UNIVERSIDADE DE SÃO PAULO

DEPARTAMENTO DE FÍSICA – FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO

LUIS FELIPE SANTOS MENDES

Caracterização estrutural e dinâmica da proteína de estruturação e compactação do complexo de Golgi (GRASP) em solução

Ribeirão Preto

2018

Luis Felipe Santos Mendes

Caracterização estrutural e dinâmica da proteína de estruturação e compactação do complexo de Golgi (GRASP) em solução

Tese apresentada ao Departamento de Física da FFCLRP-USP como parte das exigências para a obtenção do título de doutor em ciências

Área de concentração: Física Aplicada à Medicina e Biologia

Orientador: Prof. Dr. Antonio Jose da Costa Filho

Ribeirão Preto

2018

Ficha Catalográfica

Mendes, Luis Felipe Santos

Caracterização estrutural e dinâmica da proteína de estruturação e compactação do complexo de Golgi (GRASP) em solução, Ribeirão Preto, 2018. 154 p.

Tese de Doutorado apresentada à Faculdade de Filosofia Ciências e Letras de Ribeirão Preto/USP – Área de concentração: Física Aplicada à Medicina e Biologia

Orientador: Costa-Filho, Antonio José.

1. GRASP 2. Secreção não convencional de proteínas 3. Espectroscopia 4. Proteínas Intrinsicamente Desordenadas 5. Molten Globule

Nome: MENDES, Luis Felipe Santos

Título: Caracterização estrutural e dinâmica da proteína de estruturação e compactação do complexo de Golgi (GRASP) em solução

Tese apresentada ao Departamento de Física da FFCLRP-USP como parte das exigências para obtenção do título de Doutor em Ciências (Física Aplicada à Medicina e Biologia). Área de concentração: Física Aplicada à Medicina e Biologia

Aprovado em: ___/___/___

Banca Examinadora

Prof. Dr. _____ Instituição: _____

Julgamento:_____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento:_____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento:_____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento:_____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento:_____ Assinatura: _____

Janus

Universidade de São Paulo

ATA DE DEFESA

Ata de defesa pública de Tese do(a) Senhor(a) Luis Felipe Santos Mendes no Programa: Física Aplicada à Medicina e Biologia, do(a) Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo.

Aos 07 dias do mês de fevereiro de 2018, no(a) Bloco B2 - sala 220 realizou-se a Defesa de Tese do(a) Senhor(a) Luis Felipe Santos Mendes, apresentada para a obtenção do título de Doutor intitulada:

"Caracterização estrutural e dinâmica da proteína de estruturação e compactação do complexo de Golgi (GRASP) em solução"

Após declarada aberta a sessão, o(a) Sr(a) Presidente passa a palavra ao candidato para exposição e a seguir aos examinadores para as devidas arguições que se desenvolvem nos termos regimentais. Em seguida, a Comissão Julgadora proclama o resultado:

Nome dos Participantes da Banca	Função	Sigla da CPG	Resultado
Antônio José da Costa Filho	Presidente	FFCLRP - USP	APROVADO
Glaucius Oliva	Titular	IFSC - USP	APROVADO
Carlos Henrique Inacio Ramos	Titular	UNICAMP - Externo	Aprovado
Fabio Ceneviva Lacerda de Almeida	Titular	UFRJ - Externo	APROVADO
Richard Charles Garratt	Suplente	IFSC - USP	APROVADO

videoconferência

Resultado Final: APROVADO

Parecer da Comissão Julgadora *

O Prof. Glaucius Oliva participou por videoconferência.

Eu, Cesar Ponim Brites Cesar Brites, lavrei a presente ata, que assino juntamente com os(as) Senhores(as). Ribeirão Preto, aos 07 dias do mês de fevereiro de 2018.

Glaucius Oliva

Carlos Henrique Inacio Ramos

Fabio Ceneviva Lacerda de Almeida

Richard Charles Garratt

Antônio José da Costa Filho
Presidente da Comissão Julgadora

* Obs: Se o candidato for reprovado por algum dos membros, o preenchimento do parecer é obrigatório.

A defesa foi homologada pela Comissão de Pós-Graduação em 01/03/2018 e, portanto, o(a) aluno(a) faz jus ao título de Doutor em Ciências obtido no Programa Física Aplicada à Medicina e Biologia.

Confere com o original
09/03/2018
Rossana Minghetti Ribeiro
Técnica p/ Assuntos Administ
Nº USP 6582950

Presidente da Comissão de Pós-Graduação

Profa. Dra. Sonia Regina Fusion
Presidente da Comissão de Pós-Graduação

This thesis is dedicated to my family, for all their love, incentive and support. There is no bigger love than this that I feel for you.

Acknowledgments

To do my PhD in the Physics Department of FFCLRP-USP was undoubtedly one of the best and unique experience I could have ever had in my life. I would like to thank God for giving me the strength to go through all the hard and easy challenges in the process and helped me to get many insights out of nowhere. As a scientist, I should not believe in something or someone I cannot prove the existence but what I feel is much stronger and clearer for me than even the sensation of cold and heat. I will never deny my faith and this has only made me stronger with time.

I am especially grateful for all the love, dedication, incentive, support and several other things that my parents have (done) for me since ever. I would never be in this position without your unquestionable love, support, and patience. Hope that this thesis and all the things that I have conquered so far might make you proud of me, despite all the hard work, nights without sleeping (in particular my father working overnight in Petrobras) and headaches that I often give you. I feel that I am getting really far in my career and just like Newton once said, I truly believe that I could only see and go this far because during my entire life I was standing on the shoulders of two giants. Thanks for being my giants, love you both. I am also especially grateful for the friendship, patience and love of my little sister Lidiane. Thanks for helping me not getting crazy during my infinite crises. You were a great roommate and friend. Hope to be a person as kind and good as you one day. Still working on this, it's hard than do science. *"I lived a million miles of memories on that road; with every step I take I know that I'm not alone. You take the home from the boy, but not the boy from his home… Who says you can't go home? (John Bon Jovi, 2005)"*

I am extremely thankful to my supervisor, Antonio Jose da Costa Filho (Jabah), for his guidance, support, advices and suggestions in all moments during my entire PhD time. Thanks for trusting me and for letting me have the unique opportunity of working under your supervision over the last 8 years of my scientific life. Thanks also for the several beers, whiskies and the friendship. Your support was essential for my growing as a person, scientist, and, of course, for the success of this PhD thesis. I would also like to thank all the actual and former members of the LBM. All of you have an especial part in this thesis.

I am thankful to Prof. Dr. Anthony Watts for giving me the unique opportunity of working with you and his group at the University of Oxford-UK. It was such an outstanding experience that I cannot measure or simply define using ordinary words. I would also like to give an especial thank to the Watts group for the help and friendship during this amazing year.

I am thankful to Prof. Dr. Christina Redfield for guiding me into the solution NMR world. Thanks for being so patient, coming to the NMR machine with me and sharing a little of your infinite knowledge.

I am thankful to Gunnar Jeschke and Alex Smirnov for my short internships in the ETH Zurich and North Caroline State University, respectively. For sure, these times in Switzerland and USA contributed enormously to my scientific maturation. Thanks also for all the members of both groups.

I am also thankful to Mariana (Gatinha) for the company, friendship and for being on and by my side from the beginning until the very end of this Ph.D thesis. Our relationship had several stages and I am thankful for each Mariana that you were in my life (the colleague, the friend and the girlfriend). You made and make me a better person and professional. Hope to have you still by my side making me better for as long as you can handle. *"If you are lucky enough to find love, remember it is there and don't throw it away (Stephen Hawking)"*

I am especially grateful to Dr. Luis Guilherme Mansor Basso and Dr. Assuero Faria Garcia for being my friends and indirect co-supervisors during these years. Without you, guys, this thesis would be nothing more than a stamp collection. I believe I have now several good qualities as a professional that I copied-and-pasted from you both. Hope to be as good scientist as you are. I am also thankful to Dra. Patricia Suemy Kumagai for the SRCD data collection presented in chapter 2. It is always an honor to work and have some beer with you.

Coming to people that had a special importance for me during these years outside the lab. You all had a huge contribution!!! Science is much more than the work inside the wet lab or on the computer. Without all of you, I would have got crazy in the very beginning. I know that there are several names and I don't want to forget anyone, so I will avoid a list. If you were by my side during my PhD (and if you are one of these people you will know that you are a part of this), just know that you are important and I am thankful.

I would like to thank Tequila for being by my side during this time. You did not choose it, I did, but you still look okay with this hehehehehe There is nothing more pure and truthful than the friendship of an animal. Jurema was not direct by my side but soon she will be in the water castle she deserves and Tequila will have his entire garden to eat without any limits.

I am extremely thankful to FAPESP for my grants during my PhD and the BEPE internship. CNPq, BBSRC and CAPES are also acknowledge.

“Science knows no country, because knowledge belongs to humanity, and is the torch which illuminates the world.”

(Louis Pasteur, Free Lance of Science (1960) by René Jules Dubos, Ch. 3 "Pasteur in Action")

“Imagine there's no countries... No need for greed or hunger, a brotherhood of man”

(John Lennon, 1971)

Let’s connect ourselves through knowledge

Resumo

O complexo de Golgi é um organela responsável pela recepção de carga sintetizada no retículo endoplasmático e por subsequente modificações pós-traducionais, classificação e secreção. Uma família de proteínas chamada *Golgi Reassembly and Stacking Proteins* (GRASP) é essencial para o correto empilhamento das cisternas e conexões laterais das pilhas do complexo de Golgi, uma estruturação necessária para manter essa organela funcionando corretamente. A estrutura das GRASPs é composta de duas regiões principais: uma extensão N-terminal formado por dois domínios PDZ conectados por um loop (domínio GRASP) e uma região C-terminal não conservada, rica em resíduos de serina e prolina. Embora existam algumas estruturas cristalográficas resolvidas para o domínio N-terminal, é surpreendente notar que não havia nenhuma informação na literatura sobre a construção inteira de um GRASP, ou mesmo um estudo detalhado sobre os PDZs no N-terminal em solução, que é a principal região funcional dessa proteína. Usando um modelo de GRASP em sua construção completa, fomos capazes de detectar a coexistência de estruturas secundárias regulares e grandes quantidades de regiões desordenadas. A estrutura é menos compacta do que uma proteína globular e a alta flexibilidade estrutural torna o seu núcleo hidrofóbico mais acessível ao solvente. GRASPs coexistem em um conjunto conformacional dinâmico numa escala de tempo característico de μs-ms. Nossos resultados indicam um comportamento incomum da GRASP em solução, similar à de uma classe de proteínas intrinsicamente desordenadas colapsadas conhecidas como glóbulos fundidos. Nós relatamos também as propensões de transição estrutural do tipo desordem-ordem para uma proteína glóbulo fundido nativa, induzidas pela presença de diferentes miméticos de condições celulares especificas. A mudança na constante dielétrica do meio (como as experimentadas próximas à superfície da membrana biológica) é o principal modulador estrutural, capaz de induzir múltiplas transições desordem-ordem na GRASP, sugerindo um comportamento muito distinto quando em condições que imitam a vizinhança da superfície da membrana em comparação com os encontrados quando livre em solução. Outros fatores de enovelamento, tais como o *molecular crowding*, contra-ions, pH e a fosforilação exibem efeitos menores (ou nenhum) na estrutura secundária e/ou estabilidade da GRASP. Este é o primeiro estudo focado na compreensão das transições desordem-ordem em uma estrutura do tipo glóbulo fundido sem que houvesse a necessidade de qualquer condição desnaturante. Em relação aos PDZs que formam o domínio GRASP, observamos que as GRASPs são formadas por um PDZ1 mais instável e flexível e um PDZ2 muito mais estável e estruturalmente bem comportado. Mais do que isso, muitas das regiões instáveis encontradas no PDZ1 estão no predito bolsão de ligação, sugerindo uma promiscuidade estrutural dentro desse domínio que se correlaciona com a promiscuidade funcional de interação com múltiplos parceiros proteicos. É apresentado nesta tese a primeira caracterização estrutural de uma GRASP em sua forma completa, o primeiro modelo de como as GRASPs (ou qualquer proteína em forma de glóbulo fundido) pode ser modulada estruturalmente pela célula durante diferentes funcionalidades e o primeiro trabalho na comunidade provando que a estabelecido ideia de que ambos os PDZs são estruturalmente equivalentes não é completamente correta.

Palavras chaves:

1. GRASP 2. Secreção não convencional de proteínas 3. Espectroscopia 4. Proteínas intrinsicamente desordenadas 5. *Molten Globule*

Abstract

The Golgi complex is an organelle responsible for receiving synthesized cargo from the endoplasmic reticulum for subsequent post-translations modifications, sorting and secretion. A family of proteins named Golgi Reassembly and Stacking Proteins (GRASP) is essential for the correct assembly and laterally tethering of the Golgi cisternae, a necessary structuration to keep this organelle working correctly. The GRASP structure is mainly composed of two regions: an N-terminal formed by two PDZ domains connected by a short loop (GRASP domain) and a non-conserved C-terminal region, rich in serine and proline residues. Although there are now a few crystal structures solved for the N-terminal domain, it is surprising to notice that no information is currently available regarding a full-length protein or even about dynamic and structural differences between the two PDZs in solution, which is the main functional region of this protein. Using a full-length GRASP model, we were capable of detecting the coexistence of regular secondary structures and large amounts of disordered regions. The overall structure is less compact than a regular globular protein and the high structural flexibility makes its hydrophobic core more accessible to solvent. GRASP coexist in a dynamic conformational ensemble of a μs-ms timescale. Our results indicate an unusual behavior of GRASP in solution, closely resembling a class of collapsed intrinsically disordered proteins called molten globule. We report here also the disorder-to-order transition propensities for a native molten globule-like protein in the presence of different mimetics of cell conditions. Changes in the dielectric constant (such as those experienced close to the membrane surface) seem to be the major factor capable of inducing several disorder-to-order transitions in GRASP, which seems to show very distinct behavior when in conditions that mimic the vicinity of the membrane surface as compared to those found when free in solution. Other folding factors such as molecular crowding, counter ions, pH and phosphorylation exhibit lower or no effect on GRASP secondary structure and/or stability. This is the first study focusing on understanding the disorder-to-order transitions of a molten globule structure without the need for any mild denaturing condition. Regarding the PDZs that form the GRASP domain, we observed that GRASPs are formed by a more unstable and flexible PDZ1 and much more stable and structurally well-behaved PDZ2. More than that, many of the unstable regions found in PDZ1 are in the predicted binding pocket, suggesting a structural promiscuity inside this domain that correlates with the functional promiscuity of interacting with multiple protein partners. This thesis presents the first structural characterization of a full-length GRASP, the first model of how GRASPs (or any molten globule-like protein) can be modulated by the cell during different cell functionalities and the first work in the community proving that the established idea that both PDZs are structurally equivalent is not completely right.

Keywords:

1. GRASP 2. Unconventional Protein Secretion 3. Spectroscopy 4. Intrinsically Disordered Proteins 5. Molten Globule

List of Figures

List of tables

List of manuscript published during the Ph.D.

Mendes LFS, Garcia AF, Kumagai PS, de Morais FR, Melo FA, Kmetzsch L, Vainstein MH, Rodrigues ML, Costa-Filho AJ. New structural insights into Golgi Reassembly and Stacking Protein (GRASP) in solution. *Sci Rep*. 2016 Jul 20;6:29976.

Mendes LFS, Basso LGM, Kumagai PS, Fonseca-Maldonado R, Costa-Filho AJ. Disorder-to-order transitions in the molten globule-like Golgi Reassembly and Stacking Protein. *BBA general subjects. 2018 Apr;1862(4):855-865.*

Micheletto MC, **Mendes LFS***, Basso LGM, Fonseca-Maldonado RG, Costa-Filho AJ. Lipid membranes and acyl-CoA esters promote opposing effects on acyl-CoA binding protein structure and stability. *Int J Biol Macromol*. 2017 Sep;102:284-296.

Basso LGM, **Mendes LFS***, Costa-Filho AJ. The two sides of a lipid-protein story. *Biophys Rev*. 2016 Jun;8(2):179-191.

Alvarez N, Noble C, Torre MH, Kremer E, Ellena J, Peres de Araujo M, Costa-Filho AJ, **Mendes LFS**, Kramer MG, Facchin G. Synthesis, structural characterization and cytotoxic activity against tumor cells of heteroleptic copper (I) complexes with aromatic diimines and phosphines. *Inorganica Chim Acta*, v. 466, p. 559-564, 2017.

Fonseca-Maldonado R, Meleiro L, **Mendes LFS**, Alves LF, Carli S, Morero LD, Basso LGM, Costa-Filho AJ, Ward RJ. Lignocellulose binding of a Cel5A-RtCBM11 chimera with enhanced β-glucanase activity monitored by electron paramagnetic resonance. *Biotechnol Biofuels*. 2017;10:269

*** Shared first authorship**

List of manuscript in revision or preparation

Sergey M, Micheletto M, **Mendes LFS**, Costa-Filho AJ, Smirnov A, Smirnova T. A HYSCORE and Continuous Wave Q-band EPR Investigation of Substrate Binding to Chlorocatechol 1,2-Dioxygenase: an Allosteric Regulation by Native Amphiphilic Molecules. J Phys Chem B. 2017

Mendes LFS, Redfield C, Judge P, Watts A, Costa-Filho AJ. Why would nature give two PDZs for the Golgi Reassembly and Stacking Protein? *In preparation*

Citadini AP, **Mendes LFS***, Basso LGM, Lima JF, Costa-Filho AJ. Ion-induced conformational changes in human S100A12 calcium binding protein: A site-directed spin labeling study. *In preparation*

Molina GM, **Mendes LFS**, Fuzo C, Costa-Filho AJ, Ward RJ. Temperature affects substrate-binding dynamics of a GH11 xylanase. *In preparation*

Alvarez N, **Mendes LFS**, Kramerc G, Torre MT, Costa-Filho AJ, Ellena J, Facchin G. Development of Copper(II)-diimine-iminodiacetate mixed ligand complexes as potential cytotoxic agents. *In preparation*

*** Shared first authorship**

Summary

Chapter 1 – Introduction and scope of the thesis

The eukaryotic systems have several unique features when compared to the prokaryotic, with the presence of an extensive internal biomembrane system as the most prominent one [1]. Several different organelles, such as the Golgi complex and the Endoplasmic Reticulum (ER), fulfill very specific roles inside the eukaryotic cell and work independently in a general way [1]. Of course, this compartmentalization of functions allowed an increase in the overall efficiency and made possible the dramatically increase in cell volume [1]. Still, all the organelles require specific molecules that are synthesized elsewhere to arrive in time in order to sustain the correct functionalities that the cell needs to stay alive and be functional [1]. Because of this, the cell evolved to provide a very efficient transport system that could feed all the different organelles, send membrane proteins to the plasma membrane and active cargo to the extracellular environment [1]. In this context, the transport of proteins deserves a special attention since these are the biomolecules with the largest array of functions, including signaling processes, enzymatic catalyzes, initiator of several biological processes, transport of biomolecules across the membrane, synthesis of other biomolecules, among several others [1].

