\pi$ fields were constrained using a Gaussian prior with a $\sigma = 1\%$ and centered at 15 Hz or 30 Hz. The bound state chemical shift (Ω_{B}) was restricted using a Gaussian prior with $\sigma = 5\%$ and centered at the experimental ^{15}N chemical shifts obtained from Oliveira et al. (2016). We used a Gaussian prior at the bound VirB9$^{\mathrm{Ct}}$ population (p_{B}), which was centered at $p_{\mathrm{B}} = 0.935$ with $\sigma = 0.015$. This initial guess was estimated

from the ratios between the unbound (A-state) and bound (B-state) peak volumes ($\frac{V_B}{V_A}$) in the ^{15}N CEST spectra, considering that $\frac{p_B}{p_A} \approx \frac{V_B}{V_A}$ and that $p_A + p_B = 1$. Flat priors were used for the free parameters, which were allowed to fluctuate within the following ranges: $1.0 \leq k_{ex}^{b} \leq 150$ s^{-1}, $0.8 \leq p_B \leq 1.0$, $-500 \leq \Delta\Omega \leq 500$ Hz, $0.001 \leq R_A^1 \leq 20$ s^{-1}, $0.001 \leq R_B^1 \leq 20$ s^{-1}, $1 \leq R_A^2 \leq 200$ and $1 \leq R_B^2 \leq 50$ s^{-1}, where A refers to the unbound state, and B to the bound state. MCMC analysis were carried out only for a few fitted residues due to the high computational cost, using 100 chains of 2500 points each, with a burning of 2000 points. Chemical Shift Differences (CSD) between bound and unbound VirB9^{Ct} were calculated using Eq 2.1 [91]:

$$\text{CSD} = \sqrt{\frac{1}{2}[\delta\text{H}^2 + (0.14\delta\text{N}^2)]}, \tag{2.1}$$

where δH and δN are the differences in the ^{1}H and ^{15}N resonances of the bound [16] and unbound VirB9^{Ct} at 35°C.

2.5 Circular Dichroism

Circular Dichroism (CD) spectra were recorded using a Jasco J720 (Jasco, Japan) equipment at 25°C, in 1 mm path-length quartz cells. Far UV-CD spectra were recorded from 190 to 260 nm using integration step of 1 nm, scanning speed of 50 nm/min and accumulation of 6 scans. Samples were prepared at 10 μM protein concentration in 20 mM sodium phosphate pH 7.26 with 10 mM NaCl. A spectrum of the buffer was recorded with the same acquisition parameters as the protein sample and subtracted as blank. The deconvolution of the VirB9^{Ct} and VirB7^{Nt} CD spectra was carried out with the online server of DICHROWEB using the CDSSTR method and dataset # 7, that contains globular and denatured protein spectra [92, 93, 94] . Denaturation experiments were performed in a Jasco J-815 spectrometer in a 1 mm path-length cell. Far UV-CD spectra were recorded as a function of temperature within the 190-260 wavelength range, using integration steps of 1 nm, scanning speed of 100 nm/min

and accumulation of 3 scans. The temperature was varied from 25°C to 95°C in steps of 5°C. Samples of the unbound VirB9$^{\mathrm{Ct}}$ and the 1:1 VirB9$^{\mathrm{Ct}}$-VirB7$^{\mathrm{Nt}}$ complex were prepared at 24-30 μM in 20 mM sodium phosphate pH 7.26 and 10 mM NaCl. Temperature denaturation by CD was also monitored at specific wavelengths at 215 nm for changes in β-type structures and 233 nm for changes in signals arising from exciting coupling effects of aromatic residues. The denaturation curves were fitted using Eq 1.97 and Bayesian statistics as described in section 2.8. We used an in-house python program to model our data to Eq 1.97 and obtain values for the parameters T_{m} and ΔH. The fitted free parameters were ΔH, T_{m}, b and m which are the intercept and slope that describes the signal of the native protein (S_N) shown in Eq 1.98, and S_D which is the signal of the denatured protein, in the denaturation curve. We used 300 chains and 1500 points each in our MCMC analysis, with a burning of 500 points. In addition, we used flat priors for the free parameters, in which $20 \leq \Delta H \leq 100$, $343°\mathrm{C} \leq T_{\mathrm{m}} \leq 363°\mathrm{C}$, $5.0 \leq b \leq 30$, $-0.1 \leq m \leq 0.0$ and $-8 \leq S_D \leq -1$.

2.6 Fluorescence spectroscopy

All fluorescence spectra were recorded using a Hitachi F-7000 FL spectrofluorometer in the laboratory of Prof. Iolanda M. Cuccovia.

2.6.1 ANS Fluorescence

The fluorescence spectra of 1-anilino-8-naphthalenesulfonic acid (ANS) were recorded using using an excitation wavelength of 375 nm, emission wavelength from 400 nm to 650 nm, and temperature of 25°C. A 2 mM ANS stock solution was prepared in 50 mM NaCl and 20 mM sodium acetate buffer (pH 5.5). Stock protein and peptide solutions were prepared in same buffer. The ANS concentration was estimated using the molar extinction coefficient at 350 nm of 4950 M^{-1} cm^{-1} [95]. Samples consisting in 10 μM of protein and/or peptide plus 250 μM of ANS in 50 mM NaCl and 20 mM sodium acetate (pH 5.5), were pre-incubated for 30

minutes before fluorescence measurements.

2.6.2 Trypthophan Fluorescence

The intrinsic tryptophan fluorescence of VirB9^{Ct}, VirB7^{Nt}, and of VirB9^{Ct}:VirB7^{Nt} complexes was recorded using excitation wavelenght at 280 or 295 nm, and emission scan from 280 nm to 450 nm. The concentration of all protein samples was 10 μM in 20 mM sodium phosphate pH 7.5 and 10 mM NaCl. The fluorescence of VirB9^{Ct}-5OHW was monitored with excitation wavelengths at 280, 295, 310 and 320 nm. Samples of VirB9^{Ct}-5OHW were prepared in 20 mM sodium phosphate pH 7.5 and 10 mM NaCl.

2.6.3 Stopped flow kinetics

Kinetic experiments to monitor the interaction between VirB9^{Ct} and VirB7^{Nt} were carried out using a Applied Photophysics stopped-flow spectrofluorometer in the laboratory of Prof. Iolanda M. Cuccovia. The reaction course was followed measuring the Trp fluorescence emission at 350 nm, with excitation wavelenght at 280 nm. Samples were prepared in 20 mM sodium acetate buffer pH 5.5, containing 50 mM of NaCl. The experiments in pseudo-first order conditions were carried out with 1 μM of VirB9^{Ct} and increasing the VirB7^{Nt} peptide concentration from 10 up to 40 μM. The experiments under reverse pseudo-first order conditions were carried out by keeping the concentration of VirB7^{Nt} constant at 1 μM, while varying the concentration of VirB9^{Ct} from 10 up to 40 μM. Each trace was fitted to either single or double exponential functions using the proData software (Applied photophysics) to extract the observed rate constant (k_{obs}) at each concentration. Each k_{obs} corresponds to the average k_{obs} value from 3-5 fitted traces. The behavior of k_{obs} as a function of protein or peptide concentration was fitted to a one-step or to a two-step binding mechanism (Eqs 1.7, 1.15, or 1.17), to extract the values of the fundamental rate constants. To this end, we developed an in-house program written in python that uses Bayesian statistics as described in section 2.8. Briefly, k^1_{obs} and k^2_{obs}, obtained at 25 °C

as a function of [L], were simultaneously fitted to the CS mechanism (discussed in the subsection 3.3.2). Fittings were carried out setting k_{r}, $k_{-\mathrm{r}}$, k_{on} and k_{off} as free parameters. MCMC sampling was carried out with 400 chains of 1000 points each, and a burning of 300 points. In addition, we used flat priors for the free parameters, in which $0.001 \leq k_{\mathrm{r}} \leq 10.0$, $0.001 \leq k_{-\mathrm{r}} \leq 10.0$, $0.001 \leq k_{\mathrm{on}} \leq 10.0$ and $0.001 \leq k_{\mathrm{off}} \leq 10.0$. The kinetic data obtained at 35°C in excess of VirB7$^{\mathrm{Nt}}$ and in excess of VirB9$^{\mathrm{Ct}}$ were used to simultaneously fit the behavior of k_{slow} (k^2_{obs}) and k_{fast} (k^1_{obs}) as a function of $[L]$ to the CS or to the IF mechanism (subsection 3.3.2). We set k_{r}, $k_{-\mathrm{r}}$, k_{on} and k_{off} as free parameters in the CS model, while k_{on}, k_{off}, k_{f} and k_{uf} were set as free parameters in the IF model. MCMC sampling was carried out with 400 chains of 1500 points each, and a burning of 500 points. In addition, we used flat priors for the free parameters, in which $-2.0 \leq k_{\mathrm{r}} \leq 2.0$, $-2.0 \leq k_{-\mathrm{r}} \leq 2.0$, $-2.0 \leq k_{\mathrm{on}} \leq 2.0$, $-2.0 \leq k_{\mathrm{off}} \leq 2.0$, $-1.0 \leq k_{\mathrm{f}} \leq 2.0$ and $-2.0 \leq k_{\mathrm{uf}} \leq 2.0$. The fitting of the combined CS-IF mechanism at 35°C, discussed in the subsection 3.3.2, was carried out by simultaneously fitting k^2_{obs} obtained excess of VirB7$^{\mathrm{Nt}}$ to the CS fast binding approximation (Eq 1.13), as well as k^1_{obs} and k^2_{obs} obtained in excess of VirB9$^{\mathrm{Ct}}$, i.e. reverse pseudo-first order condition, to the IF model (Eq 1.15). We set k_{r}, $k_{-\mathrm{r}}$, k_{on}, k_{off}, k_{f} and k_{uf} as free parameters. The value of $K_{\mathrm{d}}^{\mathrm{app}}$, extracted from ITC at 35°C, was also given as input for the fitting as a Gaussian prior with a value centered at 0.723 μM and a confidence interval of 0.082 μM. MCMC sampling was carried out with 200 chains of 3000 points each, and a burning of 2000 points. In addition, we used flat priors for the free parameters, in which $0.001 \leq k_{\mathrm{r}} \leq 10.0$, $0.001 \leq k_{-\mathrm{r}} \leq 10.0$, $0.001 \leq k_{\mathrm{on}} \leq 10.0$, $0.001 \leq k_{\mathrm{off}} \leq 10.0$, $0.001 \leq k_{\mathrm{f}} \leq 10.0$ and $0.00001 \leq k_{\mathrm{uf}} \leq 2.0$.

2.7 Thermodynamic experiments

2.7.1 Isothermal Titration Calorimetry

Isothermal titration calorimetry experiments were performed on a VP-ITC micro calorimeter (MicroCal incorporated) at the laboratory of Prof. Iolanda M. Cuccovia. Protein and

peptide samples were prepared in the same buffer that consisted in 20 mM sodium acetate pH 5.5 and 50 mM NaCl. All measurements were performed in triplicate, at temperatures of 35°C unless otherwise stated, with a control experiment injecting VirB9$^{\mathrm{Ct}}$ in the cell filled with buffer solution. Titrations of the peptide with the protein were carried out with 20 μM of VirB7$^{\mathrm{Nt}}$ peptide in the calorimeter cell and 200 μM of VirB9$^{\mathrm{Ct}}$ in the syringe using injections of 5 μl. Similarly, measurements for the binding reaction between VirB9$^{\mathrm{Ct}}$-5OHW and VirB7$^{\mathrm{Nt}}$ were performed with 16 μM of VirB7$^{\mathrm{Nt}}$ peptide in the calorimeter cell and 160 μM of VirB9$^{\mathrm{Ct}}$-5OHW in the syringe using injections of 5 μl. The OriginLab7.0 software was used for data analysis using the standard supplier software for K_{d} and ΔH estimation.

A set of ITC measurements of VirB9$^{\mathrm{Ct}}$-VirB7$^{\mathrm{Nt}}$ binding interaction were performed with protein sample prepared at 150μM and peptide prepared at 10, 15 and 20 μM in buffer 20 mM sodium acetate pH 5.5 and 50 mM NaCl. The resulted isotherms, that crossed each other at the same inflexion point, were simultaneously fitted to the one-site binding model to obtain an accurate value of $K_{\mathrm{d}}^{\mathrm{app}}$. For this, we used an in house python program based on the code described by Cardoso and Rivera, et al (2020) [90]. Briefly, the stoichiometry of the complex was constrained at $n = 1$ with a narrow Gaussian prior with $\sigma = 0.01$, while the peptide concentrations were allowed to vary as free parameters. We considered the dissociation constant K_{d}, and the molar binding enthalpy (ΔH) as free parameters. As mentioned by Cardoso and Rivera, et al (2020) [90], we included a set of nuisance parameters involved in the theory, among which are the protein concentration in the syringe [P], the ligand concentration in the cell $[\mathrm{L}]_{\mathrm{tot}}$ and the heat offset Q_0. The calibration of the protein and ligand concentration values were performed in the global fitting of the ITC binding curves. For the MCMC analysis we used 100 chains with 4000 points each, and a burning of 2500 points. In addition, we used flat priors for the free parameters, in which $0.01 \leq K_{\mathrm{d}} \leq 100$ and $-40 \leq \Delta H \leq 2.0$. For more details see the work of Cardoso and Rivera, et al (2020) [90], `https://github.com/jodarie/CalBayS_v1`.

2.7.2 Differential Scanning Calorimetry

Differential scanning calorimetry experiments were performed on a VP-DSC micro-calorimeter (MicroCal incorporated) at the laboratory of Prof. Iolanda M. Cuccovia. All samples were prepared in the same buffer, which consisted in 20 mM sodium acetate pH 5.5 and 50 mM NaCl. The protein concentration of the unbound VirB9$^{\mathrm{Ct}}$, and of the VirB9$^{\mathrm{Ct}}$:VirB7$^{\mathrm{Nt}}$ complex was 300 μM, at the 1:1 molar ratio. All measurements were performed in triplicate and were reproducible. The two protein systems showed some degree of reversibility in the transition, once the system was cooled down after the first run. The denaturation curves were integrated in order to determine the value of ΔH, using the `scipy` funtion `integrate_romberg` that estimates the definite integral using the Romberg's method. We used the Nelder-Mead method as optimization algorithm via the `minimize` function in python.

2.8 Data analysis based on Bayesian inference

The data analysis and in-house python codes used in this thesis were done in collaboration with Dr. Jose David Rivera Echeverri. The parameter estimations of the model given the experimental data were carried out using Bayesian inference. This powerful approach allows us to add prior (extra) information taken from other sources, which improves our predictions. It allows us to determine the probability of a given parameter from the model (posteriori), based on the probability of the data given a theoretical hypothesis (likelihood). In other words, the Bayes theorem tells us that the prior information about the hypothesis is updated according to the knowledge of the observed data via the likelihood function, then

$$\mathcal{P}(H(\mathbf{\Theta})|D) \propto \mathcal{L}(D|H(\mathbf{\Theta})) \times pr(H(\mathbf{\Theta})), \tag{2.2}$$

where $\Theta = (\theta_1, \theta_2, ..., \theta_n)$ is the parameter vector of the model, $\mathcal{P}(H(\mathbf{\Theta})|D)$ is known as the posteriori probability, $H(\mathbf{\Theta})$ is the hypothesis or the model, D represents the observed

data, $\mathcal{L}(D|H(\mathbf{\Theta}))$ is the likelihood function and $pr(H(\mathbf{\Theta}))$ is the prior probability. The posteriori probability is a proportional relationship in which the proportionality constant is called evidence, and it is related to the probability of the observed data. In practice we are interested in constraining the model parameters based on the posteriori distribution, then the proportional factor (i.e. evidence) is independent of the parameters, so it is simply a constant for our purposes.

In order to constrain the model parameters, i.e. to find the parameter values that best fit the model with the observed data, we marginalize the posteriori probability,

$$\mathcal{P}(\theta_i|D) = \int_{\Omega} \mathcal{P}(H(\mathbf{\Theta})|D)\mathrm{d}\theta_1\mathrm{d}\theta_2...\mathrm{d}\theta_{i-1}\mathrm{d}\theta_{i+1}...\mathrm{d}\theta_n, \tag{2.3}$$

so as to obtain the probability distribution function of the parameter θ_i given the observed data, i.e. the marginal distribution, $\mathcal{P}(\theta_i|D)$. The best fit value of θ_i is considered as being the median of the marginal distribution, and the 1σ interval the confidence interval (i.e. the interval defined by the 16th and 84th percentiles), which we assume to be the θ_i uncertainty.

The posteriori probability distribution function was sampled using the Markov Chain Monte Carlo (MCMC) method developed by Daniel Foreman-Mackey, David W. Hogg, Dustin Lang and Jonathan Goodman "emcee: The MCMC Hammer", [96], which was implemented in the `emcee` python package,`https://emcee.readthedocs.io/en/stable/` . In order to optimize the computing time cost, we used the maximum of the posteriori as initial guess in the MCMC algorithm. The maximum of the function was determined using the `differential_evolution` function, a population-based meta-heuristic search algorithm, that optimizes a problem by iteratively improving a candidate solution based on an evolutionary process as described by Storn et al. (1997) [97]. The differential evolution method is implemented in the `scipy` python package, `https://scipy.github.io/devdocs/index.html`.

Chapter 3

Results and discussion

3.1 Structural characterization of unbound VirB9^{Ct}

3.1.1 Secondary and tertiary structures

Previous NMR observations about VirB9^{Ct} suggested that it is only partially folded in the free form, becoming well folded upon binding to the N-terminal tail of VirB7 (VirB7^{Nt}) (residues 24-51) [16] [11]. We wondered whether VirB9^{Ct} could be a molten globule in the unbound state, and whether such molten globule could be structurally characterized in atomic detail. To obtain information about the secondary structure content of VirB9^{Ct} we carried out measurements of CD spectroscopy both in the absence (Fig. 3.1 panel A) and in the presence of VirB7^{Nt} (Fig. 3.1, panel B). The CD spectrum of VirB7^{Nt} indicates that it behaves as a random coil (Fig. 3.2, panel A), which is consistent with the expected for short peptides in aqueous solution, and with previous measurements of backbone ^{15}N relaxation rates made with the full-length VirB7 [11]. Curiously, the unbound VirB9^{Ct} CD spectrum was similar to that of the complex (Fig. 3.1,panel A and B), showing a negative band at 218 nm characteristic of peptide bond transitions associated to β type structures [98], and consistent with the β-sandwich structure of VirB9^{Ct} in the bound state [16]. A second band observed at 232-233 nm with positive ellipticity probably arises from contributions

of aromatic residues, that seems to be significant in proteins with low helical content [99]. Positive CD bands in the 225-230 nm region may result from interactions of aromatic groups (Phe, Tyr, Trp) with neighboring peptide groups or between themselves, giving rise to exciton coupling effects [99, 100, 101, 93]. VirB9^{Ct} contains 1 Trp, 8 Tyr and 6 Phe residues, so its aromatic residue content is significant. Particularly, the Tyr residues are located very close to each other in the VirB9^{Ct} three-dimensional structure, what could perhaps explain the exciton coupling effect observed at 233 nm. We used the software Dichroweb [92, 93] to deconvolute the CD spectra from the unbound VirB9^{Ct} and VirB7^{Nt}, to obtain their content of secondary structure (Fig. 3.2). According to this analysis, VirB7^{Nt} contains 14% of β type content, 58% of disordered structure and 16% of turns. On the other hand, the unbound VirB9^{Ct} contains 37% of β type structure in comparison with the 47% of β type content of the VirB7^{Nt}-bound protein. This result is expected given that VirB9^{Ct} folds and become more stable upon binding to VirB7^{Nt} [16]. Indeed, we did not expect the unbound and bound VirB9^{Ct} forms to present similar proportions of β type content. On the contrary, we though that the unbound VirB9^{Ct} would display a high content of disordered structure and few secondary structure elements, since our initial hypothesis was that it had a molten globule-like structure. Thus, VirB9^{Ct}, both in the free and in the bound states, display similar secondary structure content in terms of β-type structures.

The hydrophobic probe 8-Anilino-1-naphthalenesulfonic acid (ANS) has been extensively used to identify exposed hydrophobic sites in partially folded proteins. Binding of ANS to hydrophobic patches is accompanied by a strong increase in its fluorescence, with a shift of the spectrum towards shorter wavelengths [102, 103]. Therefore, we used the ANS as a probe to compare the amount of solvent exposed hydrophobic surface in the unbound and the bound states of VirB9^{Ct}. Fig. 3.3 shows the fluorescence spectra of ANS acquired in the absence and presence of VirB7^{Nt}, VirB9^{Ct}, and the complex. The addition of VirB7^{Nt} does not affect the ANS fluorescence (Fig. 3.3, panel B). In contrast, addition of VirB9^{Ct} resulted in a small but significant increase in the ANS fluorescence intensity, accompanied

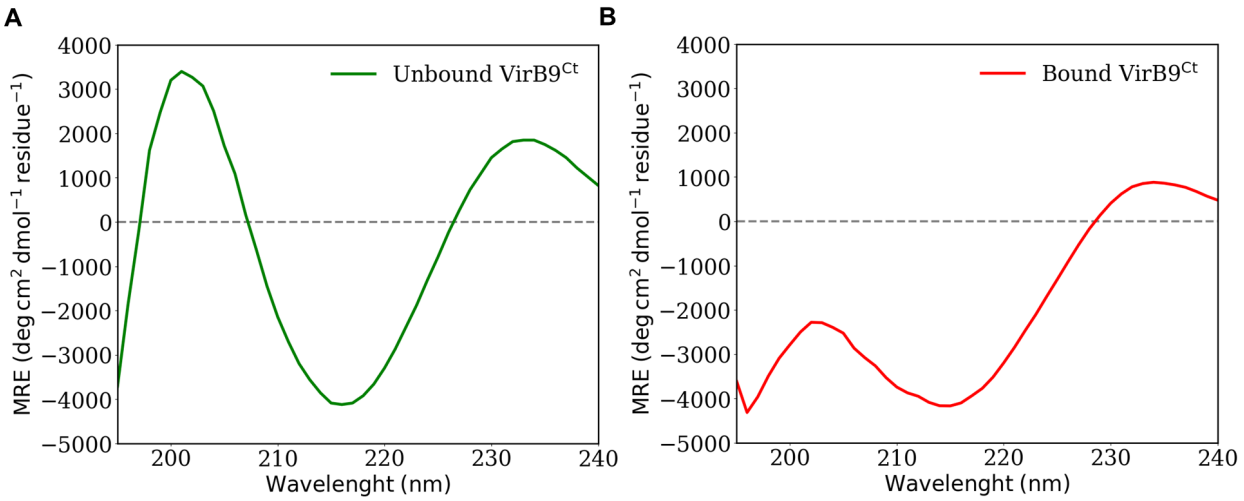


Figure 3.1: CD spectra of (**A**) the unbound VirB9^{Ct} and (**B**) the VirB9^{Ct}-VirB7^{Nt} complex. As a note of caution, CD signals below 200 nm are low resolved

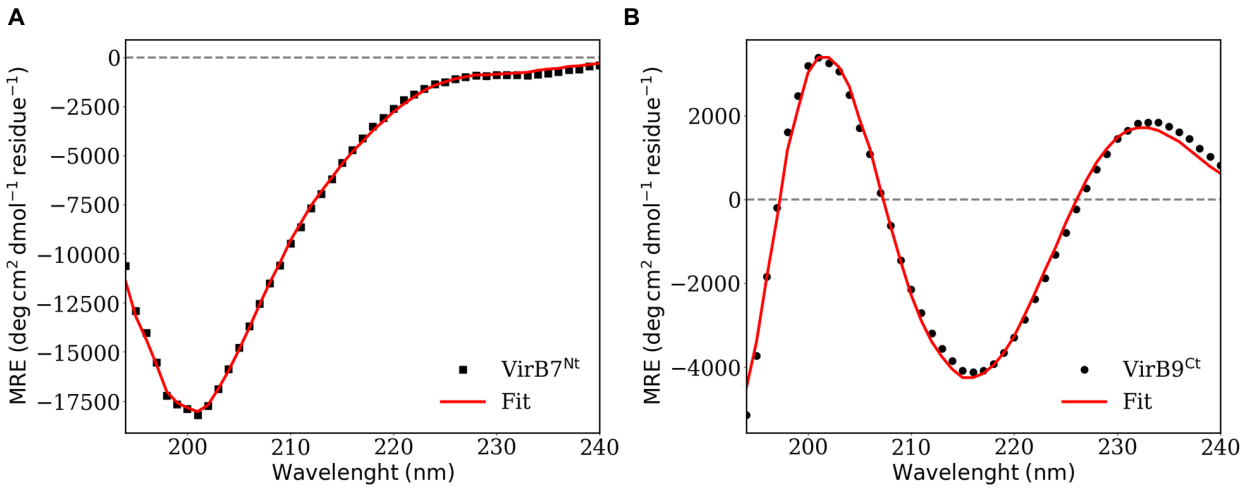


Figure 3.2: Deconvolution of the CD spectrum from (**A**) VirB7^{Nt} and from (**B**) VirB9^{Ct} using the Dichroweb online server [94]. As a note of caution, CD signals below 200 nm are low resolved

by a blue-shift of the maximum wavelength, while a smaller effect was observed upon the addition of the complex (Fig. 3.3, panel B). These observations suggest that VirB9^{Ct} exposes to the solvent a slightly larger hydrophobic surface in the free state than when in complex.

DSC denaturation experiments showed reversible thermal transitions at a melting temperature $T_m \approx$ 68-70°C for the unbound VirB9^{Ct} and for the complex (Fig. 3.3). This finding was somehow unexpected, since VirB7^{Nt} binding was thought to "stabilize" VirB9^{Ct}, which should be reflected in an increase in the T_m [104, 105, 106, 107]. A significantly higher enthalpy of unfolding (area under the transition curve) was observed for the thermal denaturation experiment of the complex (29.57 kcal/mol) in comparison with the free protein (10.42 kcal/mol) (Fig. 3.3). Two possible explanations could be proposed: it is possible that during thermal denaturation of the complex, both the unfolding of the protein (VirB9^{Ct}) and the release of VirB7^{Nt} occurs within a single thermal transition, accounting for an increase in ΔH due to the coupling between unfolding and dissociation equilibrium [108, 109, 110]. Alternatively, the lower ΔH of unfolding of the unbound state could arise from an ensemble of VirB9^{Ct} conformations, from which poorly folded conformations will not necessarily contribute to a significant change in heat capacity, diminishing the contribution to the enthalpy [111, 112]. The fact that VirB9^{Ct} displays the same T_m as the complex, is consistent with the observation that peptide binding does not lead to a significant increase in the secondary or tertiary structure contents of VirB9^{Ct}.

Since both the complex and the free VirB9^{Ct} undergo thermal denaturation at the same temperature (T_m), then the large ΔH of unfolding observed for the complex would imply a large ΔS of unfolding. Similarly, the small ΔH for the unfolding of the unbound VirB9^{Ct} implies a small ΔS of unfolding. Considering that the final denatured states are equivalent, the low ΔS of unfolding observed for VirB9^{Ct} must result from the high entropy of the unbound sate [113, 112, 111]. Indeed, NMR experiments showed previously that VirB9^{Ct} is more flexible in the free state than when in complex with VirB7^{Nt} [16] [11].

To confirm that the observed DSC transitions were associated with the thermal denat-

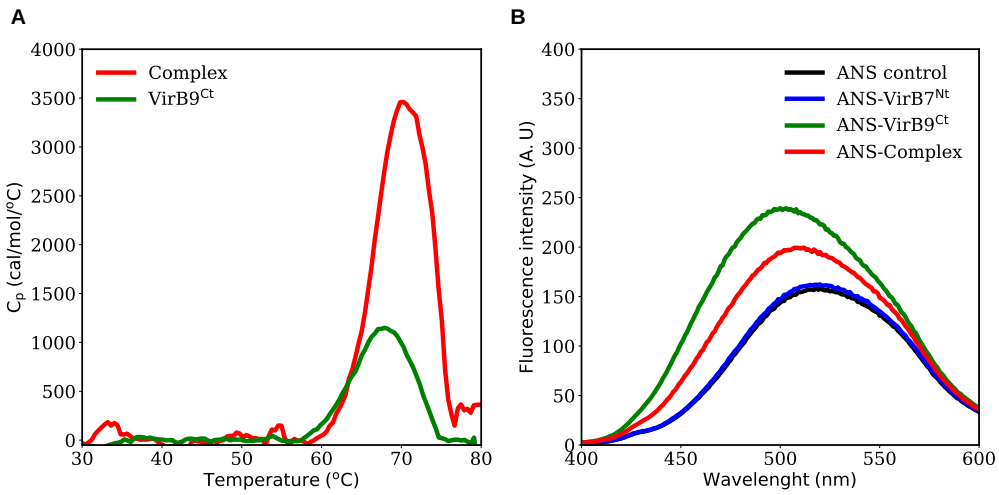


Figure 3.3: Comparison of the (**A**) DSC transition curves obtained for the thermal denaturation of unbound VirB9^{Ct} (green) and bound VirB9^{Ct}(red) and (**B**) the ANS fluorescence measured in the absence (black), or presence of VirB7^{Nt} (blue), VirB9^{Ct} (green) or the VirB7^{Nt}:VirB9^{Ct} complex (red).

uration of the free protein and the complex, we investigated the effect of the temperature increase on the secondary structure content of the unbound VirB9^{Ct} and the complex. To this end, CD spectra were recorded as a function of temperature in the range from 25 up to 95 °C. The denaturation process of the unbound VirB9^{Ct} and the complex showed to be reversible as indicated by the little loss of signal observed after cooling down (see Fig. 3.4, top right and left). Both the complex and the unbound VirB9^{Ct} displayed similar denaturation patterns, where it was observed a drop in ellipticity of the shoulder at 233 nm while a minor ellipticity change was noted at 218 nm (Fig. 3.5, panel A). Thus, we chose to measure the ellipticity at 233 nm as a function of temperature in order to calculate the T_m of the two systems.

Both the complex and the unbound VirB9^{Ct} CD spectra showed a clear decrease in ellipticity at 233 nm as the temperature was increased (Fig. 3.5, panel B). Since a positive CD band at this position reflects interactions between aromatic groups in close proximity

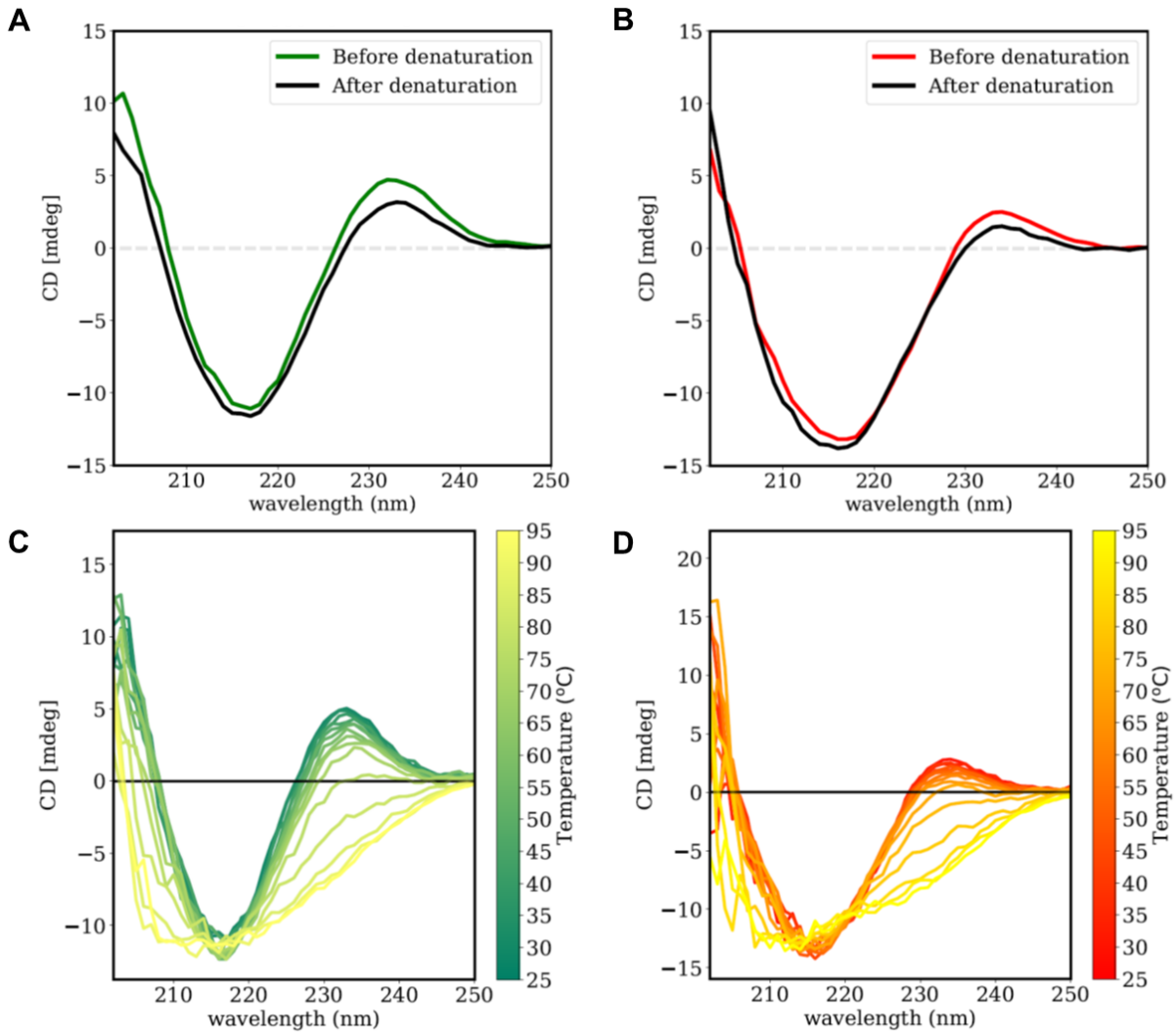


Figure 3.4: CD spectra showing in (**A**) and (**B**) the reversibility of the thermal denaturation, and in (**C**) and (**D**) the thermal denaturation of the unbound VirB9^{Ct} (green) and bound VirB9^{Ct} (red), respectively.

in the structure, these changes suggest a loss in the tertiary structure as a function of temperature. In agreement with the DSC data, the thermal denaturation monitored at 233 nm revealed a reversible transition with the same T_{m} for both the complex and the unbound VirB9$^{\mathrm{Ct}}$, (Fig. 3.3, panel A). Fitting of the CD denaturation curves (see panels C and D of Fig. 3.5) to Eq. 1.97, yielded a $\Delta H \approx 57$ kcal/mol for both unbound VirB9$^{\mathrm{Ct}}$ and complex, and a $T_{\mathrm{m}} \approx 78.40$°C (see the parameters fitting correlations in Fig. 3.6, panel A and B), which is a little bit higher than that obtained by DSC (70°C) [114]. The reason for the slightly different T_{m} values obtained by DSC and CD is unclear, but it should be considered that DSC measures changes experienced by the whole system (protein and solvent) that probably are not immediately translated into structural changes detectable by CD [115]. As observed in panels A from Fig. 3.5, the signal at the 215 nm band from VirB9$^{\mathrm{Ct}}$ arising from β structure content is poorly affected, indicating preservation of secondary structure in the denatured state.