The Conventional Secretory Pathway

The conventional secretion is the cellular process present in every organism that delivers soluble proteins to the extracellular space and membrane proteins to the plasma membrane, in a ER/Golgi dependent way [2,3]. In the very beginning of the electron microscopy development, the scientists were also starting to introduce several methods for cell immobilization in surfaces and apply analytical ultracentrifugation to fractionalized different organelles. These methods allowed Albert Claude, George Palade e Christian de Duve to give precious data regarding the transport of biomolecules inside the cell [4,5,6]. These notable scientists showed for the first time that secretory proteins were first synthesized by the ribosome in the ER lumen and posteriorly transported to the Golgi complex where they were packed and sorted to their final destination [2]. Of course, this was a unique

finding and the three of them were laureate in 1974 with the Nobel Prize in Medicine or Physiology. Getting further insights about these processes has been probably one of the most intense study targets since then. In 1999, Günter Blobel was laureate with the Nobel Prize in Medicine or Physiology for his discovery that proteins targeted to the conventional secretory pathway possess an intrinsic amino acid peptide that is used as a guide signal by the secretory machinery to not only identify that this protein needs to be secreted, but also which is its final destination [7,8]. In the initial work of George Palade, it was suggested that small vesicles which are first formed in the ER surface, are the ones that mediate the biomolecule transport towards the Golgi [2] (Figure 1). These vesicles were found to be coated with structural proteins that are responsible for their arrange and function, and that is why they are now called COat Protein complex (COP), where the COPII is the responsible for the anterograde transport from the RE to the Golgi and COPI for the retrograde one [9,10,11] (Figure 1). In 2013, James E. Rothman, Randy W. Schekman e Thomas C. Sudhof won the Nobel Prize in Medicine or Physiology for their work involving the discovery of the machinery that regulates the vesicle traffic inside the cell, with a special attention to the formation, structure and fusion processes of the COPs [12]. This pathway is described by the Royal Swedish Academy of Sciences as the *"The Major Transport System in our Cells".*

As discussed before, Günter Blobel showed that secreted proteins have an intrinsic signal in their peptide sequence that drives them to the secretory pathway. We now know that a protein targeted to the ER needs to have a Signal Peptide (SP) in the N-terminus [1,8]. The SP is usually formed by 13-36 residues, with the N-terminus of this peptide containing at least one positively charged residue, a very hydrophobic region in the middle of the sequence and a C-terminus composed by amino acids with short side-chain [1]. When the messenger RNA (mRNA) binds to the ribosome, the translation process begins and the protein chain is prolonged until the SP is recognized by the Signal peptide Recognition Protein (SRP), a soluble riboprotein of approximately 300 kDa that can be found dispersed in the cytoplasm [13]. The protein translation is then paused and the Ribosome-mRNA-SP-SRP complex translocates to the ER surface, being recognized by the SRP receptor [14,15]. The interaction of the SRP with the SRP receptor promotes the exchange of GDP for GTP in the α subunit of the receptor, increasing its affinity for the SRP and leading to the release of the complex ribosome-mRNA-SP [13,14]. Concomitant with this process, the ribosome binds to other components of the ER membrane

that are collectively named *translocon* [16]. The *translocon* is then responsible for the transfer of the translated SP-protein to the ER lumen, where the protein translation is now continued [7] (Figure 1).

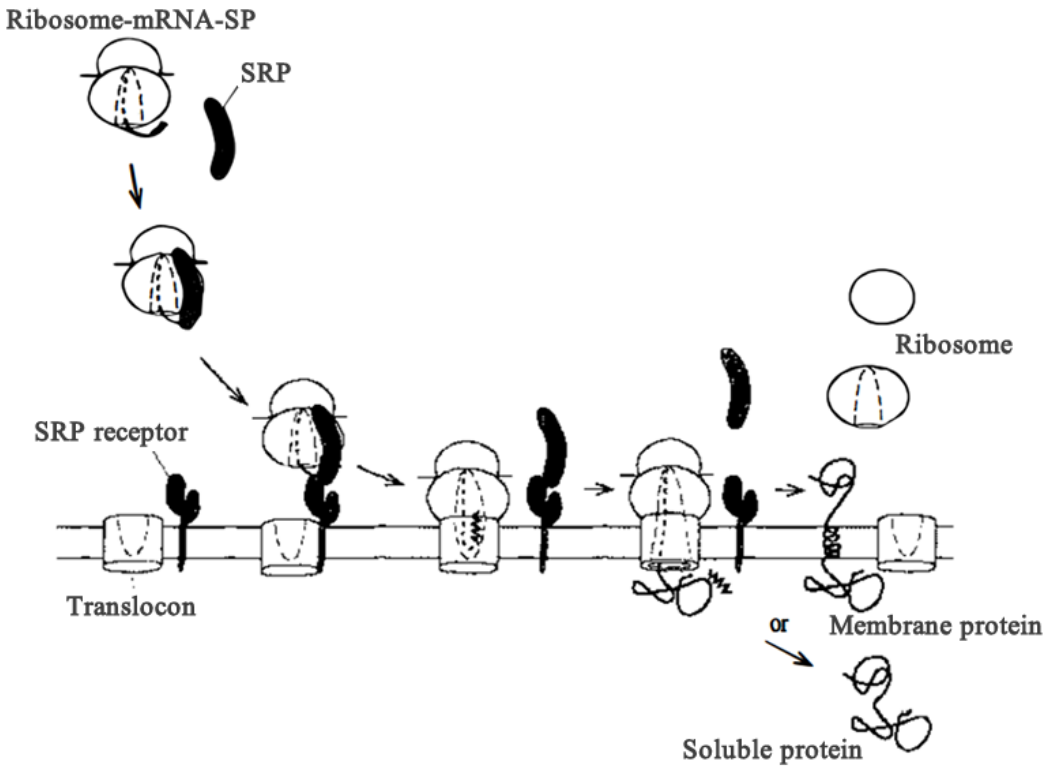


Figure 1: A model of how an SP-contained protein (membrane-associated or a soluble one) is first translated by the ribosome until the SP is water-exposed and recognized by the SRP. The Ribosome-mRNA-SP-SRP is then transported to the ER surface where is recognized by the SRP receptor and the protein translation can take place inside the ER lumen. Figure adapted from reference [17].

Once inside the ER lumen, the protein acquires its final folding, oligomers can be formed and a great fraction of these proteins are glycosylated in a very specific ER way, rich in mannose oligosaccharides [18]. Posteriorly, the secreted protein or complex is transported to sites known as *exit sites of the ER* (in mammalians) or *ER transitional sites* (tER in yeast and Drosophila) [19]. Both sites are characterized by the presence of the secretory vesicles COPI and COPII [10]. The next step for most of the secreted proteins is the Golgi complex, where the proteins are submitted to additional post-translation modifications [1]. One of the most characteristic modifications in the Golgi, at least in mammalian cells, is the rearranging of the oligosaccharide chain initiated in the ER, where different enzymes catalyze the cleavage and linkage of different saccharides in the protein structure, building up a complex known as N-glycan [20,21]. This additional change can be probed easily in the wet-lab using

a method known as endoH-method, since the first change promoted by the Golgi is the addition of the N-Acetylglucosamine saccharide, making this protein insensible to the presence of endoH.

The next step is the packing and sorting of these proteins to their final destination. In the classical secretory pathway, the transport and the fusions of these vesicles are mediated by "*soluble N-ethylmaleimide-sensitive fusion protein (NSF) accessory protein (SNAP) receptors"* (SNARES) [22]. There are two different classes of SNARES, known as target SNARES (tSNARES) and vesicular SNARES (vSNARES), that are exclusive located in heterotypic membranes [1,22]. One tSNARE, Syntaxin 5 (also known as Sed5 in yeast), is essential for the anterograde and retrograde transport [23,24] since the knockout of this gene was shown to block the entire classical secretory pathway [25].

Evidence of the unconventional secretory pathway

Although the general aspects of the conventional secretory pathway is now well-known, it has been shown over the last 20 years that some proteins can reach the extracellular environment, or the membrane surface, through pathways different than those discussed in the last section. Since this was only recently discovered, it is not known how they really work yet, or how many different pathways are. That is why they are collectively called Unconventional Protein Secretion (UPS) pathways. However, some "coincidences" between them started to rise and are now helping to define the panorama in a less superficial way. Since there is no consensus on how to describe all the processes observed so far yet, we are going to use in this thesis the one proposed by *Catherine Rabouille* [26] because is the most complete one in our opinion. Before discussing these processes, it is important to emphasize that there is only one "coincidence" for all the different processes, which is cellular stress [26,27]. It appears that stress is the main trigger for most UPS mechanisms, and it can be caused either by nutrient starvation, mechanical stress, high temperature, ER stress or others [26].

The stress-dependent UPS is divided into four types. The type one (UPS 1) corresponds to the direct protein or peptide translocation across the membrane by forming a local pore [28,29]. The second one (UPS 2) involves peptides that are translocated across the plasmatic membrane by members of the ABC transporters family, a mechanism that is still very poorly understood [26]. The third one (UPS 3)

involves soluble proteins with specific cytosolic functionality, which lacks SP and are transported to the extracellular space in an autophagosomal and endosomal-dependent way. The fourth mechanism (UPS 4) relates to SP-containing membrane proteins that are synthesized inside the ER but transported to the plasmatic membrane bypassing the Golgi complex. The fourth pathway is the only one that has a special name and is now known as the Golgi Bypass.

Even though the UPS 3 and 4 seem very different from each other at first sight, they share a second "coincidence" besides cellular stress, which is the direct involvement of a Golgi-protein named Golgi Reassembly and Stacking Protein (GRASP). As it will be discussed soon, this protein has a structural functionality of keeping the Golgi arrangement and shape. How can a Golgi-protein be involved in secretory pathways that do not involve the Golgi complex?

Unconventional secretion of soluble proteins (UPS 3)

Before start discussing the UPS 3, a brief introduction about a fungi drug known as brefeldin A will be given. As it was discussed before, the transport of vesicles between the ER and the Golgi is mediated by the COPs. The COPI is constituted by seven different proteins and its assembly is dependent on the GTPase ARF1 protein [30]. Brefeldin A is a drug that affects directly the ARF1 functionality, disturbing the correct assembly of the COPI and leading to a still not understood phenomenon of fusion between the ER and the Golgi [31]. What is already known is that the presence of brefeldin A blocks the conventional secretion pathway due to its direct interaction with ARF1. The effect caused by this drug is reversed after washing.

The number of soluble proteins that are now known to be secreted even though they do not possess an SP and are independent of the classical secretory machinery (like being insensitive to the presence of brefeldin A or the Syntaxin 5) is rapidly increasing [32,33,34,35]. One of the most interesting examples of UPS 3 is the secretion of Acyl CoA Binding Protein (ACBP) in *Dictyostelium* [36,37].

The *Dictyostelium* development under cell stress (nutrient starvation) is characterized by the formation of two different cell phases known as prespore and mature sorocarp (fruiting body), where the spores are released. The UPS of ACBP is triggered somehow and this protein is cleaved in the extracellular space by a serine-protease (Tag C) generating a peptide that induces the sporulation process [38]. Interestingly, ACBP is a small soluble protein (~10 kDa) that lacks an SP and, even though its secretion is independent of the classical secretory machinery, it depends on the Golgi protein GRASP. Similar results were also observed for ACBP in different organisms like *S. cerevisiae* and *Pichia pastoris* [39,40].The details of this secretion process are still not known, but it is already a consensus in the community that it is triggered by stress and needs the autophagy machinery. Autophagy was also identified to be important, although still very unappreciated, in the cytokine IL-1β biogenesis and secretion in mammalian cells, through a UPS pathway that also requires the GRASP participation [41]. These results suggest that the autophagosomes formation, or the machinery involved in it, and GRASPs are necessary for the UPS of soluble proteins via type 3 pathway. However, the exact role played by GRASPs in these processes remains unknown. It was also suggested that multi-vesicular bodies would have a participation in the UPS 3, but this is still a matter of intense debate [42], especially because both MVB and autophagosomes are normally directed to the lysosome for degradation. Of course, it was proposed that both kinds of vesicles might fuse directly with the plasma membrane instead [27], but it is not a consensus yet.

It is important to highlight that during the development of this thesis, Yoshinori Ohsumi was laureate with the Nobel Prize in Medicine or Physiology in 2016 due to his notable work and discovery of the key mechanisms that regulate the autophagy system.

The Golgi Bypass phenomenon (UPS 4)

As the name suggests, the UPS 4 involves pathways of secretion that bypass the Golgi complex. Membranous proteins that possess an SP are translated and go through all the post-translational modification of the ER but are secreted independently of the COPII assembly or the Golgi SNAREs [3,26]. They also have the oligosaccharide chains characteristic of the ER, therefore being EndoH sensitive, and their secretion is insensitive to the treatment with brefeldin A. The determinants that drive

the UPS 4 are still not known but it has been suggested that it can be used to speed up the transport and the membrane surface expression of membrane proteins, regulation of the protein activity by avoiding the post-translational modifications of the Golgi and even a kind of protection to the Golgi itself [26].

One good example of a protein that is secreted by the Golgi bypass pathway is the αPS1 integrin in *Drosophila*. The αPS1 is an integral membrane protein that has an N-terminus SP and is expresseed inside the lumen of the ER [25,43]. Rabouille and collaborators [43] have decided to study the phenotype changes caused by the knockout of the single GRASP gene in *Drosophila.* At that moment, the involvement of GRASP in the UPS type 3 of ACBP was known already. The authors observed that the absence of GRASP leads to defects in the epithelium integrity, including occasional loss of the contacts between cells. They suggest that the GRASP has a role in the epithelium integrity and adhesion, and more than that, they were capable to observe a phase in the *Drosophila* growth where there is a GRASP and αPS1 colocalization in the contact zones between cells, and the absence of GRASP disrupt the αPS1 accumulation in these zones. However, this phenotype was only observed in a *Drosophila* growth phase known as 10B and before this phase, the secretion of αPS1 is sensitive to brefeldin A and the presence of Syntaxin 5. The same does not happen in the phase 10B, suggesting that at this time the secretion is independent of the Golgi complex but still dependent on GRASP.

A second classical example of UPS 4 involves the *Cystic Fibrosis Transmembrane conductance Regulator* (CFTR) [44]. The cystic fibrosis is a recessive genetic disorder characterized by the abnormal transport of Cl and K ions across the epithelium [45]. The most common genetic mutation observed in this pathology is the single depletion of phenylalanine 508 of CFTR (ΔF508-CFTR) which does not totally compromise its functionality but completely abolish the transport to the plasmatic membrane, keeping this protein inside the ER lumen [44]. Even thought ΔF508-CFTR is still functional, it is posteriorly degraded by the degradation machinery of the ER [44,46]. I*n vitro* and *in vivo* data have shown that the ΔF508-CFTR secretion could be triggered via a GRASP-dependent UPS by means of ER stress or GRASP overexpression (44). The interaction between GRASP and ΔF508-CFTR is essential for this process to happen and the transgenic overexpression of GRASP in ΔF508-CFTR containing rats recovers the healthy phenotype without any apparent toxicity [44].

The Golgi Reassembly and Stacking Protein (GRASP)

The Golgi complex functions as a factory in which proteins received from the ER are further processed and sorted for transport to their eventual destinations: lysosomes, the plasma membrane, or secretion [47]. Morphologically the Golgi is composed of flattened membrane-enclosed cisternae and associated vesicles, which can be also laterally linked, building the Golgi Ribbon (Figure 2) [47]. In mammalian cells, during mitotic division, the contiguous Golgi stack and ribbon are first vesiculated and then partitioned into the daughter cells by the mitotic spindle [47]. The reassembly of these vesicles into a mature Golgi complex occurs in the end of the mitotically period [47].

The Golgi reassembly after mitosis times is sensitive to the N-ethylmaleimide (NEM – a reactive alkyne against thiol groups and largely used to modify cysteine in proteins and peptides) action, which prevents the correct Golgi cisternae stacking [48]. The search for the main NEM target led to the discovery of a protein with an apparent molecular mass of 65 kDa named GRASP65 [48]. In the manuscript describing the GRASP65 discovery, the authors also showed that this is a protein that mainly localizes in the *cis* face of the Golgi complex and is double anchored to the membrane by using a G2-Myristoylation and interacting with a coiled-coil Golgin known as GM130 [48].

Later on, a second report described the discovery of a GRASP65 paralogue in mammalian, with an apparent molecular mass of 55 kDa, and by this reason named as GRASP55 [49]. Although GRASP55 and GRASP65 have very similar amino acid sequences [49], they do not colocalize *in vivo* [48,49], with GRASP55 being presented mostly in the *trans* and *medial* faces of the Golgi complex [49]. Besides, GRASP55 do not interact with GM130 and, instead, interact with another coiled-coil Golgin known as Golgin45 [49].

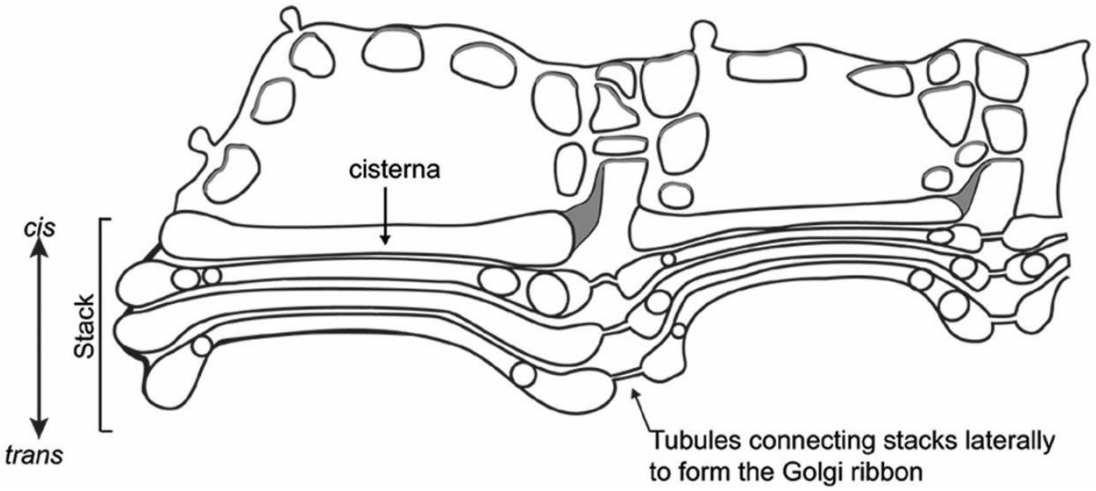


Figure 2: representation of the mammalian Golgi complex with the cisternae stacked and laterally connected, building the Golgi ribbon. The Golgi complex is also polarized, with the cis surface facing the ER. Figure adapted from [63].

GRASPs are not directly involved in the general conventional protein secretion [50,51,52], even though they appear to have a role in the secretion of some proteins with a C-terminal valine motif, including CD8-α [53], Frizzled [54] and TGF-α [55]. Their efficient transport is thought to involve GRASP65 chaperoning ER-to-Golgi transit and GRASP55 chaperoning intra-Golgi trafficking [53]. GRASP is part of the Golgi matrix and was shown to be responsible for aiding an efficient retention of p24 cargo receptors and other membrane proteins in the Golgi apparatus [56].

The single depletion of GRASP65 in mammalian cells leads to a reduction in the number of cisternae per stack, without completely abolishing the stack organization characteristic of this organelle [48]. The double GRASP55 and GRASP65 depletion completely disrupt the Golgi organization [57,58], showing that both proteins can have a redundant function in Golgi structuration. The Golgi disassembly is also one of the critical steps during apoptotic times and it was shown that GRASP65 is sensitive to the Caspase proteolysis [59]. The expression of Caspase 3-resistant forms of GRASP65 significantly reduce the Golgi fragmentation and leads to a delay of the apoptotical process, suggesting that the GRASP65 cleavage is essential during apoptotic times [59].

The correct Golgi assembly, coupled with the action of the GRASP family, was also observed to be of medical relevance. It was shown a correlation between β-amyloid peptides accumulation in the extracellular environment and the Golgi fragmentation caused by GRASP phosphorylation in Alzheimer disease [60]. In this process, the GRASP phosphorylation inhibition recovers the healthy

Golgi phenotype and leads to a significant decrease of the β-amyloid peptides secretion, suggesting that the Golgi and GRASPs could be interesting targets in the pharmacologic strategies against Alzheimer disease [60]. Furthermore, it has been also shown that GRASP65 is a critical downstream target for human immunodeficiency virus (HIV) to acquired immune deficiency syndrome (AIDS) progression in macrophages by inhibition of GRASP65-mediated cisternal linking [61,62].

Although GRASPs are essential for the correct Golgi structuration in mammalian cells and several organisms, the fact that any obvious GRASP homologs have not been found in plants so far, although they have their cisternae perfectly stacked, suggests that GRASPs might be involved in additional functions [63]. Moreover, only 40% of the cisternae were found to be organized into stacks in *S. cerevisiae,* although it has one homolog of GRASP [63]. It was discussed before the GRASP involvement in UPS for several cargos, showing that GRASPs have a very dynamic set of functions inside the cell.