In summary, even though the unbound VirB9$^{\mathrm{Ct}}$ displayed greater ANS fluorescence intensity than the complex, this difference is small. Overall, the CD and the DSC data suggest that VirB9$^{\mathrm{Ct}}$ retains most of the secondary and tertiary structures of the bound state, despite having its hydrophobic interior a bit more exposed to the solvent. This is opposite to what was expected for a molten globule or a partially folded protein [116]. Hence, despite being highly flexible, VirB9$^{\mathrm{Ct}}$ seems to be folded in the free state.

3.1.2 Three-dimensional structure and dynamics

Backbone resonance assignment of unbound VirB9$^{\mathrm{Ct}}$ at 7°C

Qualitatively, the ^{1}H-^{15}N HSQC spectrum of unbound VirB9$^{\mathrm{Ct}}$ at 35°C indicates that the free protein is highly flexible, presenting conformational exchange processes at intermediate exchange rates in the NMR time scale. In contrast, the spectrum of VirB9$^{\mathrm{Ct}}$ in complex is well resolved, characteristic of folded proteins [16]. Since the NMR spectra of the unbound and the bound VirB9$^{\mathrm{Ct}}$ samples were considerably different, we became surprised to find that

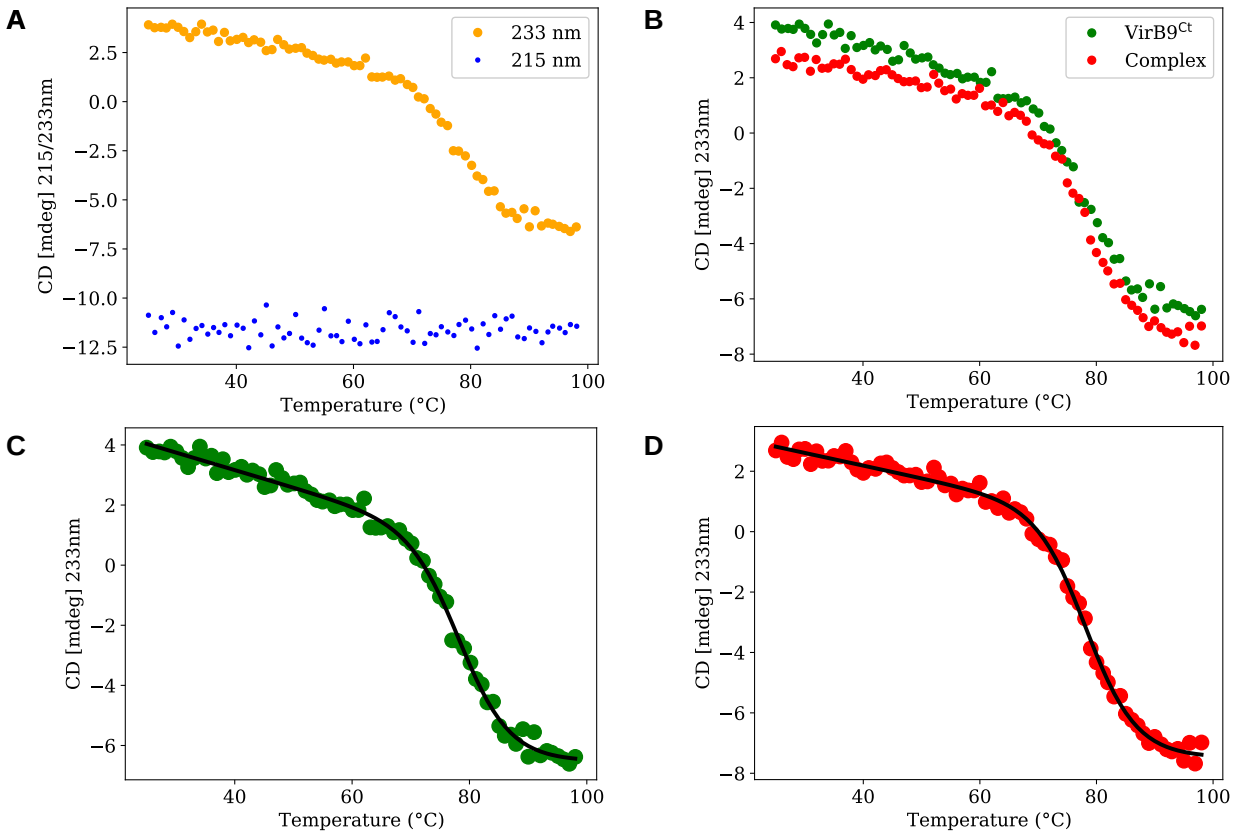


Figure 3.5: The ellipticity measured as a function of temperature (**A**) of unbound VirB9^{Ct} at 233 and 215 nm. (**B**) of unbound (green) and bound (red) VirB9^{Ct} at 233 nm. (**C**) The denaturation CD curves shown in B and their fitting to Eq. 1.97 shown as black lines.

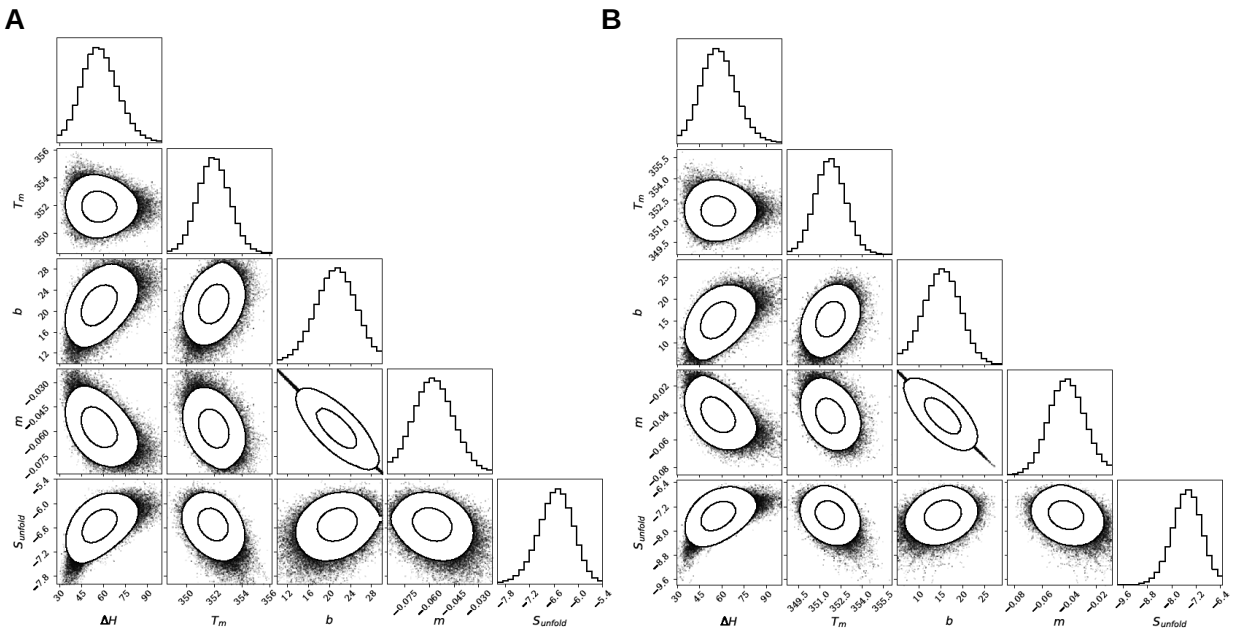


Figure 3.6: Triangle plots showing the correlations between the parameters adjusted to fit the CD denaturation curves from the unbound VirB9^{Ct}(**A**) and the bound VirB9^{Ct} (**B**) to Eq. 1.97.

these two states have similar secondary structure contents as indicated by CD spectroscopy. Thus, we aimed to determine the secondary structure propensities of the unbound VirB9^{Ct} in more detail by NMR spectroscopy. To assign the backbone resonances of the protein, we lowered the temperature to restrict its dynamics. The NMR spectrum of the unbound VirB9^{Ct} recorded at 7°C is qualitatively consistent with a rigid protein (Fig. 3.7, panel A). Prof. Denize Favaro (IQ-UNICAMP) analyzed a set of multidimensional triple resonance NMR experiments recorded with the unbound VirB9^{Ct} sample at 7°C. She identified the backbone resonances (^{1}H, ^{15}N, ^{13}Cα, ^{13}CO) of 70 of the VirB9^{Ct} amino acids (Fig. 3.7A and 3.7B) **(listed in Appendix C)**, except for residues located in β-strands β1 and β2, which form the VirB7^{Nt} binding site (Fig. 3.7, panel B), and whose cross peaks seem to be missing.

The assigned backbone chemical shifts of the unbound VirB9^{Ct} at 7°C were used as input for TALOS [117] to predict the secondary structure content of the protein. As observed in Fig. 3.8, the β strands β_3 to β_8 are preserved in the unbound protein, as well as the short helical propensity in the inter-strand loop between β_3 and β_4 [16]. Since we do not have assignments for residues in β_1 and β_2, we do not have secondary structure information for this region. Curiously, we observed in Fig. 3.8 that an additional α-helix could be formed in the inter-strand loop between β_4 and β_5 in the unbound VirB9^{Ct}, which is absent in the structure of the bound state.

To compare the secondary structure propensities of the unbound and the bound states of VirB9^{Ct} at the same temperature, 35°C, and to obtain information about the residues in the binding interface, we attempted to assign the backbone resonances at higher temperatures.

Backbone resonance assignments of unbound VirB9^{Ct} at 35°C

The assignments of the ^{1}H-^{15}N cross peaks at low temperature (Fig. 3.7) were transferred to the spectrum recorded at 35°C using a temperature series of VirB9^{Ct} ^{1}H-^{15}N HSQCs, starting at 7°C and going up to 35°C in 3°C steps (Fig. 3.9). Using this approach, we transferred almost all of the assignments obtained at 7°C to the spectrum recorded at 35°C.

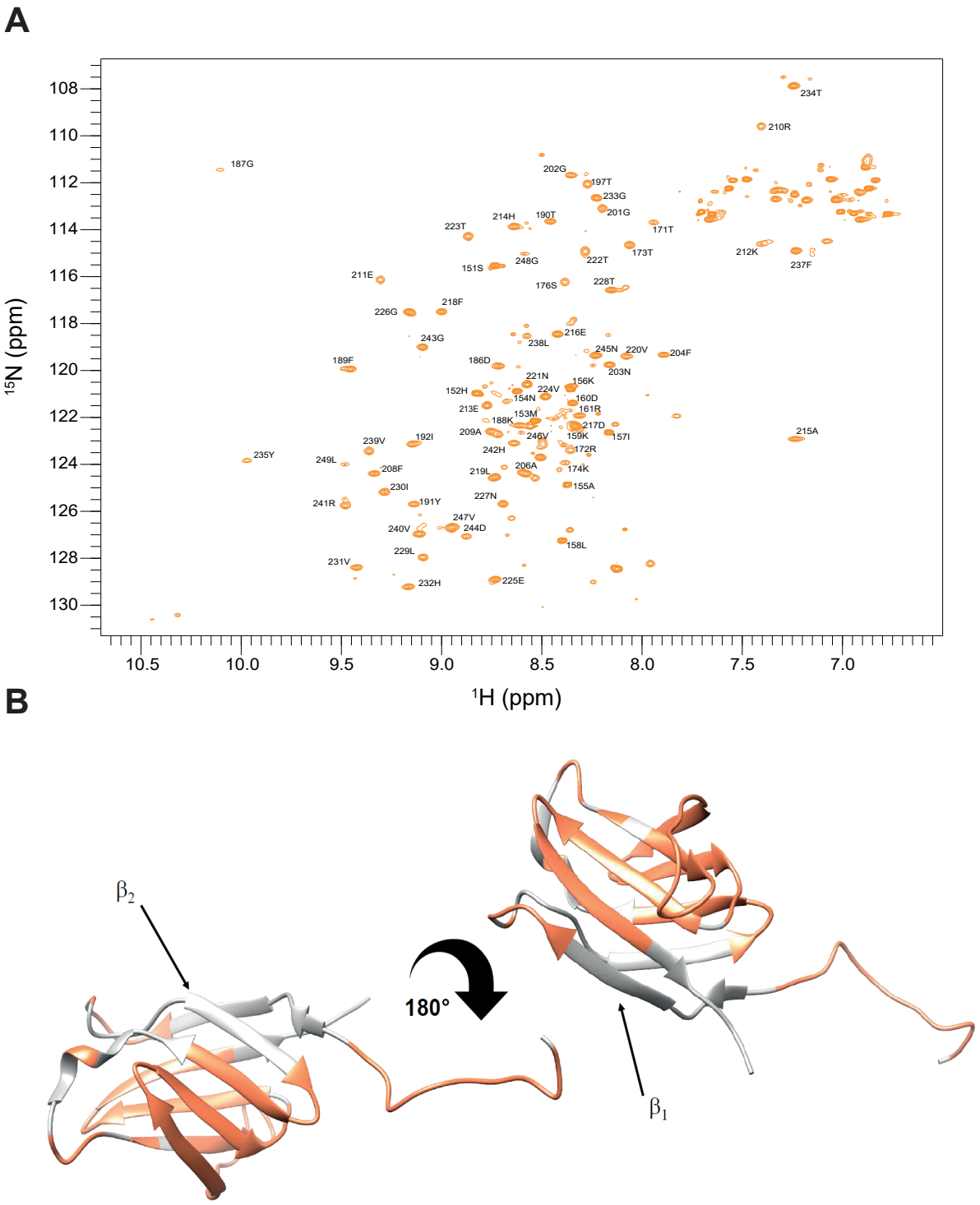


Figure 3.7: ^{1}H-^{15}N HSQC spectrum of the unbound VirB9^{Ct} at 7°C. The assigned peaks are indicated. (**A**) Mapping of the assigned residues on the structure of VirB9^{Ct} in complex (**B**).

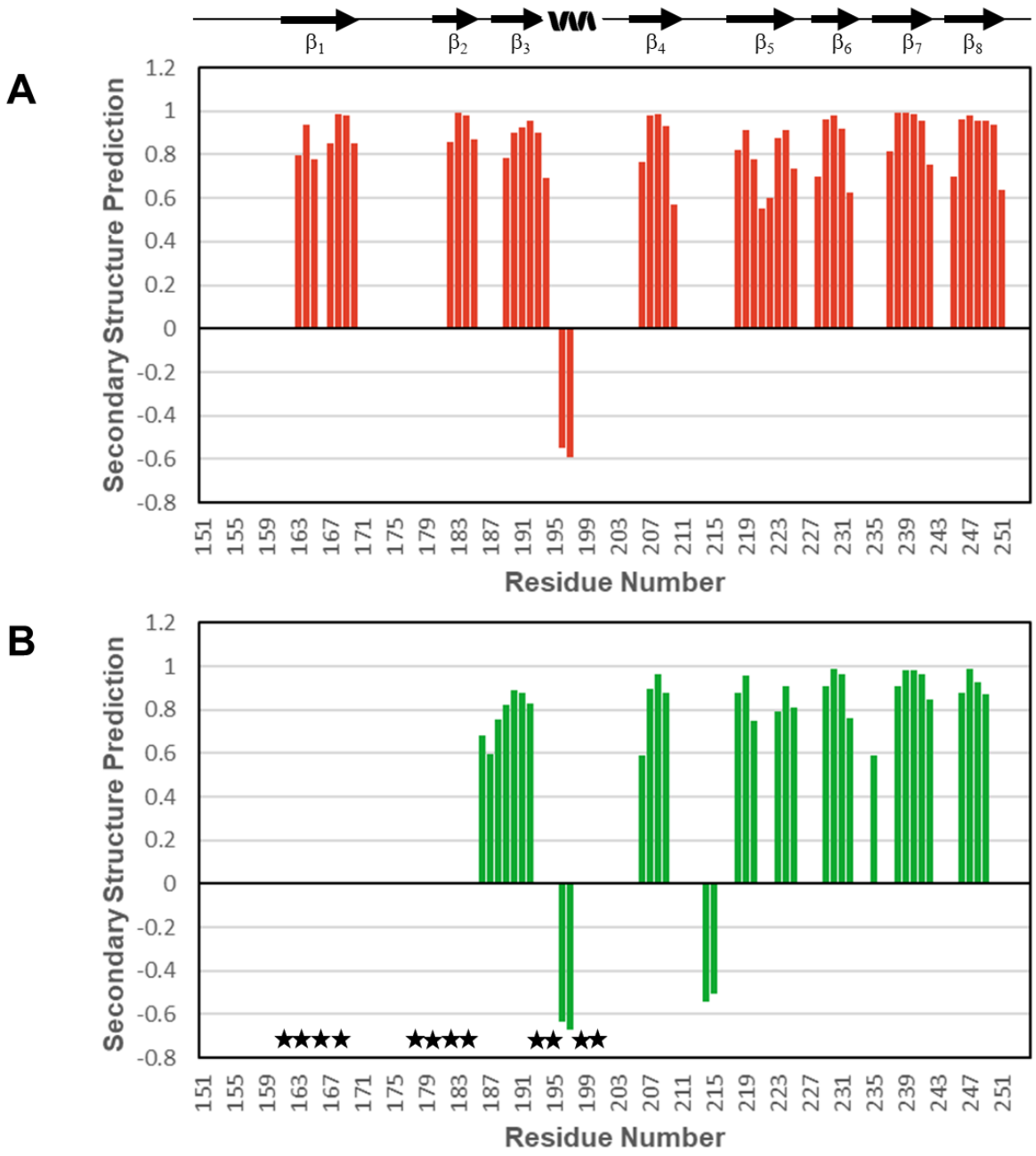


Figure 3.8: Talos predicted secondary structure for (**A**) bound VirB9^{Ct} using the assignments from VirB9^{Ct} in complex at 35°C [16] and (**B**) unbound VirB9^{Ct} using the resonance assignments obtained at 7°C. The secondary structure of VirB9^{Ct} in the complex is shown on top. The black stars highlight the regions of unbound VirB9^{Ct} that were not assigned.

Some of the assignments could not be transferred because the signals became overlapped or were lost in the noise level as the temperature was raised. The assigned cross peaks in the spectrum of the unbound VirB9^{Ct} at 35°C are shown in Fig. 3.10.

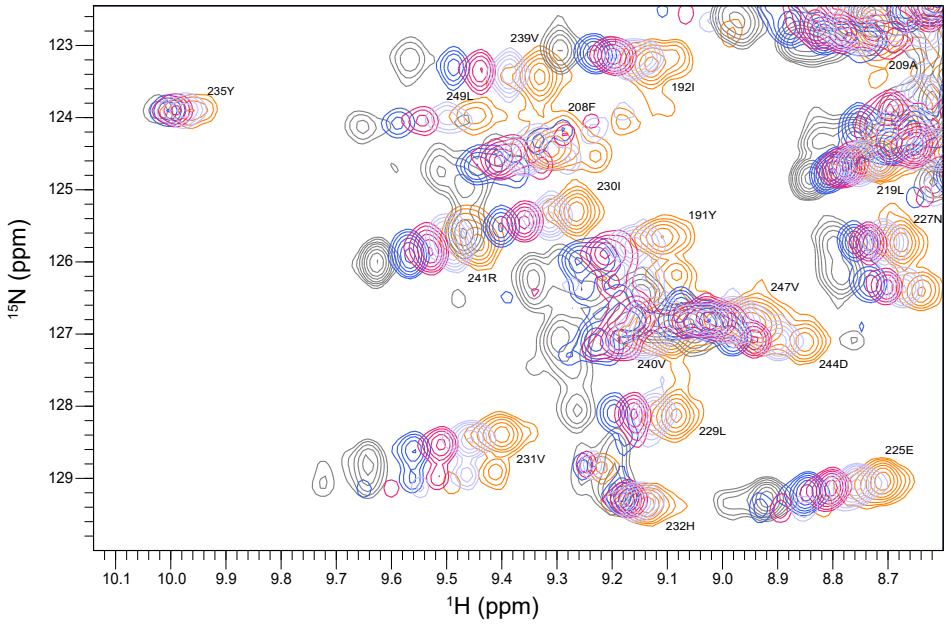


Figure 3.9: Assignment transfer of unbound VirB9^{Ct} resonances to higher temperatures. The colors indicate the temperature in which each spectrum was recorded: orange corresponds to 280 K, lilac 286 K, magenta 292 K, blue 298 K and gray 308 K.

To obtain additional assignments at 35°C we prepared a NMR sample of the complex at the substoichiometric 1:06 (VirB9^{Ct}:VirB7^{Nt}) molar ratio, and recorded a ZZ-exchange ^{15}N-HSQC experiment. To identify the exchange cross peaks we compared the ZZ spectrum with the HSQC spectra of the unbound and the bound VirB9^{Ct} (Fig. 3.12). As a control, we looked for correlations among cross peaks already assigned with help of the temperature series. Using this approach, we assigned additional 12 cross peaks of the unbound VirB9^{Ct} as shown in Table 3.1 and Figs. 3.11 and 3.13. Most of them are located at or near the binding site with VirB7^{Nt} (Fig. 3.11).

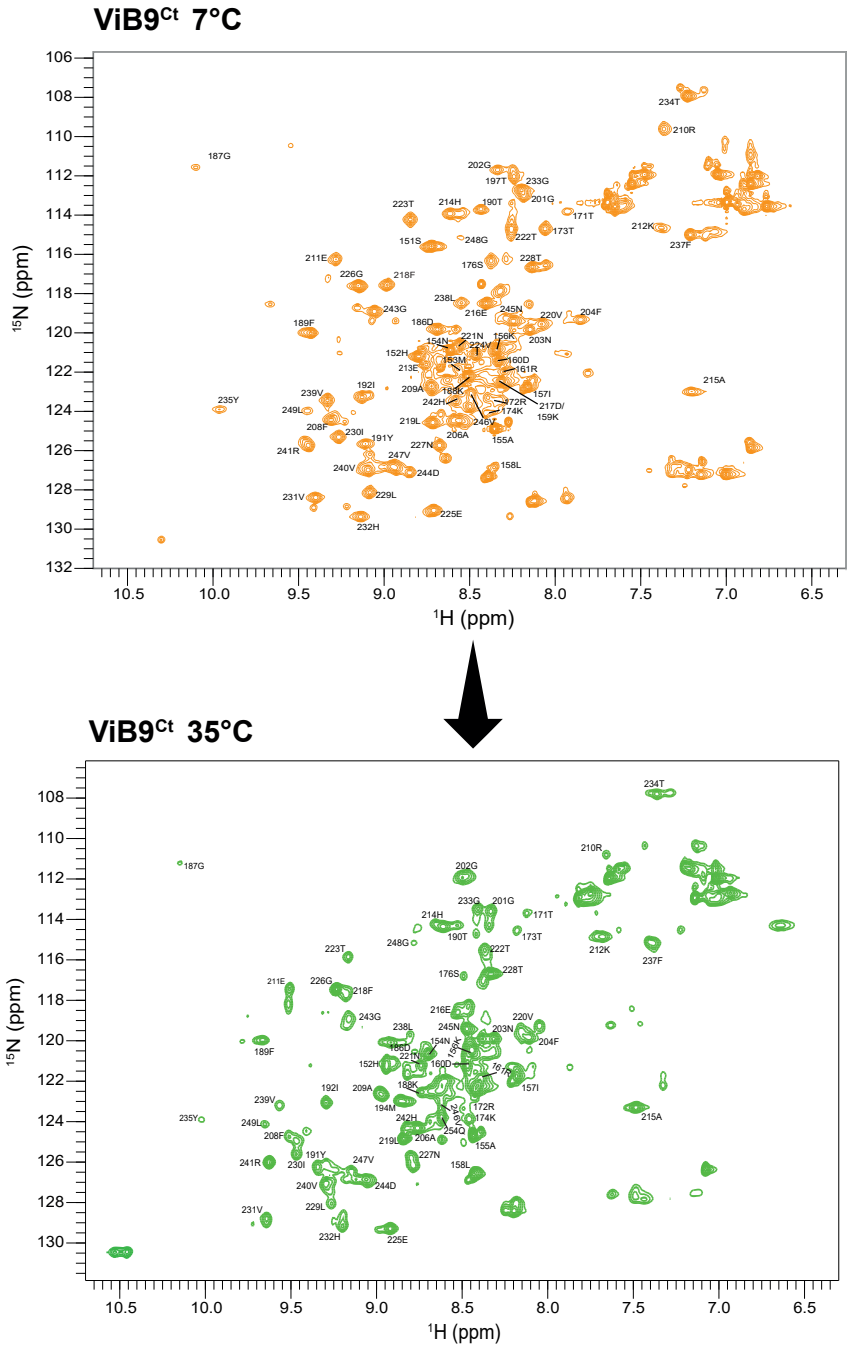


Figure 3.10: Assignment of unbound VirB9^{Ct} at 7°C (spectrum in orange) and assignments that could be transferred to 35°C (spectrum in green).

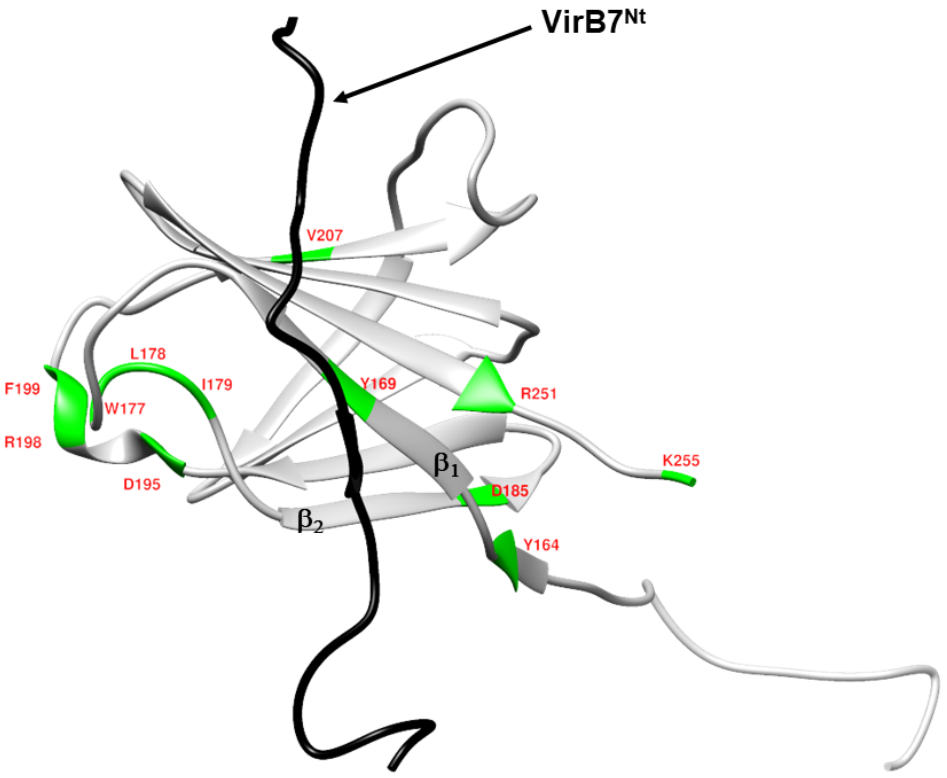


Figure 3.11: The residues of the unbound VirB9^{Ct} that were assigned by the analysis of the ZZ exchange spectrum are colored green. The structure of VirB9^{Ct} is shown in grey, while VirB7^{Nt} is shown in black.

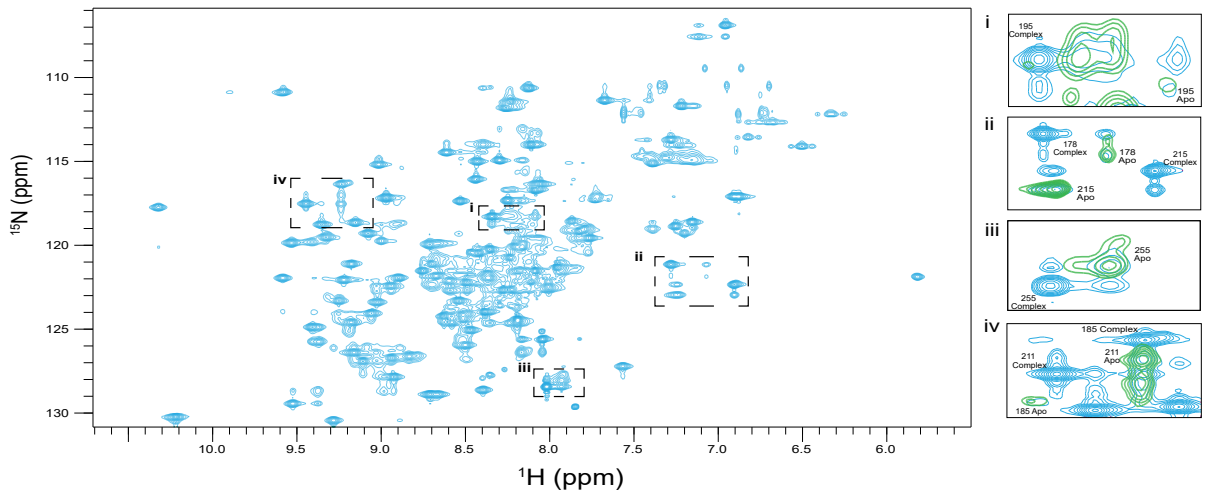


Figure 3.12: ZZ exchange spectrum recorded at 35°C on a sample of ^{15}N-VirB9^{Ct}:VirB7^{Ct} at 1:06 molar ratio (shown in light blue). The black squares on the right are zoomed-in regions showing some of the assigned residues and the superposition with the unbound VirB9^{Ct} ^{1}H-^{15}N-HSQC spectra (in green).

Table 3.1: Assignments of unbound VirB9^{Ct} obtained by the ZZ-exchange analysis at 35°C.

Residue number	^{1}H δ (ppm)	^{15}N δ (ppm)
195	8.09586	118.7364
251	7.91738	127.81049
164	9.31115	121.90374
177	7.62078	121.09657
178	7.05866	121.91286
199	7.37857	118.97314
185	9.51281	118.47092
255	7.92818	128.11445
198	7.2545	118.1891
179	7.08988	117.78742
207	7.84351	120.24858
169	8.65484	119.87029

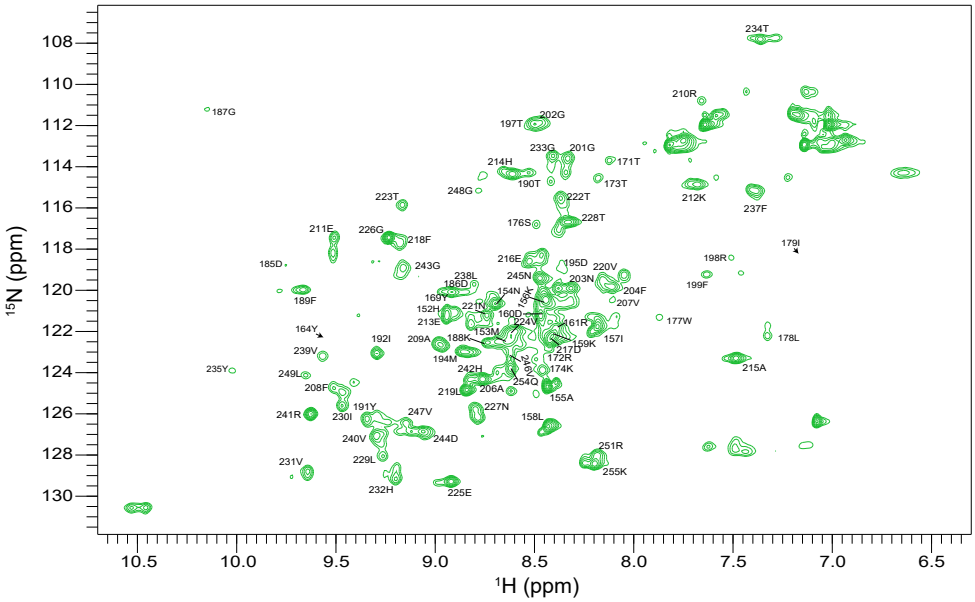


Figure 3.13: ^{15}N HSQC spectrum of the unbound VirB9^{Ct} showing all the assigned residues obtained by temperature transfer from 7 to 35°C and by ZZ-exchange at 35°C.