The primary sequence of the GRASP is formed by two distinct putative domains in the N-terminal region that belong to a class of protein/peptide interaction domains known as PDZs [48,49,53,63]. Because these domains are responsible for the GRASP main functionality of "grasping" opposite membranes (or cisternae, in this case), they are now jointly called as GRASP domain. The GRASP domain lacks the typical βββαββαβ secondary structure organization of eukaryotic PDZ domains, with each binding groove been formed by the final, rather than the second, β-strand within the fold (Figure 3). Interestingly, this is a pattern commonly observed in prokaryotic PDZs [64].

Besides the GRASP domain, GRASPs have also a C-terminus that is usually larger than the N-terminus, with an apparent regulatory function and rich in serine and proline residues [63], the reason why this region is also called SPR (Serine and Proline Rich) domain. By the time this thesis was written, no full-length structure of a GRASP had been reported in the literature. Furthermore, the only report describing structural behaviour of a full-length GRASP is the one presented in this thesis. However, some crystal structures of the GRASP domain were already released including GRASP55 [PDB ID 3RLE and 4KFW, 64] and GRASP65 (PDB ID 4KFV,65). Interestingly, the structures suggest that the GRASP domains is composed of two structurally similar PDZ domains (Figure 3) [64]. This aspect will

be discussed in details in chapter 4 where we show for the first time that the PDZs, although being similar in terms of their crystal structures, are very different from each other in solution.

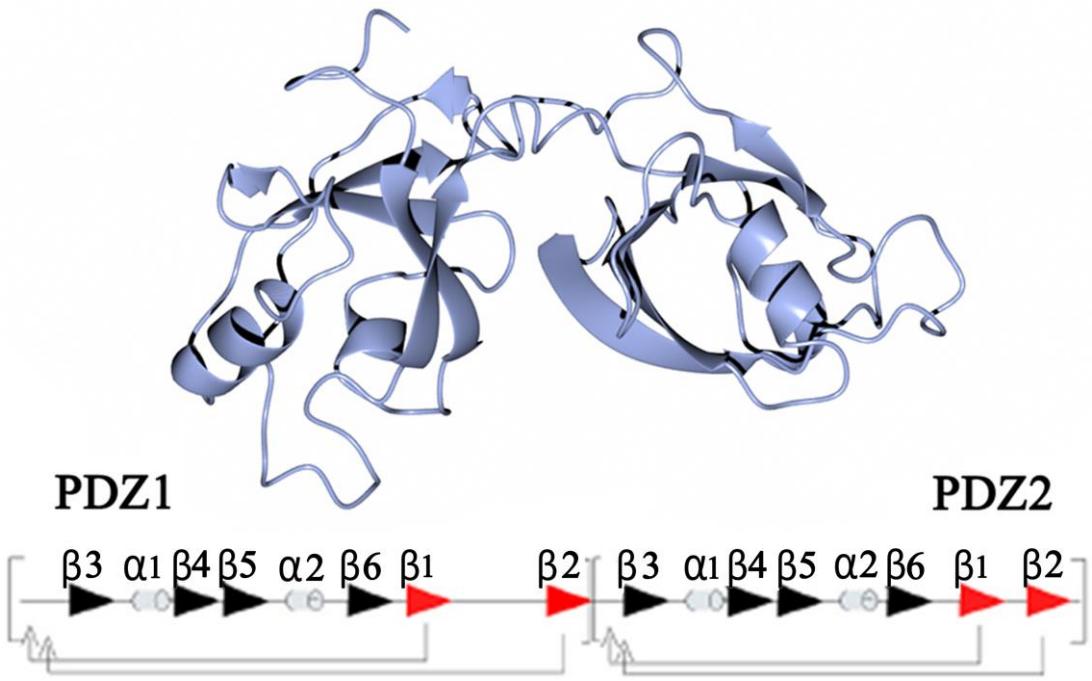


Figure 3: Crystallography structure of GRASP55 GRASP domain. Shown below the structure is the arrangement of secondary structure across the 3D one. Notice that there is a circular permutation of *β2* responsible for the PDZ binding pocket formation. Figure adapted from [64].

Some attempts to solve the molecular details of the GRASP65/55 interaction with the associated Golgins GM130/Golgi45 were published recently in the literature [66,67]. Fen hu *et al* observed that GM130 binds to GRASP65 at two distinct sites concurrently, using both PDZ1and PDZ2 domains, contradicting what was previously reported regarding GRASP65 use of the PDZ2 only for the interaction with GM130 [66, PDB ID. 4REY]. However, those reports contain serious limitations that were somehow neglected by the community, especially because they employed a strategy where only a peptide of the GM130 C-terminus was used in the complex formation. Moreover, only the GRASP domain was used. Whether the attachment of the N-terminus part of the GM130 peptide in the PDZ1 was an artifact of the truncated version (which clearly seems to be the case, since it would not exist if a full-length GM130 was used) was not discussed at all. The same strategy was used with GRASP55 and Golgin45 and a similar conclusion was reached [67, PDB ID. 5H3J].

GRASP oligomerization is regulated by phosphorylation in the SPR domain [63] and it was suggested before that there is an internal peptide of approximately 20 residues inside the PDZ2 structure that is capable of interacting directly with the PDZ1 of a second GRASP, bridging the dimer structure [68]. It has been shown that phosphorylation in the serine 189 of GRASP55 is capable of breaking the dimer structure because there is a conformational change in the GRASP internal ligand that prevented its insertion into the PDZ binding pocket [69]. Even though this phosphorylation was shown to affect GRASP dimerization *in vivo* [68], the issue of whether this conformational change would be the correct model is still controversial because the authors used an isolated GRASP55 domain that was predominantly a monomer in solution.

The most accepted model of how GRASPs act in the stacking of the Golgi cisternae is by grasping the cisternae together using the dimerization, where one GRASP is located in one cisterna and interact with a second GRASP in a opposite cisterna (trans interaction) [50,70]. The nom-functional GRASP dimerization in the same cisterna (cis interaction) is avoided by the dual anchoring to the membrane [71]. A model of how this could happen is presented on chapter 3.

Cryptococcal meningitis

The acquired immunodeficiency syndrome (AIDS) is the most advanced stage of HIV infection and is a viral disease that affects the human immunological system by destroying the CD4 T Lymphocytes [72]. Because it strongly affects the immunological system, the patients become much more susceptible to infectious opportunist diseases (IOD) in advanced cases, which usually do not compromise healthy people [73]. Despite the fact that the most advanced antiretroviral therapy has helped to control most of the AIDS symptoms, when it comes to availability of such therapy, most of the poor regions in the world still lack proper medications. Furthermore, the late diagnosis in most of these third world countries has contributed to a large spread of deaths due to IOD.

The cryptococcal meningitis is a fungus disease caused by the genus *Cryptococcus spp*. and one of the most spread IOD, especially when compared with other meningitis caused by *Streptococcus*

pneumoniae or *Neisseria meningitidis* [74,75]. It is observed an alarming mortality rate of 13-44% of AIDS patients in undeveloped countries, with more than 500.000 deaths per year only in the Sub-Saharan Africa [74]. This number is higher than those attributed to other sexually transmitted diseases, hepatitis B, Dengue and tuberculosis, for instance (Figure 4) [74].

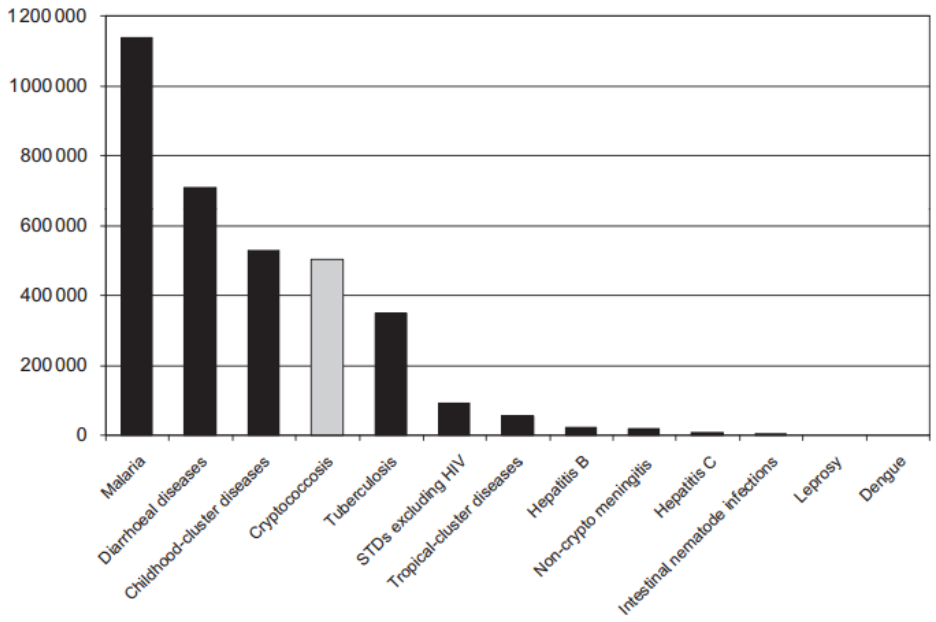


Figure 4: IOD-associated deaths in AIDS patients. Highlighted in grey is the number of deaths due to cryptococcal meningitis (also known as crypyococcosis). Figure adapted from [74].

One of the key factors that contribute for this alarming situation is the poor availability and high costs associated with the most efficient antifungal treatment of first line (Amphotericin B). This is added to the difficulties in the control and monitoring of the associated toxicity of this treatment, which usually leads to a problematic increase in the intracranial pressure [76,77]. It is clear that we still lack an efficient treatment against cryptococcal meningitis and this is an urgent issue.

GRASP from *Cryptococcus neoformans* (CnGRASP) as a GRASP55/65 model and a possible target against cryptococcal meningitis

There are more than 30 different species associated with the *Cryptococcus* genus, but only two of them (*Cryptococcus neorfomans and Cryptococcus gattii*) are responsible for nearly all the humans

and animal infections [74]. The *C. neoformans* fungus can be found in the soil at nearly anywhere around the world and the infection starts by the inhalation of the spores. Symptoms associated with this disease, such as pulmonary infection, are rare in nom-immunosuppressed patients [75].

The *C. neoformans* external morphology is peculiar (Figure 5). The plasmatic membrane is surrounded by an extensive polysaccharide capsule that resembles gram-negative bacteria and is composed mainly by two complex polysaccharides: glucuronoxylomannan (GXM, ~90%) and galactomannan (GXMGal, ~10%). The external capsule and its key component GXM are the main virulence factors associated to *C. neoformans* [*78*].

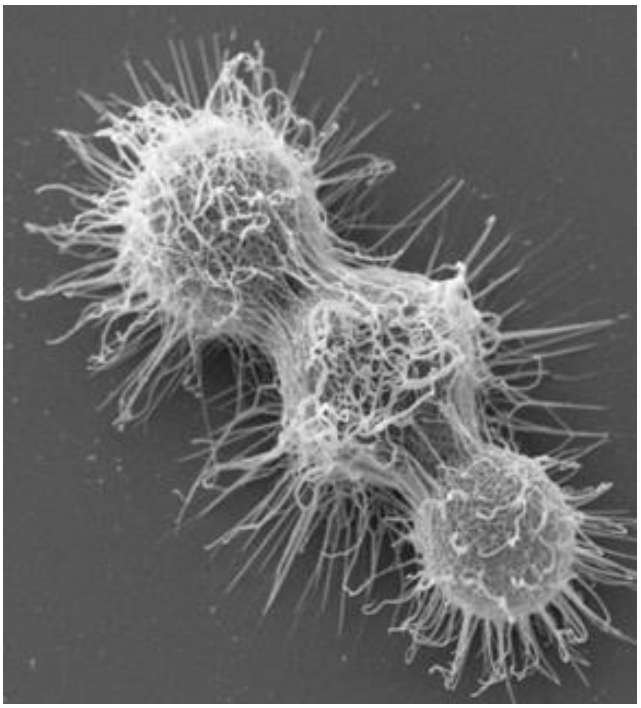


Figure 5: Image of a single C. neoformans cell obtained using electron microscopy. Figure obtained from the web server of the University Albert Einstein, College of Medicine (http://www.einstein.yu.edu/labs/arturo-casadevall/default.aspx, accessed in 2017).

The capsule assembly requires the intracellular polysaccharide syntheses [79,80], followed by their secretion to the extracellular space for further incorporation into the surface of the plasma membrane [81]. The capsule expression is the most important restriction to the *C. neoformans* virulence [82]. The *C. neoformans* treatment with brefeldin A leads to a partial inhibition of the capsule assembly, suggesting that the Golgi is necessary for the synthesis and/or secretion of the GXM [80,83]. However, the fact that the capsule expression was still happening after treatment with brefeldin A suggests that

the capsule components might have reached the outside space by unconventional means. Based on this assumption, our collaborator in FIOCRUZ-RJ Marcio L. Rodrigues and his group decided to test the hypothesis that a GRASP-dependent UPS pathway might have a role in GXM secretion. They observed that *C. neoformans* cell silenced with the GRASP gene end up with significantly smaller capsules, making them less virulent and more easily associated with macrophages [84]. Although there is a clear GRASP involvement in the GXM secretion, it is still not clear by what means it happens. However, the clear reduction of the C. *neoformans* virulence after GRASP depletion is of great pharmacological interesting, validating GRASP as a promising target against cryptococcal meningitis.

Besides, it has been observed that, in *C. neoformans*, ACBP is also secreted by UPS pathways during the yeast-to-hypha transition, a cellular adaptation induced by stress, and is followed by a deep change in the fungus morphology that goes from a single cell to a filament arrange [85,86]. As it was discussed in the UPS type 3 section, the ACBP UPS is also GRASP-dependent [85]. These results suggest that the GRASP from *C. neoformans* (CnGRASP) can also be a good model for GRASPs in the UPS.

However, although CnGRASP is an interesting target against cryptococcal meningitis, why should someone start investigations using this GRASP if the main goal is to unravel GRASP structure and functionality? Shouldn`t the mammalian GRASP55/65 be better models? This is a true and valid concern but we did not decide to start with CnGRASP without a good reason. Unlike the GRASP55/65, Rodrigues' group successfully purified the CnGRASP in a full-length form and show that this gene codified a protein that is functional in an *in vivo* assay [84]. That is why, even after 20 years since the discovery of GRASP65 by Barr *et al* in 1997 [48], the results presented in chapter two were the first ones describing structural features of a full-length GRASP, still being the only one in the literature thus far.

Intrinsically Disordered Protein (IDPs) and their functions – the present and the future of the "Unstructured Biology"

During the evolution of life on our planet, several unique and striking moments shaped life in the way we know it now. Until a certain time of the evolution, at about 580 million years ago, most of the life systems were unicellular and grouped themselves to form small colonies [87]. At 70/80 million years ago, the evolution rate increased by an order of magnitude (according to terms previously related to the extinction of species) and life started to resemble the one we know now [88]. This fast appearance of the divisions, which have persisted until today, was such an outstanding event that received a special name and is now known as the Cambrian Explosion [89].

If we consider all the limitations and respecting the inherent differences, we could say that the break of the classical structural biology paradigm that says "one structure-one function" and the still exponentially increase of a new class of proteins that lack a well-folded tertiary structure, might be the "Cambrian Explosion" of the structural biology in the 21st century. Before that, the idea and importance of flexibility in many biological functions were several times "rediscovered" until a particular moment where it became clear that intrinsically disordered proteins (IDPs) are not only exceptions to a rule, but a unique class of high representative amount of proteins [90,91,92]. The most well-accepted definition of IDPs, as described by Uversky *et al* [93], is that the "intrinsically disordered" term is based on the notion that the corresponding protein (or protein region) is biologically active, despite the fact that it is formed by a dynamical conformational ensemble, either extended or collapsed, at a secondary and/or tertiary structure level. At this point is important to emphasize that to be an IDPs does not mean to be unfolded or to exist as an extended theoretical random coil. The IDP structure is not a random organization in solution and that is why only the term IDPs will be used in this thesis instead of other nomenclatures, such as intrinsically unfolded proteins, natively unfolded or even loopy [94]

A great number of IDPs are well known to be structurally promiscuous, positioned at the center of the cell interactome, where they act as hubs for interactions with multiple protein partners [90]. The disordered-to-order transition is probably a unique characteristic of IDPs, allowing them to adapt to several different situations [90,93] and is what usually allow these hubs to be so promiscuous for the protein-protein interaction [93]. It has been already shown that hub complexes are essential for any living organism and the deletion of a hub is fatal [95]. Besides, several IDPs involved in human pathogeneses are also hubs and examples of these disordered hubs with hundreds of protein partners are p53, α-synuclein, BRCA1, XPA, p21, p27, estrogen receptor and many others [93,94]. There are

also IDPs capable of forming protein complexes known as “fuzzy” where they can adopt more than one conformation within the same partner, in a dynamic or static way [96,97].Figure 6 shows some of the functions already associated with IDPs and well-folded protein, where we can clearly observe that many of them cannot be described without the notion of disordered structure. For more details about this fascinated world of IDPs, there are a great number of excellent revisions in the literature [90,91,94,98] and books devoted only to this class of proteins [99].

One natural and predictable question that someone who starts working with an unknown protein can do is why IDPs are disordered in physiological conditions. Of course, this is not random and the “gene” for the disorder propensity is encoded somehow in the amino acid sequence. In one of the pioneering and most interesting reports about IDPs, Uversky *et al* showed that IDPs are disordered because they are composed of a great number of charged and polar amino acids but with a very small amount of hydrophobic ones [100]. A small number of hydrophobic residues can disturb the correct hydrophobic collapse, which is the main driven force for protein folding. Interestingly, it is possible to check whether a protein has propensity for being an extended IDP by calculating the mean charge and hydrophobicity from the protein sequence [100]. This will be discussed in more details in chapter 2.

The degree of structural rearrangements observed in IDPs is very high. The presence of IDPs in several human pathogens cannot be ignored as well [99], the same with their still intriguing and promising role as targets against several diseases, which has been completely unexplored so far. Additional experimental and computational data will help understanding the still not known roles played by IDPs in the cell, a future that seems shining brightly for this new field of knowledge that I like to call Unstructured Biology.

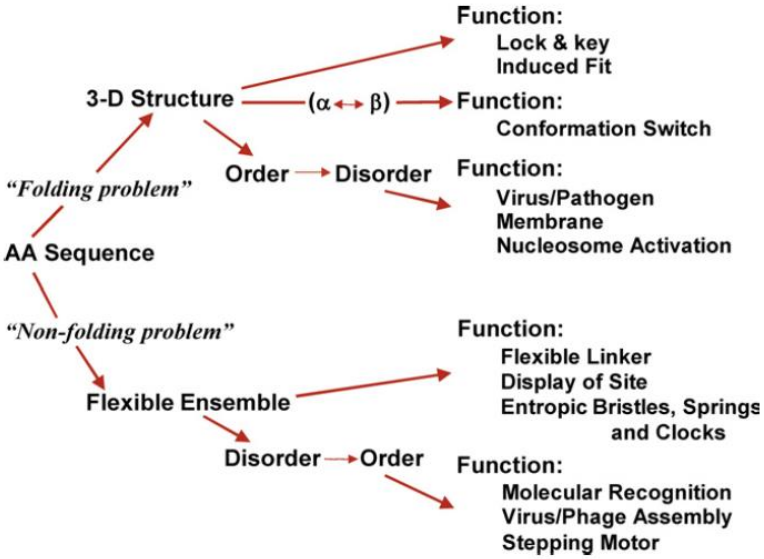


Figure 6: Several biological functions that require a well-folded structure ("folding problem") and intrinsically disordered proteins or regions ("non-folding problem"). Figure adapted from [93].