Backbone resonance assignments at 35 °C obtained using CEST

To obtain additional backbone ^{15}N and ^{1}H resonance assignments, and to assign the ^{13}CO resonances of VirB9^{Ct} in the unbound state, we recorded Chemical Exchange Saturation Transfer (CEST) experiments of the complex in the presence of a slight excess of free VirB9^{Ct}. To this end, a ^{15}N-CEST experiment was recorded on a sample of VirB9^{Ct} containing sub-stoichiometric VirB7^{Nt} concentrations at the 1:0.7 molar ratio (VirB9^{Ct}:VirB7^{Nt}). The ^{15}N-CEST profile of some of the residues showed two dips, corresponding to two VirB9^{Ct} exchanging states (Fig. 3.14). However, VirB9^{Ct} in complex does not undergo conformational exchange at the millisecond time scale on its own, as showed by previous NMR data [16]. Interestingly, after further addition of VirB7^{Nt} up to the 1:0.8 molar ratio, we noticed that the smaller dip decreased or disappeared (Fig. 3.14, black line), indicating that it corresponded to the free VirB9^{Ct} in slow exchange equilibrium with the bound state (larger dip).

To obtain ^{13}CO chemical shifts, which are more informative about secondary structure than ^{15}N chemical shifts, we recorded a ^{13}CO-CEST experiment (3.15, top) on a ^{13}C/^{15}N doubly labeled complex sample prepared at the 1:0.7 (VirB9^{Ct}:VirB7^{Nt}) molar ratio (Fig. 2.3). Furthermore, to obtain ^{1}HN chemical shifts for the free state, we attempted to record a ^{1}HN-CEST experiment on a ^{15}N-labeled sample of VirB9^{Ct} in complex with unlabeled VirB7^{Nt}, prepared at the same molar ratio. Unfortunately, the ^{1}HN-CEST experiment did not show any dip due to exchange (3.15, bottom). We suspect that the reason for the lack of chemical exchange contribution in this experiment was the fact that we had to decrease the magnetization exchanging period (T_{ex}) from 400 ms used in the ^{15}N and in the ^{13}CO-CEST experiments, to 150 ms in the ^{1}HN-CEST experiment. This was necessary to compensate for the very fast relaxation of the $2H_zN_z$ two spin order magnetization, compared with the longer longitudinal relaxation rate of ^{13}CO spins. The shorter exchanging period was probably not sufficient to the allow the build up of exchanging magnetization.

The assigned chemical shifts of the ^{15}N and ^{13}CO spins, obtained using CEST exper-

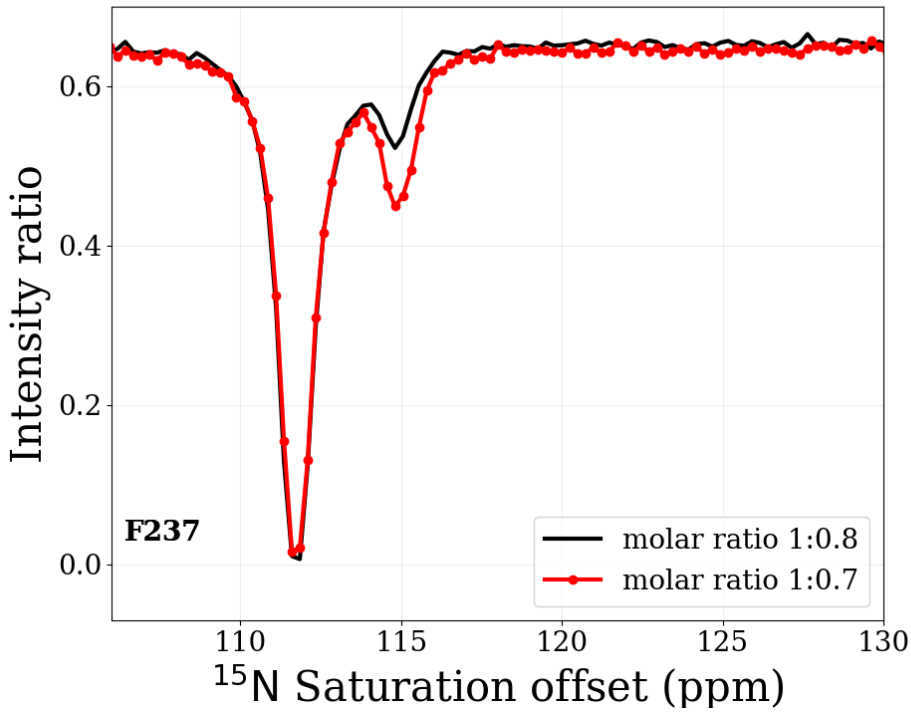


Figure 3.14: ^{15}N CEST profile of F237. The experiment was recorded with a sample of VirB9^{Ct} containing sub-stoichiometric VirB7^{Nt} concentrations at the 1:0.7 molar ratio (red line) and 1:0.8 molar ratio (black line).

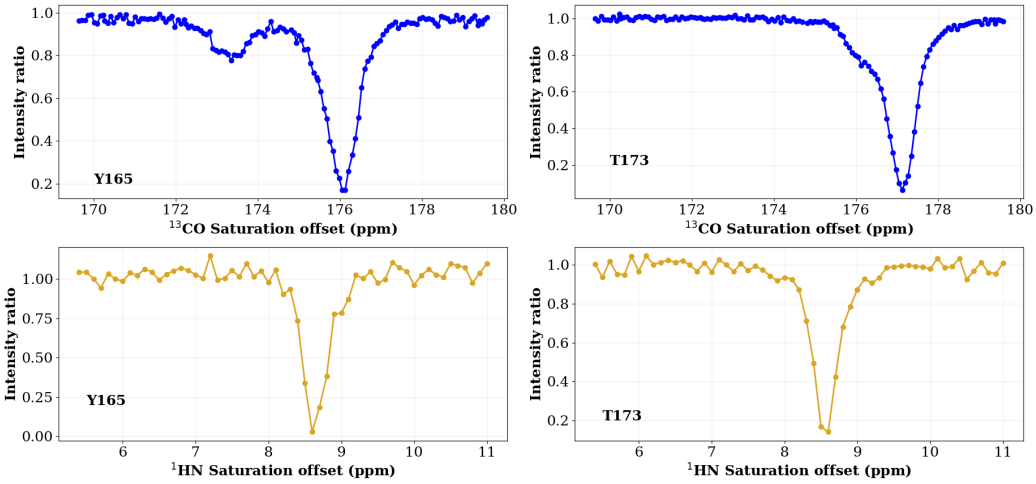


Figure 3.15: CEST experiments recorded at 35°C on a sample of ^{15}N-VirB9^{Ct}:VirB7^{Nt} at 1:0.7 molar ratio, showing in blue the ^{13}CO-CEST profiles (**Top**) and in yellow the ^{1}HN-CEST profiles (**Bottom**) of chosen VirB9^{Ct} residues.

iments, as well as random coil ^{15}N and ^{13}CO chemical shifts predicted using the software ncIDP [118], were used to calculate secondary chemical shifts and, hence, to obtain information on the secondary structure content of the unbound VirB9^{Ct}. The same analysis was carried out using the chemical shifts of the bound VirB9^{Ct} obtained by Oliveira and co-workers [16]. A comparison of the ^{15}N and ^{13}CO secondary chemical shifts of the bound and unbound VirB9^{Ct} is shown in Fig. 3.16. It is worthy noting that β-strand β_1 displays smaller ^{13}CO and ^{15}N secondary chemical shifts in the unbound state, suggesting that it is unfolded. A similar tendency is observed for β_2, however, the magnitude of the difference is not as large, suggesting that this strand is partially folded in the free state (Fig. 3.16). Overall, the other β-strands seem to be preserved in the unbound state.

The assignments for the ^{1}H-^{15}N cross peaks in the HSQC spectrum of the unbound VirB9^{Ct} were compared to the assignments of the bound protein at the same temperature. Chemical shift perturbations (CSP) were calculated using Eq. 2.1, [91]. This analysis shows

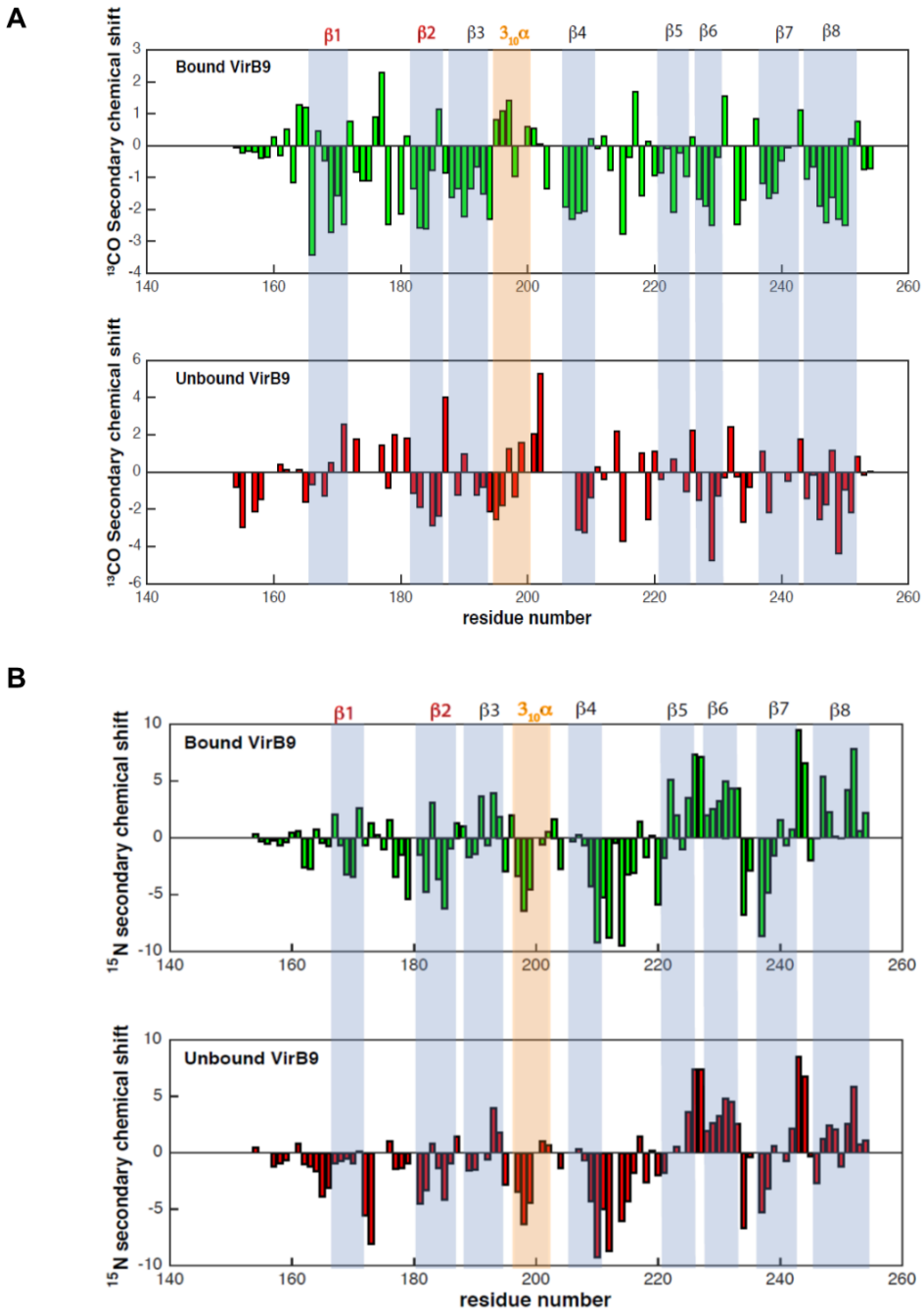


Figure 3.16: Secondary Chemical Shifts (SCS) calculated using the random coil chemical shifts predicted with ncIDP (`https://st-protein02.chem.au.dk/ncIDP/`) [118] and the ^{15}N and ^{13}CO-CEST chemical shift data for unbound VirB9^{Ct}, shown in green, and VirB9^{Ct} in complex, shown in red.

that the residues most affected by the presence of VirB7^{Nt} were located in the β_1, in the β_1-β_2 inter-strand loop, β_5, β_4-β_5 loop, β_7, and residue V247 in β_8 (Fig. 3.17A and 3.17B). A close inspection of these regions shows that β-strands β_1, β_8 and β_7 form an antiparallel β-sheet where VirB7^{Nt} lies on, which explains the significant changes in chemical shift. Curiously, the β_5 and the β_4-β_5 loop are located far away from the VirB7^{Nt} binding site. It is interesting that TALOS indicated that residues in in the β_4-β_5 loop have a tendency to assume helical conformation (Fig. 3.8), suggesting that this loop undergoes a major structural change upon complex formation.

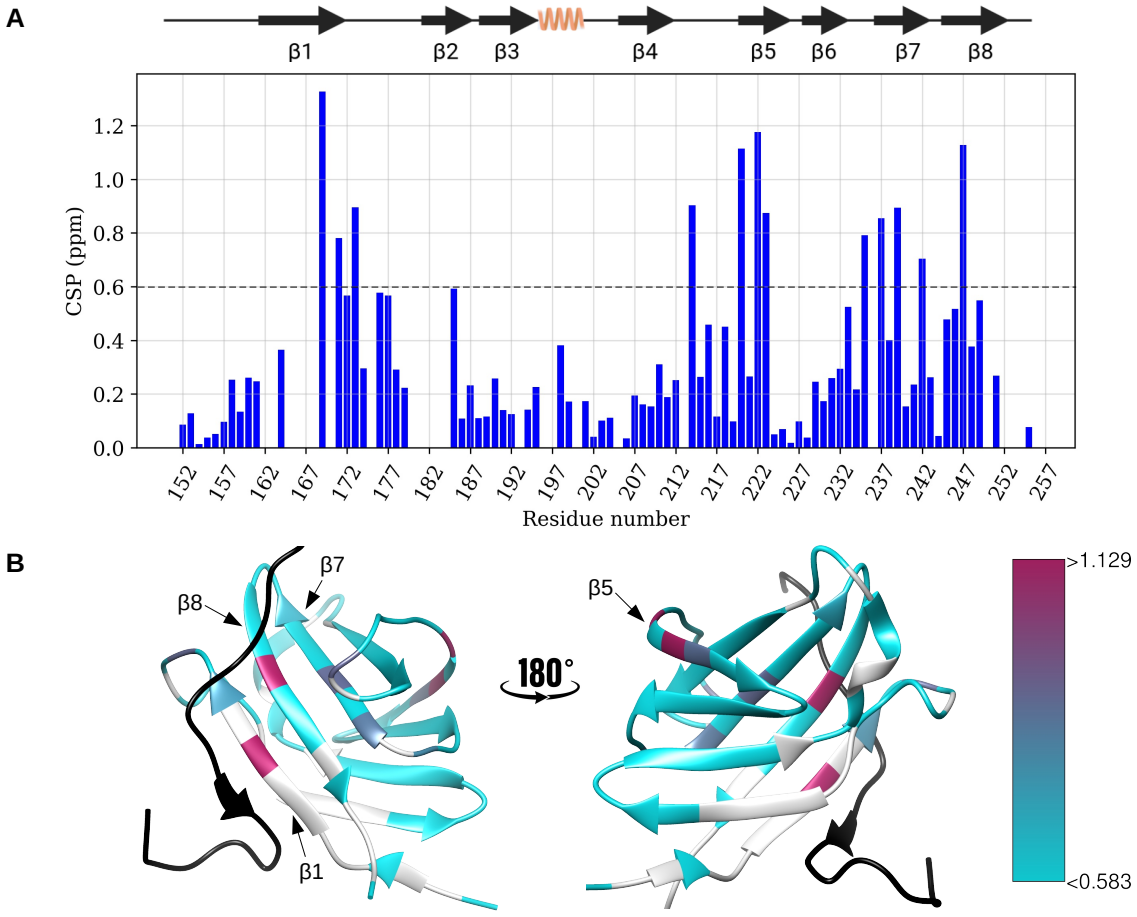


Figure 3.17: Chemical shift perturbation calculated using Eq. 2.1 between unbound and bound VirB9^{Ct} (**A**) and its mapping on the structure of VirB9^{Ct} in complex (**B**).

Fitting the CEST profiles to the Bloch-McConnell equation

We fitted the ^{15}N CEST profiles obtained with the 1:0.7 sample to the Bloch McConnell equation (Eq. 1.83). Two CEST experiments were recorded using saturation frequencies $\gamma B_1/2\pi$ of 15 and 30 Hz. The two CEST profiles were fitted simultaneously to decrease the degenerescence of the free parameters. Only residues displaying two dips were fitted as exemplified for I179 and R241 in Fig. 3.18 (Appendix C). From the fitting we could extract the value of the binding equilibrium exchange rate k_{ex}^{b} ($k_{ex}^{b} = k_{on}[L] + k_{off}$) (Eq. 1.5), the population of the bound state (p_B), as well as the chemical shift of the unbound state for each fitted residue (Fig. 3.18). Fitted k_{ex}^{b} values varied from ~ 10 s^{-1} to ~ 160 s^{-1}, and p_B varied between 0.83 and 0.97. A close inspection of the fitted CEST profiles showed that the highest k_{ex}^{b} were from residues whose CEST profiles did not show a clear separation between the dips corresponding to the unbound and bound VirB9^{Ct} states. For example, R241 (Fig. 3.18D) whose CEST profile displayed a small shoulder corresponding the unbound state, yielded the highest k_{ex}^{b}. In contrast, residues displaying a clear separation between the two dips, as I179 (Fig. 3.18A), yielded the lowest values of k_{ex}^{b}. Indeed, the analysis of the correlations between the parameters distributions (triangle plots) that were obtained using MCMC (2.8), indicated that residues with few separation between the two dips, such as R241 or D168 (see D168 CEST profile in Appendix C), presented increased correlations between parameters (Appendix C). This means that any change in one parameter will affect the value of the other, reflecting a compensation in the fit which can mislead the best fit values (Appendix C). In contrast, residues with well defined dip separation, such as I179 or Y165 (Fig. 3.18, Appendix C), presented weaker correlations between parameters (Appendix C).

Despite simultaneously fitting two CEST curves with well separated dips, we observed that k_{ex}^{b} and p_B tend to be highly correlated in all residues. Therefore, we used a prior information of the value of p_B to increase the accuracy of k_{ex}^{b}. The initial guess for p_B was obtained from the intensity ratios between the NMR signals from VirB9^{Ct} Ala215 and Thr234 bound and unbound states in the CEST experiment. To obtain accurate k_{ex}^{b} values

we chose a cut-off of $|\Delta\Omega| > 160$ Hz $\sim$ 2 ppm to select the best residues to analyse. The selected residues are highlighted by a gray circle in Fig. 3.18 panel B and the zoomed in region in panel C. In that logic, the best value of k_{ex}^{b} is $\approx$ 15-20 s^{-1}, while p_B is $\approx$ 90-93% of the total VirB9^{Ct} population. This analysis also showed that our VirB7^{Nt} concentration was underestimated.

To identify the regions with the highest ^{15}N chemical shift difference between the bound and unbound states, we mapped the $\Delta\Omega$ distribution (applying the $|\Delta\Omega| > 2$ ppm cutoff) on the NMR structure of VirB9^{Ct} in complex. Similarly, we also mapped the k_{ex}^{b} distribution. As observed in Fig. 3.19 panel A, the VirB9^{Ct} regions displaying the greatest ^{15}N chemical shift differences correspond to the β-sheet over which VirB7^{Nt} lies on, and the β_1 and β_2, which is consistent with the VirB7^{Nt} binding site. The distribution of k_{ex}^{b} values within the structure of VirB9^{Ct} follows a similar pattern.

3.2 The thermodynamics of the VirB9^{Ct}-VirB7^{Nt} complex

In protein-ligand interactions the binding energy depends on critical aspects of the intermolecular interface, such as the interface size, residue composition, the structural and thermodynamic properties of the solvent and the flexibility of the interacting partners [119, 73, 120]. The formation of hydrogen bonds and van der Waals interactions during binding contribute favorably to the binding enthalpy, while the desolvation of non-polar and polar groups contributes favorably to the binding entropy. On the other hand, folding of disordered and flexible regions contributes unfavorably to the binding entropy as conformational degrees of freedom are reduced upon binding [121]. Therefore, in coupled folding and binding interactions as those observed in IDPs/IDRs, the energetics of folding also contributes to the measured thermodynamics. Here we used isothermal titration calorimetry (ITC) to investigate the thermodynamics of the VirB9^{Ct}-VirB7^{Nt} complex formation, in view

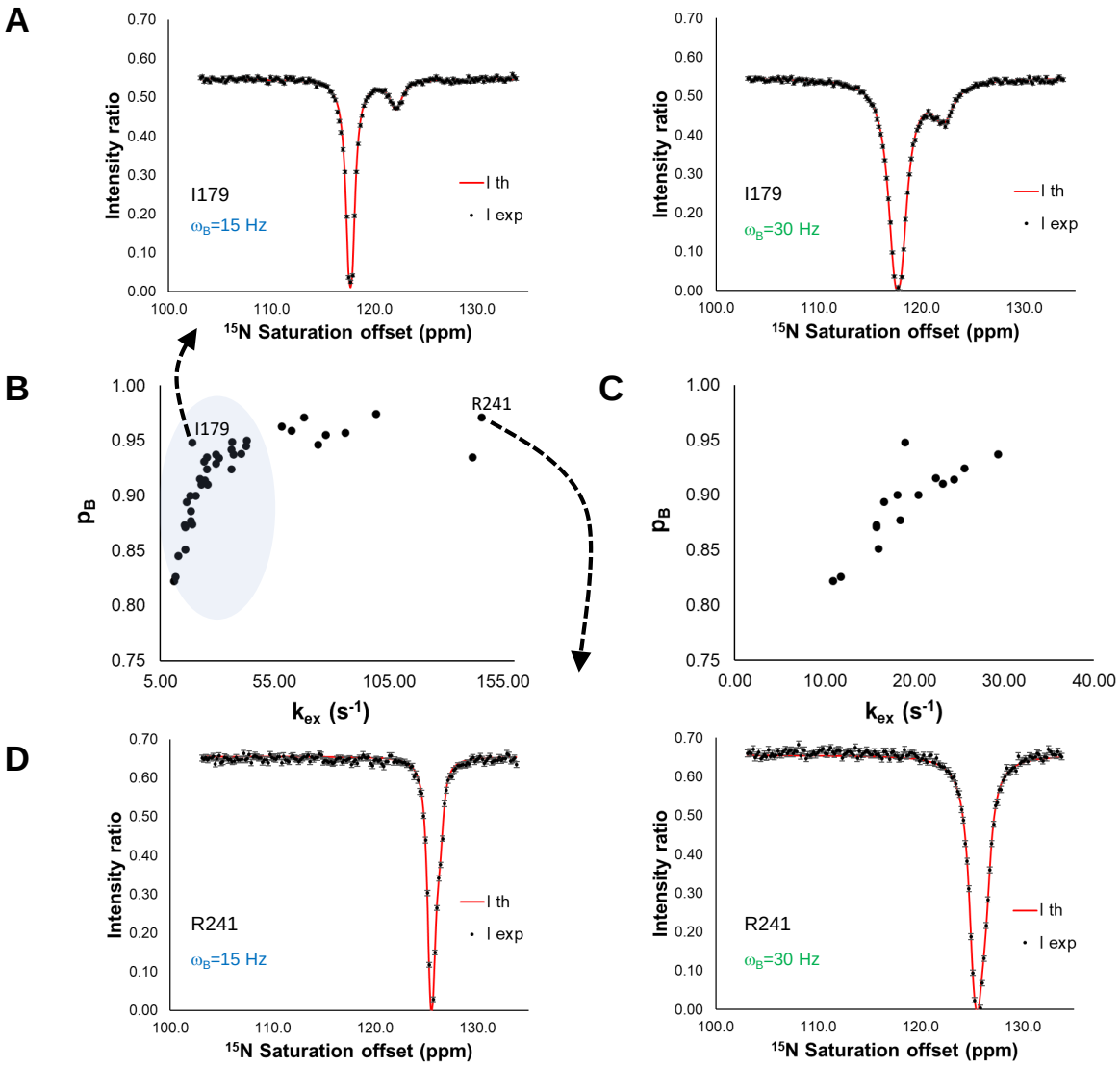


Figure 3.18: Fitting of the ^{15}N-CEST profiles to the Bloch McConnell equation (1.83) showing in (**A**) and (**D**) the fitting of I179 and R241 at 15 and 30 Hz, respectively. Plot of p_{B} vs k_{ex} obtained from the fitting (**B**) and an update of this plot using residues displaying a $|\Delta\Omega| > 2$ ppm (**C**).

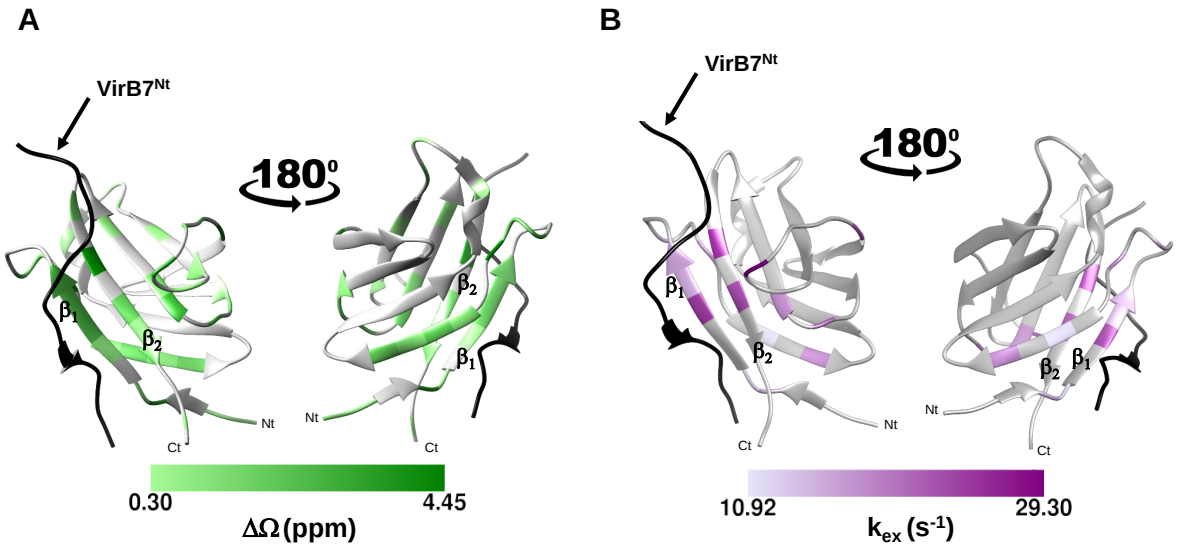


Figure 3.19: Color coding the values of k_{ex}^{b} (**A**) and $|\Delta\Omega|$ (**B**) with the $\Delta\Omega > 2$ ppm cutoff on the structure of VirB9^{Ct} in complex.

that the two proteins are intrinsically flexible and become structured upon binding. We first wanted to measure a more accurate value of the binding affinity (K_d^{app}) for the VirB9^{Ct}-VirB7^{Nt} binding interaction. For this, we performed three ITC binding measurements at 35°C with different ligand concentration within the cell in such a way that the isotherms will cross in a mutual point just in the inflexion point, which determines the stoichiometry of the reaction. Subsequently, these three isotherms were fitted simultaneously to calculate the value of K_d^{app}. The reason of this strategy is because we have noticed that there is a significant error in the determination of the binding affinities by ITC due to the uncertainties in the measurement of protein concentration using spectroscopic techniques.

The K_d value obtained by techniques such as ITC is an overall measure of the binding affinity, that include the contributions of conformational rearrangements to the value of K_d (see Background theory). Having said that, we refer to the binding affinity measured by ITC as K_d^{app} that makes reference to the apparent value of K_d measured by ITC, and we refer to the intrinsic binding affinity as K_d^{int} defined as the ratio k_{off}/k_{on}. As observed in Fig. 3.20

panel A, the experimental isotherms were measured at 35°C at 10, 15 and 20 μM of ligand concentration. Panel B of Fig. 3.20 shows the simultaneous fitting of these three curves to a one binding site model, using an in-house Python program described by Cardoso and Rivera et al. (2020) [90]. Our data yielded a value of $K_{\mathrm{d}}^{\mathrm{app}} = 0.7$ μM using this approach, which is ~ 1.5 times the value of the $K_{\mathrm{d}}^{\mathrm{app}}$ reported in the literature by Olivera et al. (2016), who had measured a value of 1 μM at 35°C [16].

In addition, we measured the binding energetics of complex formation at three different temperatures, 10, 25 and 35°C, as shown in Fig. 3.21, panel A. The values for the thermodynamic parameters at the three temperatures are shown in table 3.2, as well as in Fig. 3.21 panel B. The binding interaction was exothermic at the three temperatures and yielded a $\Delta G \approx -9$ kcal/mol (see Fig. 3.21, panel B). Because there is a linear relationship between the change in enthalpy and the change in entropy, thus favorable changes in binding enthalpy are compensated by opposite changes in entropy and vice versa, which results in small changes in binding affinity over a range of temperatures [122]. This is the so-called enthalpy-entropy compensation and it appears to be a general consequence of the formations and disruptions of weak non-covalent interactions [122, 73]. In systems where binding is coupled to folding as often experienced by IDPs and IDRs, the enthalpy-entropy compensation is even more evident given the lost of conformational freedom upon binding. The conformational rearrangement experienced by the protein is coupled to a favorable enthalpic component due to the formation of hydrogen bonds and van der Waals interactions within the disordered domain or region, and to an unfavorable entropic contribution arising from the loss of conformational degrees of freedom associated to folding [121]. For example, we can consider the binding thermodynamics of the HIV-1 envelope glycoprotein gp120, which contains intrinsically disordered regions. The binding of gp120 to the human CD4 receptor or to the MAb 17b antibody in the disordered regions is accompanied by folding, while the binding of the MAb b12 antibody to an structured region of gp120 do not induce conformational rearrangements [121, 123, 124, 125, 126]. The binding thermodynamics of

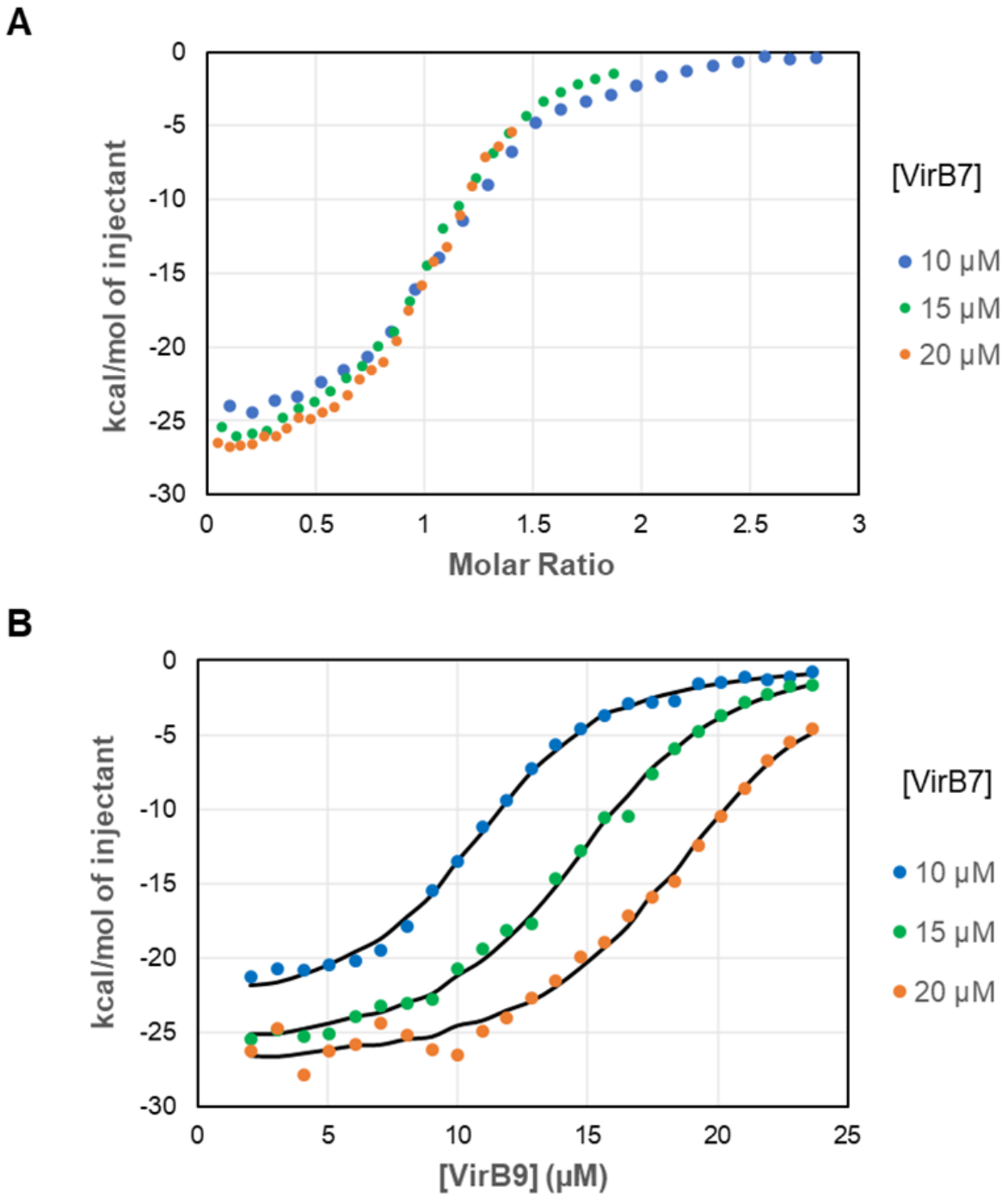


Figure 3.20: Calorimetric titrations of VirB7^{Nt} with VirB9^{Ct} at 35°C. Three experiments were performed with different VirB7^{Nt} concentrations (**A**) and their simultaneous fitting to the one-site binding model (**B**) using the method described by Cardoso and Rivera et al. 2020 [90].