Scope of the thesis

As it was discussed throughout this Introduction, GRASPs seem to constitute a very dynamic family of proteins with central participation in the Golgi complex structuration and functionality, despite its unusual role in UPS. Moreover, structural information is still very restricted to the GRASP domain, with no information regarding the full-length structure or the structural behavior of this protein when not in a packed crystallographic environment. The main goal of this thesis was to provide, for the first time, structural information of a full-length GRASP and its dynamics in solution. In chapter 2, we present the first characterization of a full-length member of the GRASP family, where we observed that this protein has a very peculiar structural behavior in solution, closely resembling a class of collapsed IDPs called molten globules. These data inserted GRASP in the family of collapsed IDPs and opened a whole new area for GRASP studies. In chapter 3, we show that several cell conditions can affect the structure of our GRASP model, in particular, the changes in physicochemical parameters of the solution induced by the biological membrane. The structural plasticity of GRASP is discussed and a model for GRASP function in the cell is proposed. We also observed that high PEG concentrations and dehydration, conditions usually used in the crystallographic trials or induced by the crystal packing, could severely affect the GRASP structure, especially by inducing several disorder-to-order transitions. In chapter 4, we challenge the established idea that both PDZs of the GRASP domain are structurally the same, despite their distinct functions observed *in vivo*. By using high-resolution solution NMR and synchrotron radiation circular dichroism, we show that GRASPs are formed by two asymmetric PDZs domains, with PDZ1 being more unstable and flexible, while PDZ2 is much more stable and structurally well-behaved. More than that, many of the unstable regions found in PDZ1 are in the predicted binding pocket for ligands/partners, suggesting a structural promiscuity inside this domain that correlates with the functional promiscuity of interacting with multiple protein partners. In summary, our work is the first characterization of a full-length GRASP and the first to suggest that in solution the structure and dynamics of the PDZs are not similar.

Chapter 2 - New structural insights into Golgi Reassembly and Stacking Protein (GRASP) in solution

Abstract: Among all proteins localized in the Golgi apparatus, a two-PDZ (PSD95/DlgA/Zo-1) domain protein plays an important role in the assembly of the cisternae. This Golgi Reassembly and Stacking Protein (GRASP) has puzzled researchers due to its large array of functions and relevance in Golgi functionality. We report here a biochemical and biophysical study of the GRASP55/65 homolog in *Cryptococcus neoformans* (CnGRASP). Bioinformatic analysis, static fluorescence and circular dichroism spectroscopies, calorimetry, small angle X-ray scattering, solution nuclear magnetic resonance, size exclusion chromatography and proteolysis assays were used to unravel structural features of the full-length CnGRASP. We detected the coexistence of regular secondary structures and large amounts of disordered regions. The overall structure is less compact than a regular globular protein and the high structural flexibility makes its hydrophobic core more accessible to solvent. Our results indicate an unusual behavior of CnGRASP in solution, closely resembling a class of intrinsically disordered proteins called molten globule proteins. To the best of our knowledge, this is the first structural characterization of a full-length GRASP and observation of a molten globule-like behavior in the GRASP family. The possible implications of this and how it could explain the multiple facets of this intriguing class of protein are discussed.

Based on the manuscript published in Scientific Reports 2016

DOI: 10.1038/srep29976.

Introduction

The Golgi apparatus is a highly dynamic organelle responsible for sorting out proteins and other biomolecules to the cell surface and to the extracellular milieu [101]. The Golgi assembles in its characteristic pile structure, which is important for the correct post-translational modifications of proteins [101,102]. The structural organization of the cisternae into stacks and their lateral connection, building the Golgi ribbon, require a family of proteins called Golgi ReAssembly and Stacking Proteins (GRASP). GRASP is a caspase-3 substrate, whose cleavage contributes to Golgi fragmentation during apoptosis [59]. Two homologues in vertebrates have been previously described (GRASP55 and GRASP 65) [48,49] and their functions have been associated to Golgi phosphorylation-regulated assembly/disassembly [103,104], protein secretion [50], Golgi remodelling in migrating cells [52], among others [63]. There is only one gene for GRASP in lower eukaryotes [63]. Although the gene has been apparently lost in the Plantae, plants conserve the regular Golgi shape structured in stacks [63], which implies that GRASPs are likely involved in other vital functions in the cell.

GRASPs can be dramatically regulated not only in a limited physical space, such as in between the cisternae of the Golgi complex, but also in the cytosol. This family of proteins must be able to correctly interact with itself in a *trans* orientation (thus preventing the *cis* oligomerization) [71] to be accessible to the phosphorylation/dephosphorylation machinery in the cell cycle [105] and to proteases during apoptotic times [59]. Furthermore, GRASP is involved in the direct interaction with a large number of partners/ligands in the conventional and unconventional secretion pathways [44,56]. Among those, the unconventional secretion of a major capsular component and the most important virulence factor [106] of the basidiomycetes fungus *Cryptococcus neoformans*, the polysaccharide glucuronoxylomannan (GXM), has proven to be GRASP-dependent [84]. The yeast-like pathogen *C. neoformans* is the principal causative agent of cryptococcal meningitis, an opportunistic disease that kills about half million people every year just in the sub-Saharan Africa [107,108]. Fungal infections are a threat to human health. About 1.2 billion people worldwide are estimated to suffer from a fungal disease [109]. Of these, over 1.35 million people are estimated to die [110]. The need for new antifungals is thus clear, which stimulates studies on the structural characterization of molecular targets, such as GRASP, regulating fungal virulence and/or physiology.

Structural information about GRASP is still incipient. The crystal structures of the N-terminal domain from human GRASP55 (64), *Rattus norvegicus* GRASP55 and *Rattus norvegicus* GRASP65 (65) have been only recently reported. Two-PDZ domains, which form the so-called GRASP domain, and a non-conserved serine and proline rich domain, usually larger than the GRASP domain, compose the overall structure (63). These structural data together with several *in vivo* studies aiming at functional elucidation show that GRASPs are involved in a large set of functions that likely includes a great number of interacting partners. However, the behavior of a full-length GRASP is still obscure, since its biophysical properties in solution are still unknown. As a first step towards unravelling GRASP behavior, we report here a biochemical and biophysical study of *C. neoformans* GRASP (CnGRASP) in solution. Bioinformatic analysis, static fluorescence and circular dichroism spectroscopies, calorimetry, small angle X-ray scattering, solution nuclear magnetic resonance, size exclusion chromatography and proteolysis assays were used to probe structural features of the full-length CnGRASP. Our results indicate an unexpected behavior in solution and provide information that can affect the way one thinks GRASP accomplishes its plethora of functions.

Methods

Bioinformatics

PHYRE2 (111) and I-TASSER (112) were used for CnGRASP structure prediction, posteriorly refined using FG-MD (113). The prediction of disordered segments was performed using VSL2B (114), VL3 and VLXT (115), and Ronn (116) methods. SymPRED (117), PredictProtein (118), JPred (119) and SSPro (120) were used for the secondary structure prediction. The protein charge under physiological pH was estimated using Protein Calculator v3.4 (*http://protcalc.sourceforge.net/; accessed in 2015*).

Protein expression and purification

E. coli Rosetta (DE3) cells harboring the pETSUMO-CnGRASP vector were grown at 310 K in LB medium containing kanamycin (50 μg/mL) and chloramphenicol (34 μg/mL), and expression was induced by isopropyl-β-D-thiogalactopyranoside (0.5 mM) for 16 hours at 291 K. The cells were

harvested by centrifugation and the pellets stored at 251 K. Cells were resuspended in 25 mM HEPES, 150 mM NaCl, 1 mM CHAPS, 10% V/V Glycerol, pH 7.4 (buffer A) and lysed by sonication. The insoluble material was removed by centrifugation (14,000xg for 25 minutes, in a temperature of 8°C). CnGRASP was loaded onto a Ni-NTA superflow column (QIAGEN). The column was washed and the 6His-SUMO tag was removed using 2 mg of purified recombinant ULP-1 protease (for 3 hours, 283 K). CnGRASP (without 6His-SUMO tag) was collected in the flow-through fraction, concentrated by centrifugation in a Sorvall RC 6 plus centrifuge (Thermo Scientific) and the remaining contaminants were removed by size exclusion chromatography onto a Superdex 200 10/300 GL gel filtration column (GE Healthcare Life Sciences). For protein concentration, an Amicon Ultra-15 Centrifugal Filter with a NMWL of 10 kDa (Merck Millipore) was used.

Circular Dichroism (CD)

Far-UV (190-250 nm) CD experiments were carried out in a Jasco J-815 CD Spectrometer (JASCO Corporation, Japan) equipped with a Peltier temperature control and using a quartz cell with a path length of 1 mm s. The experimental parameters were: scanning speed of 50 $nm \cdot min^{-1}$, spectral bandwidth of 1 nm, response time of 0.5 s and with separate measurements prior the average in order to avoid instability contributions in the final spectra. CnGRASP was dialyzed in 10 mM of Sodium Phosphate buffer, 10 mM NaCl, pH 8.0 and a final concentration of 7 μM was used. The software CDSSTR (132,133) with an appropriate database available at DICHROWEB (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) (134) was used for spectral deconvolution. Chemical stability experiments were performed in 25 mM HEPES, 10 mM NaCl, 1 mM CHAPS, 1mM DTT, 10% V/V Glycerol, pH 7.4 and increasing urea concentration. All the experiments were temperature controlled, fixed in 293 K. The synchrotron radiation circular dichroism (SRCD) spectra of CnGRASP (19 μM) were obtained at the UV-CD12 beamline located at the ANKA Synchtrotron (Germany). Three scans were collected over the wavelength range from 280 to 180 nm with 0.5 nm step size, 1.5 s dwell time in a 10 μm pathlength demountable Suprasil quartz cell (Hellma Ltd, UK); the baseline for each sample consisted of the buffer and all components present in the sample other than the protein. The effect of temperature was assessed by measuring SRCD spectra over the temperature

range 20 to 85°C, at 5°C intervals, with an equilibration time of 5 min at each step. The spectra were processed using CDTools software (121).

Steady-State Fluorescence Spectroscopy

Fluorescence was monitored using a Hitachi F-7000 spectrofluorimeter equipped with a 150 W xenon arc lamp. The excitation and emission monochromators were set at 2.5 nm slit width in all experiments. The protein concentration was 10 μM. The chemical unfolding was performed using a fixed concentration of 1-anilino-8-naphthalenesulfonic acid (ANS – 250 μM) in order to probe the exposure of hydrophobic regions of the protein upon urea addition. The excitation wavelength was set at 360 nm and the emission spectrum was monitored from 400 up to 650 nm. For tryptophan fluorescence experiments, the selective tryptophan excitation wavelength was set at 295 nm and the emission spectrum was monitored from 310 up to 400 nm. Fluorescence quenching using the water-soluble acrylamide as a quencher was done at 293 K and analyzed using the Stern-Volmer model (154).

Size-exclusion Chromatography

Size-exclusion chromatography was carried out using a Superdex 200 (HR 10/30) coupled to an *Äkta purifier* system (GE Healthcare). Apparent molecular mass and hydrodynamic radius (R_h) were obtained from plots of K_{av} vs. log (R_h) or log (MM) (molecular mass), where K_{av} is the partition coefficient. Standard proteins used were from the gel filtration calibration kit LMW and HMW (GE Healthcare) with the exception of human S100A12 that was purified as reported in the literature (122). In order to estimate the correct molecular dimension, for these experiments, CHAPS were dropped off in the protein preparation.

Limited proteolysis

Proteolysis sensitivity was assessed using bovine Trypsin (Sigma) in a molecular ratio (protein/trypsin) of 20/1 and 100/1. Conalbumin and Ovoalbumin (GE Life Science) were used as models of well-

structured proteins for direct comparison. All reactions were checked by comassie stained 15% SDS-PAGE.

Small Angle X-Ray Scattering (SAXS)

SAXS data were measured at the small-angle X-Ray scattering beamline of the National Synchrotron Light Laboratory (Campinas, Brazil) (123). Samples were measured at the wavelength of 1.488 Å with detector/sample distance of 1,500 mm. Concentrations of purified CnGRASP of 0.5 mg/mL, 1 mg/mL and 2 mg/mL were used. The protein solutions and buffer were exposed in time-varying frames to monitor radiation damage and beam stability. Possible artefacts from buffer scattering were subtracted after the data being normalized to the intensity of the incident beam and corrected for detector response. The data analysis was carried out using ATSAS package (124), more specifically the softwares GNOM (125) and PRIMUS (126).

Differential Scanning calorimetry (DSC)

DSC experiments were done in a Nano-DSC II from Calorimetry Sciences Corporation, CSC (Lindon, Utah, USA). Samples were degassed under vacuum for 10 min before use and the data subtracted from buffer contributions. Scans were recorded from 10-95°C at an average heating rate of 20°C/hour, under pressure of 3 atm. Protein concentrations of 25, 50 and 100 µM were used. For comparison to a well-structured protein, a solution of 50 µM of Conalbumin (GE Life Science) was used as a control.

Ultracentrifugation

A sucrose gradient ranging from 10 to 35% in a 25 mM Hepes/NaOH, 150 mM NaCl, pH 7.4 was prepared. The sucrose percentage was estimated by the difference in the diffraction indices. Samples of purified CnGRASP (less than 1 mg/mL) were loaded onto the gradient and centrifuged using a vertical rotor (Hitachi P65VT3) in a Hitachi 55P-72 ultracentrifuge (100,000*g* for 4 h at 4°C). Thyroglobulin, Conalbumin, BSA (GE Life Science) and Carbonic Anhydrase (Sigma Aldrich) were used as standard markers. The protein identification was carried out using absorbance at 280 nm and SDS-PAGE.

Dynamic light scattering (DLS)

The CnGRASP DLS experiments were performed using a Nano-ZS dynamic light scattering system (Malvern Instruments Ltd, Malvern, UK). This system employs a 633 nm laser and a fixed scattering angle (173°). The experiments were carried out at 20°C in a 25 mM Hepes/NaOH, 150 mM NaCL, 1 mM β-Mercaptethanol, 10% glycerol, pH 7.4 buffer. Before any measurement, the sample was centrifuged at 11,000xg for 5 minutes at room temperature, and subsequently loaded into a quartz cuvette prior to measurement.

Nuclear magnetic resonance (NMR)

^{15}N-CnGRASP was expressed in minimum medium supplemented with ^{15}N labelled ammonium chloride (Sigma Aldrich) and purified using the same procedure for the unlabelled sample. The protein was dissolved in 10% deuterium oxide in phosphate buffer pH 7.4 at a final concentration of 180 µM for NMR measurements. Glycerol (5% V/V) and beta-mercaptoethanol (5mM) were present in the sample. NMR experiments were conducted in an AVANCE III HD Bruker spectrometer (Germany) operating at 600 MHz for ^{1}H equipped with a triple resonance cryoprobe. The regular 1D spectrum was first obtained using the water suppression by excitation sculpting pulse sequence with gradients. A spectral width of 16 ppm and acquisition time of 3.4 s were set. A recycle delay of 2 seconds and delay for gradient recovery of 200 µs were used and 256 scans were recorded. For ^{1}H -^{15}N Heteronuclear Single Quantum Coherence (HSQC), the spectra were collected in increasing urea concentration (0, 2 and 4 M). For each experiment, 256 complex increments of 2048 complex data points were collected using 16 scans. The spectral widths were set to 14 ppm (^{1}H) and 32 ppm (^{15}N) and a relaxation delay of 2 seconds and delay for gradient recovery of 200 µs were used between scans.

Results

Primary sequence analyses of CnGRASP suggest multiple disordered sites

GRASPs have a well-conserved domain at their N-terminus comprising two tandem PDZ domains (PDZ-1 and PDZ-2), a classical protein-peptide interaction domain, which is responsible for GRASP homo-oligomerization and for the attachment to the Golgi membrane [127]. The C-terminus is

a Serine and Proline Rich domain (SPR domain) with a large number of phosphorylation sites in mammalian cells. The SPR domain is apparently required for GRASP regulation [127].

The CnGRASP primary sequence [84, Broad Institute Accession No. CNAG_03291, sequence at the end of this chapter] has 256 amino acids and a theoretical molecular mass of 27551.7 Da with a composition of almost 30% of bulky, or order promoting, amino acids (I, C, L, V, W, Y and F). As a matter of comparison with ordinary soluble proteins, for Conalbumin (GenBank: CAA68468.1) this value reaches 31%, Ovalbumin (GenBank: AAB59956.1) is 33% and Bovine Serum Albumin (GenBank: CAA76847.1), 34%. Those amino acids are responsible for driving the formation of the hydrophobic core in globular proteins.

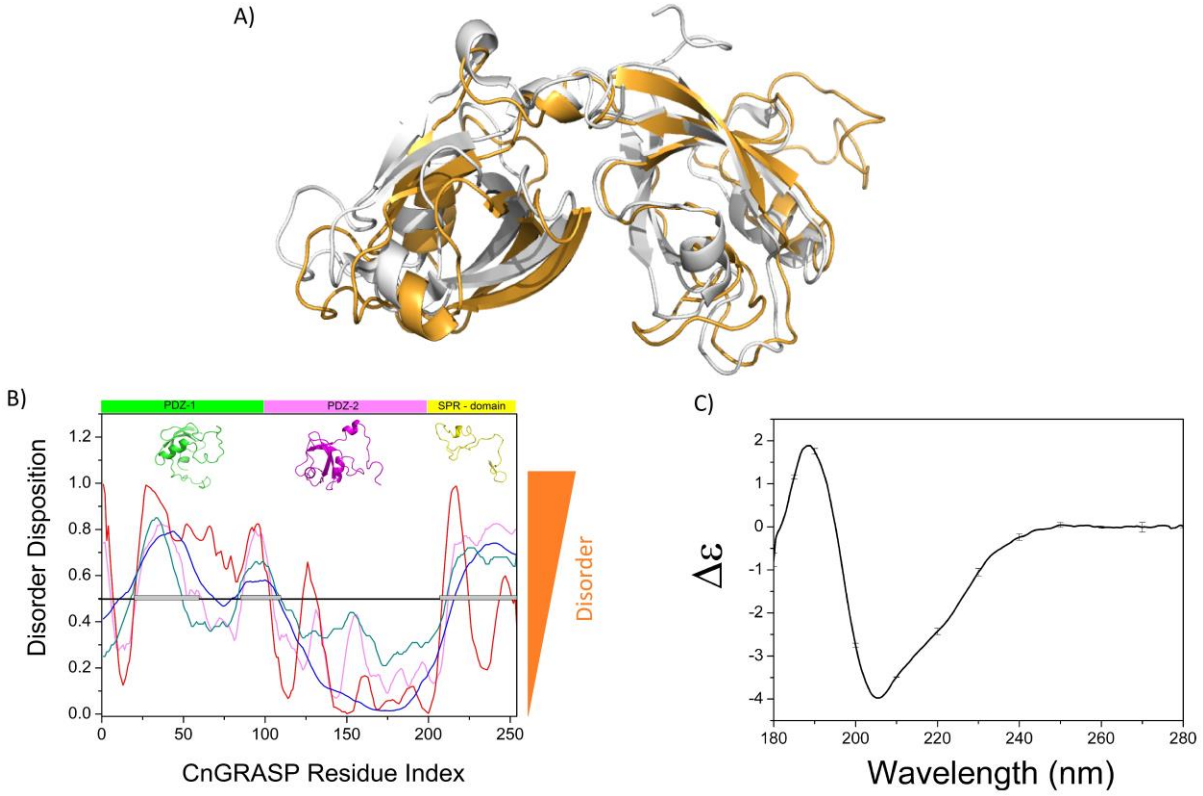


Figure 7: Secondary structure analyses. A) CnGRASP (Orange) and GRASP55 (Gray, PDB ID 3RLE) GRASP domain alignment using Pymol. B) Disorder prediction using the disorder predictors VSL2B (Magenta), VLXT (Red), VL3 (Blue) and Ronn (Green). The black line indicates the limit above which the residue has more than 50% propensity of being disordered. Represented in the upper part of the graph are the refined predicted models of CnGRASP. The green part represents PDZ-1, the magenta represents PDZ-2 and the SPR domain is colored in yellow. C) SRCD spectrum of CnGRASP in solution.