CD4 and MAb 17b are characterized by similar ΔG values with very large favorable binding enthalpies (-39.0 and -28.7 kcal/mol) and a extremely large unfavorable entropy penalty (27.3 and 17.2 kcal/mol). These large favorable enthalpy and unfavorable entropy changes is associated to processes in which binding is coupled to large conformational changes. In contrast, binding of MAb b12 to gp120 presents a completely different thermodynamic signature, with a ΔH of $\sim$ -9 kcal/mol and a binding entropy of $\sim$ -3 kcal/mol, reflecting a binding mostly driven by the enthalpic term and without inducing any conformational structuring [121, 125, 124, 126]. Similarly, by comparing our ITC data of the VirB9$^{\mathrm{Ct}}$-VirB7$^{\mathrm{Nt}}$ interaction at 35°C, we observed that the binding thermodynamics is favourable with a large and favorable change in enthalpy ($\Delta H = -28.61$ Kcal/mol) that compensates the strongly unfavorable change in entropy ($-T\Delta S = 18.48$ Kcal/mol). This is expected since the formation of the VirB9$^{\mathrm{Ct}}$-VirB7$^{\mathrm{Nt}}$ complex involves the occurrence of new contacts in the binding interface along with a decrease in VirB9$^{\mathrm{Ct}}$ and VirB7$^{\mathrm{Nt}}$ intrinsic flexibility.

In order to obtain more information about VirB9$^{\mathrm{Ct}}$ and VirB7$^{\mathrm{Nt}}$ folding upon binding interaction, we calculated the change in heat capacity (ΔC_{p}), determined by plotting the values of ΔH as a function of temperature and calculating the slope of the temperature-dependent ΔH curve with a linear regression analysis. ΔC_{p} is a parameter that measures the change in heat with temperature at constant pressure, and it is strongly correlated to the solvent-exposed surface area being buried upon complex formation, therefore it provides a link between thermodynamic data and structural information [122, 73]. During binding and folding, the burial of hydrophobic residues with the concomitant displacement of water molecules is associated with a decrease in heat capacity, and the ΔC_{p} is consequently negative and proportional to the amount of surface involved [122, 121]. A small binding heat capacity is usually indicative of no significant conformational change, while a large binding heat capacity is usually indicative of a large conformational change. When the structuring of an intrinsically disordered protein along with ligand binding involves a large number of residues, the magnitude of the apparent binding heat capacity change will be considerably

large [127, 121]. For example, the binding interactions of CD4, MAb 17b or MAb b12 to the gp120 glycoprotein, results in ΔC_{p} values of -1.3, -1.8 and -0.4 kcal/(K mol), respectively. As expected, the heat capacities changes of the systems experiencing folding coupled to binding (CD4 and MAb 17b) are about three to fourfold larger than the binding of gp120 to MAb b12, which reflects the smallest ΔC_{p} value because the dehydration is associated primarily with the binding interface [121]. Comparing to the data obtained from the binding of VirB9$^{\mathrm{Ct}}$ to VirB7$^{\mathrm{Nt}}$, we obtained a ΔC_{p} = -0.23 kcal/(K mol), which is a small value for a binding process primarily driven by burial of hydrophobic surfaces and large conformational rearrangements. This is nevertheless expected, considering that in our study we have evidence from NMR CEST data, CD spectroscopy and denaturation assays by DSC and CD that VirB9$^{\mathrm{Ct}}$ populates a bound-like conformation in the free state and that the region of VirB9$^{\mathrm{Ct}}$ involved in a significant conformational change is limited to the binding interface with VirB7$^{\mathrm{Nt}}$. On the other hand, VirB7$^{\mathrm{Nt}}$ also folds upon binding to VirB9$^{\mathrm{Ct}}$, but it forms only a small β strand that complements one of the β strands of VirB9$^{\mathrm{Ct}}$, without involving the burial of a large solvent-exposed region and consequently releasing less water molecules. The above arguments suggest that the expected contributions to ΔC_{p} are minimum even though VirB7$^{\mathrm{Nt}}$ is an intrinsically disordered protein and, as well as VirB9$^{\mathrm{Ct}}$, gets structured upon binding.

Table 3.2: Values for the thermodynamic parameters obtained in the VirB9$^{\mathrm{Ct}}$-VirB7$^{\mathrm{Nt}}$ binding isotherms at 10, 25 and 35°C.

Temperature	n	$K_{\mathrm{d}}^{\mathrm{app}}(\mu\mathrm{M})$	ΔG(kcal/mol)	ΔH(kcal/mol)	$-T\Delta S$(kcal/mol)
10°C	0.89	0.03	−9.64	−21.05	11.40
25°C	1.14	0.40	−8.79	−22.79	13.99
35°C	1.02	0.72	−8.56	−28.17	19.62

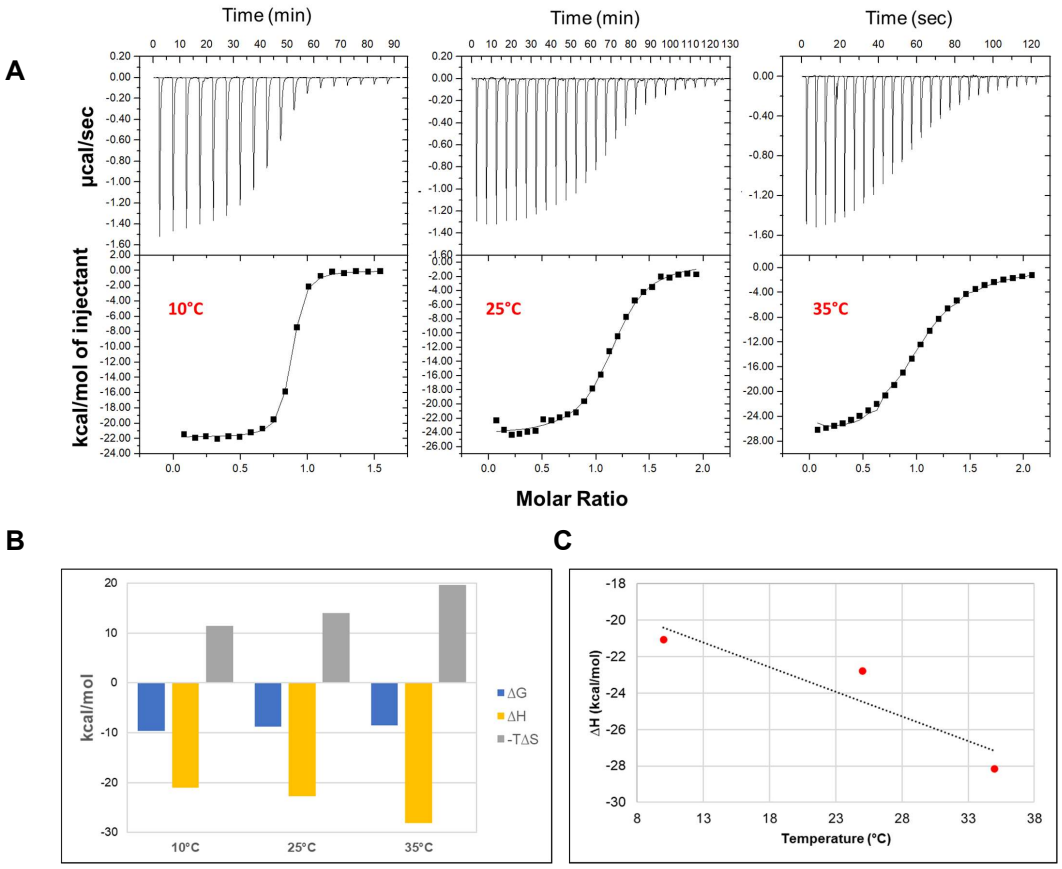


Figure 3.21: (**A**) Calorimetric titrations of VirB7$^{\mathrm{Nt}}$ with VirB9$^{\mathrm{Ct}}$ at 10, 20, and 35°C. (**B**) Thermodynamic values obtained from the fittings of the binding isotherms shown in A to a one binding site model. (**C**) Plot of ΔH as a function of temperature to determine the value of ΔC_{p}.

3.3 The VirB9^{Ct}-VirB7^{Nt} binding mechanism

3.3.1 Insights from NMR spectroscopy

Additional unbound states are in equilibrium with the bound VirB9^{Ct}

We observed that depending on the location on the structure, VirB9^{Ct} residues displayed characteristic CEST profiles reflecting the more 'dynamic' or 'structured' regions of the protein. Specifically, some residues displayed a single CEST dip, indicating that they do not experience chemical shift changes upon binding to VirB7^{Nt}, and hence they are located far away from the binding interface. Other residues displayed 2 dips indicating that they undergo large chemical shift changes upon binding and experience slow exchange equilibrium between bound and unbound states, and other residues displayed more than two dips, indicating that VirB9^{Ct} populates two or more unbound states as observed for R172 (Fig. 3.22). In all cases the smaller dips of the unbound states decreased in intensity with the addition of more VirB7^{Nt} peptide, indicating that they are in equilibrium with the bound VirB9^{Ct} (Fig. 3.22). We grouped all the residues displaying similar CEST profiles and mapped the different residue groups on the structure of VirB9^{Ct} in complex (Fig. 3.23). Curiously, residues that displayed two CEST dips, such as N245 and L249, were mainly located near the VirB7^{Nt} binding site. Residues that displayed three CEST dips tend to locate at β_1, β_2, and at the β_1-β_2 inter-strand loop. Finally, residues that did not show conformational exchange were located far from the VirB7^{Nt} binding site, most of them in the antiparallel β_2-β_3-β_5-β_6 β-sheet (Fig. 3.23). The above suggests that residues located in the VirB7^{Nt} binding site, particularly at the β_1, display complex dynamics in the free state.

Dynamics of the unbound VirB9^{Ct}

Having found that three or more unbound VirB9^{Ct} states were in equilibrium with the bound state, we decide to characterize the slow time scale dynamics of the unbound protein by CEST. To this end we carried out ^{15}N CEST experiments of VirB9^{Ct} alone. We observed

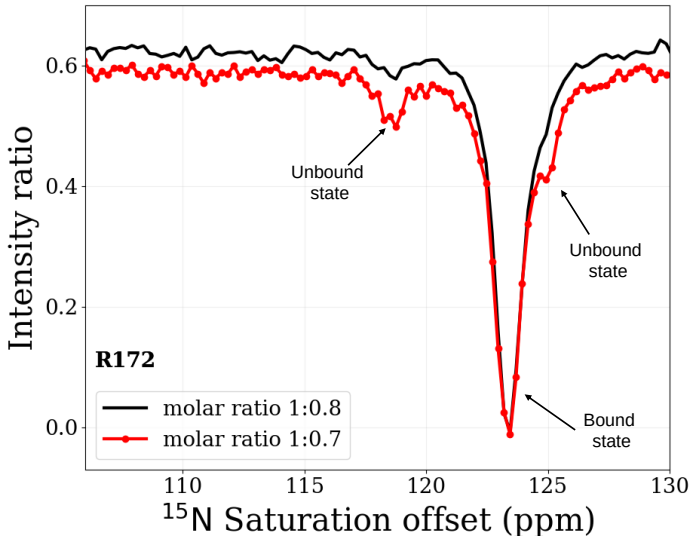


Figure 3.22: ^{15}N CEST experiment of R172 in a sample of VirB9^{Ct} containing sub-stoichiometric VirB7^{Nt} concentrations at the 1:0.7 molar ratio (red line) and 1:0.7 molar ratio (black line).

two types of behavior: on one hand some residues showed more than two dips in their CEST profiles, indicating the occurrence of slow conformational exchange at the millisecond time scale (see Fig. 3.24, top residues Y164 and S176), on the other hand other residues displayed a single CEST dip indicating the lack of conformational fluctuations at this time scale (Fig. 3.24, top residue H214). When mapping the two residue groups on the structure of the complex we did not observe a clear trend with respect to the proximity to the VirB7^{Nt} binding site. However, residues showing slow time scale dynamics tend to group in the center of the antiparallel β-sheet that forms the VirB7^{Nt} binding site (Fig. 3.24), and a few of them were found in the opposite β-sheet. Despite few exceptions, most residues located on the edges of the β strands did not show conformational exchange, suggesting that these regions are rigid, and that the main beta-sandwich structure could be preserved in the absence of VirB7^{Nt}. Curiously, the structure distribution of residues displaying complex dynamics is not the same as that observed on the CEST experiment carried out with the complex.

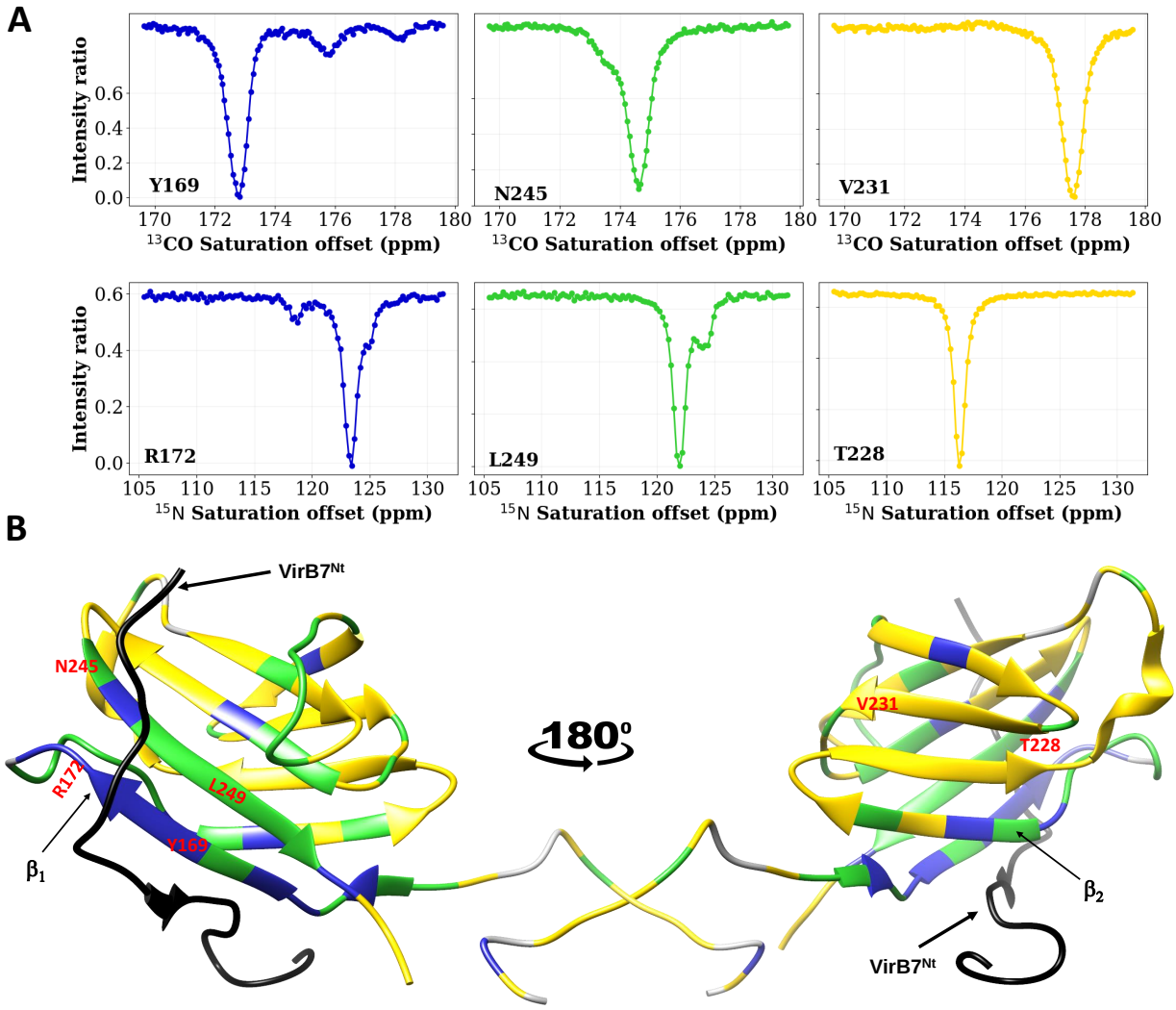


Figure 3.23: ^{13}CO and ^{15}N CEST experiments of a sample of VirB9^{Ct} containing sub-stoichiometric VirB7^{Nt} concentrations at the 1:0.7 molar ratio, showing in (**A**) the CEST profiles for selected residues of VirB9^{Ct} displaying three states (blue), two states (Lime green), one state (yellow) and in (**B**) their location on the structure of VirB9^{Ct} in complex.

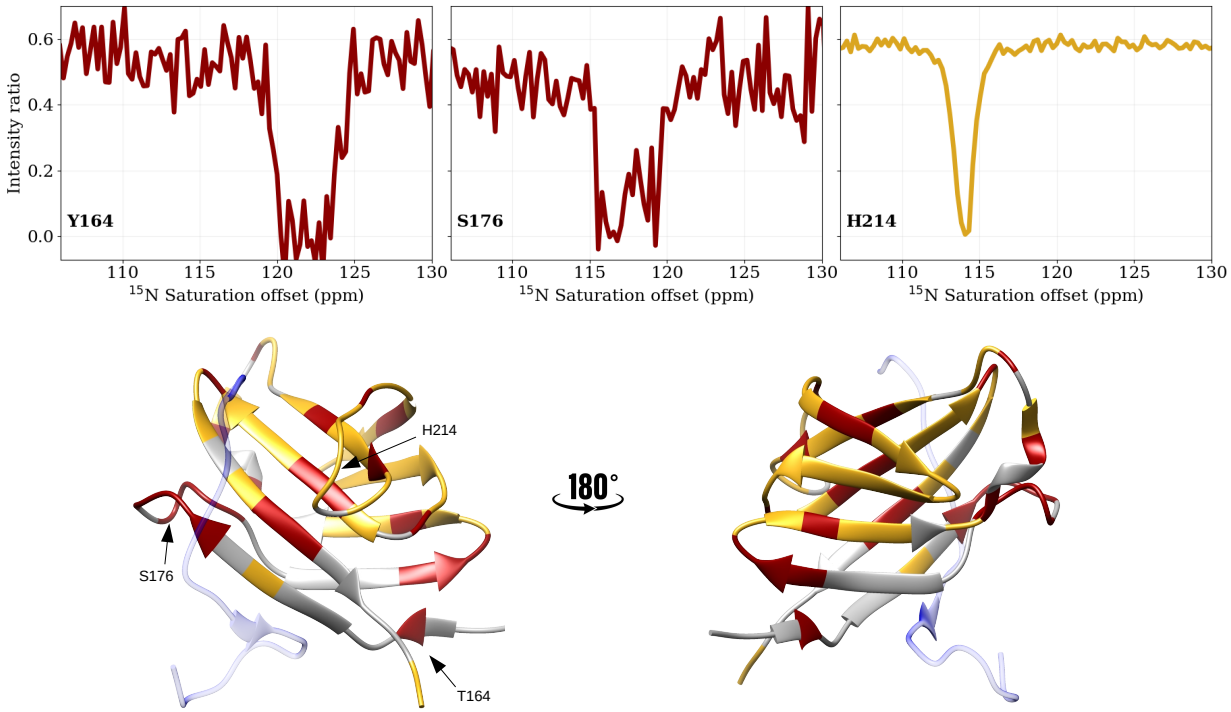


Figure 3.24: ^{15}N CEST profiles of unbound VirB9^{Ct} depicting the profiles from Y164 and S176 that display several dips, and from H214 displaying a single dip (**Top**). Residues displaying a single dip are colored yellow, and those with several dips are colored maroon on the structure of VirB9^{Ct} in complex (**Bottom**). The VirB7^{Nt} peptide is shown subtly in blue to indicate the binding site. Residues that could not be assigned at 35°C are shown in gray.

The residues that displayed complex dynamics in their CEST profiles, also showed broad line widths in the HSQC spectrum, hence their CEST profiles were highly noisy. Improvement of the signal-to-noise ratio in the CEST experiment was achieved by decreasing the temperature to 28°C. The CEST experiment carried out at lower temperature showed profiles with more defined dips compared to those observed at 35°C (see Fig. 3.25). As observed for T173 (Fig. 3.25), at 28°C we may differentiate three conformational states that were otherwise difficult to identify at 35°C. In addition, lowering the temperature slows down conformational exchange rates, leading to a shift of the equilibrium towards a predominant conformation as observed in the ^{15}N-^{1}H HSQC experiment recorded at 7°C. Similarly, the unbound states of S176 and V239 of free VirB9^{Ct} became better resolved by ^{15}N CEST at 28°C compared to 35°C (Fig. 3.25).

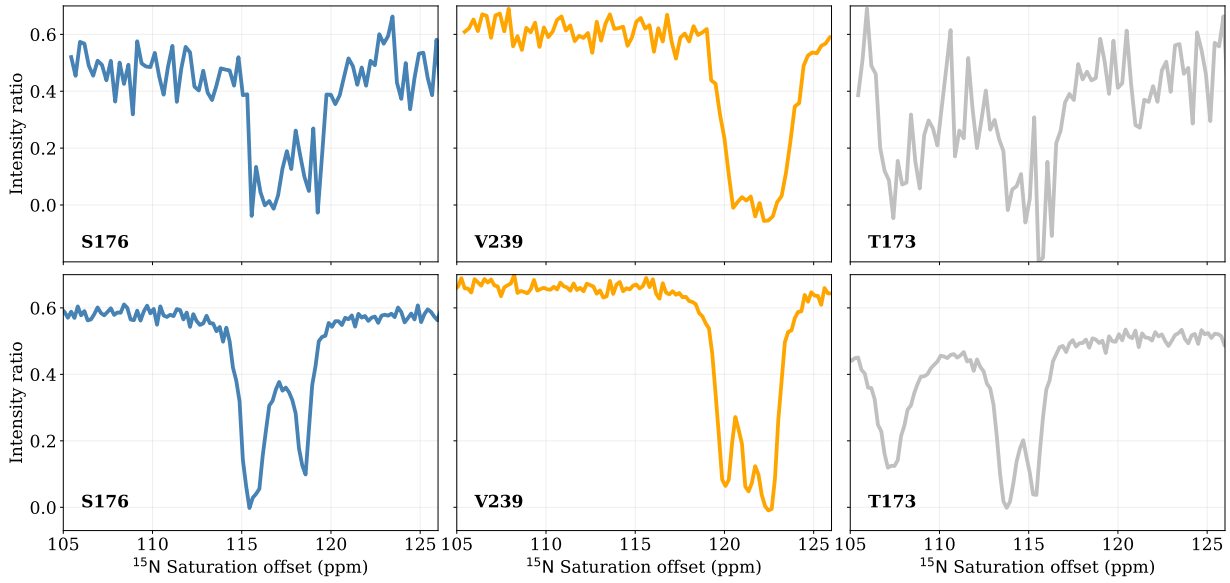


Figure 3.25: Comparison of the ^{15}N CEST profiles of unbound VirB9^{Ct} at 35 °C (**Top**) and at 28 °C (**Bottom**.)

We did not detect the exchange equilibrium between unbound and bound VirB9^{Ct} at 28 °C, only at 35 °C (Fig. 3.26). A possible explanation is that at lower temperatures, the binding affinity increases as a consequence of a higher k_{on} or a lower k_{off}, which decreases k_{ex}^{b} making the exchange too slow to be observed by CEST.

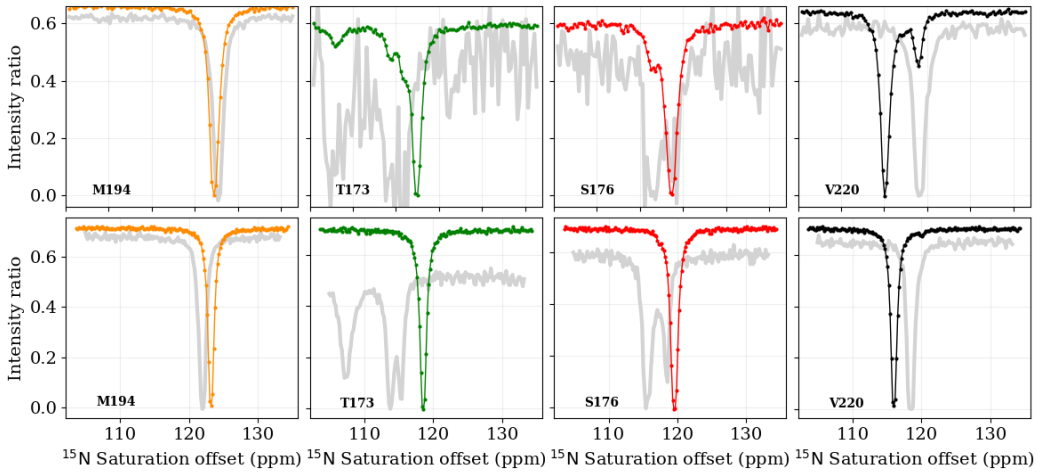


Figure 3.26: Comparison of the ^{15}N CEST profiles of unbound and bound VirB9^{Ct} at 35 °C (**Top**) and at 28 °C (**Bottom**). The ^{15}N CEST profiles of unbound VirB9^{Ct} are shown in gray, while for bound VirB9^{Ct} they are colored.

VirB9^{Ct} samples the bound conformation

In order to evaluate if the unbound VirB9^{Ct} will visit the bound conformation even in the absence of VirB7^{Nt}, we compared the ^{15}N CEST profiles obtained with samples of VirB9^{Ct} alone and samples containing sub-stoichiometric VirB7^{Nt} concentrations at 35°C. We compared the profiles of 58 VirB9^{Ct} residues for which it was possible to obtain ^{15}N CEST data with the bound and unbound samples. We clustered them according to the CEST effect observed in the presence and absence of VirB7^{Nt}. Some residues, exemplified by M194 and T228 (Figs 3.26-3.27), displayed a single CEST dip in the absence and presence of VirB7^{Nt} and at the same or nearly the same ^{15}N chemical shift, indicating that they did not experience chemical shift changes at the millisecond timescale observable by CEST. These residues will be classified as type I. They were located on the β strand edges and far from the VirB7^{Nt} binding site (Fig. 3.27). Other residues (type III) exemplified by V220 (Fig. 3.26) and F237 (Fig. 3.27), displayed a single conformational state in the absence of VirB7^{Nt} (a single CEST dip), and a single unbound state in its presence (two CEST dips). Clearly, type III residues

are relatively rigid in the unbound state, but experienced a change in their ^{15}N chemical shift upon binding to VirB7^{Nt} due to the proximity with the peptide aromatic residues, or to a change in the protein conformation. Type III residues are mainly located in the antiparallel β_1-β_4-β_7-β_8 β-sheet, over which VirB7^{Nt} lies on (Fig. 3.27). Residues classified as type II, IV, and V have in common several unbound states indicated multiple CEST dips, or a broadened CEST dip, in the ^{15}N CEST profiles of the unbound VirB9^{Ct} sample. The difference between them lies on their CEST profile observed in the presence of VirB7^{Nt}. Type V residues, such as S176 (Fig. 3.26) and Y164 (Fig. 3.27), display two or more CEST dips in the unbound sample, one of which coincides with that of the unbound state in the presence of VirB7^{Nt}. This observation is a strong suggestion that the unbound VirB9^{Ct} samples the "bound" conformation even in the absence of the peptide. Type II residues exemplified by R210 (Fig. 3.27) are mostly are located at the VirB9^{Ct} antiparallel β_2-β_3-β_5-β_6 β-sheet, opposite to the VirB7^{Nt} binding site. Despite their dynamics in the unbound state, binding to VirB7^{Nt} does not lead to a significant ^{15}N chemical shift change (Fig. 3.27). In contrast, type IV residues exemplified by V239 (Fig. 3.27) and T173 (Fig. 3.26), experienced slow conformational exchange in the unbound state but binding to VirB7^{Nt} caused a significant ^{15}N chemical shift change. Type IV residues are located in the VirB7^{Nt} binding region.

The observation that some residues display multiple conformations in the unbound state, one of which shares the same ^{15}N chemical shift as the bound state (type V), suggests that VirB9^{Ct} binds VirB7^{Nt} according to a conformational selection mechanism. On the other hand, the observation that other residues displaying one or multiple conformations in the unbound state, do experience a ^{15}N different chemical shift upon binding (types III and IV), suggests a IF mechanism. How to reconcile these two different observations? Perhaps one alternative is to consider that the ^{15}N chemical shift of the bound VirB9^{Ct} contains not only the contribution of the secondary structure of the protein, but also the contribution of the chemical environment conferred by the proximity with the peptide, such as ring current effects. An accurate analysis of CS versus IF based on VirB9^{Ct} chemical shifts, will require

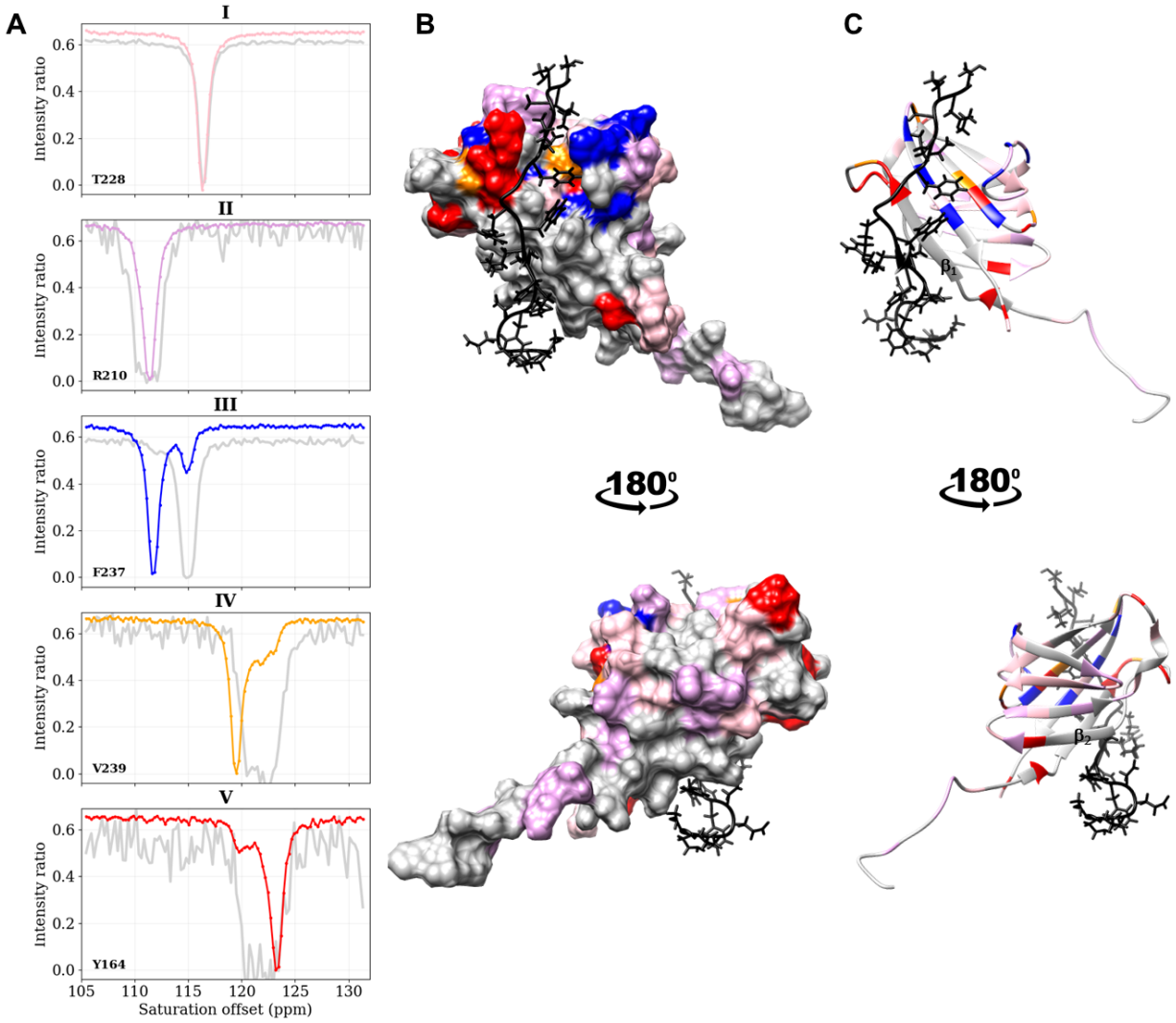


Figure 3.27: Residue classification according to their ^{15}N CEST profile in the bound and unbound states, showing the five types of VirB9^{Ct} residues (**A**) and their localization on the structure of VirB9^{Ct} in complex, shown in surface (**B**) and ribbon (**C**) representations.

the separation of these two contributions.

3.3.2 The kinetics of VirB7^{Nt} - VirB9^{Ct} association

Can we use the Tryptophan fluorescence to follow VirB7^{Nt} binding to VirB9^{Ct}?

The intrinsic tryptophan fluorescence is highly sensitive to the local environment[128]. VirB7^{Nt} and VirB9^{Ct} have each a single W residue. The VirB7^{Nt} and the VirB9^{Ct} W residues are located at approximately 17 Å from each other in the structure of the complex; hence, they experience different chemical environments. While the VirB9^{Ct} W177 is located in the β_1-β_2 inter-strand loop and exposed to the solvent, the VirB7^{Nt} W34 is buried in a pocket on the surface of one of the two antiparallel β-sheets from VirB9^{Ct}, the one composed by strands β_1-β_8-β_7-β_4 (Fig.3.28) [16]. It is noteworthy that the substitution W34A abolished binding of VirB7^{Nt} to VirB9^{Ct}, preventing the T4SS assembly and causing loss of bacterial killing activity via the T4SS$_{XAC}$ [16]. Therefore, W34 is essential for the interaction with VirB9^{Ct}.

Considering that the indole ring fluorescence is sensitive to the surrounding environment, conformational changes, ligand binding and protein denaturation, could change the emission spectra of tryptophan [128]. Thus, the tryptophan fluorescence could be an excellent tool to monitor the interaction between VirB7^{Nt} and VirB9^{Ct}. Indeed, binding of VirB7^{Nt} to VirB9^{Ct} leads to an increase in the tryptophan fluorescence as shown in Fig. 3.29.

Given that W34 (VirB7^{Nt}) is located at the complex interface, while W177 (VirB9^{Ct}) is located further away (Fig. 3.28), we considered that W34 should be the best probe to monitor the binding interaction. Thus, to selectively excite one of the two tryptophan residues, we replaced W177 by the analogue 5-hydroxy-tryptophan (5OHW) (Fig. 3.30-B). The advantage of using 5OHW is that it can be specifically excited at 310-320 nm (Fig. 3.30-C) even in a tryptophan background environment [83]. To incorporate 5OHW in VirB9^{Ct}, we expressed this protein in a Trp-auxotropic *E.coli* strain in the presence of 5OHW as tryptophan source. To test whether the W177 substitution for the 5OHW

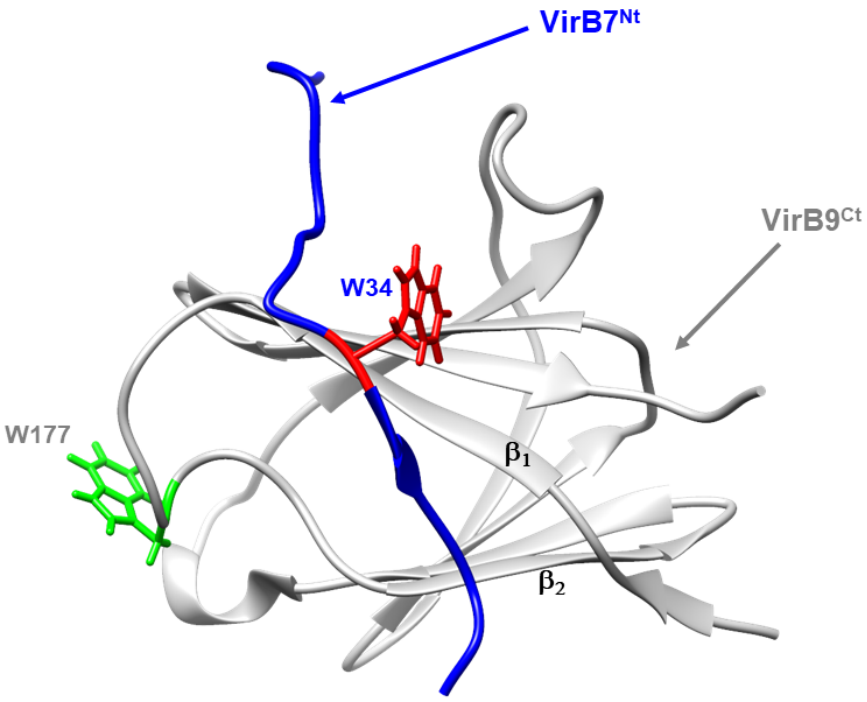


Figure 3.28: NMR solution structure of the complex showing VirB9^{Ct} in gray and VirB7^{Nt} in blue. The tryptophan residues of VirB9^{Ct} (W177) and VirB7^{Nt} (W34) are colored green and red, respectively.

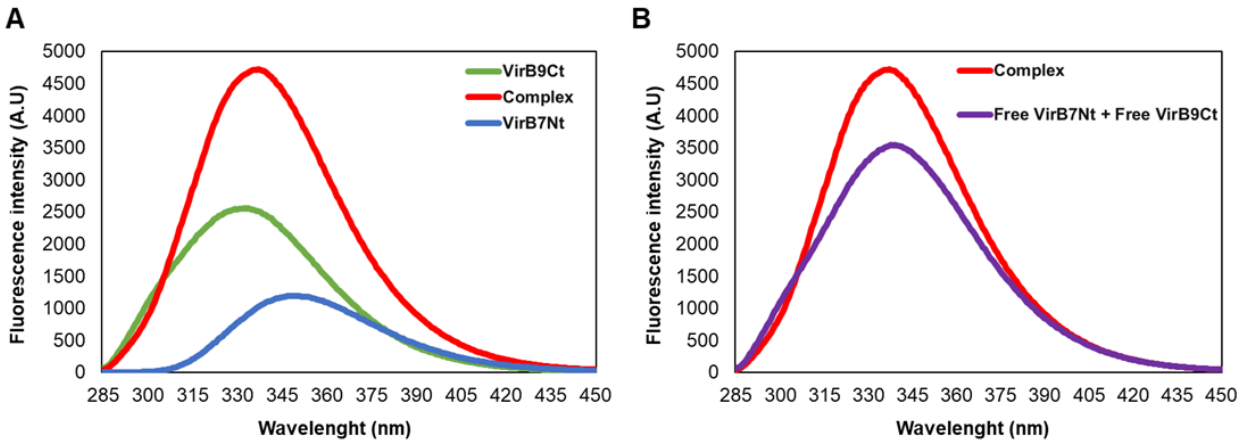
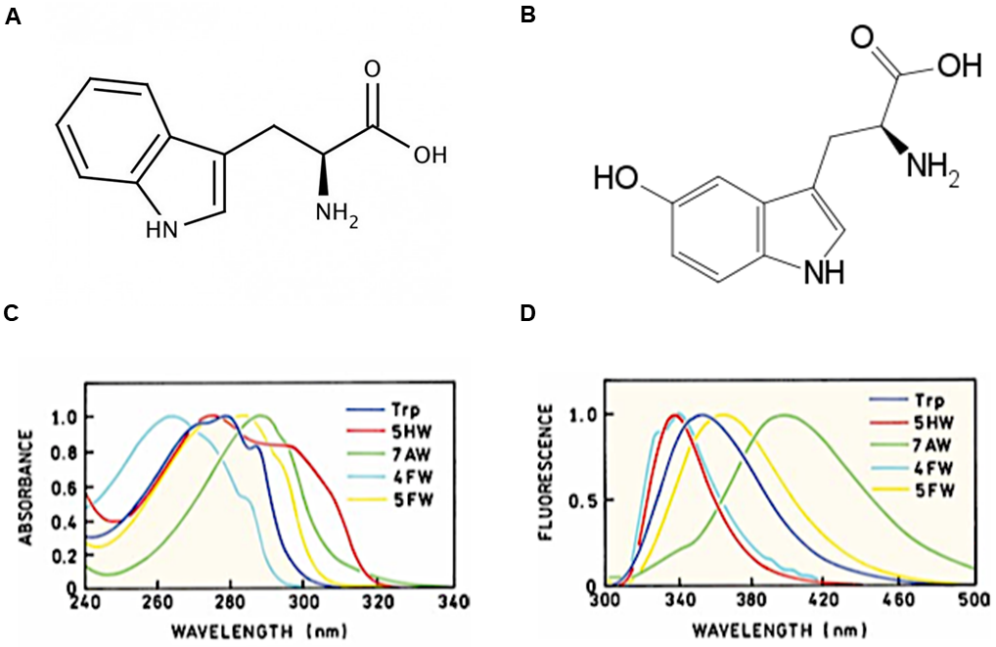


Figure 3.29: The tryptophan fluorescence spectrum changes when VirB7^{Nt} binds to VirB9^{Ct} (**A**) Superposition of the fluorescence spectra from W34 in VirB7^{Nt} (blue), W177 in VirB9^{Ct} (green), and from the complex (red). (**B**) Comparison of the fluorescence spectrum of the complex (red) with the sum of the fluorescence spectra of VirB7^{Nt} and VirB9^{Ct} (purple).

analogue affected the structure of VirB9^{Ct} or the affinity for VirB7^{Nt}, we performed CD and ITC measurements, and compared with the same measurements done with the wild type protein (Fig.3.31). VirB9^{Ct}-5OHW displayed a CD spectrum similar to that of unbound VirB9^{Ct}. The dissociation equilibrium constant (K_d) obtained by ITC was 0.72, consistent with that obtained for the wild type VirB9^{Ct} binding to VirB7^{Nt} (Fig.3.31 and Table 3.2).

To verify the incorporation of 5OHW into VirB9^{Ct}, the fluorescence emission spectra of the protein was recorded using a range of excitation wavelengths. In contrast with VirB7^{Nt}, VirB9^{Ct}-5OHW emitted fluorescence when excited not only at 280-295 nm but also at 310-320 nm, confirming the presence of 5OHW (Fig. 3.32-A and Fig. 3.32B). The same behavior was observed for the complex formed by VirB7^{Nt} and VirB9^{Ct}-5OHW (Fig. 3.32-B and Fig. 3.32-C).

Comparison of the fluorescence spectra of the unbound VirB9^{Ct} and the VirB7^{Nt}-VirB9^{Ct} complex recorded at the same protein concentrations, showed that the incorporation of



htp

Figure 3.30: Structure of (**A**) L-tryptophan and (**B**) 5-hydroxy-tryptophan. Absorbance (**C**) and emission (**D**) spectra of the tryptophan and some of its analogues. Pictures taken from Principles of Fluorescence Spectroscopy. [128]

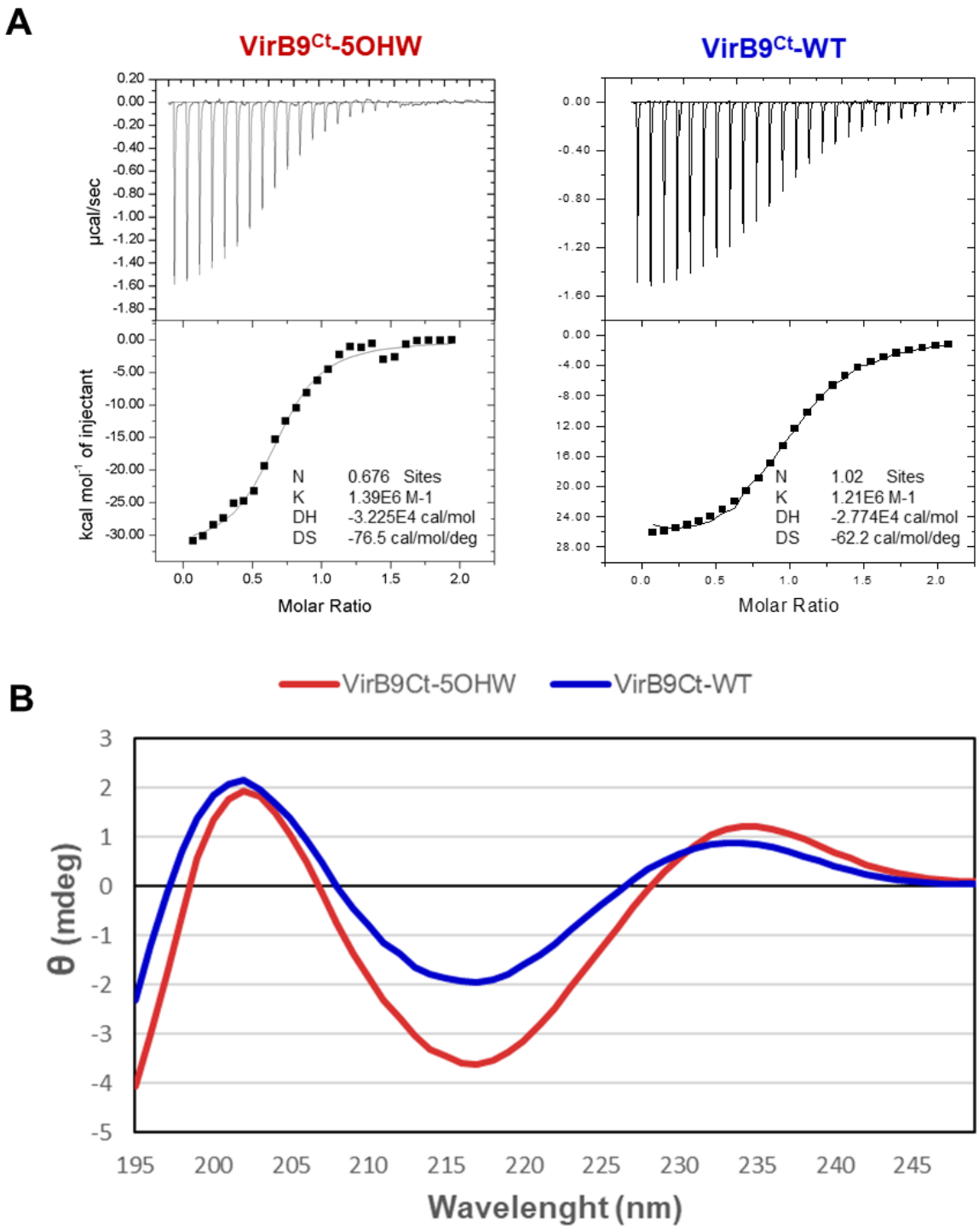


Figure 3.31: Characterization of the VirB7^{Nt} binding affinity by ITC from VirB9^{Ct}-5OHW or VirB9^{Ct} (**A**). Comparison of the CD spectrum from the unbound VirB9^{Ct} without (blue) and with 5OHW (red) incorporated at position 177 (**B**).

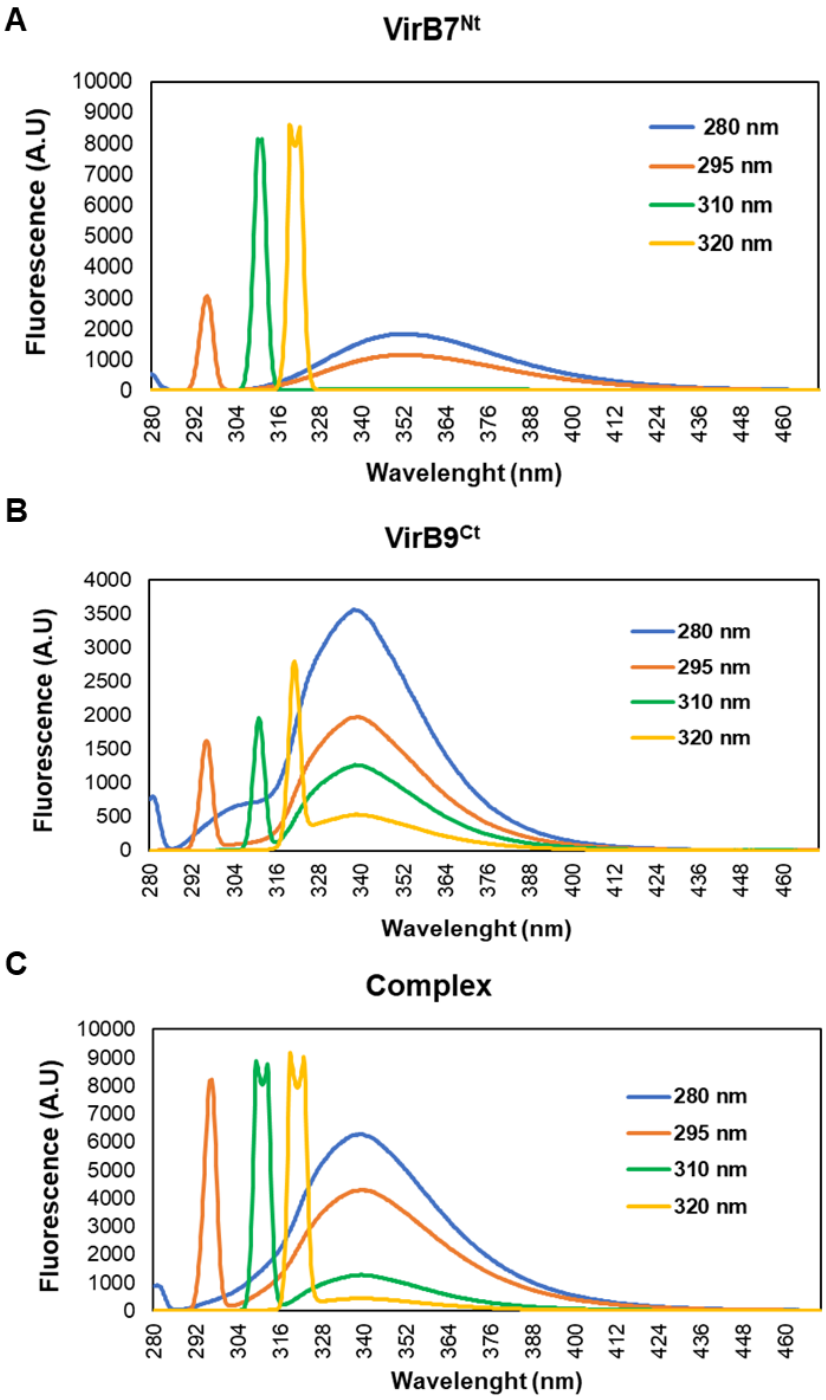


Figure 3.32: Emission fluorescence spectra of (**A**) VirB7^{Nt}, (**B**) VirB9^{Ct}-5OHW and (**C**) the VirB7^{Nt}: VirB9^{Ct}-5OHW complex using different excitation wavelengths: 280 nm (blue), 295 nm (orange), 310 nm (green), and 320 nm (yellow).

5OHW into VirB9^{Ct} allowed the selective excitation of the tryptophan at position 177. Unfortunately, the fluorescence spectrum of 5OHW at position 177 does not seem to be affected by the binding of VirB9^{Ct} to VirB7^{Nt} (Fig. 3.33), and, consequently, the 5OHW probe at this position is not a useful tool to monitor the interaction. Considering that the 5OHW fluorescence spectra is independent of the excitation wavelength, the spectrum change observed upon VirB7^{Nt} binding to VirB9^{Ct} with the λ_{ex} = 280 nm (Fig. 3.33) could be attributed to the polarity change in the surroundings of W34. The fact that VirB7^{Nt} tryptophan is affected by the presence of VirB9^{Ct}, leading to an increase in the intensity of fluorescence upon complex formation, is consistent with its location at the complex interface, while W177 is exposed to the solvent and less susceptible to the binding interaction. Therefore, we used the excitation wavelength at 280 nm to monitor the interaction of VirB7^{Nt} with VirB9^{Ct}, since the local environment of VirB7^{Nt} tryptophan is perturbed during complex formation, which can give us information about the interaction mechanism.

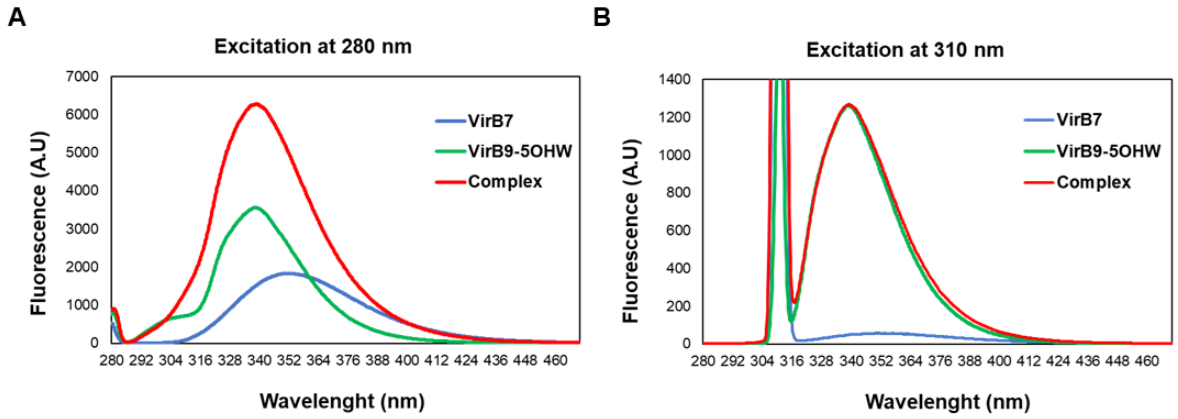


Figure 3.33: Fluorescence emission spectra of VirB7^{Nt} (blue), VirB9^{Ct}-5OHW (green) and complex (red) recorded using excitation wavelenghts at (**A**) 280 nm or (**B**) 310 nm.

The kinetics of VirB7^{Nt} - VirB9^{Ct} association at 25°C

We measured the association kinetics between VirB7^{Nt} and VirB9^{Ct} at 25°C by stopped-flow fluorescence. Upon mixing the two proteins under pseudo first order conditions for VirB7^{Nt} or VirB9^{Ct}, the time course of the fluorescence intensity change was monitored using $\lambda_{ex} = 280$nm and $\lambda_{em} = 350$nm as shown in Fig. 3.34.

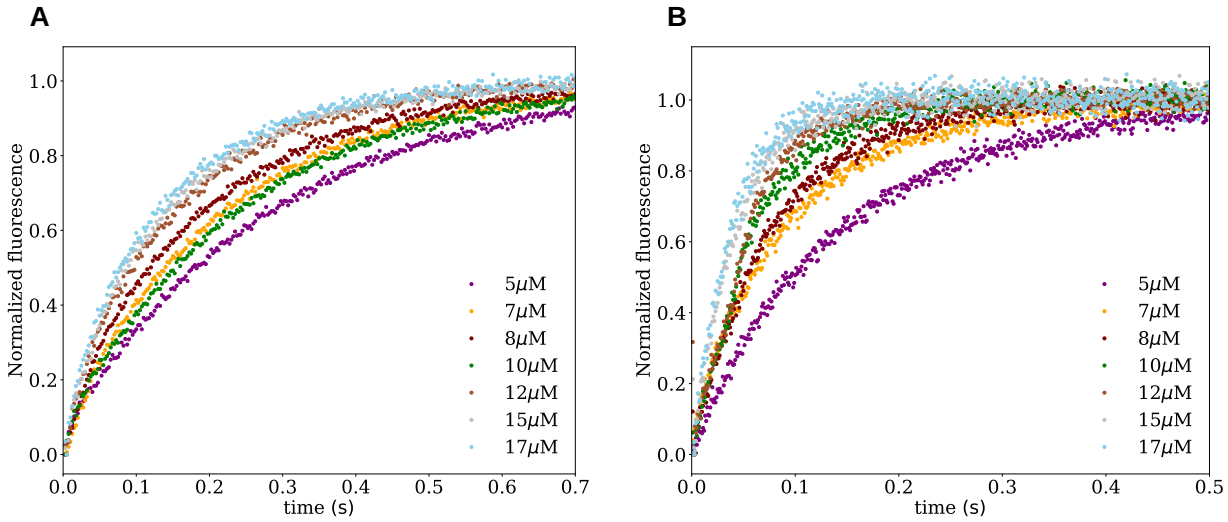


Figure 3.34: Stopped-flow kinetic traces at 25°C showing the fluorescence intensity change over the time course of the reaction. The different colors indicate the concentrations of VirB7^{Nt} in the presence of 1 μM of VirB9^{Ct} (pseudo-first order condition for VirB7^{Nt}) **(A)** or the concentrations of VirB9^{Ct} in the presence of 1 μM VirB7^{Nt} (pseudo-first order condition for VirB9^{Ct}) **(B)**. $\lambda_{ex} = 280$ nm; $\lambda_{em} = 350$ nm. The fluorescence intensities were normalized according to the value of the plateu.

The VirB9^{Ct} concentration ([P]) was kept constant, while the concentration of VirB7^{Nt} ([L]) was increased under pseudo-first order conditions (large excess of VirB7^{Nt} over VirB9^{Ct}) (Fig. 3.34A). Fitting of the experimental traces to mono and bi-exponentials (Eqns. 1.2 and 1.10) indicated that the association kinetics is biphasic under our experimental conditions (Fig. 3.35, A and B), an observation that is consistent with coupled folding and binding

events [22].

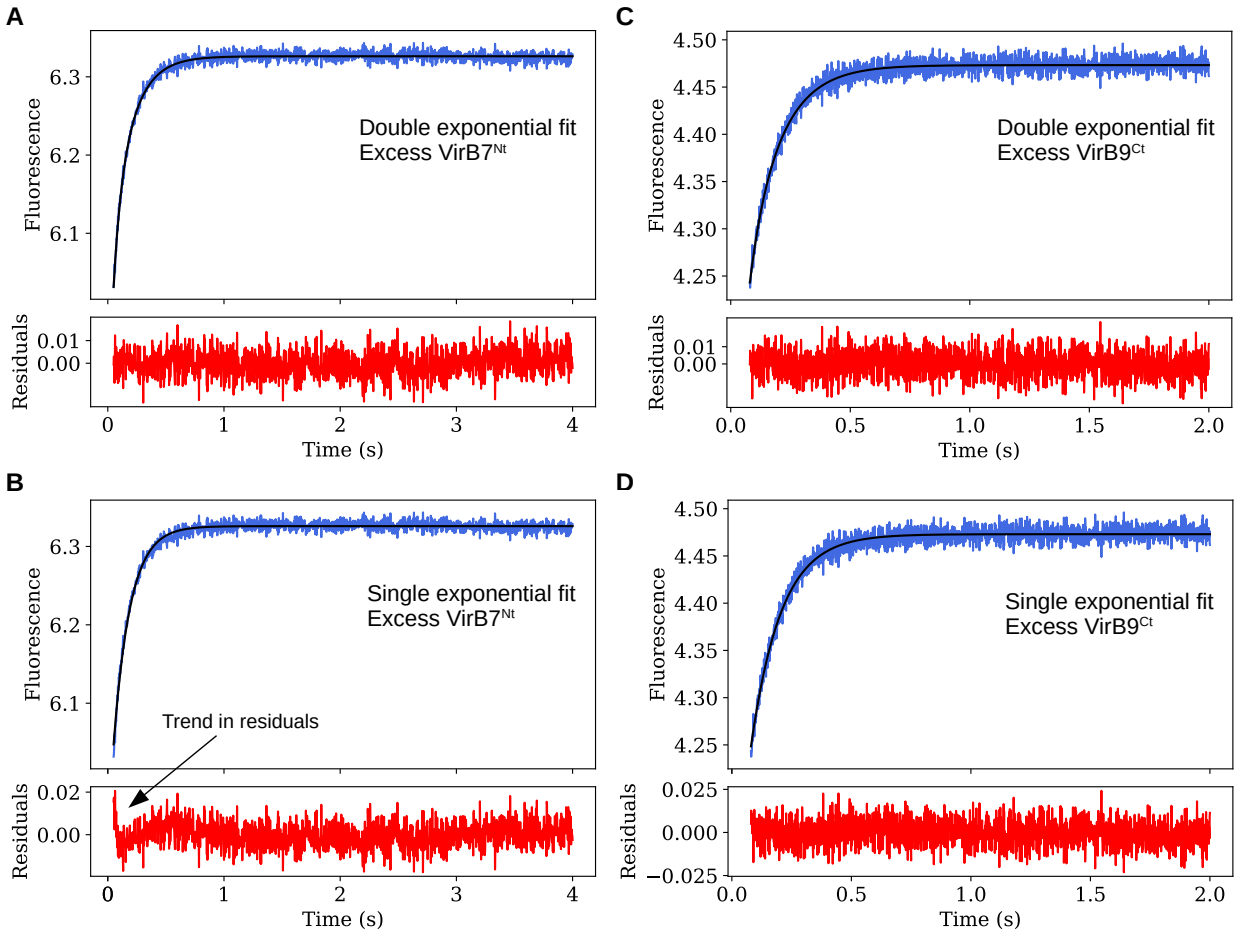


Figure 3.35: Fitting of the stopped-flow experimental curves obtained at 25°C to Eqns. 1.2 or 1.10. **(A)** and **(B)** shows the same kinetic trace under excess of $VirB7^{Nt}$, while **(C)** and **(D)** shows the same kinetic trace under excess of $VirB9^{Ct}$.

Fittings of the kinetic traces acquired at increasing $VirB7^{Nt}$ or $VirB9^{Ct}$ concentrations to Eq. 1.10 (Figs. 3.34 and 3.35), allowed us to observe the behavior of the two kinetic rate constants, k^1_{obs} and k^2_{obs}, as a function of the free ligand concentration (or the complex molar ratio). This analysis showed a ill resolved kinetic fast phase (k^1_{obs}) that increased with the ligand concentration ([L]) (Fig. 3.37-B), and a slower kinetic phase (k^2_{obs}) that increases hyperbolically with [L] (Fig. 3.37-A). This scenario, the appearance of two kinetic phases that

increase with the free ligand concentration can be equally described by the conformational-selection (CS) or induced-fit (IF) mechanisms as shown in Fig. 1.4 and discussed in the "Background theory" chapter, section Kinetics of protein-protein interactions. The discrimination between these two mechanisms is only possible for the CS case when the slow phase (k^2_{obs}) is independent or decreases with [L] (Fig. 1.4). [55].

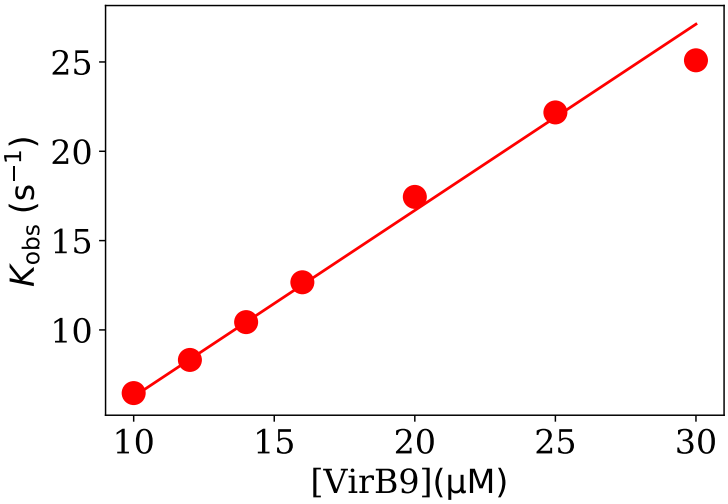


Figure 3.36: Observed rate constant (red dots) at 25°C as a function of the complex molar ratio, determined under conditions of excess of VirB9$^{\text{Ct}}$. A linear fit of this curve (shown as a red line) yielded the value of $k_{\text{on}} = 2.08 \pm 0.03\ \text{s}^{-1}\text{M}^{-1}$

As discussed in the Background theory chapter, subsection 1.1.2, the ambiguity between the CS and the IF mechanisms can be solved by performing the same kinetic experiment under reverse pseudo-first order conditions, i.e varying the concentration of VirB9$^{\text{Ct}}$ while keeping the VirB7$^{\text{Nt}}$ concentration constant. As explained earlier (see section 1.1), in the conformational selection mechanism, a pre-equilibrium of conformations is sampled by the macromolecule before ligand binding, hence, only a single kinetic phase corresponding to the binding step will be observed when the kinetics experiment is performed under conditions of excess of the macromolecule (reverse pseudo-first order conditions). In contrast, in the induced-fit mechanism, the two kinetic rate constants should display the same behavior

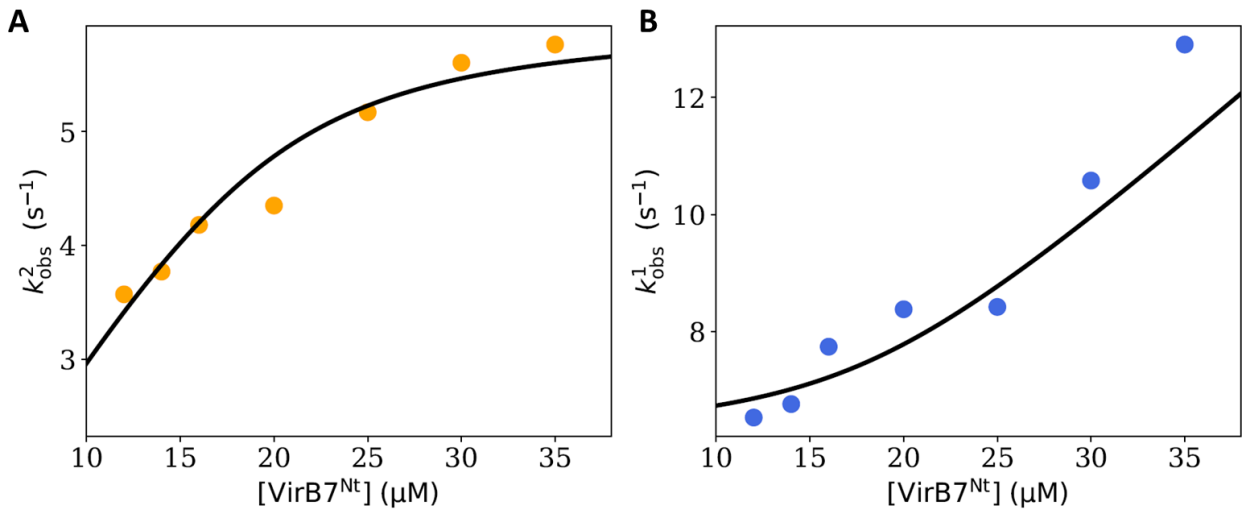


Figure 3.37: Observed rate constants at 25°C as a function of the complex molar ratio, determined under conditions of excess of VirB7^{Nt}. The slow (k^2_{obs}) and fast (k^1_{obs}) phase observed are shown in **(A)** as orange dots and **(B)** as blue dots, respectively. Simultaneous fitting of these curves (shown as continuous lines in black) using Eq. 1.17 for the CS model, yielded values for the rate constants k_{-r}, k_r, k_{off} and k_{on}, shown in table 3.3.

as a function of the ligand and the macromolecule concentrations, i.e. under the pseudo-first order or reverse pseudo-first order conditions. This statement holds true because the conformational rearrangements in the IF mechanism occur after the binding event, so there is really no difference whether VirB7$^{\mathrm{Nt}}$ or VirB9$^{\mathrm{Ct}}$ is in excess during the experiment (see Background theory, subsection 1.1.2).