The CnGRASP GRASP domain seems to have an overall structure similar to GRASP55 (Figure 7A), even though a much more pronounced quantity of disorder is apparently present. Despite this

overall structure resemblance, the analysis of the sequences of each domain showed that PDZ-1, PDZ-2, and SPR presented variable amounts of disorder-promoting amino acids (A, G, R, D, H, Q, T, K, N, M, S, E and P). PDZ-2 has the highest percentage of order promoting amino acids (40%), whereas SPR has only 23% in total length, being mainly composed of prolines, arginines and serines, a general feature within the GRASP family. It is worth noting that the amount of proline, a well-known structure breaker, is almost 8% of the total protein amino acid content and that is much larger than values found in structured proteins such as human serum albumin (4.1%), human lysozyme (1.5%) and Bos taurus Ribonuclease A (3.2%). The CnGRASP primary sequence was also analyzed using the Uversky plot (100). The mean net charge and mean hydrophobicity values place CnGRASP in the native side of this charge-hydrophobicity phase space (Figure 8A) similar to well-structured proteins, and the same behavior observed for other GRASPs (Figure 8B). These analyses indicated that CnGRASP has the necessary amino acid content to be a native and well-structured protein, although presenting non-homogeneous distribution of the bulky disorder-promoting amino acids.

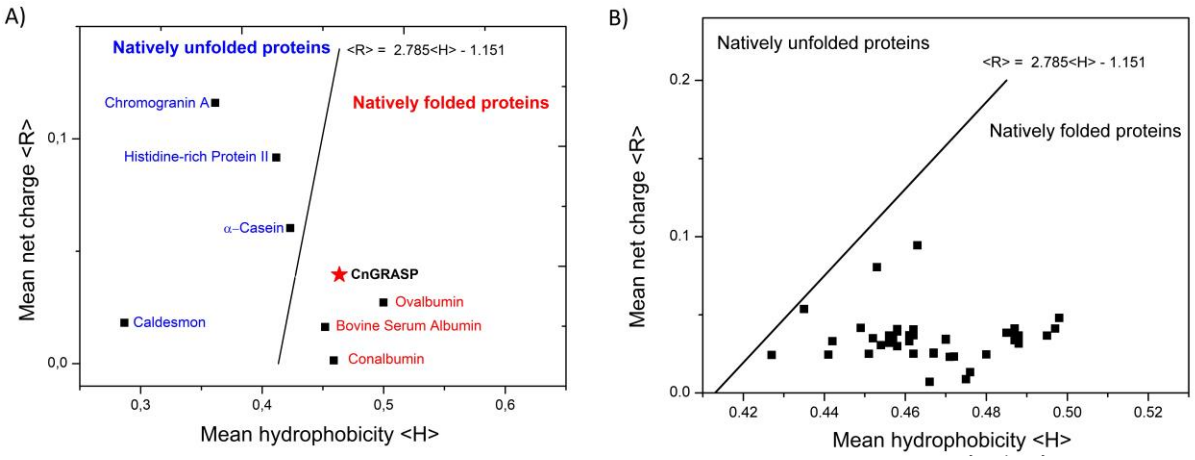


Figure 8: Uversky plot. Values of mean net charge (<R>) versus mean hydrophobicity (<H>) taken from reference [21] and/or calculated based on primary sequence collected from NCBI database (presented in the end of this chapter). The red star locates the position of CnGRASP in the plot. B) The same plot presented in A but only with GRASPs sequences presented. It is worthwhile to notice that all GRASPs tested are located in the natively folded proteins side. The list with the GRASPs sequences can be founded in the end of this chapter.

The prediction of local disorder can be performed with reasonable accuracy [128]. Analysing factors such as local amino acid composition and hydropathy, protein disorder predictors classify each residue within a sequence with a propensity to be either ordered or disordered. Through the use of a combination of such programs, including VSL2B, VLXT, VL3 and Ronn, disordered regions in CnGRASP were assessed. This analysis clearly demonstrated the high propensity of the SPR domain

to be intrinsically disordered (Figure 7B). The behaviour of the first PDZ domain is of particular interest since it is responsible for the *trans*-oligomerization in vertebrates and possibly in *C. neoformans.* PDZ-1 contains regions having a high propensity for local disorder, a behaviour that is not observed for CnGRASP PDZ-2, which remains structured over most of its sequence. The reason why PDZ-1 follows the disorder propensity observed above, and PDZ-2 does not, is still an open issue. Several reports suggest the concept of cargo-specificity in the GRASP-mediated tethering of cargo molecules [54,55]. The structural flexibility observed in intrinsically disordered proteins can confer high specificity with low affinity for molecular recognition, allowing it to interact with multiple partners [129], a feature that could be properly explored by the PDZ-1 in order to interact with the cargo proteins during functional processes.

Secondary structure analyses of CnGRASP revealed well-structured regions

Computational methods, as those used in the previous section, are very useful tools for protein structural analyses and, for CnGRASP, it could be predicted that multiple sites of disorder coexist with more structured regions. These observations were further investigated by including experimental methods that provided information on the structural organization of CnGRASP. Circular dichroism is a powerful technique for structural studies of proteins in solution (130,131). The CnGRASP Synchrotron CD spectrum in the far-UV region (Figure 7C) indicated that the protein is of the α/β type with the typical features of α-helical components (bands centered around 208 and 222 nm), which always show up more clearly in the CD spectra of proteins. The presence of β-strand contributions is suggested by the loss of resolution in those α-helical bands that likely come from the admixture of the β–associated band around 215 nm, which was not resolved in the spectrum. To quantify the secondary structure content, deconvolution algorithms can be used [132,133]. For CnGRASP, DICHROWEB [134] was employed and the results of the deconvolution are presented in Table 1.

Table 1: Deconvolution results of CnGRASP CD data using DICHROWEB

Program	Data base	α-helix	β - sheet	Turn	Disordered	NRMSD
CDSSTR	SET 3	0.22	0.24	0.22	0.32	0.021
CDSSTR	SET 6	0.19	0.19	0.18	0.44	0.027
CDSSTR	SP175	0.21	0.24	0.14	0.40	0.01
Mean value		20.7%	22.3%	18%	38.7%	

DICHROWEB utilizes three distinct algorithms to evaluate the contents of secondary elements, thus providing their mean values. On the basis of this analysis, the secondary structure content appeared to be mostly due to non-helical elements such as β-sheets (~22,3%) and to structures in the form of loops (~18%) and disordered regions (~39). It is important to notice that the predicted amount of disorder reached high values, which are consistent with the bioinformatic results and with the secondary structure prediction (Table 2).

Table 2: Secondary structure prediction content in CnGRASP using SymPRED, Jpred and SSpro

Predictors	α-helix	β-sheet	Loop
PROFsec	8%	25%	67%
Jpred	10%	28%	62%
SSpro	11%	24%	65%
Mean value	**9.7%**	**25.7%**	**65.7%**

Quaternary structure analyses of CnGRASP

Size exclusion chromatography (SEC) is a hydrodynamic technique widely used to study protein oligomerization states. A calibration curve can be prepared with known globular proteins and if the unknown protein is also globular, the partition coefficient (k) can be directly related to its molecular mass. The elution pattern of CnGRASP (Figure 9A) resulted in a k value of 0.42. From this, we can infer an apparent molecular mass in solution of 78 kDa, indicating that the protein has a quaternary

structure that seems to be a trimer in solution (Figure 9B). These results contradict those obtained for the homologous of CnGRASP in mammals and are also in disagreement with the proposed mechanism of membrane tethering in which a GRASP dimer is fundamental [63]. Barr *et al* described that the apparent molecular mass of GRASP65 measured by gel filtration on a Superose 6 HR10/30 column (Pharmacia) indicated a large discrepancy with respect to the expected mass of a dimer [135], even though GRASP65 has proved to be a dimer in solution [127]. Furthermore, Yanzhuang [136] and Barr [48] found in coexpression experiments that GRASP65 behaved as a dimer in solution. An aberrant hydrodynamic behavior is usually a very solid indication of protein disorder (99). The presence of intrinsically disordered regions increases the average volume of occupation, which would make these proteins slightly "larger" than expected.

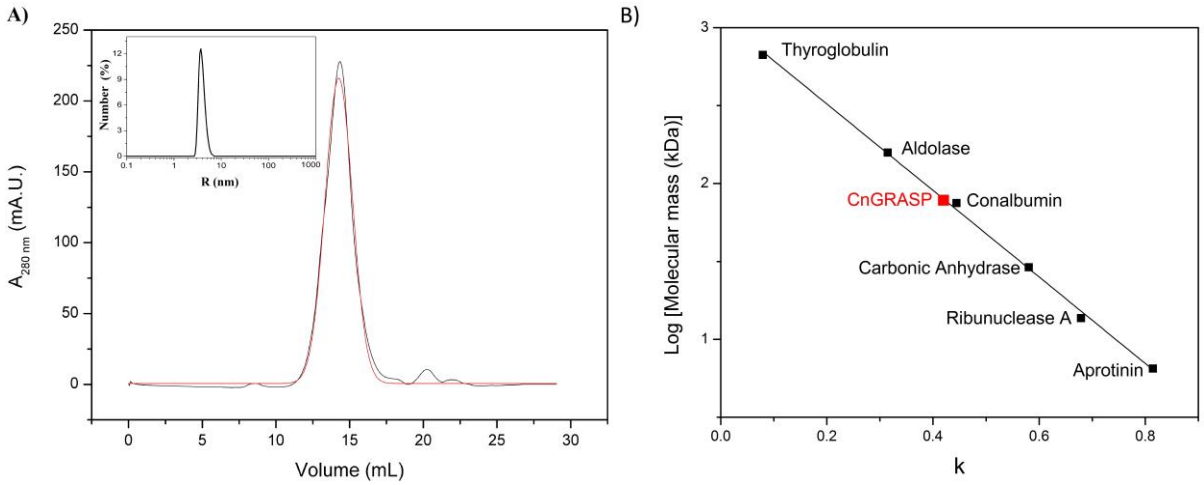
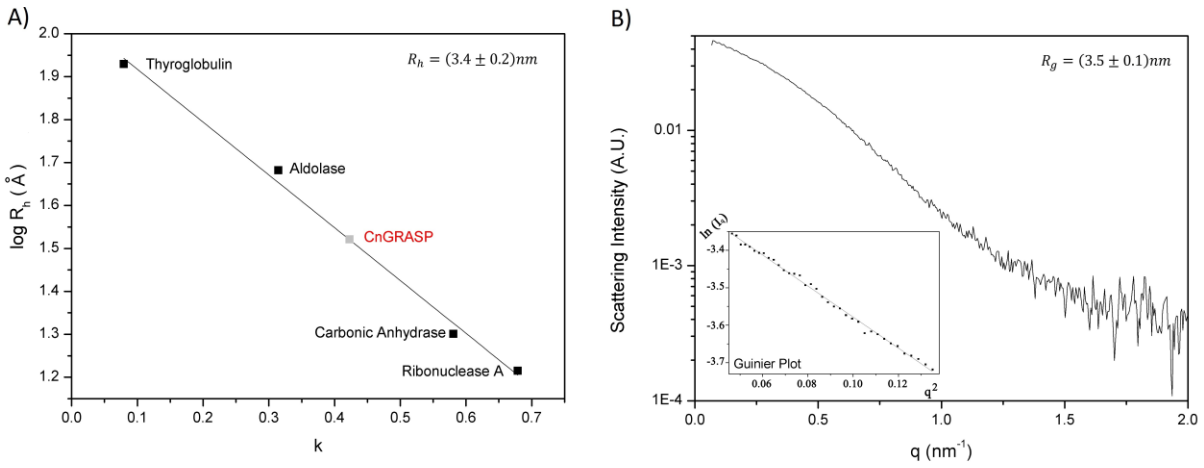


Figure 9: Determination of the apparent molecular mass of CnGRASP by size exclusion chromatography (SEC) A) Elution pattern of purified CnGRASP and a Gaussian fit (red) which suggest a homogeny elution. In the inset on left, it is a DLS result showing that the sample is monodisperse and centered in a value close to the one observed in the SEC results. The choice for the SEC value is because we could not satisfactory evaluate a viscosity coefficient for Rh determination. (Using the water viscosity value, we still rescue an Rh that gives a molecular weight of 68.9 kDa in the Siegel and Monte model, consistent with a dimer conformation). B) Partition coefficient as function of the logarithm of molecular mass, fitted with a linear function where the adjusted R-square was 0.998.

SEC can also be used to recover the hydrodynamic radius (R_h) of an unknown protein in solution [137] and, in the case of CnGRASP, resulted in the R_h value of (3.4±0.2) nm (Figure 10A). The expected R_h value for a dimer of CnGRASP, if it were a globular protein, would be close to 3 nm (Figure 10C) as calculated according to reference [138]. Instead, the R_h value calculated in Figure 10A agrees with the expected value for a dimer of a molten globule protein (Figure 10C).



C)

	M.M (kDa)	R_n (nm)	R_{mg} (nm)	R_{pmg} (nm)	R_{urea} (nm)
Monomer	27	2,38	2,67	3,36	4,57
Dimer	54	3,06	3,37	4,42	6,55
Trimer	81	3,53	3,86	5,18	8,1

Figure 10: Hydrodynamic properties of CnGRASP. A) Linear dependence of R_h as function of the partition coefficient obtained from SEC. Standard proteins were obtained from Gel Filtration Calibration Kit LMW/HMW (GE Healthcare). B) SAXS data obtained for CnGRASP and the same data in the region of Guinier approximation (inset). The range used was between 0.759 and 1.294 qR_G, values obtained using PRIMUS in ATSAS package. C) Theoretical values of molecular mass (MW) and the theoretical associated hydrodynamic radius expected for a native globular protein (R_n), molten globule (R_{mg}), pre-molten globule (R_{pmg}) and denatured protein in urea (R_{urea}). The calculated hydrodynamic radii were based on the theoretical molecular mass of each oligomer using the experimental equations in reference [138].

Nevertheless, what is a molten globule? Molten globules were initially observed and characterized as kinetic intermediates of the protein folding/unfolding [139]. Molten globules are globules that conserves a native-like secondary structure content but without the tightly packed protein interior, having a characteristic structural dynamic in a µs-ms timescale and high affinity for ANS [99]. The precise definition of molten globules is still a matter of debate. We will use in this work the same definitions used by Uversky [99,156] and classify this not as a kinetic intermediate, but as a collapsed IDP.

Another method also used to gain information on the quaternary structure of proteins is the Siegel and Monte model (140), which has been applied in oligomerization studies of a large number of proteins in solution (137-143), including IDPs (144,145). The method relies on two hydrodynamic parameters (hydrodynamic radius and sedimentation coefficient) that are combined to obtain the molecular mass of the protein in solution without the need for the globularity premise. The final equation is: $M = sN_0(6\pi\eta R_h)/(1 - \nu_2\rho)$, where M is the molecular mass, s is the sedimentation

coefficient, ρ is the solvent density, η is the viscosity of the solvent, N_0 is the Avogadro's number, R_h is the hydrodynamic radius and ν_2 is the partial specific volume of the protein. For the calculation, η was set to 0.001 $Pa.s$ (water density) and $\nu_2\rho$ equals to 0.73 (using the standard protein $\nu_2 = 0.73\ cm^3/g$ and $\rho = 1\ g/\ cm^3$). The hydrodynamic radius of CnGRASP is (3.4±0.2) nm (as found from SEC data) and the sedimentation coefficient was estimated using zone sedimentation in a sucrose gradient by comparison to standard proteins of known s (carbonic anhydrase, bovine serum albumin and conalbumin) (Figure 11). The value obtained for s was close to 4.2 S, which implies an estimated molecular mass of 60 kDa, close to the theoretical value expected for a dimer (54 kDa).

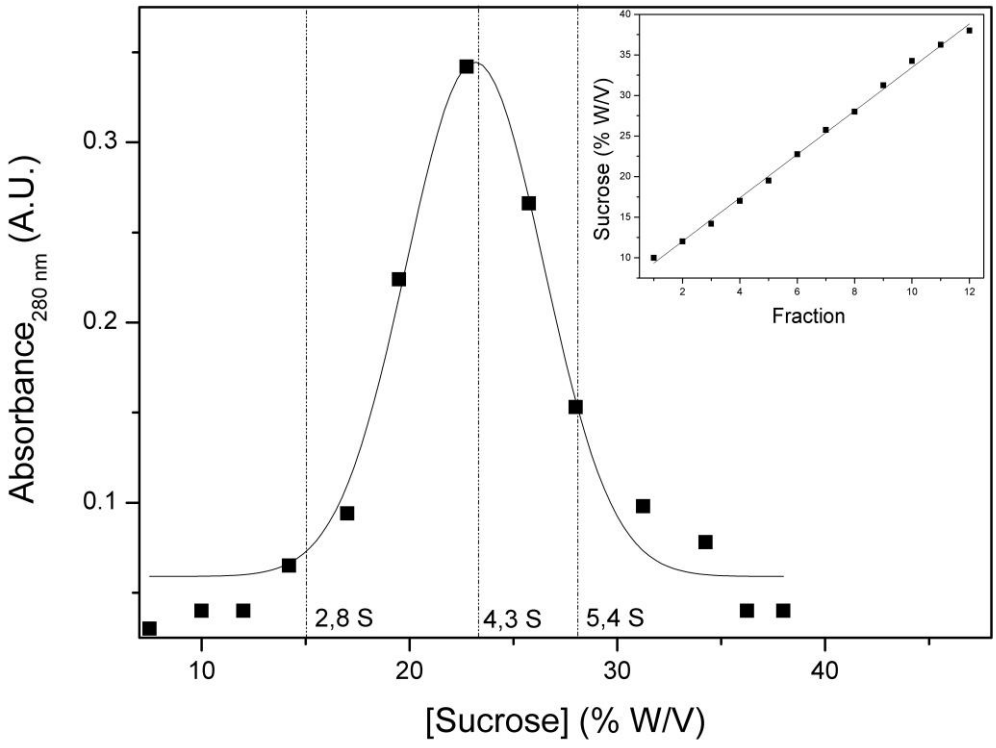


Figure 11: Elution profile of CnGRASP in a sucrose gradient after ultracentrifugation The dotted lines indicate the elution of the standard proteins. The linearity of the sucrose gradient was monitored using the diffraction index variation and is located in the inset.

We also evaluated the R_h value using DLS experiments (inset in Figure 9A). In this case, the result showed a signal centered at 3.9 nm that accounts for 99% of the measured volume. This signal has a relative polydispersity of 18%, which indicates that this peak is monodisperse and that there is one major protein population. The apparent discrepancy between the R_h values determined from SEC and DLS experiments is probably due to the use, in the DLS calculation, of a poorly defined viscosity value. The buffer used during CnGRASP preparation contains 10% V/V glycerol, thus increasing the

solution's viscosity. This makes the molecules diffuse slower and that is accounted for during the analysis as a "larger" molecule than it truly is. From this R_h value, we obtained a molecular mass of 68.9 kDa, which is more consistent with a dimer than a trimer, since the mass values determined in this sort of calculation are usually deviated to greater values. This result confirmed that the globularity deviation observed for CnGRASP is not caused by an unexpected quaternary structure as initially suggested by the estimated mass determined from our SEC experiment.

Another related measure of the size of a molecule is its radius of gyration (R_g). For CnGRASP, the value of R_g, determined using Synchrotron Small Angle X-Ray Scattering (in the Guinier approximation) was (3.5±0.1) nm (Figure 10B). This R_g value is consistent with the one obtained from the distance distribution function calculated from the experimental X-ray scattering data (3.7 nm, using GNOM, from ATSAS package) in the range of concentrations used. Furthermore, Jayaram *et al* [146] used a computational based method and structures available in the protein data bank to construct a linear function relating the values of R_g with the total number of amino acids in a protein sequence (http://www.scfbio-iitd.res.in/software/proteomics/rg.jsp, accessed in 2017). As the PDB database is mostly comprised of structured and globular proteins, the value obtained experimentally by SAXS for CnGRASP (3.5 nm) does not follow the R_g of 2.4 nm obtained using the Jayaram model, again suggesting that CnGRASP does not behave as a completely globular protein.