Therefore, to disambiguate the binding mechanism between VirB7$^{\mathrm{Nt}}$ and VirB9$^{\mathrm{Ct}}$ at 25 °C, we repeated the stopped-flow experiment under reverse pseudo-first order conditions (Fig. 3.34B). Analysis of the kinetic traces yielded a single rate constant (Fig. 3.35, C and D) that increased linearly with protein concentration ([P]) as shown in Fig. 3.36. This behavior indicates that in excess of VirB9$^{\mathrm{Ct}}$ only the binding step is observed, in agreement with a conformational-selection mechanism.

This linear kinetic curve was analysed quantitatively using Eqn. 1.7, yielding a value of $k_{\mathrm{on}} = 2.08 \pm 0.03\ \mathrm{s^{-1}M^{-1}}$, assuming the simplest model involving just a binding step (Scheme 1.5). The intercept in the "y" axis corresponds to the value of k_{off}, which in our fit yielded a value of -4.157 $\mathrm{s^{-1}}$. As we know, a negative value for a rate constant does not make sense, moreover we can not trust in this value of k_{off} since the intercept would be experimentally accurate only if it was possible to follow the reaction in the range where the protein concentration in excess (in this case VirB9$^{\mathrm{Ct}}$) is comparable to the apparent equilibrium dissociation constant $K_{\mathrm{d}}^{\mathrm{app}}$, which for our system is around $\sim 0.4 - 0.7\ \mu\mathrm{M}$. Therefore, in this case the value of k_{off} is poorly estimated from the intercept of the straight line, and we should perform displacement experiments in order to measure an accurate value of k_{off} (see reference [50], pages 30-33).

We may estimate the value of the other rate constants of the CS model by inspecting both phases in excess of VirB7$^{\mathrm{Nt}}$ at their limiting values. According to Table 1.1 subsection 1.1.2, in the CS model, the upper limit of the slowest phase ($k^2_{\mathrm{obs}}(\infty)$) unequivocally defines k_{r}, but the lower limit depends on the value of k_{off} and the sum of $k_{\mathrm{r}} + k_{-\mathrm{r}}$ [54]. Since our slowest relaxation k^2_{obs} increases with [L], it implies the condition of $k_{-\mathrm{r}} + k_{\mathrm{r}} > k_{\mathrm{off}} < k_{\mathrm{r}}$

and, therefore, the lower limit for $k^2_{\rm obs}(0)$ tends to $k_{\rm off}$ [55]. Regarding $k^1_{\rm obs}$, the upper limit $k^1_{\rm obs}(\infty)$ grows linearly with [L] as $k_{\rm on}$[L] and the lower limit $k^1_{\rm obs}(0)$ approaches the value of $k_{-\rm r} + k_{\rm r}$ (See Table 1.1 in the subsection Conformational-selection in the general case) [54, 55]. In our case and accordingly to these limits, we can estimate that $6.0 \leq k_{\rm r} \leq 7.0$, $0.0 \leq k_{\rm off} \leq 3.0$ and $0.0 \leq k_{-\rm r} \leq 1.0$ since $k_{-\rm r} + k_{\rm r} \approx 7.0$.

To further confirm or assumptions regarding the values of the kinetic constants, we used Eq. 1.17 to model our kinetic curves in excess of ligand to the CS model. Except for the value of $k_{\rm on}$, the fitted values are in agreement with the expected based on the analysis of the limiting values (see Table 3.3). However, we must have in mind that the fast phase in excess of VirB7$^{\rm Nt}$ was low resolved and thus we can not have certainty on the values obtained by the fitting of these curves. However, we can have a general idea of the relative values for the rate constants. Particularly, the value of $k_{\rm off}$ should be around 3.0 $\rm s^{-1}$ according to the inferior limit analysis of $k^2_{\rm obs}$, but our fitting to the CS model yielded a value of 0.42 $\rm s^{-1}$, which is considerably different. However, about the value of $k_{\rm off}$ we can certainly infer that it should be lower than $k_{\rm on}$ and it should be in the range of $0.42\,\rm s^{-1} < k_{\rm off} < 3.0\,\rm s^{-1}$. In contrast, we do have accuracy in the value of $k_{\rm on}$ determined by linear regression analysis of the experimental $k_{\rm obs}$ obtained in excess of VirB9$^{\rm Ct}$.

We may calculate the intrinsic $K^{\rm int}_{\rm d}$ value for the interaction between VirB7$^{\rm Nt}$ and VirB9$^{\rm Ct}$ at 25°C through Eq. 1.31, and using the value of $K^{\rm app}_{\rm d}$ from ITC at 25°C ($\sim 0.4\mu$M). Our calculations indicate that at conditions where $k_{\rm r} \gg k_{-\rm r}$ the value of $K^{\rm app}_{\rm d} \approx K^{\rm int}_{\rm d}$. Thus, $K^{\rm int}_{\rm d} \approx 0.4\,\mu$M and therefore the value of $k_{\rm off}$ defined as $K^{\rm int}_{\rm d} = k_{\rm off}/k_{\rm on}$ is $\sim 0.76 - 0.83\,\rm s^{-1}$, which is within the range expected for $k_{\rm off}$.

Recalling our NMR CEST data at 28°C, we see that in the absence of VirB7$^{\rm Nt}$ we observed exchange between defined states of unbound VirB9$^{\rm Ct}$, while in the presence of the peptide at 1:07 molar ratio, we couldn't see exchange between the bound and unbound states of VirB9$^{\rm Ct}$, and only the bound state was observed. Analyzing together with our kinetic data at 25°C, we can infer now why we didn't observed exchange in our CEST experiments at 25°C in the

Table 3.3: Kinetic rate constants obtained for the interaction between VirB7^{Nt} and VirB9^{Ct} according to the CS using the kinetic data at 25°C. [1]This value was extracted from linear fitting of the relaxation in excess VirB9^{Ct} with the one-binding step model (for details see the main text). The values highlighted in red are poorly estimated by the fitting

Model	T(°C)	$k_r^{\mathbf{P}^*\rightarrow\mathbf{P}}$(s^{-1})	$k_{-r}^{\mathbf{P}\rightarrow\mathbf{P}^*}$(s^{-1})	k_{on}(10^{-6} s^{-1}M^{-1})	k_{off}(s^{-1})
CS	25	≈ 6.00	≈ 0.38	≈ 0.57	≈ 0.42
Lock and key	25	$--$	$--$	2.08[1]	-4.16

presence of VirB7^{Nt} but did at 35°C. For instance, we saw that the value of $K_d^{app} \approx K_d^{int}$ at 25°C is lower than at 35°C (Table 3.2), which translates into an increase in complex affinity, perhaps by the lower value of k_{off} compared to k_{on} at this temperature, preventing conformational exchange from being seen by CEST NMR measurements. Moreover, as a consequence of temperature, k_r (rate constant for the $\mathbf{P}^* \rightarrow \mathbf{P}$ transition) is greater than k_{-r} (rate constant for the $\mathbf{P} \rightarrow \mathbf{P}^*$ transition) at 25°C (see Table 3.3), which means that almost all molecules of VirB9^{Ct} are in the active P state ready to bind ligand ($P^* \rightleftharpoons P + [L] \rightleftharpoons PL$), making the switch to the P* state even harder to occur. Similarly, we can notice that the time constant ($1/(k_{-r} + k_r)$) of the exchange between unbound VirB9^{Ct} states is ≈ 0.157 s, which is within the millisecond time scale for events observable by CEST. The above assumptions shows consistency between our thermodynamic, kinetic and NMR data, and support our conclusions so far.

The kinetics of VirB7^{Nt} - VirB9^{Ct} association at 35°C

To evaluate the VirB7^{Nt}:VirB9^{Ct} association mechanism at the same conditions that were used for NMR spectroscopy, we repeated the stopped-flow kinetic measurements at 35°C.

We mixed the two proteins and monitored the change in fluorescence as a function of time under pseudo-first order conditions as shown in Fig. 3.38.

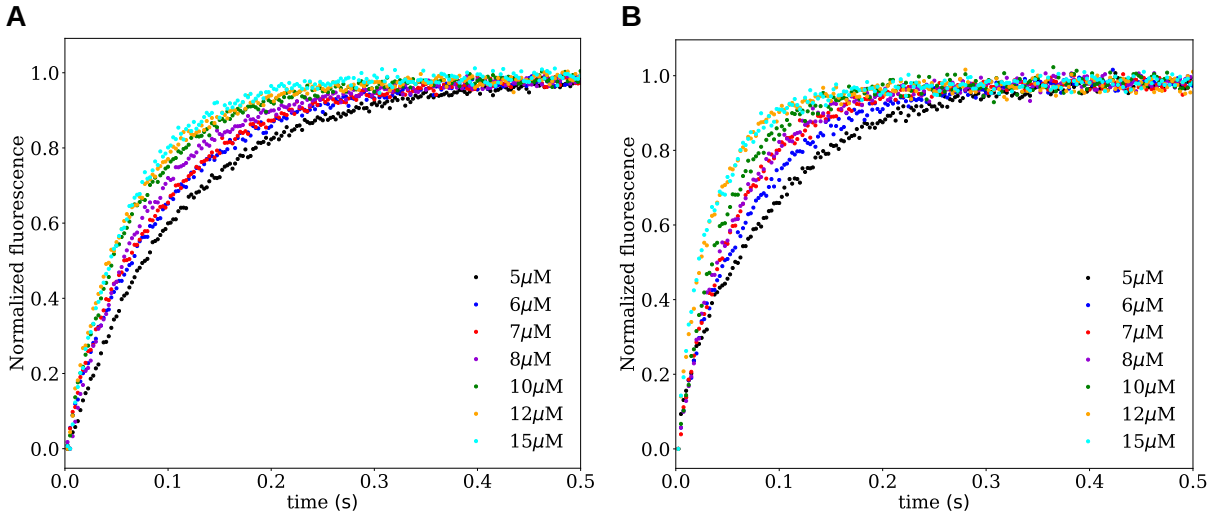


Figure 3.38: Stopped-flow kinetic traces recorded at 35°C showing the change in tryptophan fluorescence intensity over the time course of the binding reaction. The different colors indicate the excess concentration of VirB7$^{\mathrm{Nt}}$ **(A)** or VirB9$^{\mathrm{Ct}}$ **(B)**.

The kinetics was biphasic under conditions of excess of VirB7$^{\mathrm{Nt}}$ (Fig. 3.39A and 3.39B), which suggests that the interaction mechanism involves at least two steps [39, 55, 22]. In this experiment we observed that the faster phase k^1_{obs} increases linearly with the increase in ligand concentration [L], while the slower phase k^2_{obs} decreases hyperbolically with [L] (Fig. 3.40A). The decreasing behavior of k^2_{obs} with the free ligand concentration is a kinetic signature of the conformational-selection mechanism as described in Background theory, section 1.1. The existence of a conformational VirB9$^{\mathrm{Ct}}$ pre-equilibrium is in agreement with the analysis of the unbound VirB9$^{\mathrm{Ct}}$ dynamics by NMR at 25 and 35°C as described previously. However, in contrast with the behavior observed at 25°C (Fig. 3.36), at 35°C the kinetics was biphasic in excess of the macromolecule (VirB9$^{\mathrm{Ct}}$) (Fig. 3.39C and 3.39D). It was characterized by a fast relaxation k^1_{obs} that increases linearly with [P], and a slow relaxation k^2_{obs} that increases with the protein concentration until saturation i.e. hyperbolically)

(Fig. 3.40-B). The observation of a second phase in excess of VirB9^{Ct} was unexpected.

Considering that the behavior in excess of VirB7^{Nt} is consistent with a conformational selection mechanism (Fig. 3.40A), the only possibility to explain the additional kinetic phase observed under excess of protein is to assume the existence of a third event other than the VirB9^{Ct} pre-equilibrium and the binding to VirB7^{Nt}.

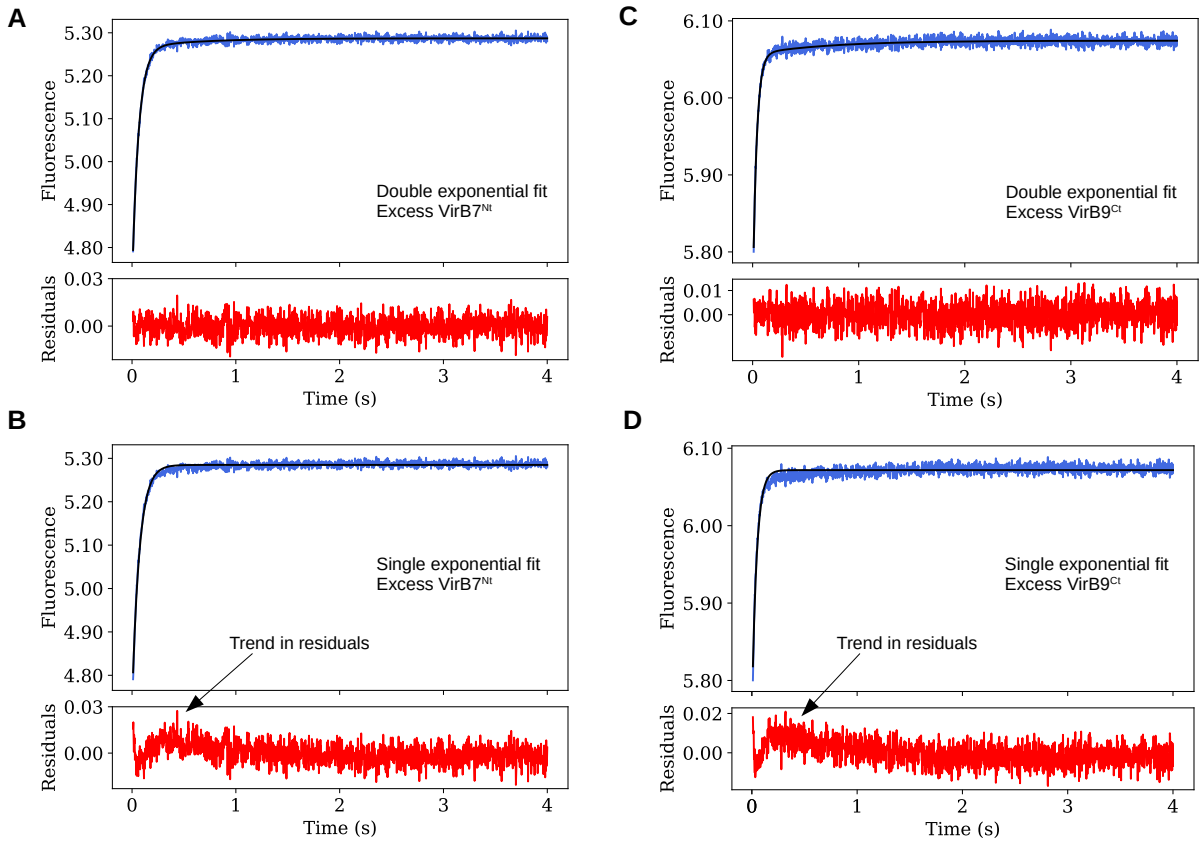


Figure 3.39: Modeling of the stopped-flow experimental curves obtained at 35°C using double and single exponential fits. **(A)** and **(B)** shows the same kinetic trace under conditions of excess of VirB7^{Nt}, while **(C)** and **(D)** shows the same kinetic trace under conditions of excess of VirB9^{Ct}.

When k^1_{obs} and k^2_{obs} both increase with [P], the kinetics may follow either the CS or the IF mechanism (Background theory, section 1.1) [55]. Therefore, we first tested whether the

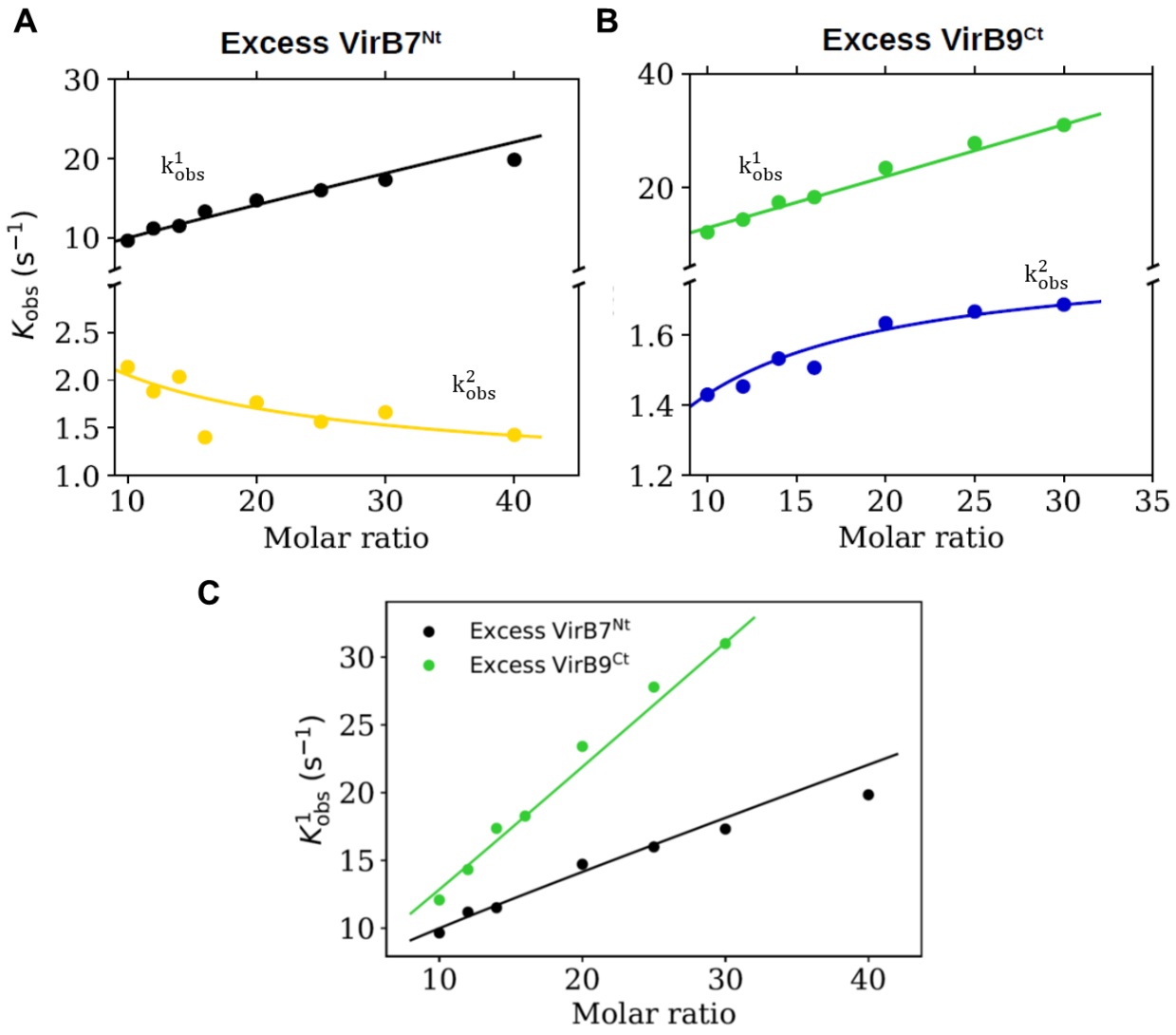


Figure 3.40: Observed rate constants at 35°C under (A) Excess of VirB7^{Nt} showing the slow phase in yellow and the fast phase in black. The lines correspond to fittings of the experimental data to the CS model using Eq. 1.17. The fitted parameters were $k_r = 0.95 \pm 0.11$ s^{-1}, $k_{-r} = 3.15 \pm 0.62$ s^{-1}, $k_{on} = 0.76 \pm 0.03$ $s^{-1}\mu M^{-1}$, and $k_{off} = 4.12 \pm 0.58$ s^{-1} **(B) Excess of VirB9^{Ct}** showing the slow phase in blue and the fast phase in green. The lines represent fittings to the IF model using Eq. 1.15. The fitted parameters were $k_{on} = 1.85 \pm 0.02$ $s^{-1}\mu M^{-1}$, $k_{off} = 3.20 \pm 0.29$ s^{-1}, $k_f = 1.42 \pm 0.04$ s^{-1} and $k_{uf} = 0.43 \pm 0.09$ s^{-1}. **(C)** Comparison between the fast phases k^1_{obs} obtained at 35°C.

CS model could give reliable values for the kinetic rate constants through a fitting of the behavior of k^1_{obs} and k^2_{obs} as a function of the VirB9^{Ct} concentration to Eq. 1.17. In this case, the observed slow phase (k^2_{obs}) could be due to a pre-equilibrium of VirB7^{Nt} conformations. The conformational exchange rate constants obtained by this analysis were $k_r = 1.84\ s^{-1}$ and $k_{-r} = 2.90\ s^{-1}$, which are not realistic considering that VirB7^{Nt} behaves as a random coil. Indeed, typical rates of exchange between different peptide conformations would be too fast (sub-nanosecond time scale) compared to the k_r and k_{-r} values obtained in the fit. Furthermore, the time scale of the stopped-flow experiment is in the milliseconds range. Therefore, even if a CS event on VirB7^{Nt} does exist, we would not be able to detect it by stopped-flow fluorescence measurements.

We next considered whether the IF mechanism could explain the increase in k^2_{obs} under excess of VirB9^{Ct} (Fig. 3.40B). The kinetic data obtained in excess of VirB9^{Ct} fitted well to the IF mechanism using Eq. 1.15, and we obtained reliable values for the kinetic rate constants as shown in Fig. 3.40B. However, if the system followed an IF mechanism, then the slow phase k^2_{obs} should have increased under the excess of both the ligand and the macromolecule. But k^2_{obs} decreases under conditions of excess VirB7^{Nt} (Fig. 3.40A), which immediately rules out the possibility of an IF mechanism. So, where and why does this IF event come from? We next thought in the possibility of having an additional step belonging to a conformational rearrangement of the complex, which would explain the IF event observed in excess of protein. However, we still wonder why this step was not observed in conditions of excess ligand. A possible explanation is that in excess of ligand, this IF event is low resolved due to the greater number of events being observed at this condition.

In fact, the comparison of the fast phases obtained at 35°C gave us the first clue that supported the above hypothesis. It turns out that, the behavior of the fast phases (k^1_{obs}) in excess of VirB7^{Nt} and in excess of VirB9^{Ct} should coincide assuming that they reflect the same binding event. However, a comparison of these two curves showed that they differ (Fig. 3.40C), indicating that perhaps there is an additional event contributing to the

mechanism. A close inspection of the fast relaxation in excess of VirB7^{Nt} showed that this curve has a tendency to lie down hyperbolically at high ligand concentrations (Fig. 3.40C). A plausible explanation is that there exists a third event that takes place in a timescale close to k^1_{obs} in conditions of excess VirB7^{Nt}, resulting in a contaminated phase (k^1_{obs}) that displays a distorted curve. This additional unresolved event relates well with the observed phase in excess of VirB9^{Ct} as a slow phase. The appearance of a third event in excess of ligand and the analysis of the slow phase in excess of macromolecule are both consistent with a conformational rearrangement step of the complex, as previously speculated. Therefore, the third event must be an IF step, i.e. a conformational change of the complex that takes place after ligand binding.

In summary, fluorescence stopped flow experiments carried out at 35°C are consistent with the view that VirB7^{Nt} binding to VirB9^{Ct} occurs through a first CS step in which VirB7^{Nt} selects pre-existing VirB9^{Ct} conformation(s), followed by a final rearrangement of the complex towards the bound conformation. The first complex will be hereafter called "encounter complex".

The kinetics of VirB9^{Ct}-VirB7^{Nt} association at 35 °C is consistent with a CS-IF mechanism

Analysis of the fluorescence stopped flow data at 35°C suggested that binding of VirB9^{Ct} to VirB7^{Nt} occurs through a combination of CS and IF mechanisms, in which a pre-existing VirB9^{Ct} conformation binds to VirB7^{Nt}, leading to the formation of an initial encounter complex, which subsequently folds into the final complex conformation. Therefore, we need a more complex model that may explain our experimental data at 35°C. So far, we know the following about the binding kinetics of VirB9^{Ct} to VirB7^{Nt} at 35°C:

- The CS event involves a pre-equilibrium between conformations of unbound VirB9^{Ct} (P and P*)
- The IF event comes after ligang binding, thus the conformational rearrangement occurs

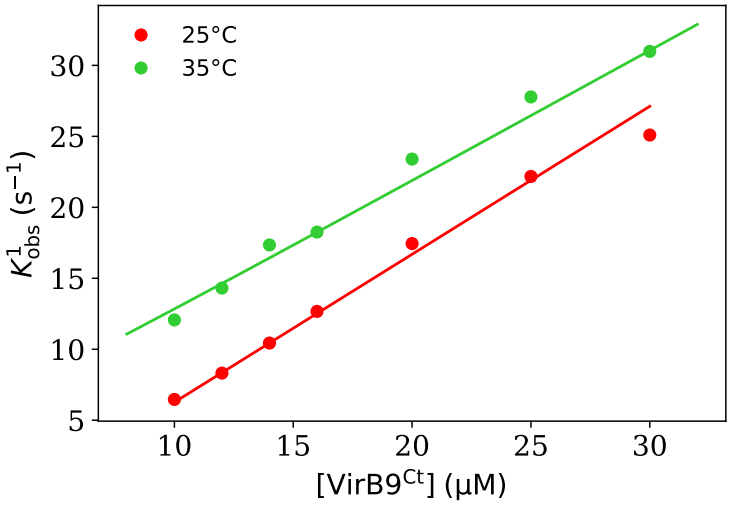


Figure 3.41: Comparison of the fast phases k^1_{obs} obtained in excess of VirB9$^{\text{Ct}}$ at 25°C and 35°C showing that they are parallel to each other, and consequently sharing the same value of k_{on} at both temperatures.

in the encounter complex, which would then fold into the final bound-conformation of the complex ($C^* \Longrightarrow C$).

- Since at 35°C we observed an IF step that was not observed in the kinetics at 25°C, therefore we may hypothesize that at higher temperatures the P^* specie is also available for ligand binding, which is not the bound-like conformation of VirB9$^{\text{Ct}}$ (P). This would explain why an additional IF step is necessary to rearranged the encounter complex formed by the P^* specie (C^*) to reach the final conformation of the complex (C).

Considering the above assumptions, we propose that the interaction mechanism between VirB9$^{\text{Ct}}$ and VirB7$^{\text{Nt}}$ at 35°C can be described by the following reaction scheme 3.1:

$$P \underset{k_{-r}}{\overset{k_r}{\rightleftharpoons}} P^* + L \underset{k_{\text{off}}}{\overset{k_{\text{on}}[L]}{\rightleftharpoons}} C^* \underset{k_{uf}}{\overset{k_f}{\rightleftharpoons}} C, \tag{3.1}$$

where the C^* specie is the encounter complex and the specie C is the final complex

conformation. This reaction scheme may be used to calculated an apparent dissociation constant between VirB7$^{\mathrm{Nt}}$ and VirB9$^{\mathrm{Ct}}$, $K_{\mathrm{d}}^{\mathrm{app}}$, as described in Background theory, section 1.1.3 (Eqn. 1.37):

$$K_{\mathrm{d}}^{\mathrm{app}} = \frac{(1 + [k_{-\mathrm{r}}/k_{\mathrm{r}}])}{(1 + [k_{\mathrm{f}}/k_{\mathrm{uf}}])} K_{\mathrm{d}}^{\mathrm{int}}, \tag{3.2}$$

where the intrinsic $K_{\mathrm{d}}^{\mathrm{int}} = k_{\mathrm{off}}/k_{\mathrm{on}}$, while $K_{\mathrm{d}}^{\mathrm{app}}$ is obtained from equilibrium experiments such as ITC. The value of $K_{\mathrm{d}}^{\mathrm{app}}$ contains the contributions from the conformational changes to the intrinsic dissociation constant of the complex as shown above.

We may assume the rapid equilibrium approximation ($k_{\mathrm{off}} + k_{\mathrm{on}}[L] \gg k_{-\mathrm{r}} + k_{\mathrm{r}}$), according to which binding occurs in a faster timescale compared to conformational changes, to interpret the VirB7$^{\mathrm{Nt}}$ - VirB9$^{\mathrm{Ct}}$ association kinetics at 35°C. Under this approximation the kinetic phases may be attributed individually to specif events such as binding, and conformational rearrangements [54]. Support for this assumption comes from the fact that the binding exchange rate at the lowest VirB7$^{\mathrm{Nt}}$ concentration used in the stopped-flow experiment, is approximately 5 times greater than the conformational rearrangement exchange rate. Taking the values of the kinetic constants shown in the legend of Figs. 3.40A and 3.40B, and using the lowest VirB7$^{\mathrm{Nt}}$ concentration in the stopped-flow experiments, 10 μM, the binding exchange rate may be computed as $k_{\mathrm{ex}}^{\mathrm{b}} = k_{\mathrm{on}}[\mathrm{VirB7^{Nt}}] + k_{\mathrm{off}} = 21.7\mathrm{s}^{-1}$, while the conformational rearrangement exchange rate is $k_{\mathrm{ex}} = k_{\mathrm{r}} + k_{-\mathrm{r}} = 4.1\mathrm{s}^{-1}$.

To analyze the kinetics of VirB7$^{\mathrm{Nt}}$ - VirB9$^{\mathrm{Ct}}$ association assuming the combined mechanism described by the scheme 3.1 at 35°C, we did the following approach: we modeled the data in excess of VirB9$^{\mathrm{Ct}}$ using the IF general equations (Eq. 1.15), and the slow phase k_{obs}^2 in excess of VirB7$^{\mathrm{Nt}}$ assuming the rapid equilibrium approximation for the CS mechanism (Eq. 1.13) [55] as observed in Fig. 3.42. This way, both events CS and IF will be modeled by different $\mathrm{k}_{\mathrm{obs}}^2$ (the $\mathrm{k}_{\mathrm{obs}}^2$ in excess of ligand for CS and the $\mathrm{k}_{\mathrm{obs}}^2$ in excess of protein for IF), but sharing the same binding event given by the fast phase $\mathrm{k}_{\mathrm{obs}}^1$ obtained in excess of protein (green curve of Fig. 3.40). The reliability of this fast phase describing the binding event at 35°C was further validated by comparing the fast phases (k_{obs}^1) obtained in excess

of VirB9^{Ct} at 25 and 35 °C. This comparison showed that the two curves are almost parallel to each other, indicating that they reflect binding events that share the same value of k_{on} at both temperatures.(Fig. 3.41). In our combined model, the K_d^{app} obtained from the analysis of the ITC data was used as input (prior) information for the fitting. The obtained values for the kinetic rate constants are shown in Table 3.4.

Table 3.4: Kinetic rate constants obtained for the interaction between VirB7^{Nt} and VirB9^{Ct} at 35°C. The kinetic data was analyzed assuming the CS model in rapid equilibrium approximation in excess of VirB7^{Nt}, and the IF model in excess of VirB9^{Ct}.

Model	T(°C)	$k_r^{\mathbf{P}\rightarrow\mathbf{P}^*}(s^{-1})$	$k_{-r}^{\mathbf{P}^*\rightarrow\mathbf{P}}(s^{-1})$	$k_{on}(s^{-1}\mu M^{-1})$	$k_{off}(s^{-1})$	$k_f(s^{-1})$	$k_{uf}(s^{-1})$
CS-IF	35	1.17 ± 0.13	3.42 ± 1.12	1.85 ± 0.02	3.06 ± 0.27	1.65 ± 0.06	0.21 ± 0.08

It is noteworthy that k_{on} at 35 and 25 °C are nearly the same (Table. 3.4 and 3.3). This observation is in agreement with the fact that at both temperatures the kinetic curves of k_{obs}^1 in excess of VirB9^{Ct} are nearly parallel to each other (Fig. 3.41), and thus they share a similar slope (value of k_{on}). In contrast, k_{off} increased from ≈ 0.42 s^{-1} at 25°C to 3.06 s^{-1} at 35°C, explaining the higher K_d^{int} at 35°C compared to 25°C (Table 3.5). Therefore, the lower value of k_{off} at 25°C leads to the greater intrinsic affinity between VirB7^{Nt} and VirB9^{Ct} at this temperature relative to 35°C. The rate constants for VirB9^{Ct} conformational transitions

Table 3.5: Experimental values of K_d^{app} obtained by the ITC technique and calculated values of K_d^{int} using Eq. 3.2 at 25°C and 35°C.