Although both R_g and R_h can be used as parameters to assess the degree of compactness of proteins, they refer to different aspects of the molecular structure. The value of R_g is related to the geometrical dimensions of the molecule, whereas R_h is related to the hydrodynamic properties. A constant known as the Q-factor is an indicative of molecular conformation (147,148). The Q-factor is given by $Q = {R_g}/{R_h}$, being equal to 0.775 for a compact sphere or close to 0.8 for globular proteins, and 1.51 for a theoretical random coil [147]. For proteins with pre-molten globule conformation, the values are close to 1, and for molten globule they are between 0.7 and 0.93, both being closer to the values observed for compact proteins [147]. The values determined above for CnGRASP yielded a Q-factor equal to (1.03±0.09), which is closer to what is expected for a protein either in a molten globule or in a pre-molten globule conformation.

The information needed to distinguish between molten globule and pre-molten globule can be extracted from the SAXS data seen in the Kratky plot [128], whose shape is sensitive to the conformational state of the scattering particles and to their degree of flexibility [128,135,149]. The scattering curve in the Kratky plot has a characteristic maximum when the protein is compact, such as in the globular or in the molten globule states, while this maximum is absent in the pre-molten globule and unfolded states [150,151]. A very useful pattern recognition scheme is presented by Rambo and Tanier [152] and following such scheme we determined that the shape of the SAXS data in the Kratky plot for CnGRASP (Figure 12) is compatible with the data from a partially folded protein containing a compact and a disordered domain, which is exactly the expected structural arrangement for CnGRASP (the GRASP domain linked to the SPR domain) [152]. We can then conclude that rather than behaving as a pre-molten globule protein, CnGRASP has a molten globule-like shape.

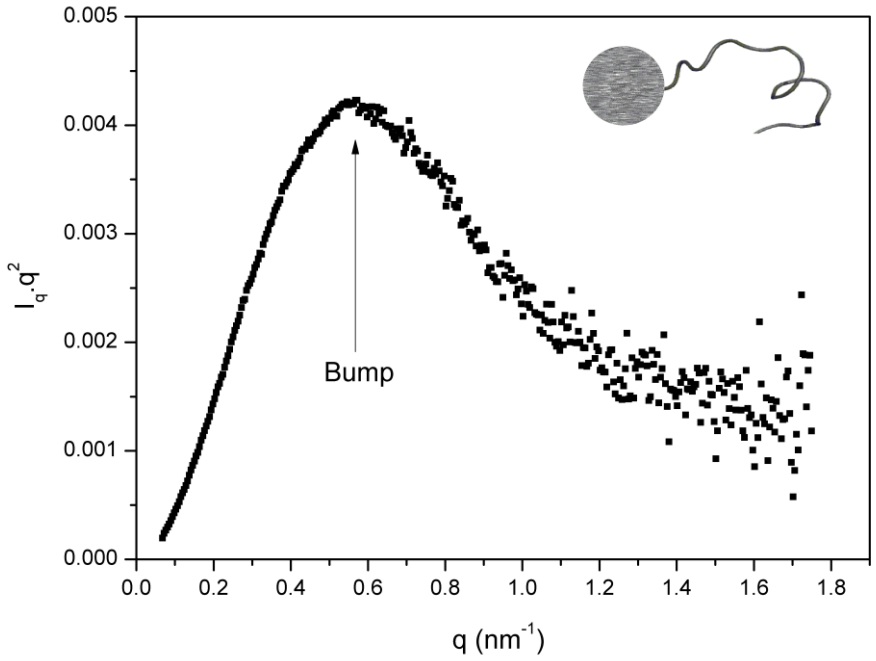


Figure 12: Solution behavior of CnGRASP. SAXS data viewed in a Kratky plot for CnGRASP. The bump observed for CnGRASP at q value close to 0.6 nm^{-1} is indicative of a compact structure. The model of a compact structure plus a coil one (located in the inset) was based on reference [153].

Fluorescence experiments using the tryptophan residues in CnGRASP were also used to strength the conclusions drawn from the SAXS data in the Kratky plot. Tryptophan amino acids can be selectively excited using a monochromatic radiation of 295 nm. CnGRASP has four native tryptophan residues that should all be in similar structural environments if CnGRASP is indeed a molten-globule

protein, as suggested by the results presented above, since they are located in diferent positions according to the molecular model presented in figure 7A. Similar environments in a highly flexible structure will lead to high tryptophan accessibility to the solvent [154]. For solutions of free tryptophan in water, the wavelength of maximum emission in the fluorescence spectrum occurs around 350 nm (354 nm in Figure 13A) [154]. Tryptophans located in the hydrophobic cluster of proteins show a blue shift in that wavelength to values around 320 nm [154]. CnGRASP has an emission maximum close to 344 nm (Figure 13A), indicating that the tryptophans are neither fully exposed to the solvent nor in a very hydrophobic environment, therefore implying a certain degree of structuration of CnGRASP. Moreover, all tryptophans apparently have the same high accessibility to the solvent based on the linear Sterm-Volmer plot obtained from a quenching experiment using acrylamide, but still not the same as in an unfolded state like in 7 M urea (B). Tryptophan residues in a hydrophobic environment (fluorescing at 344 nm) and still with high and similar solvent accessibility (Stern-Volmer plot) is consistent with molten globules, mainly because of their less compact and still structured conformations and high structural flexibility [155].

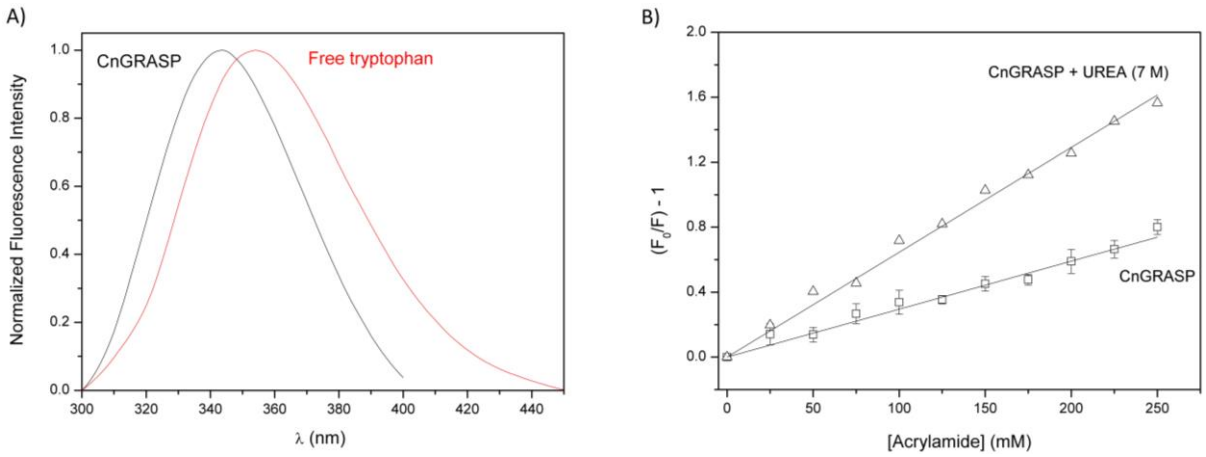


Figure 13: Hydrophobic core accessibility using tryptophan fluorescence. A) Normalized fluorescence emission of CnGRASP and a solution of 1 mM Tryptophan in the same buffer solution, using an excitation radiation of 295 nm. B) Stern-Volmer plot of CnGRASP and a denatured state condition (with 7 M urea) using acrylamide as a collisional quencher.

Low cooperativity of unfolding

The free energy landscape for protein folding (or unfolding) can dramatically change when structured proteins and IDPs are compared [156]. A plateau due to the high flexibility of the structure can represent the free energy pattern for IDPs. This leads to a monotonic decrease of protein structure

content and thus to a low cooperativity in the protein fold/unfold transition induced by a denaturant agent [156]. CnGRASP followed this behavior as assessed by chemical and thermal denaturation experiments performed using CD spectroscopy (Figure 14 A-C). The CD values at 190 nm and 222 nm are typically associated with helical structures and their temperature variation showed a low cooperative thermal unfolding (Figure 14 B). Moreover, there was no detected gain in protein secondary structure induced by the temperature increase, a feature previously observed for fully extended IDPs [91,99,148]. The thermal unfolding is irreversible, as can be observed in Figure 14B, since the SRCD spectrum measured after the system was allowed to cool down is different compared to the initial spectrum. The low cooperativity in the thermal induced unfolding was also confirmed using DSC with conalbumin as a control (Figure 14D).

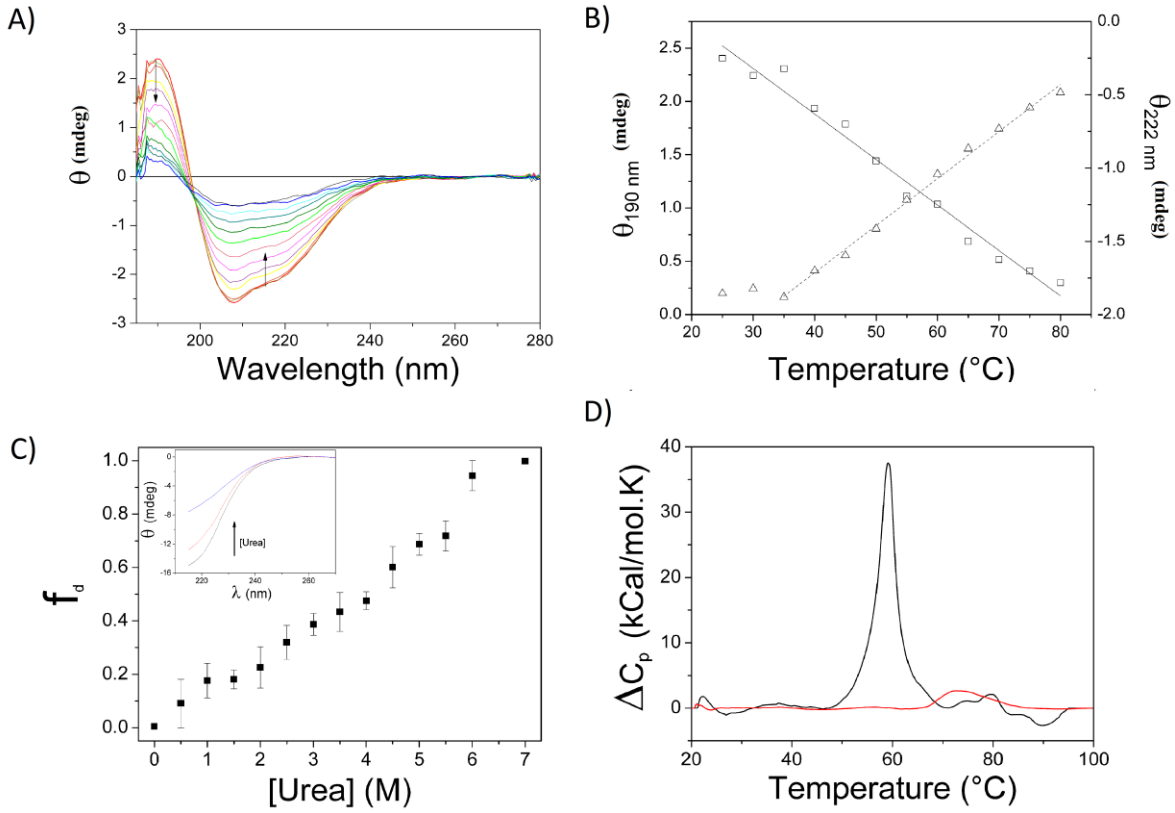


Figure 14: CnGRASP denaturation profile under thermal and chemical perturbation. A) Thermal induced unfolding monitored using SRCD. The arrows indicate the direction of spectral changes upon temperature increases. B) Plots of the peak magnitudes at 190 (square) and 222 nm (triangle) during the heating process. C) Urea-induced unfolding of CnGRASP monitored by far-UV. CD. The unfolding fraction f_d was calculated from the CD intensity at 220 nm (inset). D) Thermal unfolding of CnGRASP (red line) and Conalbumin (black line) monitored by DSC

The chemical unfolding was investigated by means of urea titration experiments. It is possible to observe a monotonic decrease of the ellipticity at 220 nm (Figure 14C), which has a linear instead of

the sigmoid-like behavior usually observed for well-structured proteins, indicating a low cooperativity also during chemical unfolding as probed by CD. An equivalent experiment performed using static fluorescence and monitoring the fluorescence of 8-anilino-1-naphthalenesulfonic acid (ANS) led to the same conclusion (Figure 15). ANS is a fluorescent probe widely used in the identification of accessible hydrophobic sites in proteins [157,158]. The change in the fluorescence pattern upon increasing urea concentration reflects changes in the microenvironment of ANS from a hydrophobic medium (interior of the protein - structured) to a polar environment (exposed to the solvent after denaturation) (Figure 15A). The shift in the wavelength of maximum emission occurred in a low cooperativity way (Figure 15B), indicating that the exposure of the hydrophobic core also followed the behavior observed for the secondary structures during chemical denaturation. High fluorescence intensity in the beginning is a result of high ANS affinity (Figure 15A), also a property observed for proteins in the molten globule state [99]. Altogether, these results suggest that the secondary structure elements in CnGRASP are not well stabilized by tertiary contacts such that the protein loses them in a poorly cooperative way when unfolded by a denaturant agent. The high ANS affinity shows that CnGRASP structure is flexible enough to allow access to its partially hydrophobic interior, which is in agreement with the tryptophan fluorescence data.

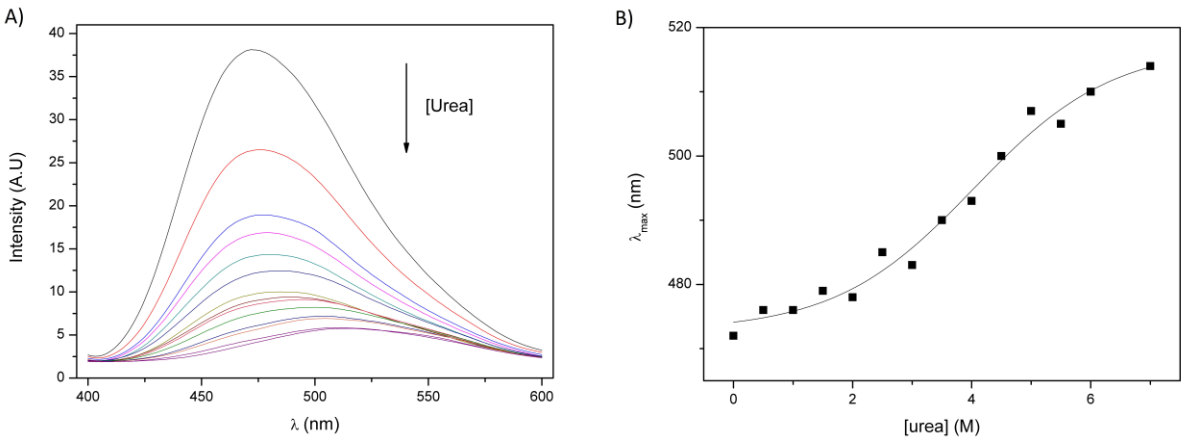


Figure 15: Urea induced unfolding of CnGRASP monitored using the ANS fluorescence. A) The decrease in fluorescence intensity is an indicative of the ANS microenvironment change, during urea titration, from a non-polar environment (hydrophobic core) to a polar environment (water exposed). The high fluorescence intensity in the beginning is an indicative of high ANS affinity. B) Variations of the wavelength of maximum emission intensity as a function of urea concentration, again showing a poorly cooperative unfolding process.

High structural flexibility

The structural flexibility of CnGRASP was further investigated by enzymatic digestion, which has been a widely used method to elucidate local disorder and flexibility in proteins. Fontana *et al* showed that local flexibility and not only solvent exposure is the determinant factor for an efficient proteolytic activity [128**,**159**]**, whereas stable secondary structures usually have a protective effect against proteolytic cleavage [99]. The proteolytic assay was performed with CnGRASP and the structured proteins conalbumin and ovalbumin as controls. Trypsin is a protease that cleaves peptide chains mainly at the carboxyl side of the amino acids lysine or arginine, except when they are linked to proline [160]. All the proteins tested have comparable multiple sites predicted to be cleaved by trypsin, allowing then the identification of cleavage sites solely based on the flexibility of the protein substrate. The results of the procedures are outlined in Figure 16. Three incubation times with trypsin (0, 1, and 16 h) were used for each protein tested. Aliquots of each sample were then collected and visualized in SDS-PAGE (Figure 16A). It was possible to notice the high sensitivity of CnGRASP to proteolysis since no band was seen on the gel after 1 h of incubation with trypsin. This observation validated our results concerning structure ordering and flexibility. Well-structured proteins can resist to proteolysis for long periods as observed with conalbumin and ovalbumin. Conalbumin generated a low content of protein digests with significant amounts of intact bands even after 16 h of incubation. Similarly, ovalbumin showed no signs of susceptibility to proteolysis. To assess CnGRASP hydrolysis in more detail, a 5-fold higher ratio of CnGRASP:Trypsin was tested. Even though protein degradation occurred slower than observed under the conditions described before, it was already impossible to detect the presence of intact CnGRASP by SDS-PAGE after 30 minutes (Figure 16 B).