Temperature	$K_d^{app}(\mu M)$	$K_d^{int}(\mu M)$
25°C	0.40	0.37
35°C	0.72	1.63

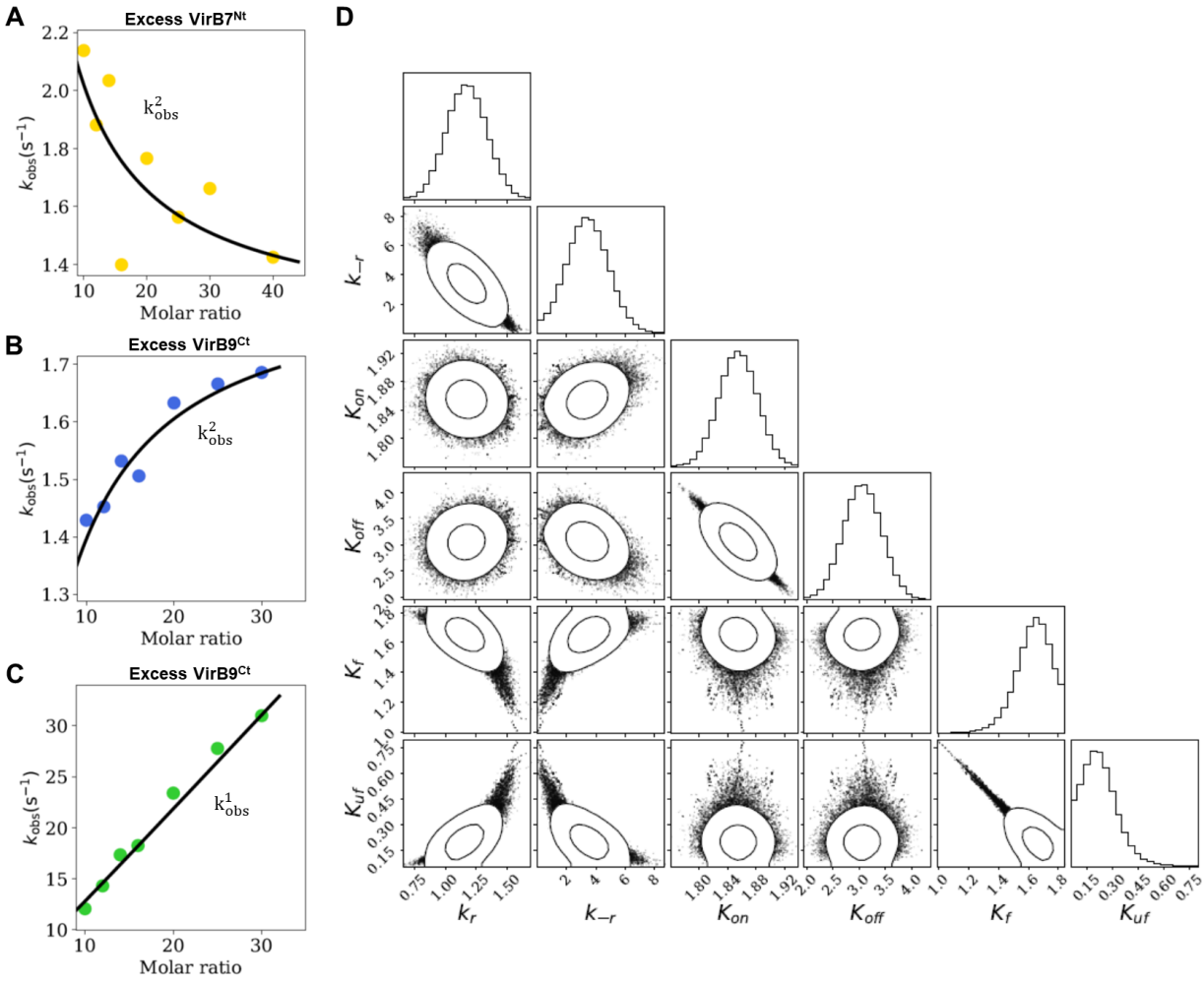


Figure 3.42: Simultaneous fit of the observed rate constants obtained in excess of VirB9^{Ct} (**B and C**) and the slow phase obtained in conditions of excess VirB7^{Nt} (**A**) according to the CS-IF combined model using Eq. 1.15 and Eq. 1.13. The value for K_d^{app} at 35°C from ITC experiments was used as Gaussian prior in the fitting, with a value centered at 0.723 μM and a confidence interval of 0.082 μM. (**D**) Triangle plot showing the correlations between the fitted paramaters to the CS-IF combined mechanism

.

were also affected by temperature. In this case, k_{r} is the rate constant for the $\mathbf{P} \rightarrow \mathbf{P}^{*}$ transition and $k_{-\mathrm{r}}$ corresponds to the rate constant for the $\mathbf{P}^{*} \rightarrow \mathbf{P}$ transition. Thus, the pre-equilibrium given by $k_{\mathrm{r}}^{\mathbf{P} \rightarrow \mathbf{P}^{*}}$ and $k_{-\mathrm{r}}^{\mathbf{P}^{*} \rightarrow \mathbf{P}}$ is re-distributed as observed in Table 3.4), such that the equilibrium constant between the states P* and P changed from $K_{\mathrm{eq}}^{\mathbf{P}^{*} \rightarrow \mathbf{P}} = 15.0$ at 25°C ($\mathrm{P}^{*} \rightleftharpoons \mathrm{P}$) to $K_{\mathrm{eq}}^{\mathbf{P}^{*} \rightarrow \mathbf{P}} = 2.92$ at 35°C ($\mathrm{P} \rightleftharpoons \mathrm{P}^{*}$). This indicates that there is a redistribution of the unbound VirB9$^{\mathrm{Ct}}$ conformational states as a function of temperature, but consistent with the greater preference for populating the bound-like conformation of VirB9$^{\mathrm{Ct}}$ (P specie) at 35°C. Overall, these observations suggest that at 25°C the conformational pre-equilibrium of VirB9$^{\mathrm{Ct}}$ is shifted towards the conformation that best resemble the 'bound' state. However due to the conformational re-distribution of species of unbound VirB9$^{\mathrm{Ct}}$ at 35°C, now the P* specie is also available for ligand binding. Therefore, at 35°C both species P* and P bind to VirB7$^{\mathrm{Nt}}$ forming an encounter complex in the case of the P* specie (C*) and the final complex in the case of the P specie (C). Furthermore, the lower affinity at 35 °C could be explained by the energetic cost of the IF event of the complex ($\mathrm{C}^{*} \Longrightarrow \mathrm{C}$) observed at this temperature.

Chapter 4

Final Conclusions

We found interesting features in our study about the structure of VirB9^{Ct} and its binding mechanism to VirB7^{Nt}. Evidence coming from the CD data deconvolution suggests that the unbound VirB9^{Ct} presents similar secondary structure content compared to bound VirB9^{Ct}. Furthermore, ANS fluorescence and DSC denaturation experiments suggested that, although the unbound VirB9^{Ct} presents a bit greater solvent exposed surface area than the bound form, its thermal denaturation occurred at same melting temperature that the complex, indicating that both proteins undergo similar thermal unfolding processes and, thus, they might share similar tertiary structure content. The NMR spectra of unbound VirB9^{Ct} at 7°C showed that the protein is more rigid at lower temperatures, and analysis of the backbone resonances indicated that the secondary structure content seems to be preserved with respect to bound VirB9^{Ct}, the only exceptions being β_1 and β_2, which could not be assigned. Secondary chemical shift analysis using ^{15}N and ^{13}Co chemical shifts obtained from CEST suggested that β_2 is partially formed, in contrast to β_1 that seems to be disordered. As expected, the most affected residues by the presence of VirB7^{Nt} were located on the β-sheet binding interface with the peptide. Overall, the present data indicate that, despite displaying a complex dynamics in the absence of the peptide, the unbound VirB9^{Ct} is similar to the bound state in terms of secondary and tertiary structures, with significant differences at

the β_1 and β_2, which form the VirB7^{Nt} anchoring point.

Quantitative analysis of the ^{15}N CEST experiments indicated that $k_{ex}^{b} \approx 25 - 30\ s^{-1}$, which is in the ms time scale in agreement with the time scale of the stopped-flow experiment. The dynamics of unbound VirB9^{Ct} at the millisecond time scale was analyzed qualitatively by CEST. We found that temperature considerably alters the stabilization of unbound VirB9^{Ct} states. Furthermore, analysis of the CEST experiments carried out at 35°C suggested that the unbound VirB9^{Ct} samples the bound state in the absence of VirB7^{Nt}, which supports the conformational selection-mechanism.

The confirmation of our hypothesis regarding the binding mechanism came from our stopped flow kinetic experiments. The kinetic data obtained at 25°C in excess of the peptide yielded two k_{obs} that increased with the peptide concentration, which may be explained by both the CS and the IF models. However, the k_{obs} obtained in excess of VirB9^{Ct} increased linearly with protein concentration, which suggest that only the binding event is being observed under this condition, supporting a CS event and ruling out the IF mechanism. The complexity of the binding mechanism became evident at higher temperatures. The kinetic data measured at 35°C showed that in excess of VirB7^{Nt} the binding mechanism is governed by a CS event since the observed k_{obs}^{2} decreased with ligand concentration. This kinetic relaxation that can be only explained by the CS mechanism. However, at reverse pseudo first order conditions (in excess of VirB9^{Ct}), an additional phase (k_{obs}^{2}) that increased with protein concentration was observed, suggesting an additional CS or IF step. Comparison of the obtained k_{obs} at 25°C and k_{obs}^{1} in excess of VirB9^{Ct} at 35°C revealed that the kinetic curves are highly parallel to each other, which indicates that the value of k_{on} is nearly the same at both temperatures. Therefore, the idea that the on-forward binding event is conserved at both temperatures, and the detailed analysis of our kinetic curves at 35°C, led us to consider the possibility of a combined mechanism. Finally, we propose that at 35°C the binding event is preceded by a conformational pre-equilibrium between conformations of unbound VirB9^{Ct}, followed by a conformational rearrangement of the encounter complex

(C*) towards the final complex conformation (C). Fitting the kinetic data at 25°C and 35°C to the CS and CS-IF combined model, respectively, confirmed the assumptions described above.

It is evident that at 25°C the conformational equilibrium of unbound VirB9^{Ct} is shifted towards the formation of the P species, which binds to VirB7^{Nt} through a conformational-selection mechanism. The scene is totally different at higher temperatures. At 35°C, the conformational equilibrium of unbound VirB9^{Ct} is re-distributed, and now the P* specie, which is not the"preferential" binding form of the protein, will also bind to VirB7^{Nt}, even though the P specie is still more abundant at 35°C. The encounter complex C* formed by ligand binding to the P* specie, would subsequently rearrange towards the final complex conformation C. Therefore, the pathway followed by VirB7^{Nt} to bind to VirB9^{Ct} is highly dependent on the availability and distribution of the species P and P*, and hence, depends on temperature. The internal pre-equilibrium of unbound VirB9^{Ct} will always exist, but at lower temperatures, the predominant pathway is a conformational-selection mechanism by binding preferentially to the P specie that forms the final complex C. In contrast, at higher temperatures, an IF pathway is also activated due to binding of the P* specie to VirB7^{Nt}. Therefore, at 35°C, both mechanisms work in concert to ensure the formation of the final complex C, either by the binding of the P or P* species to VirB7^{Nt}. The above discussion can be better understood by looking at Fig. 4.1.

Finally, measurements of heat capacity change upon binding, are consistent with the view that VirB9^{Ct} does not undergo a major conformational change upon binding to VirB7^{Nt}. The peptide rather shifts the VirB9^{Ct} equilibrium towards the bound-like conformational state. This gives supports to our discussed conclusions and is consistent with a weak coupling, a term given to systems in which it is considered that the conformational change can be driven by ligand binding and by temperature, and that the binding at different temperatures proceeds with different population distributions, as observed in the VirB9^{Ct}-VirB7^{Nt} association.

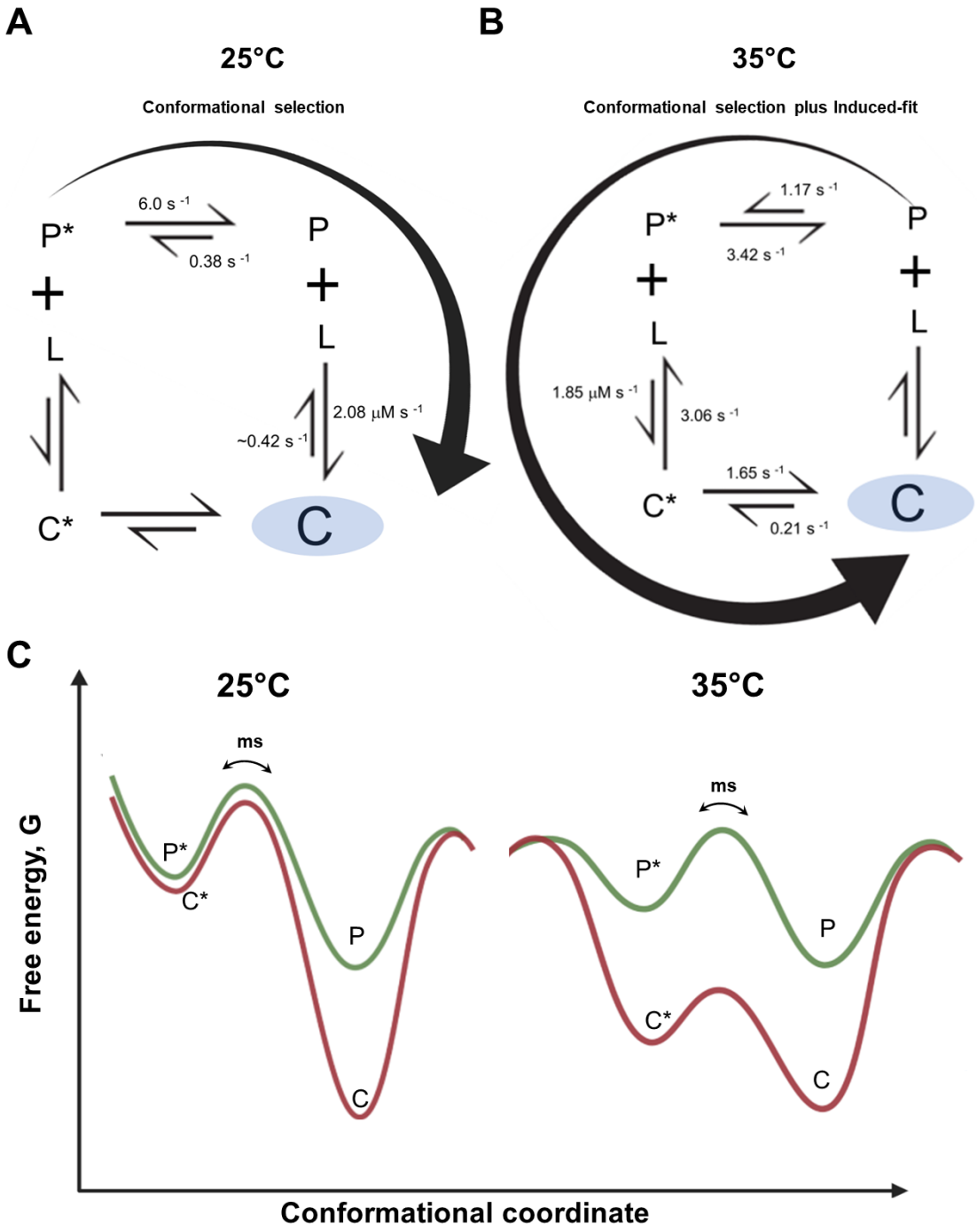


Figure 4.1: Scheme that illustrates the proposed binding mechanism between VirB9^{Ct} and VirB7^{Nt} and the preferential pathways at 25 (**A**) and 35°C (**B**). Conformational energy landscape at 25 and 35°C of the unbound VirB9^{Ct} in green and the complex in reddish maroon (**C**).

Appendix A

Two-state kinetic system

Calculation of the $\lambda_{1,2}$ eigenvalues of the kinetic matrix of a two-site exchanging system:

$$\begin{bmatrix} \frac{dA}{dt} \\ \\ \frac{dB}{dt} \end{bmatrix} = \begin{bmatrix} -k_{AB} & k_{BA} \\ \\ k_{AB} & -k_{BA} \end{bmatrix} \begin{bmatrix} A \\ \\ B \end{bmatrix} \tag{A.1}$$

$$\begin{bmatrix} -C & D \\ \\ C & -D \end{bmatrix} \tag{A.2}$$

$$\begin{vmatrix} -C-\lambda & D \\ \\ C & -D-\lambda \end{vmatrix} = (C+\lambda)(D+\lambda) - CD = 0, \tag{A.3}$$

therefore:

$$\lambda_1 = 0 \tag{A.4}$$

and

$$\lambda_2 = -C - D \tag{A.5}$$

Appendix B

Three-state kinetic system

The IF and the CS mechanisms may be described assuming a three-site chemical exchange kinetic model:

$$\mathrm{A} \underset{k_{\mathrm{BA}}}{\overset{k_{\mathrm{AB}}}{\rightleftharpoons}} \mathrm{B} \underset{k_{\mathrm{CB}}}{\overset{k_{\mathrm{BC}}}{\rightleftharpoons}} \mathrm{C} \tag{B.1}$$

whose kinetic matrix is given by:

$$\mathcal{M} = \begin{bmatrix} -k_{AB} & k_{BA} & 0 \\ k_{AB} & -(k_{BA}+k_{BC}) & k_{CB} \\ 0 & k_{BC} & -k_{CB} \end{bmatrix}. \tag{B.2}$$

The solution of a system of 3 differential equations is given by the sum of three exponential functions with argument λ_i.

$$x(t) = A_0 \exp(-\lambda_0 t) + A_1 \exp(-\lambda_1 t) + A_2 \exp(-\lambda_2 t) \tag{B.3}$$

where λ_i are the eigenvalues of the matrix $\mathcal{M}$ associated to the system. The eigenvalues may be calculated by solving the following equation:

$$|\mathcal{M} - \lambda\mathcal{I}| = 0 \tag{B.4}$$

where $\mathcal{I}$ is the 3x3 identity matrix and $|\cdot|$ is the determinant of the matrix. It follows then that:

$$\begin{vmatrix} -k_{AB} - \lambda & k_{BA} & 0 \\ k_{AB} & -(k_{BA} + k_{BC} + \lambda) & k_{CB} \\ 0 & k_{BC} & -(k_{CB} + \lambda) \end{vmatrix} = 0 \tag{B.5}$$

$$-(k_{AB} + \lambda)[(k_{BA} + k_{BC} + \lambda)(k_{CB} + \lambda) - k_{BC}k_{CB}] + k_{BA}k_{AB}(k_{CB} + \lambda) = 0 \tag{B.6}$$

which leads us to the equation:

$$\lambda\left[\lambda^2 + (k_{AB} + k_{BA} + k_{BC} + k_{CB})\lambda + (k_{CB}k_{BA} + k_{AB}k_{BC} + k_{AB}k_{CB})\right] = 0 \tag{B.7}$$

notice that one of the λ_i values is zero, the other two non-zero eigenvalues are given by the general quadratic equation. Therefore,

$$\begin{aligned} \lambda = &\frac{1}{2} - (k_{AB} + k_{BA} + k_{BC} + k_{CB}) \pm \\ &\frac{1}{2}\sqrt{(k_{AB} + k_{BA} + k_{BC} + k_{CB})^2 - 4(k_{CB}k_{BA} + k_{AB}k_{BC} + k_{AB}k_{CB})}] \end{aligned} \tag{B.8}$$

Let's analyze the term inside the square root:

$$(k_{AB} + k_{BA} + k_{BC} + k_{CB})^2 - 4(k_{CB}k_{BA} + k_{AB}k_{BC} + k_{AB}k_{CB}) \tag{B.9}$$

We can re-write the above expression as:

$$(k_{AB}+k_{BA})^2+(k_{BC}+k_{CB})^2+2(k_{AB}+k_{BA})(k_{BC}+k_{CB})-4(k_{AB}(k_{BC}+k_{CB})+k_{CB}k_{BA}) \tag{B.10}$$

To simplify the expression we can add and subtract the term $k_{BC}k_{BA}$:

$$(k_{AB}+k_{BA})^2+(k_{BC}+k_{CB})^2+2(k_{AB}+k_{BA})(k_{BC}+k_{CB})-4(k_{AB}(k_{BC}+k_{CB}) + k_{CB}k_{BA}+k_{BC}k_{BA}-k_{BC}k_{BA}) \quad \text{(B.11)}$$

Rearranging the above expression and replacing in Eq. B.8 we get that:

$$-\lambda_{1,2}=\frac{1}{2}\left[(k_{AB}+k_{BA}+k_{BC}+k_{CB})\pm\sqrt{(k_{AB}+k_{BA}-k_{BC}-k_{CB})^2+4k_{BC}k_{BA}}\right] \quad \text{(B.12)}$$

Finally, the eigenvalues are $\{0,\ -\lambda_1,\ -\lambda_2\}$, where $-\lambda_1$ and $-\lambda_2$ are given by Eq. B.12. The larger eigenvalue $-\lambda_1$ is determined by the sum in Eq. B.12, while the smallest eigenvalue $-\lambda_2$ is determined by the subtraction in Eq. B.12. Therefore, for a three-state system, the kinetics of approach to equilibrium depend on two exponentials, each associated to and observed rate constant defined by $-\lambda_1$ and $-\lambda_2$ [55].

Accordingly to expression B.12, the eigenvalues for the IF mechanism represented by scheme 1.8 are given by:

$$-\lambda_{1,2}^{\mathrm{IF}}=\frac{1}{2}\left(k_{-r}+k_r+k_{\mathrm{off}}+k_{\mathrm{on}}[\mathrm{L}]\pm\sqrt{(k_{\mathrm{off}}+k_{\mathrm{on}}[\mathrm{L}]-k_{-r}-k_r)^2+4k_rk_{\mathrm{off}}}\right). \quad \text{(B.13)}$$

Similarly, the eigenvalues for the CS mechanism represented by scheme 1.9 are given by:

$$-\lambda_{1,2}^{\mathrm{CS}}=\frac{1}{2}\left(k_{-r}+k_r+k_{\mathrm{off}}+k_{\mathrm{on}}[\mathrm{L}]\pm\sqrt{(k_{\mathrm{off}}+k_{\mathrm{on}}[\mathrm{L}]-k_{-r}-k_r)^2+4k_{-r}k_{\mathrm{on}}[\mathrm{L}]}\right) \quad \text{(B.14)}$$

Appendix C

Complementary NMR data

C.1 Unbound VirB9 assignment at 7°C

Residue number	^{1}H NMR (ppm)	^{15}N NMR (ppm)
232	9.1642	129.21137
225	8.73047	128.89951
231	9.42324	128.39454
229	9.09257	127.9599
158	8.39718	127.24415
240	9.11252	126.96627
244	8.94615	126.67786
241	9.47943	125.75548
191	9.13817	125.6902
227	8.69278	125.68437
230	9.28444	125.18044
155	8.37154	124.86934
219	8.73195	124.54686
208	9.33626	124.38964
206	8.57991	124.39192
239	9.36031	123.44058
172	8.35701	123.39073
246	8.49397	123.16529
192	9.14335	123.13823
242	8.63892	123.09618
215	7.23433	122.91502
157	8.16512	122.63853
209	8.75237	122.60563
188	8.55725	122.37505
159	8.32753	122.38669
153	8.53566	122.14915
161	8.31183	121.92019
213	8.77299	121.49607
160	8.34569	121.38497
224	8.48023	121.11465
152	8.82494	120.97559
154	8.62399	120.88471
156	8.35479	120.79514
221	8.57258	120.59906
189	9.4553	119.94194
186	8.71605	119.81218
203	8.16173	119.75484
220	8.07826	119.3969
245	8.22853	119.35603
204	7.89197	119.33234

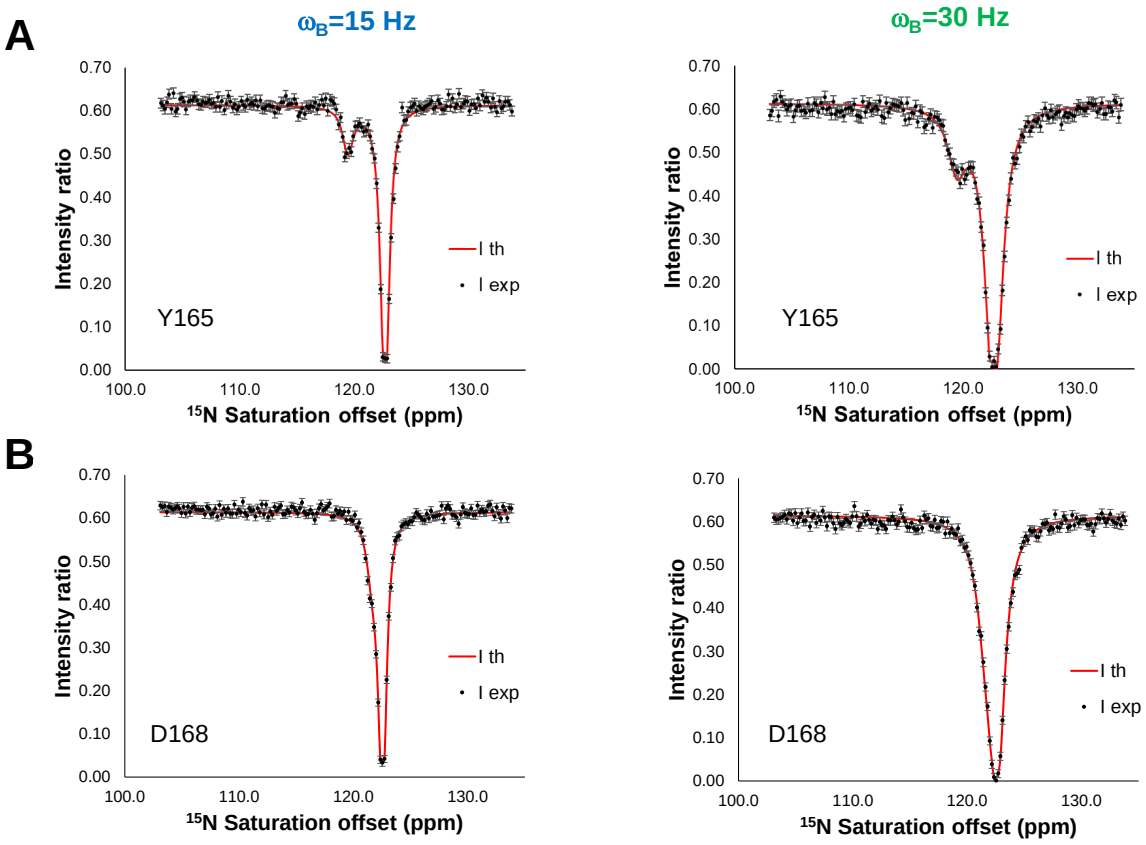
243	9.09394	119.00424
238	8.57421	118.5232
216	8.42072	118.44748
226	9.16125	117.50282
218	8.99893	117.49452
228	8.15275	116.5673
176	8.38509	116.22944
211	9.30484	116.13394
151	8.73275	115.52823
222	8.28197	114.93188
237	7.23306	114.89671
173	8.06174	114.65224
223	8.86792	114.28083
214	8.63725	113.86843
190	8.45737	113.64314
201	8.19745	113.10651
233	8.22625	112.65399
197	8.27078	112.06082
202	8.35324	111.67828
234	7.24463	107.87323
235	9.97002	123.83034
187	10.1009	111.45167
210	7.40542	109.59119
171	7.94181	113.69419
212	7.40259	114.61716
247	8.95248	126.61503
217	8.33714	122.34806
248	8.58863	115.02993
249	9.48208	124.00114
174	8.38283	123.93075

C.2 Fitting of CEST profiles to the two-state Bloch McConnel equation

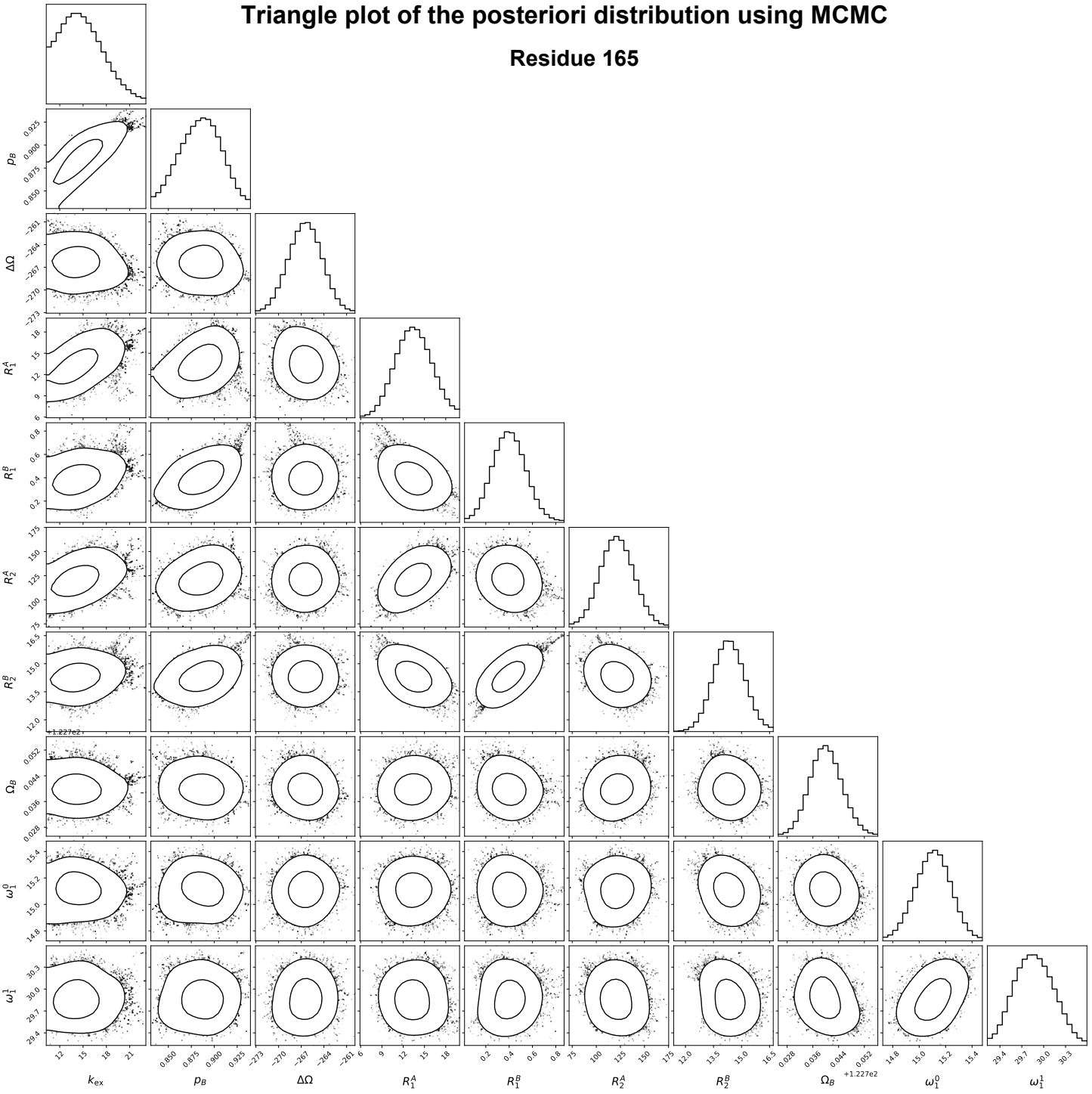
The highlighted residues in red indicates those presenting a $\Delta\Omega \approx 2$ ppm

res	kex	pB	ΔΩ	R1A	R1B	R2A	R2B	ωB_fit	ω1_fit_1
158	85.9	0.957	65.35	5.079	1.209	52.815	4.653	126.151	15.378
160	36.459	0.949	-97.732	2.775	1.202	128.11	12.254	122.782	16.317
161	74.009	0.946	-70.018	10.15	0.812	151.971	7.162	123.309	15.407
162	18.959	0.874	78.235	11.491	0.591	190.033	7.391	120.871	15.587
165	15.773	0.873	-265.072	20	0.001	152.707	10.361	122.74	15.492
166	10.917	0.822	-187.1	20	0.045	122.045	10.8	122.488	16.337
168	77.302	0.955	-74.342	20	0.517	181.982	10.348	122.543	15.214
169	24.435	0.914	202.366	11.406	0.585	121.788	10.901	118.344	15.123
170	15.808	0.871	202.275	20	0.157	88.845	11.317	123.923	16.48
171	16.639	0.894	-210.505	20	0.44	149.848	10.001	116.673	17.548
176	15.987	0.851	-187.977	19.633	0.001	163.135	13.011	119.326	16.206
177	12.939	0.845	149.741	20	0.128	147.574	20.665	119.716	15.871
178	99.473	0.974	91.576	1.475	1.256	174.096	11.578	121.853	16.451
179	18.95	0.948	360.231	4.916	1.352	198.022	17.754	117.718	15.132
182	24.154	0.931	111.821	19.925	0.542	74.667	10.924	119.507	17.173
183	11.816	0.826	-175.762	18.172	0.001	197.566	10.254	125.656	15.482
184	18.334	0.886	112.966	12.139	0.407	197.408	13.141	121.395	15.677
185	18.45	0.877	163.937	16.901	0.029	94.388	10.684	117.022	15.224
198	30.649	0.934	73.959	15.164	0.681	129.467	10.692	117.74	15.358
201	25.461	0.935	134.192	19.826	0.568	69.92	19.957	114.562	15.598
202	141.626	0.935	23.109	10.934	0.534	151.096	8.948	112.4	15.42
204	62.308	0.959	84.754	19.925	0.583	162.518	7.617	119.797	15.354
207	68.022	0.971	73.112	6.898	0.888	53.563	9.333	120.19	16.287
210	36.201	0.924	-62.761	19.867	0.001	155.451	11.417	112.05	14.8
212	29.569	0.929	80.546	12.215	0.454	74.42	11.021	114.304	15.097
214	29.296	0.937	268.905	11.094	0.755	73.032	10.627	111.269	15.238
216	36.926	0.937	114.577	6.993	0.638	90.35	8.613	117.33	15.118
218	42.75	0.95	-101.522	19.823	0.405	81.844	11.508	119.427	15.516
220	22.371	0.915	317.491	7.98	0.635	69.117	9.814	115.746	15.005
232	40.549	0.938	73.048	16.156	0.368	85.56	10.951	128.558	15.135
233	30.549	0.934	-135.205	9.278	0.555	153.406	7.003	115.692	17.083
235	20.424	0.9	188.036	20	0.236	163.452	9.691	121.787	16.681
237	18.099	0.9	260.725	4.654	0.744	57.647	10.554	112.375	15.45
238	36.238	0.942	135.872	18.031	0.572	145.627	9.544	119.917	16.756
241	145.621	0.971	72.43	0.223	1.08	48.376	10.261	125.541	15.497
245	42.454	0.945	131.591	6.298	0.773	112.036	9.992	117.956	15.319
247	25.568	0.924	-357.214	6.699	0.777	87.943	10.754	131.086	15.329
249	23.133	0.91	178.804	7.261	0.645	109.887	10.652	122.586	15.03
250	25.863	0.91	-81.676	20	0.001	186.015	10.112	122.751	15.543
254	58.283	0.963	-79.614	2.144	1.369	100.848	7.865	124.596	14.519

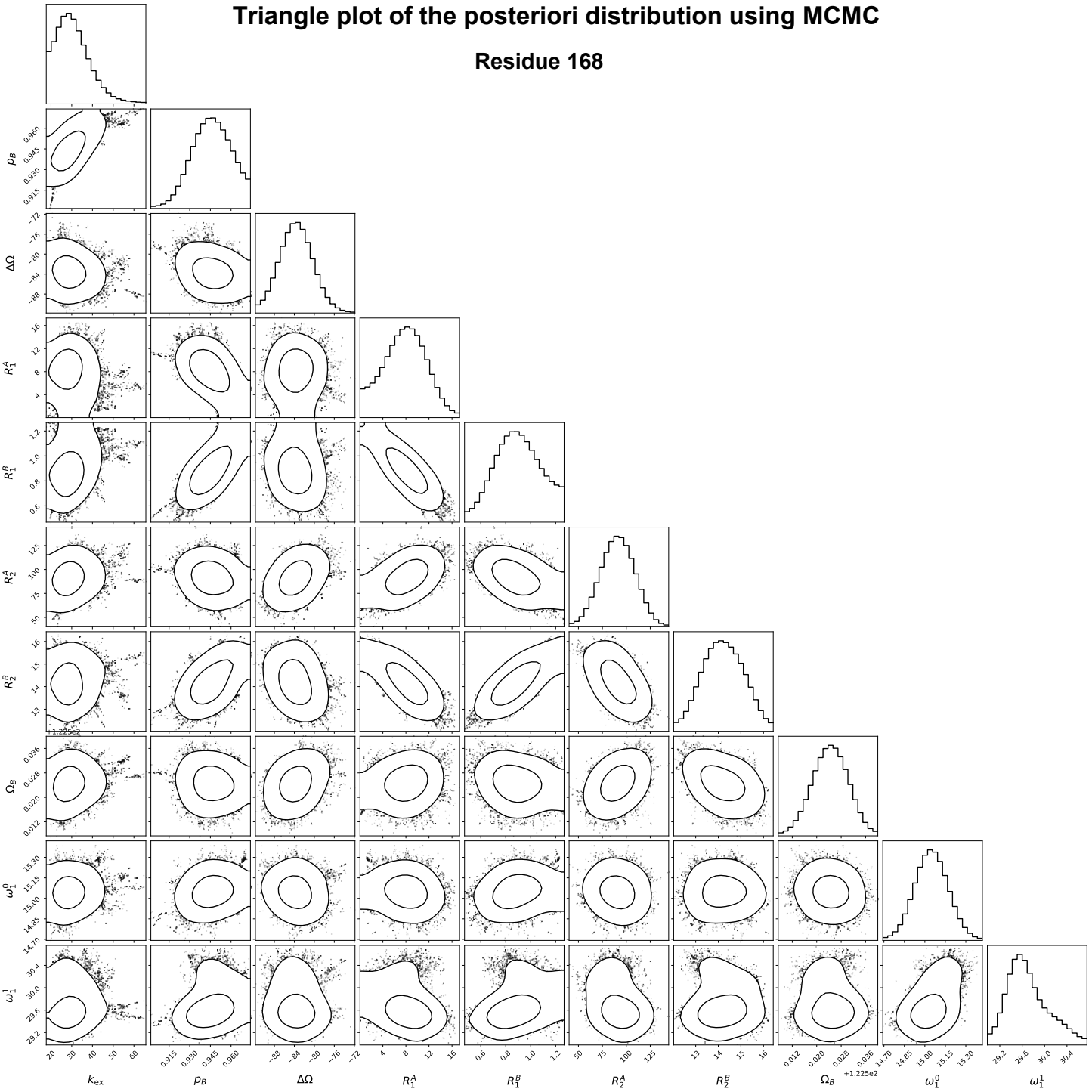
C.3 CEST profiles fitting for Y165 and D168



C.4 Triangle plots



C.5 Triangle plots



Bibliography

[1] E. R. Green and J. Mecsas, "Bacterial secretion systems: An overview," *Virulence Mechanisms of Bacterial Pathogens*, pp. 213–239, 2016.