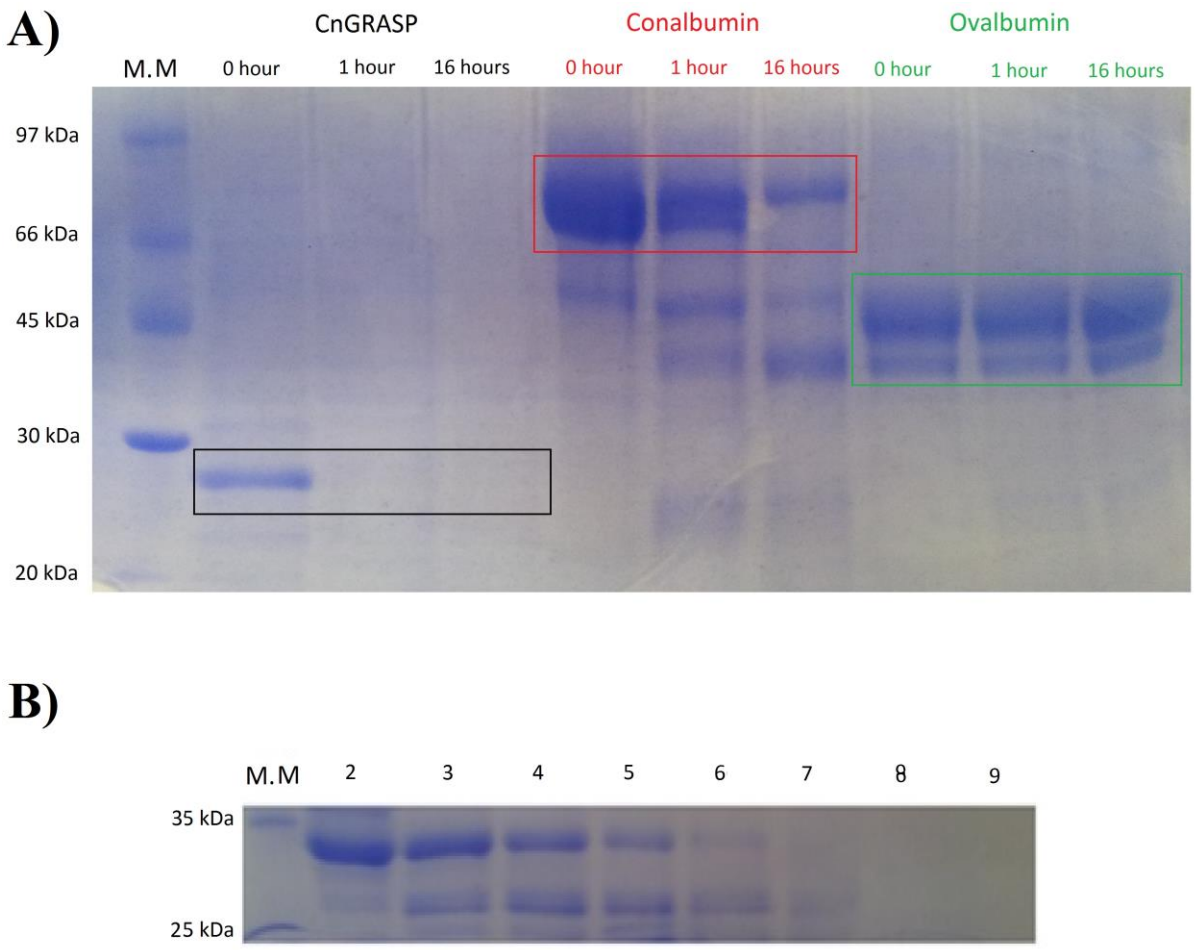


Figure 16: Proteolytic assay with Trypsin at room temperature. A) The protein:trypsin ratio was 20:1 and the collection times were 0, 1 and 16 hours after adding trypsin to the solution. B) The protein:trypsin ratio was 100:1 and the collection times were 0, 10, 15, 20, 25, 30 and 60 minutes (lanes 2-8) after adding trypsin to the solution. A protein weight standard was applied in lane 1. The reaction was stopped by adding SDS-PAGE loading buffer and heating to 95°C for 10 minutes

NMR studies prove the molten globule-like behavior of CnGRASP

Molten globule structures have a very particular behavior when NMR is used to measure the signals arising from the amide groups [161,162]. The conformational fluctuations allowed by the lack of stable tertiary contacts create a fluctuating ensemble of structures that inter-convert in a millisecond to microsecond time scale [161]. In this time scale, the amide resonances are very broad and difficult to be detected in the ^{1}H-^{15}N HSQC spectrum [161,163]. CnGRASP ^{1}H-^{15}N HSQC spectra are shown in Figure 17. It is possible to observe a very characteristic pattern of intrinsically disordered proteins: the presence of few peaks (around 30) collapsed in a very small region (close to 1 ppm) (Figure 17A). Since the assignment of the peaks has not been performed yet for CnGRASP, we could just speculate that those few resonance peaks would come from the SPR domain and/or other regions predicted as intrinsically disordered in the protein (see above). In an indirect CD result, we measured the CD

spectrum of the full-length CnGRASP and subtracted from it the CD spectrum of the isolated GRASP domain (Figure 18 A-B). The subtracted spectrum, presumably corresponding to the spectrum of the SPR domain, had the usual poliproline-2-like CD shape, characteristic of extended IDPs [130,144]. This result supports the conclusions drawn from our NMR data regarding the origin of the peaks observed in the ^{1}H-^{15}N HSQC as being from the SPR domain.

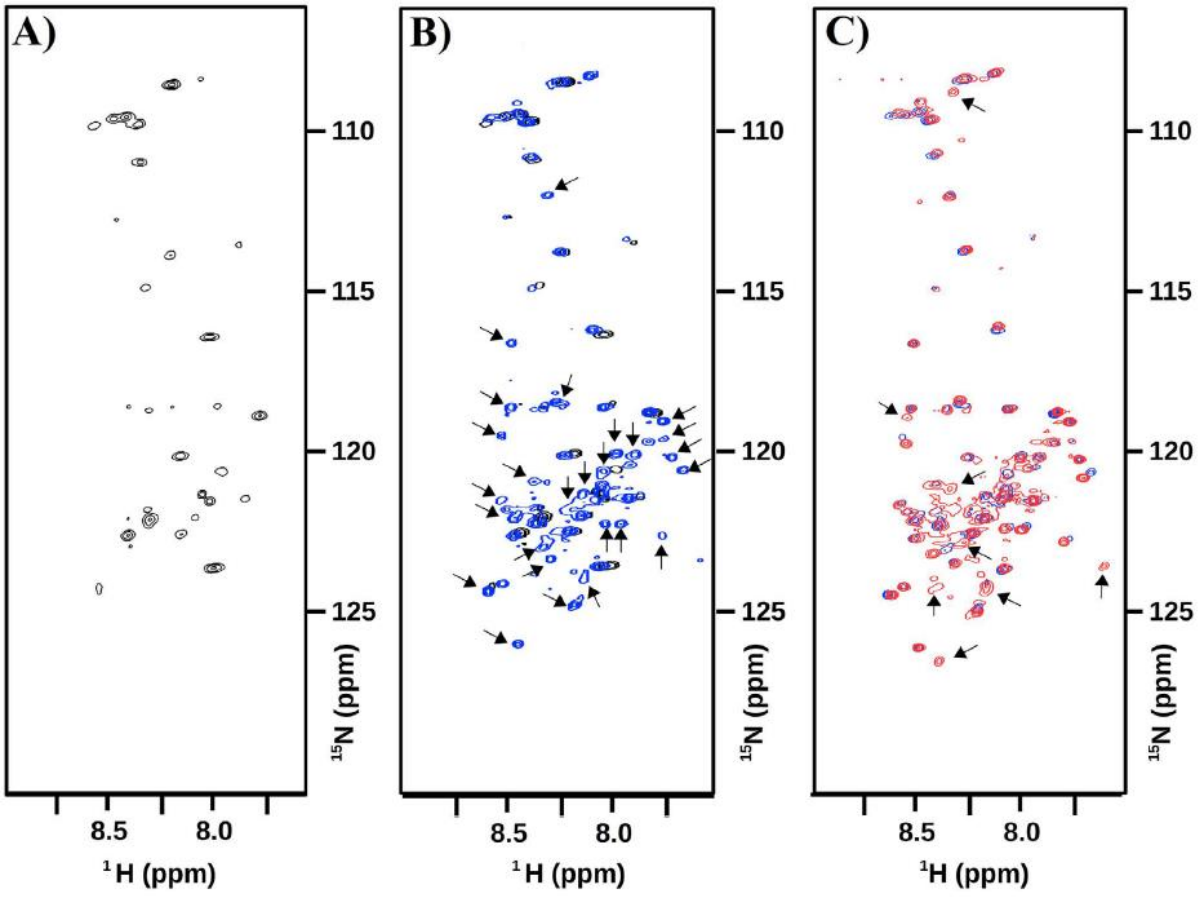


Figure 17: NMR data analyses of CnGRASP. A) ^{15}N-^{1}H HSQC spectrum of ^{15}N-labeled CnGRASP at 20°C. B) Superposition of ^{1}H-^{15}N HSQC CnGRASP spectra (red) with the same sample plus 2 M Urea (black). C) Superposition of ^{1}H-^{15}N HSQC CnGRASP spectra plus 2 M area (black) with the same sample with 4 M Urea (red). The black arrows indicate the new signals that start to appear.

It has been previously observed that it is possible to play with the molten globule conformation equilibrium by disturbing the sample with chaotropic agents or with changes in the temperature [161,163,164]. For molten globules, when the concentration of a chaotropic agent, such as urea, is increased, the resonance lines start to appear in the spectra, but still with low chemical shift dispersion [161]. When a 2 M urea solution is added to the GRASP sample (Figure 17B), the number of resonance lines seen in the spectrum is increased, and, at 4 M urea (Figure 17C), peaks are still appearing similarly to what has been observed for other molten globule structures, such as α-lactoglobulin [163] and

staphylococcal nuclease [165]. These data strongly indicate that full-length CnGRASP do have a molten-globule like behavior. However, the particular condition where all the resonance peaks can be observed is still to be determined.

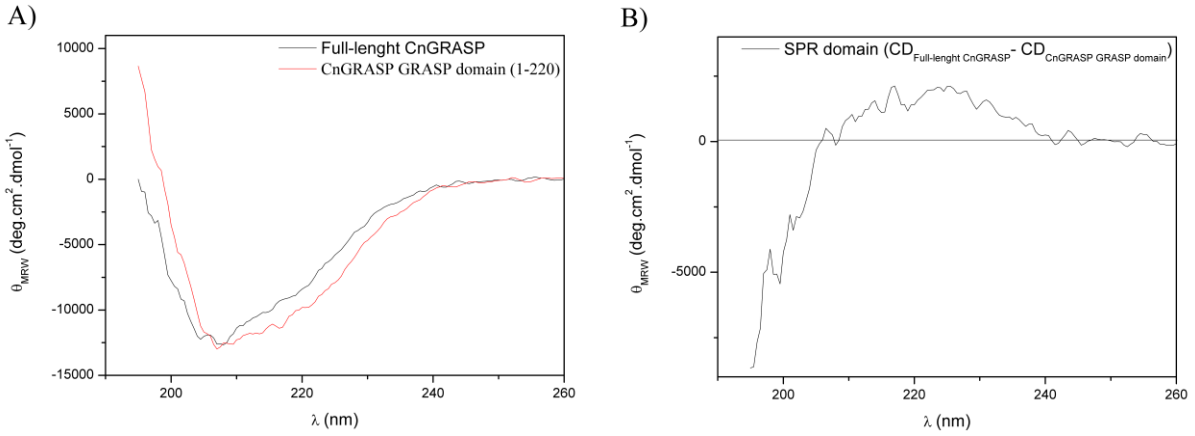


Figure 18: CD analyses of CnGRASP domains. A) CD signal of full-length CnGRASP and from our GRASP domain construction (1-220), prepared in the same way as the native one. The signals are normalized in θ_{MRW} units. B) The theoretical SPR signal obtained from the CD signals presented in A. We supposed that the GRASP domain and the SPR domain are independent structures, so they have independent CD signals that, together, gives the full-length CD one. This cannot be used as a definitive prove, but as indication, together with our other results, that the SPR domain has the full intrinsically disordered patter

Tryptophan fluorescence and local ordering

Three out of the four tryptophans present in CnGRASP are located in the PDZ-2 domain (see the additional figure in the end of this chapter), which is presumably the most ordered domain of the protein. If this hypothesis is true, PDZ-2 can go through a structural transition in a more cooperative way during chemical denaturation. In fact, tryptophan fluorescence analysis showed that under increasing urea concentration, the wavelength of maximum emission red-shifted to values around 352 nm, which are characteristic of free tryptophan in solution (Figure 13 and Figure 19). Furthermore, the decrease of the fluorescence anisotropy during urea denaturation showed that there is a relevant difference of tryptophan immobilization before and after denaturation (Figure 19). These results suggest that, in the PDZ-2 of CnGRASP, there is a hydrophobic core, but with high solvent accessibility. The cooperativity observed in the chemical denaturation proved the existence of local ordered structures, likely around the tryptophan residues in the PDZ-2 domain. Even though the whole structure seems to be flexible, there are well-structured regions within the protein that can go through a transition in a more cooperative way. Besides that, it is possible to see that the transition is not highly cooperative and this may seem

the reason why it is not resolved in other denaturation data. We should be able to observe a more cooperative transition of the whole protein if the PDZ-1 followed the same PDZ-2 behavior. However, this is not seen in our data for full-length CnGRASP and the reason why we cannot observe it is an indicative that CnGRASP has a PDZ structural asymmetry, a result that correlates well with the disordered prediction and will be explored in details in chapter 4.

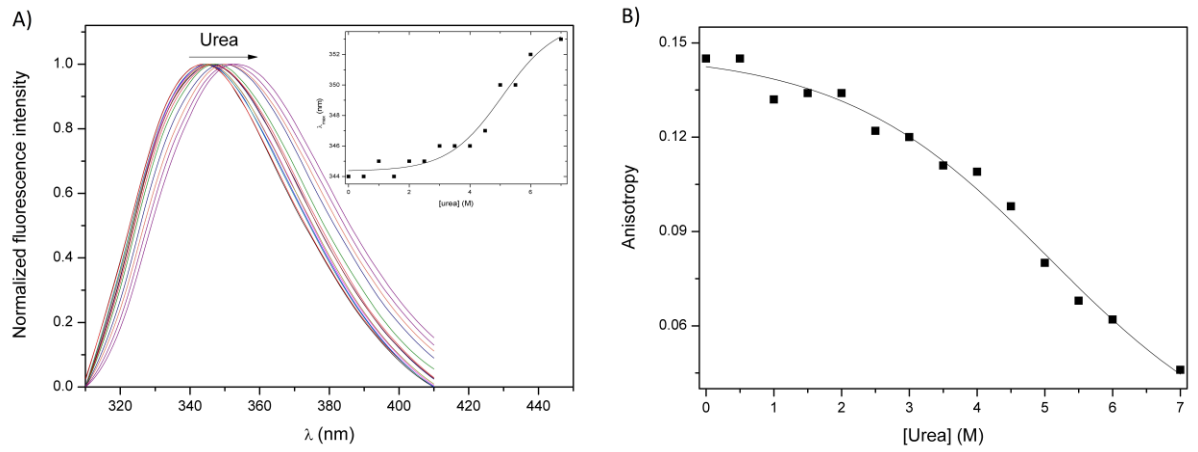


Figure 19: Tryptophan exposure and indirect local secondary structure identification using urea induced unfolding. A) Static fluorescence emission as a function of urea concentration. The wavelength of maximum emission intensity is plotted in the inset as a function of the urea variation. The data was fitted to a sigmoid-like function. B) Static fluorescence anisotropy change during unfolding.

Discussion and conclusions

The behavior of full-length GRASPs, especially in solution, is still poorly understood. A GRASP orthologue has been described as a regulator of polysaccharide export, through non-conventional mechanisms, and related to pathogenesis in the neuropathogen *C. neoformans* [84]. Here, a series of biochemical and biophysical methods were used to assess the dynamic behavior in solution of *C. neoformans* GRASP. CnGRASP apparently has a mixture of substantial content of secondary structures with a significant amount of disordered regions as assessed by four disorder predictors and by experimental data from circular dichroism, NMR and fluorescence. In particular, the SPR domain appears as a highly disordered portion of the protein in all predictions and among all species (Figure 20), thus suggesting that the intrinsic high flexibility associated with disorder could well correlate with its role as a phosphorylation site within the protein sequence. Interestingly, PDZ-1 domain of

CnGRASP also presented high disorder propensity, which could suggest an asymmetric role in functional processes, such as oligomerization, as further discussed in this section.

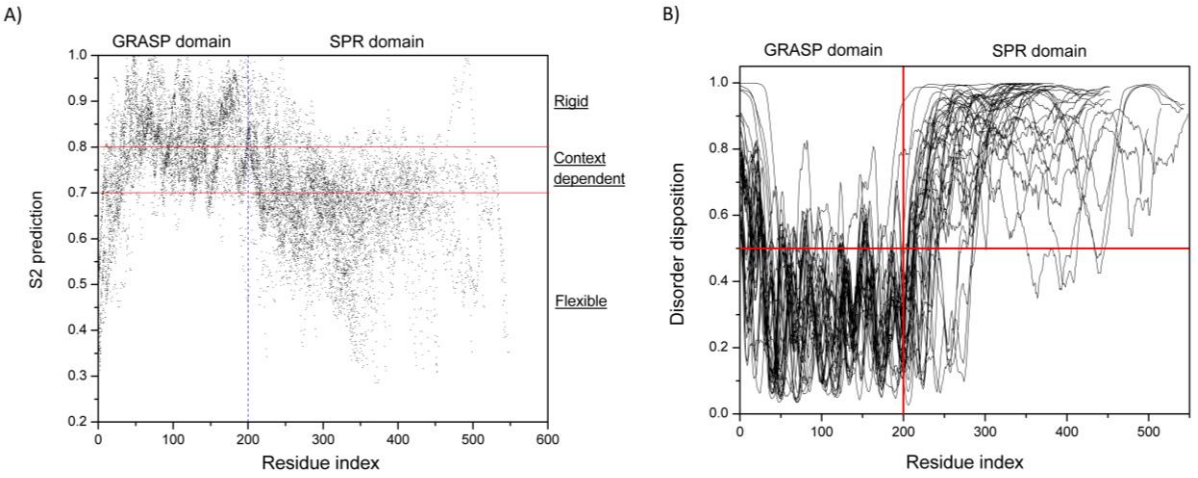


Figure 20: Disorder prediction along GRASPs sequence. A) Backbone flexibility prediction at the residue-level in the form of backbone N-H S^2 order parameter values using DynaMine. Zone propensity corresponding to a rigid, flexible and context dependent limit states are detached. B) Disorder prediction using VSL2 predictor. It is worthwhile to notice that the SPR domain has a high disordered/flexibility propensity along all GRASPs tested, a tendency not observed for the GRASP domain.

The hydrodynamic properties of CnGRASP pointed to an aberrant behavior of the protein in solution based on the results from SEC experiments that suggested CnGRASP would be a trimer, which was totally incompatible with the need for a dimer for correct function of the protein and in disagreement with data available in the literature. This flawed prediction from SEC was likely due to the use of globular proteins as controls and the correct behavior was only recovered when the Siegel-Monte model, which does not require the globular assumption, was used (prediction of a molecular mass of 60 kDa for CnGRASP, a value compatible with the expected dimer in solution). Therefore, the unusual migration in the reported experiments was not due to the formation of higher-order oligomers but only to the existence of parts of the protein structure that do not follow globular-like patterns.

Furthermore, SAXS and fluorescence results indicated that CnGRASP is a collapsed protein, but it lacks the typical globular behavior of many well-structured proteins. Profiles of protein denaturation do not follow the usual cooperative transition observed for well-structured proteins, as suggested by the results from different techniques (CD, SRCD, fluorescence and DSC) obtained during thermal and chemical CnGRASP unfolding. The high proteolysis sensitivity together with the SAXS Kratky plot profile showed that the protein presents high flexibility, which probably results in

substantial solvent accessibility as also inferred from our quenching of fluorescence data. Overall, these results indicated that CnGRASP behaves rather as a native molten globule than as a regular globular protein in solution. The ^{1}H-^{15}N HSQC CnGRASP data confirmed this conclusion by showing an increasing number of resonances upon urea addition, therefore following a pattern previously reported for other molten globules. It is important to notice that the disordered/flexibility observed does not mean that CnGRASP is not functional. The CnGRASP gene was proved to translate a functional protein [84] and can form dimers in solution, so flexibility does not seem to interfere in these processes. It may well be that the flexibility of the protein structure can facilitate the information propagation, making the GRASP domain more accessible to regulation.

The lack of cooperativity observed in the unfolding of CnGRASP when methods that are based on the observation of physicochemical parameters monitoring the overall protein structure suggest that the degree of disorder in CnGRASP plays a crucial role in preventing the structure as a whole to unfold as a regular globular protein. However, our bioinformatic results indicated that disorder-promoting residues are heterogeneous distributed along the protein primary sequence. Moreover, when fluorescence is used to monitor a localized region of the protein, such as the tryptophan residues' vicinities, a more cooperative unfold is observed, which, in agreement with the theoretical predictions, indicated the presence of ordered regions in the protein structure. Interestingly, both the computational predictions and the Trp fluorescence results suggested that the more ordered portions would be located in the PDZ-2 of the PDZ domain (three of the four Trp residues in CnGRASP are located in the PDZ-2). CnGRASP structure would be formed by more disordered SPR and PDZ-1 and more ordered PDZ-2 domains. This asymmetric structural (or unstructural) behavior observed for PDZ-1 and PDZ-2 could then be responsible for the correct formation of the needed dimers during the stacking of Golgi cisternae. It has been proposed, for human GRASPs, that during the stacking process PDZ-1 interacts with the Golgi membrane through a myristoylation of a glycine residue in its N-terminal, whereas PDZ-2 binds to its receptor GM130 located on the cisternae surface. However, it has been recently shown that the GM130 C-terminal peptide can bind concomitantly with both PDZ-1 and PDZ-2, although the mode of binding involves the interaction of distinct regions of the peptide with each one of the PDZs (66). This, again, suggests that each PDZ may have different roles in physiologically related phenomena, therefore in agreement with our conclusion on functional asymmetry of the PDZs. Moreover, during non-

conventional secretion and Golgi assembly, PDZ-1 and PDZ-2 bind to different partners as previously suggested [63,166]. In a recent report, it has been shown that even the SPR domain can have protein partners [167]. The control of which PDZ (1 or 2) or SPR binds (or interacts) to a determined partner could be performed by means of the asymmetric order/disorder behavior of the CnGRASP domains.