[2] T. R. Costa, C. Felisberto-Rodrigues, A. Meir, M. S. Prevost, A. Redzej, M. Trokter, and G. Waksman, "Secretion systems in Gram-negative bacteria: Structural and mechanistic insights," *Nature Reviews Microbiology*, vol. 13, no. 6, pp. 343–359, 2015.

[3] E. Grohmann, P. J. Christie, G. Waksman, and S. Backert, "Type IV secretion in Gram-negative and Gram-positive bacteria," *Molecular Microbiology*, vol. 107, no. 4, pp. 455–471, 2018.

[4] Y. G. Li, B. Hu, and P. J. Christie, "Biological and Structural Diversity of Type IV Secretion Systems," *Microbiology Spectrum*, vol. 7, no. 2, 2019.

[5] T. R. Costa, L. Harb, P. Khara, L. Zeng, B. Hu, and P. J. Christie, "Type IV secretion systems: Advances in structure, function, and activation," *Molecular Microbiology*, vol. 115, no. 3, pp. 436–452, 2021.

[6] D. P. Souza, G. U. Oka, C. E. Alvarez-Martinez, A. W. Bisson-Filho, G. Dunger, L. Hobeika, N. S. Cavalcante, M. C. Alegria, L. R. Barbosa, R. K. Salinas, C. R. Guzzo, and C. S. Farah, "Bacterial killing via a type IV secretion system," *Nature Communications*, vol. 6, pp. 1–9, 2015.

[7] G. G. Sgro, T. R. Costa, W. Cenens, D. P. Souza, A. Cassago, L. Coutinho de Oliveira, R. K. Salinas, R. V. Portugal, C. S. Farah, and G. Waksman, "Cryo-EM structure of the bacteria-killing type IV secretion system core complex from Xanthomonas citri," *Nature Microbiology*, vol. 3, no. 12, pp. 1429–1440, 2018.

[8] A. Rivera-Calzada, R. Fronzes, C. G. Savva, V. Chandran, P. W. Lian, T. Laeremans, E. Pardon, J. Steyaert, H. Remaut, G. Waksman, and E. V. Orlova, "Structure of a bacterial type IV secretion core complex at subnanometre resolution," *EMBO Journal*, vol. 32, no. 8, pp. 1195–1204, 2013.

[9] R. Fronzes, E. Schäfer, L. Wang, H. R. Saibil, E. V. Orlova, and G. Waksman, "Structure of a type iv secretion system core complex," *Science*, vol. 323, no. 5911, pp. 266–268, 2009.

[10] V. Chandran, R. Fronzes, S. Duquerroy, N. Cronin, J. Navaza, and G. Waksman, "Structure of the outer membrane complex of a type IV secretion system," *Nature*, vol. 462, no. 7276, pp. 1011–1015, 2009.

[11] D. P. Souza, M. O. Andrade, C. E. Alvarez-Martinez, G. M. Arantes, C. S. Farah, and R. K. Salinas, "A component of the Xanthomonadaceae type IV secretion system combines a VirB7 motif with a N0 domain found in outer membrane transport proteins," *PLoS Pathogens*, vol. 7, no. 5, 2011.

[12] G. Waksman, "From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective," *EMBO reports*, vol. 20, no. 2, pp. 1–16, 2019.

[13] H. H. Low, F. Gubellini, A. Rivera-Calzada, N. Braun, S. Connery, A. Dujeancourt, F. Lu, A. Redzej, R. Fronzes, E. V. Orlova, and G. Waksman, "Structure of a type IV secretion system," *Nature*, vol. 508, no. 7497, pp. 550–553, 2014.

[14] P. Khara, L. Song, P. J. Christie, B. Hu, P. Yuan, and S. J. Hultgren, "¡i¿in situ¡/i¿ visualization of the pkm101-encoded type iv secretion system reveals a highly symmetric atpase energy center," *mBio*, vol. 12, no. 5, pp. e02465–21, 2021.

[15] B. Hu, P. Khara, and P. J. Christie, "Structural bases for F plasmid conjugation and F pilus biogenesis in Escherichia coli," vol. 117, no. 3, 2020.

[16] L. C. Oliveira, D. P. Souza, G. U. Oka, F. d. S. Lima, R. J. Oliveira, D. C. Favaro, H. Wienk, R. Boelens, C. S. Farah, and R. K. Salinas, "VirB7 and VirB9 Interactions Are Required for the Assembly and Antibacterial Activity of a Type IV Secretion System," *Structure*, vol. 24, no. 10, pp. 1707–1718, 2016.

[17] "Intrinsically disordered protein," *Journal of Molecular Graphics and Modelling*, vol. 19, no. 1, pp. 26–59, 2001.

[18] V. N. Uversky and A. K. Dunker, "Understanding protein non-folding," *Biochimica et Biophysica Acta - Proteins and Proteomics*, vol. 1804, no. 6, pp. 1231–1264, 2010.

[19] P. Tompa, "Intrinsically unstructured proteins," *Trends in Biochemical Sciences*, vol. 27, no. 10, pp. 527–533, 2002.

[20] J. Yang, M. Gao, J. Xiong, Z. Su, and Y. Huang, "Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding," *Protein Science*, vol. 28, no. 11, pp. 1952–1965, 2019.

[21] H. J. Dyson and P. E. Wright, "Intrinsically unstructured proteins and their functions," *Nature Reviews Molecular Cell Biology*, vol. 6, no. 3, pp. 197–208, 2005.

[22] S. L. Shammas, M. D. Crabtree, L. Dahal, B. I. Wicky, and J. Clarke, "Insights into coupled folding and binding mechanisms from kinetic studies," *Journal of Biological Chemistry*, vol. 291, no. 13, pp. 6689–6695, 2016.

[23] "Intrinsically disordered proteins: Modes of binding with emphasis on disordered domains," *Open Biology*, vol. 11, no. 10, 2021.

[24] R. Pancsa and P. Tompa, "Structural Disorder in Eukaryotes," vol. 7, no. 4, 2012.

[25] Z. Obradovic, C. Garner, A. Keith, and J. Brown, "Intrinsic Protein Disorder in Complete Genomes," vol. 171, pp. 161–171, 2000.

[26] B. Xue, A. K. Dunker, V. N. Uversky, B. Xue, A. K. Dunker, and V. N. Uversky, "Orderly order in protein intrinsic disorder distribution : disorder in 3500 proteomes from viruses and the three domains of life," vol. 1102, 2012.

[27] Z. Peng, J. Yan, X. Fan, M. J. Mizianty, B. Xue, K. Wang, G. Hu, V. N. Uversky, and L. Kurgan, "Exceptionally abundant exceptions : comprehensive characterization of intrinsic disorder in all domains of life," pp. 137–151, 2015.

[28] L. M. Iakoucheva, C. J. Brown, J. D. Lawson, and A. K. Dunker, "Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins," vol. 2836, no. 02, pp. 573–584, 2002.

[29] J. Liu, N. B. Perumal, C. J. Oldfield, E. W. Su, V. N. Uversky, and A. K. Dunker, "Intrinsic disorder in transcription factors," vol. 45, no. 22, pp. 6873–6888, 2008.

[30] C. A.Galea, Y. Wang, S. G. Sivakolundu, and R. W. Kriwacki, "Regulation of Cell Division by Intrinsically Unstructured Proteins : Intrinsic," pp. 7598–7609, 2008.

[31] P. E. Wright and H. J. Dyson, "Intrinsically disordered proteins in cellular signalling and regulation," vol. 16, no. January, 2015.

[32] L. M. Iakoucheva, P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, and A. K. Dunker, "The importance of intrinsic disorder for protein phosphorylation," *Nucleic Acids Research*, vol. 32, no. 3, pp. 1037–1049, 2004.

[33] V. Uversky, "Introduction to Intrinsically Disordered Proteins (IDPs)," *Chemical Reviews*, vol. 114, pp. 6557–6560, 2014.

[34] E. A. Weathers, M. E. Paulaitis, T. B. Woolf, and J. H. Hoh, "Reduced amino acid alphabet is sufficient to accurately recognize intrinsically disordered protein," vol. 576, pp. 348–352, 2004.

[35] P. Romero, Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker, "Sequence Complexity of Disordered Protein," vol. 48, no. August 2000, pp. 38–48, 2001.

[36] R. Van Der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu, "Classification of intrinsically disordered regions and proteins," *Chemical Reviews*, vol. 114, no. 13, pp. 6589–6631, 2014.

[37] S. Lise and D. T. Jones, "Sequence Patterns Associated With Disordered Regions in," vol. 150, no. June 2004, pp. 144–150, 2005.

[38] A. BHATTARAI and I. A. EMERSON, "Dynamic conformational flexibility and molecular interactions of intrinsically disordered proteins," *J Biosci*, vol. 0123456789, pp. 1–17, 2020.

[39] L. Mollica, L. M. Bessa, X. Hanoulle, M. R. Jensen, M. Blackledge, and R. Schneider, "Binding mechanisms of intrinsically disordered proteins: Theory, simulation, and experiment," *Frontiers in Molecular Biosciences*, vol. 3, no. SEP, 2016.

[40] M. Fuxreiter, "Fuzziness in Protein Interactions—A Historical Perspective," *Journal of Molecular Biology*, vol. 430, no. 16, pp. 2278–2287, 2018.

[41] G. G. Hammes, Y. C. Chang, and T. G. Oas, "Conformational selection or induced fit: A flux description of reaction mechanism," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 33, pp. 13737–13741, 2009.

[42] J. B. Stiller, R. Otten, D. Häussinger, P. S. Rieder, D. L. Theobald, and D. Kern, "Structure determination of high-energy states in a dynamic protein ensemble," no. February 2021, 2022.

[43] R. B. Best, "Computational and theoretical advances in studies of intrinsically disordered proteins," *Current Opinion in Structural Biology*, vol. 42, pp. 147–154, 2017.

[44] H. J. Dyson and P. E. Wright, "Perspective : the essential role of NMR in the discovery and characterization of intrinsically disordered proteins," *Journal of Biomolecular NMR*, vol. 73, no. 12, pp. 651–659, 2019.

[45] R. Konrat, "NMR contributions to structural dynamics studies of intrinsically disordered proteins," *Journal of Magnetic Resonance*, vol. 241, no. 1, pp. 74–85, 2014.

[46] E. B. Gibbs, E. C. Cook, and S. A. Showalter, "Application of NMR to studies of intrinsically disordered proteins," *Archives of Biochemistry and Biophysics*, vol. 628, pp. 57–70, 2017.

[47] R. Schneider, M. Blackledge, and M. R. Jensen, "Elucidating binding mechanisms and dynamics of intrinsically disordered protein complexes using NMR spectroscopy," *Current Opinion in Structural Biology*, vol. 54, pp. 10–18, 2019.

[48] W. Wang, "Recent advances in atomic molecular dynamics simulation of intrinsically disordered proteins," *Phys. Chem. Chem. Phys.*, vol. 23, pp. 777–784, 2021.

[49] K. Kasahara, H. Terazawa, T. Takahashi, and J. Higo, "Studies on Molecular Dynamics of Intrinsically Disordered Proteins and Their Fuzzy Complexes: A Mini-Review," *Computational and Structural Biotechnology Journal*, vol. 17, pp. 712–720, 2019.

[50] Cliver R. Bagshaw, *Biomolecular kinetics : a step-by-step guide*. 2017.

[51] A. D. Vogt and E. Di Cera, "Conformational selection is a dominant mechanism of ligand binding," *Biochemistry*, vol. 52, no. 34, pp. 5723–5729, 2013.

[52] E. Fischer, "Einfluss der Configuration auf die Wirkung der Enzyme.," *Ber. Dtsch. Chem. Ges.*, vol. 27, pp. 2985–2993, 1894.

[53] T. D. Pollard, "MBOC technical perspective: A guide to simple and informative binding assays," *Molecular Biology of the Cell*, vol. 21, no. 23, pp. 4061–4067, 2010.

[54] E. D. Cera, "Mechanisms of ligand binding," vol. 011303, no. November, 2020.

[55] A. D. Vogt and E. Di Cera, "Conformational selection or induced fit? A critical appraisal of the kinetic mechanism," *Biochemistry*, vol. 51, no. 30, pp. 5894–5902, 2012.

[56] D. E. Koshland, "Application of a Theory of Enzyme Specificity to Protein Synthesis," *Proceedings of the National Academy of Sciences*, vol. 44, no. 2, pp. 98–104, 1958.

[57] D. E. Koshland, G. Némethy, and D. Filmer, "Comparison of experimental binding data and theoretical models in proteins containing subunits*," *Biochemistry*, vol. 5, no. 1, pp. 365–385, 1966. PMID: 5938952.

[58] J. Wyman Jr. and D. W. Allen, "The problem of the heme interactions in hemoglobin and the basis of the bohr effect," *Journal of Polymer Science*, vol. 7, no. 5, pp. 499–518, 1951.

[59] J. MONOD, J.-P. CHANGEUX, and F. JACOB, "Allosteric Proteins and Cellular Control Systems," pp. 306–329, 1963.

[60] J. P. CHANGEUX, "The feedback control mechanisms of biosynthetic L-threonine deaminase by L-isoleucine.," *Cold Spring Harbor symposia on quantitative biology*, vol. 26, pp. 313–318, 1961.

[61] J. Monod, J. Wyman, and J.-P. Changeux, "On the nature of allosteric transitions: A plausible model," *Journal of Molecular Biology*, vol. 12, no. 1, pp. 88–118, 1965.

[62] K. S. Chakrabarti, R. V. Agafonov, F. Pontiggia, R. Otten, M. K. Higgins, G. F. Schertler, D. D. Oprian, and D. Kern, "Conformational Selection in a Protein-Protein Interaction Revealed by Dynamic Pathway Analysis," *Cell Reports*, vol. 14, no. 1, pp. 32–42, 2016.

[63] A. D. Vogt, P. Chakraborty, and E. Di Cera, "Kinetic dissection of the pre-existing conformational equilibrium in the trypsin fold," *Journal of Biological Chemistry*, vol. 290, no. 37, pp. 22435–22445, 2015.

[64] A. Bah, L. C. Garvey, J. Ge, and E. Di Cera, "Rapid kinetics of Na+ binding to thrombin," *Journal of Biological Chemistry*, vol. 281, no. 52, pp. 40049–40056, 2006.

[65] L. Dahal, S. L. Shammas, and J. Clarke, "Phosphorylation of the IDP KID Modulates Affinity for KIX by Increasing the Lifetime of the Complex," *Biophysical Journal*, vol. 113, no. 12, pp. 2706–2712, 2017.

[66] R. Narayanan, O. K. Ganesh, A. S. Edison, and S. J. Hagen, "Kinetics of folding and binding of an intrinsically disordered protein: The inhibitor of yeast aspartic proteinase YPrA," *Journal of the American Chemical Society*, vol. 130, no. 34, pp. 11477–11485, 2008.

[67] J. M. Rogers, V. Oleinikovas, S. L. Shammas, C. T. Wong, D. De Sancho, C. M. Baker, and J. Clarke, "Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 43, pp. 15420–15425, 2014.

[68] J. Dogan, T. Schmidt, X. Mu, Å. Engström, and P. Jemth, "Fast association and slow transitions in the interaction between two intrinsically disordered protein domains," *Journal of Biological Chemistry*, vol. 287, no. 41, pp. 34316–34324, 2012.

[69] A. Sekhar, A. Velyvis, G. Zoltsman, R. Rosenzweig, G. Bouvignies, and L. E. Kay, "Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions," *eLife*, vol. 7, pp. 1–29, 2018.

[70] D. D. Boehr, R. Nussinov, and P. E. Wright, "The role of dynamic conformational ensembles in biomolecular recognition," *Nature Chemical Biology*, vol. 5, no. 11, pp. 789–796, 2009.

[71] R. Galletto, M. J. Jezewska, and W. Bujalowski, "Kinetics of allosteric conformational transition of a macromolecule prior to ligand binding: Analysis of stopped-flow kinetic experiments," *Cell Biochemistry and Biophysics*, vol. 42, no. 2, pp. 121–144, 2005.

[72] S. Gianni, J. Dogan, and P. Jemth, "Distinguishing induced fit from conformational selection," *Biophysical Chemistry*, vol. 189, pp. 33–39, 2014.

[73] X. Du, Y. Li, Y. L. Xia, S. M. Ai, J. Liang, P. Sang, X. L. Ji, and S. Q. Liu, "Insights into protein–ligand interactions: Mechanisms, models, and methods," *International Journal of Molecular Sciences*, vol. 17, no. 2, pp. 1–34, 2016.

[74] K. Takeuchi, K. Baskaran, and H. Arthanari, "Structure determination using solution NMR: Is it worth the effort?," *Journal of Magnetic Resonance*, vol. 306, pp. 195–201, 2019.

[75] G. S. Rule and T. K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*. Dordrecht, The Netherlands: Springer, 1 ed., 2006.

[76] P. Vallurupalli, A. Sekhar, T. Yuwen, and L. E. Kay, "Probing conformational dynamics in biomolecules via chemical exchange saturation transfer: a primer," *Journal of Biomolecular NMR*, vol. 67, no. 4, pp. 243–271, 2017.

[77] J. Keeler, *Understanding NMR Spectroscopy*. Wiley, 2011.

[78] F. Bloch, "Nuclear induction," *Physical Review*, vol. 70, no. 7-8, pp. 460–474, 1946.

[79] P. Vallurupalli, G. Bouvignies, and L. E. Kay, "Studying "invisible" excited protein states in slow exchange with a major state conformation," *Journal of the American Chemical Society*, vol. 134, no. 19, pp. 8148–8161, 2012.

[80] L. Sjöstrand, *Method Development for Thermal Stability Analysis by Circular Dichroism*. PhD thesis, 2018.

[81] N. J. Greenfield, "Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions," *Nature Protocols*, vol. 1, no. 6, pp. 2527–2535, 2007.

[82] J. Seelig and H.-J. Schönfeld, " Thermal protein unfolding by differential scanning calorimetry and circular dichroism spectroscopy Two-state model versus sequential unfolding ," *Quarterly Reviews of Biophysics*, vol. 49, 2016.

[83] W. P. Oliveira-Souza, F. Bronze, J. Broos, M. F. Marcondes, and V. Oliveira, "On the efficient bio-incorporation of 5-hydroxy-tryptophan in recombinant proteins expressed in Escherichia coli with T7 RNA polymerase-based vectors," *Biochemical and Biophysical Research Communications*, vol. 492, no. 3, pp. 343–348, 2017.

[84] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, M. Llinas, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue, "The ccpn data model for nmr spectroscopy: Development of a software pipeline," *Proteins: Structure, Function, and Bioinformatics*, vol. 59, no. 4, pp. 687–696, 2005.

[85] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax, "NMRPipe: A multidimensional spectral processing system based on UNIX pipes," *Journal of Biomolecular NMR*, vol. 6, no. 3, pp. 277–293, 1995.

[86] N. A. Farrow, O. Zhang, J. D. Forman-Kay, and L. E. Kay, "A heteronuclear correlation experiment for simultaneous determination of 15N longitudinal decay and chem-

ical exchange rates of systems in slow equilibrium," *Journal of Biomolecular NMR*, vol. 4, no. 5, pp. 727–734, 1994.

[87] S. Mori, C. Abeygunawardana, M. O. Johnson, and P. C. Van Zljl, "Improved Sensitivity of HSQC Spectra of Exchanging Protons at Short Interscan Delays Using a New Fast HSQC (FHSQC) Detection Scheme That Avoids Water Saturation," *Journal of Magnetic Resonance, Series B*, vol. 108, no. 1, pp. 94–98, 1995.

[88] P. Vallurupalli and L. E. Kay, "Probing slow chemical exchange at carbonyl sites in proteins by chemical exchange saturation transfer NMR spectroscopy," *Angewandte Chemie - International Edition*, vol. 52, no. 15, pp. 4156–4159, 2013.

[89] A. Sekhar, R. Rosenzweig, G. Bouvignies, and L. E. Kay, "Hsp70 biases the folding pathways of client proteins," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 20, pp. E2794–2801, 2016.

[90] M. V. Cardoso, J. D. Rivera, P. A. Vitale, M. F. Degenhardt, L. A. Abiko, C. L. Oliveira, and R. K. Salinas, "CALX-CBD1 Ca2+-Binding Cooperativity Studied by NMR Spectroscopy and ITC with Bayesian Statistics," *Biophysical Journal*, vol. 119, no. 2, pp. 337–348, 2020.

[91] M. P. Williamson, "Using chemical shift perturbation to characterise ligand binding," *Progress in Nuclear Magnetic Resonance Spectroscopy*, vol. 73, pp. 1–16, 2013.

[92] L. Whitmore and B. A. Wallace, "DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data," *Nucleic Acids Research*, vol. 32, no. WEB SERVER ISS., pp. 668–673, 2004.

[93] A. J. Miles, S. G. Ramalli, and B. A. Wallace, "DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data," *Protein Science*, vol. 31, no. 1, pp. 37–46, 2022.

[94] L. Whitmore and B. A. Wallace, "Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases," *Biopolymers*, vol. 89, no. 5, pp. 392–400, 2008.

[95] H. Nakatani, M. Haga, and K. Hiromi, "Kinetic studies on binding of bovine serum albumin with 1-anilino-8-naphthalene sulfonate," *FEBS Letters*, vol. 43, no. 3, pp. 293–296, 1974.

[96] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, "emcee: The MCMC Hammer," *PASP*, vol. 125, p. 306, Mar. 2013.

[97] R. Storn and K. Price, "Differential evolution â€" a simple and efficient heuristic for global optimization over continuous spaces," *Journal of Global Optimization*, vol. 11, pp. 341–359, Dec 1997.

[98] V. P. Saxena and D. B. Wetlaufer, "A new basis for interpreting the circular dichroic spectra of proteins.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 68, no. 5, pp. 969–972, 1971.

[99] R. W. Woody, "Aromatic side-chain contributions to the far ultraviolet circular dichroism of peptides and proteins," *Biopolymers*, vol. 17, no. 6, pp. 1451–1467, 1978.

[100] I. B. Grishina and R. W. Woody, "Contributions of tryptophan side chains to the circular dichroism of globular proteins: Exciton couplets and coupled oscillators," *Faraday Discussions*, vol. 99, pp. 245–262, 1994.

[101] R. W. Woody, "Contributions of tryptophan side chains to the far-ultraviolet circular dichroism of proteins," *European Biophysics Journal*, vol. 23, no. 4, pp. 253–262, 1994.

[102] G. V. Semisotnov, N. A. Rodionova, O. I. Razgulyaev, V. N. Uversky, A. F. Gripas', and R. I. Gilmanshin, "Study of the "molten globule" intermediate state in protein

folding by a hydrophobic fluorescent probe," *Biopolymers*, vol. 31, no. 1, pp. 119–128, 1991.

[103] L. Stryer, "The interaction of a naphthalene dye with apomyoglobin and apohemoglobin: A fluorescent probe of non-polar binding sites," *Journal of Molecular Biology*, vol. 13, no. 2, pp. 482–495, 1965.

[104] G. Barone, F. Catanzano, P. Del Vecchio, C. Giancola, and G. Graziano, "Differential scanning calorimetry as a tool to study protein-ligand interactions," *Pure and Applied Chemistry*, vol. 67, no. 11, pp. 1867–1872, 1995.

[105] T. Hendrix, Y. V. Griko, and P. L. Privalov, "A calorimetric study of the influence of calcium on the stability of bovine α-lactalbumin," *Biophysical Chemistry*, vol. 84, no. 1, pp. 27–34, 2000.

[106] M. González, C. E. Argaraña, and G. D. Fidelio, "Extremely high thermal stability of streptavidin and avidin upon biotin binding," *Biomolecular Engineering*, vol. 16, no. 1-4, pp. 67–72, 1999.

[107] C. M. Johnson, "Differential scanning calorimetry as a tool for protein folding and stability," *Archives of Biochemistry and Biophysics*, vol. 531, no. 1-2, pp. 100–109, 2013.

[108] M. S. Celej, G. D. Fidelio, and S. A. Dassie, "Protein unfolding coupled to ligand binding: Differential scanning calorimetry simulation approach," *Journal of Chemical Education*, vol. 82, no. 1, pp. 85–92, 2005.

[109] F. Pucci and M. Rooman, "Physical and molecular bases of protein thermal stability and cold adaptation," *Current Opinion in Structural Biology*, vol. 42, pp. 117–128, 2017.

[110] C. J. Layton and H. W. Hellinga, "Thermodynamic analysis of ligand-induced changes in protein thermal unfolding applied to high-throughput determination of ligand affinities with extrinsic fluorescent dyes," *Biochemistry*, vol. 49, no. 51, pp. 10831–10841, 2010.

[111] A. Cooper, "Protein heat capacity: An anomaly that maybe never was," *Journal of Physical Chemistry Letters*, vol. 1, no. 22, pp. 3298–3304, 2010.

[112] S. A. Shirdel and K. Khalifeh, "Thermodynamics of protein folding: methodology, data analysis and interpretation of data," *European Biophysics Journal*, vol. 48, no. 4, pp. 305–316, 2019.

[113] A. M. Stadler, E. Pellegrini, M. Johnson, J. Fitter, and G. Zaccai, "Dynamics-stability relationships in apo- and holomyoglobin: A combined neutron scattering and molecular dynamics simulations study," *Biophysical Journal*, vol. 102, no. 2, pp. 351–359, 2012.

[114] J. Sancho, "The stability of 2-state, 3-state and more-state proteins from simple spectroscopic techniques... plus the structure of the equilibrium intermediates at the same time," *Archives of Biochemistry and Biophysics*, vol. 531, no. 1-2, pp. 4–13, 2013.

[115] T. Schulthess, H. J. Schönfeld, and J. Seelig, "Thermal unfolding of apolipoprotein A-1. Evaluation of methods and models," *Biochemistry*, vol. 54, no. 19, pp. 3063–3075, 2015.

[116] E. Judy and N. Kishore, "A look back at the molten globule state of proteins: thermodynamic aspects," *Biophysical Reviews*, vol. 11, no. 3, pp. 365–375, 2019.

[117] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax, "TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts," *Journal of Biomolecular NMR*, vol. 44, no. 4, pp. 213–223, 2009.

[118] K. Tamiola, B. Acar, and F. A. Mulder, "Sequence-specific random coil chemical shifts of intrinsically disordered proteins," *Journal of the American Chemical Society*, vol. 132, no. 51, pp. 18000–18003, 2010.

[119] K. Teilum, J. G. Olsen, and B. B. Kragelund, "Globular and disordered-the non-identical twins in protein-protein interactions," *Frontiers in Molecular Biosciences*, vol. 2, no. JUL, pp. 1–6, 2015.

[120] T. Clackson and J. A. Welist, "A Hot Spot of Binding Energy in a Hormone-Receptor Interface," *Scientific Reports*, vol. 267, no. January, pp. 383–387, 1995.

[121] B. B. Kragelund and K. Skriver, "Binding Thermodynamics to Intrinsically Disordered Protein Domains," in *Intrinsically Disordered Proteins*, p. 933, 2020.

[122] R. Perozzo, G. Folkers, and L. Scapozza, "Thermodynamics of Protein – Ligand Interactions : History , Presence , and Future Aspects," *Journal of receptors and signal transductions*, vol. 24, pp. 1–51, 2004.

[123] P. D. Kwong, R. Wyatt, J. Robinson, R. W. Sweet, J. Sodroski, and W. A. Hendrickson, "Structure of an HIV gp 120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody," *Nature*, vol. 393, no. 6686, pp. 648–659, 1998.

[124] D. G. Myszka, R. W. Sweet, P. Hensley, M. Brigham-Burke, P. D. Kwong, W. A. Hendrickson, R. Wyatt, J. Sodroski, and M. L. Doyle, "Energetics of the HIV gp120-CD4 binding reaction," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 16, pp. 9026–9031, 2000.

[125] P. D. Kwong, M. L. Doyle, D. J. Casper, C. Cicala, S. A. Leavitt, S. Majeed, T. D. Steenbeke, M. Venturi, I. Chaiken, M. Fung, H. Katinger, P. W. Parren, J. Robinson, D. Van Ryk, L. Wang, D. R. Burton, E. Freire, R. Wyatt, J. Sodroski, W. A.

Hendrickson, and J. Arthos, "HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites," *Nature*, vol. 420, no. 6916, pp. 678–682, 2002.

[126] R. Pantophlet, E. Ollmann Saphire, P. Poignard, P. W. H. I. Parren, I. A. Wilson, and D. R. Burton, "Fine Mapping of the Interaction of Neutralizing and Nonneutralizing Monoclonal Antibodies with the CD4 Binding Site of Human Immunodeficiency Virus Type 1 gp120," *Journal of Virology*, vol. 77, no. 1, pp. 642–658, 2003.

[127] S. Vega, O. Abian, and A. Velazquez-Campoy, "On the link between conformational changes, ligand binding and heat capacity," *Biochimica et Biophysica Acta - General Subjects*, vol. 1860, no. 5, pp. 868–878, 2016.

[128] J. R. Lakowicz, *Principles of fluorescence spectroscopy.* 2006.