The presence of IDRs in CnGRASP can also be used to understand how the protein is able to indiscriminately interact with a large number of other proteins participating in secretion pathways. GRASP can directly interact with the C-terminal of the Transforming Growth Factor-a (TGF-a) and several members of p24 family, facilitating the conventional secretion of these proteins or their retention in the Golgi complex [55,56]. GRASP proteins are also directly related to the unconventional secretion of a large number of proteins, such as the soluble acyl-coenzime A binding protein (ACBP) in *Dictiostelium* (37) and with starvation-induced secretion of ACBP in *Saccharomyces cerevisiae* and *Pichia pastoris* (39,40). Furthermore, GRASP is involved in the Golgi bypass of αPS1 integrin during the stage 10B of *Drosophila* embryogenesis [43]. The GRASP mediated unconventional secretion pathway can be used to rescue the cell surface expression of the mutant ΔF508 in the cystic fibrosis transmembrane conductance regulator (CFTR) [44], the most common mutation related with the cystic fibrosis disease, and this rescue only happens with the correct interaction between both proteins. GRASP is apparently also required for exporting of molecules of non-protein nature. In *C. neoformans*, deletion of the CnGRASP resulted in inefficient secretion of GXM, reduced capsule formation and attenuated virulence [84]. These data indicate that GRASP is also implicated in polysaccharide secretion via unconventional pathways and, consequently, fungal virulence. Properties including structural flexibility and related promiscuity can be further used to gain insight of how this class of protein can be so dynamically functional, with particular emphases in unconventional protein secretion.

The unexpected behavior of full-length CnGRASP in solution, indicating the presence of intrinsically disordered regions (IDRs) within its structure, might be essential in understanding the plethora of functions in which the protein is involved. The results here presented suggest that intrinsic disorder must then be kept in mind when one is dealing with processes related to GRASP function.

List of GRASP sequences used in this chapter, collected from the NCBI database and in a FAST format.

>gi|657135458|gb|KEG07653.1| Golgi reassembly stacking protein, partial [Trypanosoma grayi]

>gi|151942478|gb|EDN60834.1| grasp65 (Golgi reassembly stacking protein of 65kd)-like protein [Saccharomyces cerevisiae YJM789]

>gi|52345514|ref|NP_001004805.1| golgi reassembly stacking protein 1, 65kDa [Xenopus (Silurana) tropicalis]

>gi|52345500|ref|NP_001004798.1| golgi reassembly stacking protein 2, 55kDa [Xenopus (Silurana) tropicalis]

>gi|148233199|ref|NP_001080519.1| golgi reassembly stacking protein 2, 55kDa [Xenopus laevis]

>gi|70833735|gb|EAN79237.1| Golgi reassembly stacking protein (GRASP homologue), putative [Trypanosoma brucei brucei strain 927/4 GUTat10.1]

>gi|666434722|gb|KEY82217.1| golgi family reassembly stacking protein [Aspergillus fumigatus var. RP-2014]

>gi|635510258|gb|KDE82201.1| golgi family reassembly stacking protein [Aspergillus oryzae 100-8]

>gi|119631632|gb|EAX11227.1| golgi reassembly stacking protein 2, 55kDa, isoform CRA_c [Homo sapiens]

>gi|629675408|ref|XP_007799201.1| putative golgi reassembly stacking protein [Eutypa lata UCREL1]

>gi|557727781|dbj|GAD93621.1| Golgi reassembly stacking protein, putative [Byssochlamys spectabilis No. 5]

>gi|573987781|ref|XP_006671811.1| golgi reassembly stacking protein [Cordyceps militaris CM01]

>gi|562972405|gb|ESW98139.1| Golgi reassembly stacking protein [Ogataea parapolymorpha DL-1]

>gi|512188996|gb|EPE04765.1| golgi family reassembly stacking protein [Ophiostoma piceae UAMH 11346]

>gi|477526341|gb|ENH78208.1| golgi reassembly stacking protein [Colletotrichum orbiculare MAFF 240422]

>gi|471559242|gb|EMR61688.1| putative golgi reassembly stacking protein [Eutypa lata UCREL1]

>gi|425769460|gb|EKV07952.1| Golgi reassembly stacking protein, putative [Penicillium digitatum PHI26]

>gi|358367117|dbj|GAA83736.1| golgi reassembly stacking protein [Aspergillus kawachii IFO 4308]

>gi|396483182|ref|XP_003841646.1| similar to golgi reassembly stacking protein [Leptosphaeria maculans JN3]

>gi|346320588|gb|EGX90188.1| golgi reassembly stacking protein [Cordyceps militaris CM01]

>gi|327348935|gb|EGE77792.1| golgi reassembly stacking protein [Ajellomyces dermatitidis ATCC 18188]

>gi|326469524|gb|EGD93533.1| golgi reassembly stacking protein [Trichophyton tonsurans CBS 112818]

>gi|326461673|gb|EGD87126.1| golgi reassembly stacking protein [Trichophyton rubrum CBS 118892]

>gi|320593115|gb|EFX05524.1| golgi reassembly stacking protein [Grosmannia clavigera kw1407]

>gi|325093153|gb|EGC46463.1| golgi reassembly stacking protein [Ajellomyces capsulatus H88]

>gi|320040036|gb|EFW21970.1| golgi reassembly stacking protein [Coccidioides posadasii str. Silveira]

>gi|149022191|gb|EDL79085.1| golgi reassembly stacking protein 2 [Rattus norvegicus]

>gi|83638659|gb|AAI09692.1| Golgi reassembly stacking protein 2, 55kDa [Bos taurus]

>gi|212529034|ref|XP_002144674.1| Golgi reassembly stacking protein, putative [Talaromyces marneffei ATCC 18224]

>gi|119584963|gb|EAW64559.1| golgi reassembly stacking protein 1, 65kDa, isoform CRA_c [Homo sapiens]

>gi|51259254|gb|AAH78731.1| Golgi reassembly stacking protein 2 [Rattus norvegicus]

>gi|157117495|ref|XP_001658795.1| golgi reassembly stacking protein 2 (grasp2) [Aedes aegypti]

>gi|557866393|gb|ESS69592.1| Golgi reassembly stacking protein [Trypanosoma cruzi Dm28c]

>gi|472581853|gb|EMS19568.1| golgi reassembly stacking protein 2 [Rhodosporidium toruloides NP11]

>gi|470242986|ref|XP_004355215.1| golgi reassembly stacking protein [Dictyostelium fasciculatum]

>gi|401428022|ref|XP_003878494.1| putative Golgi reassembly stacking protein (GRASP homologue) [Leishmania mexicana MHOM/GT/2001/U1103]

>gi|349804463|gb|AEQ17704.1| putative golgi reassembly stacking protein subunit, partial [Hymenochirus curtipes]

>gi|398021825|ref|XP_003864075.1| Golgi reassembly stacking protein (GRASP homologue), putative [Leishmania donovani]

>gi|389610007|dbj|BAM18615.1| golgi reassembly stacking protein 2 [Papilio xuthus]

>gi|154344180|ref|XP_001568034.1| putative Golgi reassembly stacking protein (GRASP homologue) [Leishmania braziliensis MHOM/BR/75/M2904]

>gi|353233065|emb|CCD80420.1| putative golgi reassembly stacking protein 2 (grasp2) [Schistosoma mansoni]

>gi|387849038|ref|NP_001248636.1| Golgi reassembly-stacking protein 2 [Macaca mulatta]

>gi|60302830|ref|NP_001012612.1| Golgi reassembly-stacking protein 2 [Gallus gallus]

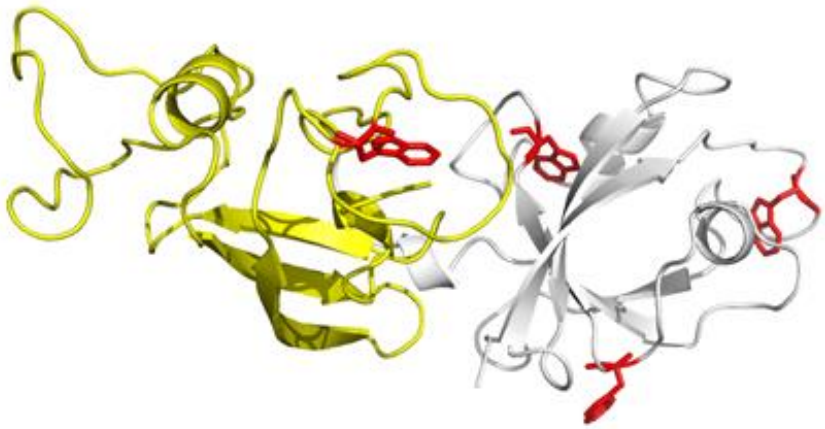
>gi|41053495|ref|NP_956997.1| Golgi reassembly-stacking protein 2 [Danio rerio]

>gi|197097294|ref|NP_001126857.1| Golgi reassembly-stacking protein 2 [Pongo abelii]

>gi|213515012|ref|NP_001133999.1| Golgi reassembly-stacking protein 2 [Salmo salar]

CnGRASP primary sequence used in this study:

MSPADGLVEPYFDYLIGVQTPSINEPTGTPGASEGGRGNSSEGVEALRPDVLGRILEENEGKQIGLRVYN
TKSQRVRDVYLVPSRAWSEEASKASGDPDAKPSLLGLSLRVCNPAHALESVYHVLDVLEGSPAEMAG
LVPWGDYVLAWSGGPLHSENDFYNLIEAHVDKPLRLFVYNADLDNLREVVLYPTRQWGGEGLIGCGI
GYGLLHRIPRPSTPPSGRASLSGADRYFESDNVRRPSMQGSFSEGGGLVGTA



Additional figure: Based on the homology model discussed in the main text, it is highlighted the possible tryptophan positions (red) on the CnGRASP sequence. PDZ 1 and PDZ2 are coloured in yellow and grey, respectively.

Chapter 3 – Disorder-to-order transitions in the molten globule-like Golgi Reassembly and Stacking Protein

Abstract

Golgi Reassembly and Stacking Proteins (GRASPs) are widely spread among eukaryotic cells (except plants) and are considered as key components in both the stacking of the Golgi cisternae and its lateral connection. Furthermore, GRASPs were also proved essential in the unconventional secretion pathway of several proteins, even though the mechanism remains obscure. It was previously observed that the GRASP homologue in *Cryptococcus neoformans* has a molten globule-like behavior in solution. We report here the disorder-to-order transition propensities for a native molten globule-like protein in the presence of different mimetics of cell conditions. Changes in the dielectric constant (such as those experienced close to the membrane surface) seem to be the major factor in inducing multiple disorder-to-order transitions in GRASP, which seems to show very distinct behavior when in conditions that mimic the vicinity of the membrane surface as compared to those found when free in solution. Other folding factors such as molecular crowding, counter ions, pH and phosphorylation exhibit lower or no effect on GRASP secondary structure and/or stability. To the best of our knowledge, this is the first study focusing on understanding the disorder-to-order transitions of a molten globule structure without the need of any mild denaturing condition. A model is also introduced aiming at describing how the cell could manipulate the GRASP sensitivity to changes in the dielectric constant during different cell-cycle periods and in unconventional protein secretion.

Based on the manuscript published on *Biochimica et Biophysica Acta* - General Subjects 2018

DOI: 10.1016/j.bbagen.2018.01.009

Introduction

The Golgi complex is a multifunctional organelle responsible for receiving a large number of proteins from the Endoplasmic Reticulum (ER) to further processing and sorting for transport to their final destination [168]. The mammalian Golgi is polarized, consisting of three particular regions: one close to the ER called *cis* Golgi network, a second called Golgi stacks (which is divided into medial and trans Golgi) and the *trans* Golgi network, where the secreted vesicles are released [1,169]. The correct assembly of the cisternae is important for proper protein glycosylation and sorting [102,170,171]. A family of proteins is responsible for keeping the arrangement of the cisternae and for correctly changing it according to the cell needs [48,49,105]. These proteins are called Golgi Reassembly and Stacking Proteins (GRASPs) and are responsible for the organization of the cisternae into stacks and ribbons [172,173], for facilitating the conventional protein secretion in some cases [54,56], for Golgi dynamics during mitotic and apoptotic periods [59], for Golgi remodeling during cell migration [52], and for unconventional protein secretion [37,44].

GRASP is a peripheral membrane protein, with an N-terminus structured into two PDZ domains exhibiting a protozoa PDZ folding, at least in vertebrates (GRASP domain) [64,65]. A second region in the GRASP structure, usually larger than the GRASP domain, consists of a non-evolutionary conserved domain that is rich in serine and proline (SPR domain) with regulatory function [63,127] and intrinsically disordered characteristics [174]. Two GRASP paralogues (GRASP55 and 65) are usually found in higher eukaryote organisms: GRASP65, which binds to GM130 C-terminus mostly in the *cis* Golgi [175], and GRASP55, which binds to golgin-45 mainly in the medial-Golgi [176]. Trans-oligomerization of GRASP55 and 65 is mitotically regulated and has complementary functions in maintaining the stacking of the cisternae's membrane [57].

In the previous chapter, we showed that the GRASP from the fungal pathogen *Criptococcus neoformans* (CnGRASP) [84] presents features usually observed for a molten globule state, even in the absence of any mild denaturing conditions [174]. That work was the

first (and until the end of 2017, still the only one) structural characterization of a full length GRASP in solution. The classification of CnGRASP as a member of the intrinsically disordered protein (IDP) family brings unexplored possibilities regarding the protein function *in vivo*. IDPs are frequently observed as main components of the cellular signaling machinery [177], commonly functioning as central hubs [178], especially because of their unique structural plasticity [156]. It has also been observed that particular Prion-forming sequences, which are especially enriched in asparagine, promote molten globule-like structures, where amyloid-nucleating contacts can be made [179,180]. It is then becoming clear that the understanding of the molten globule structural behavior in solution is of pharmacological interest.

Nevertheless, how the fungus, or any other GRASP-containing cell, could exploit this particular molten-globule feature remains elusive at this point. An appealing possibility is that different cell conditions could induce disorder-to-order transitions in GRASP allowing it to interact with multiple protein partners, thus taking different roles in cell processes. Moreover, IDPs are usually very sensitive to changes in the physicochemical parameters of the bulk solution, which raises concern about the precipitant conditions usually found in crystallization assays. Particular conditions might trap specific conformations that could not be representative of the whole ensemble.

Protein misfolding disorders such as Alzheimer disease, type 2 diabetes and inherited cataracts have been increasing their occurrence all around the world and, for most of them, there is no effective medical treatment yet [181]. Especially for the amyloidogenic diseases, misfolded intermediates, such as the molten globule and pre-molten globule states, play a key role in the amyloid formation in several human diseases [182,183,184,185]. Because these intermediate states are of transient nature and characterized by a large conformational heterogeneity, it is very challenging to study them in solution using standard techniques. Besides, so far in the literature, all the studies focusing on molten globule structures rely on artificial molten globules induced by extreme or mild denaturing conditions like low pH, the presence of chaotropic agents and/or mutations [186,187,188].

In this chapter, we investigated the effects on the GRASP structure, and consequently on a molten globule-like structure, induced by changes in the physicochemical properties of the medium without the need of any mild denaturing condition. These changes are intended to resemble particular conditions found in the cell. Using a multi-technique approach, which includes conventional and synchrotron radiation circular dichroism along with steady state and time-resolved fluorescence techniques, we showed that regions of CnGRASP structure undergo multiple disorder-to-order transitions under certain conditions. Changes in the dielectric constant seem to be the main regulator of GRASP structure, suggesting the protein undergoes a significant structural rearrangement when going from the bulk solvent to the membrane-bound state.

Materials and Methods

Protein expression and purification

Samples of CnGRASP were prepared as described previously [174, chapter 2] with a minor change in the working buffer, which now contains 25 mM Hepes/NaOH, 150 mM NaCl, 5 mM β- mercaptoethanol, 5% glycerol v/v, pH 7.4. For the molecular crowding assays, glycerol was removed prior to the experiments by extensive dialyses. The CnGRASP+Histag construct, i.e., a 6xHis-tag attached to the protein N-terminus, was obtained after a subcloning procedure to a pET28a plasmid (Novagen). The purification step was the same used for the pETSUMO GRASP but without the N-terminal tag removal procedure [174].

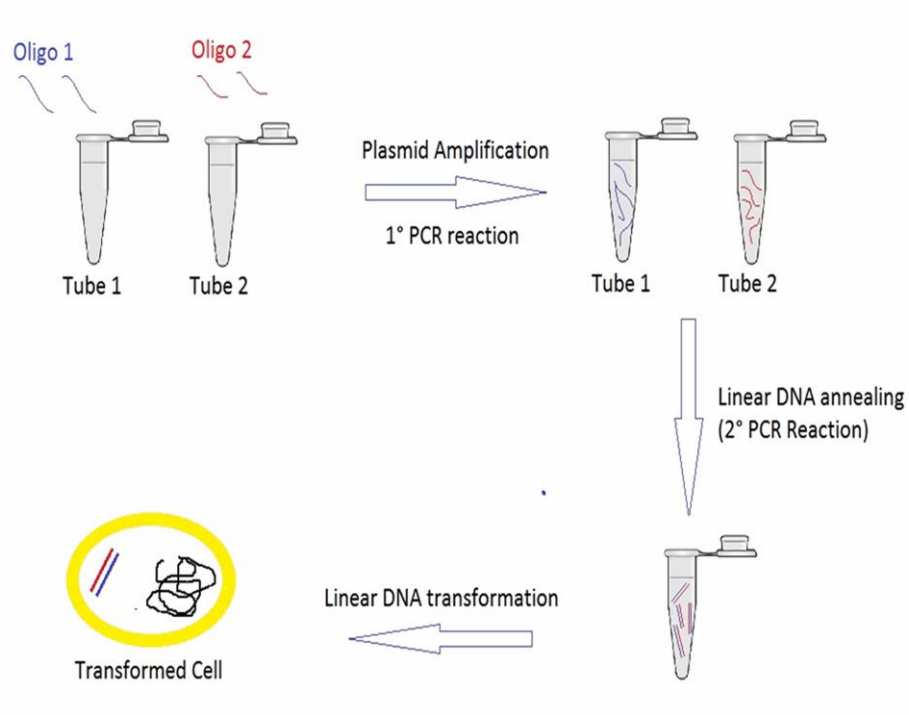


Figure 21: The 2-step method for mutant construction developed to enhance the rate of success of the quick-change protocol. The method was initially planned to increase the quick-change protocol efficiency by decreasing to zero the probability of oligonucleotide dimerization. The reaction is divided into two tubes and the standard quick-change protocol is performed but with only one oligonucleotide in each tube (Table 3). In the end of the first reaction, only one strand is amplified on each tube (Table 4). The reactions are then mixed and an "annealing step" protocol is performed, building the final linearized plasmid with the planned mutation (Table 5). The product is treated with Dpn1 and then transformed in *E.coli* cells.

The phosphomimetic mutants (T217D, S220D, S224D, S241D and S245D) were constructed using a variation of the Quick-change protocol [189] developed by us to decrease the negative impact of oligonucleotide dimers on the amplification process. The protocol consisted of two "quick-change reactions", where the traditional Quick-change conditions were used, but with only one of each oligonucleotide in each reaction tube (Table 3). After the single strand amplification was finished, both reactions were combined together and an annealing step was performed. Once this reaction was finished, the final DNA was obtained (Figure 21). The product was treated with Dpn1 to eliminate the parental non-mutated plasmid and then transformed in *E.coli* DH5α. The success of each mutation was checked by DNA sequencing. We observed a higher efficiency using this protocol to perform mutation than any other protocol tested so far. All mutants were expressed and purified using the same protocol for the native protein [174].

Table 3**:** First PCR reaction of the novel "two-steps" modified quick-change site directed mutagenesis

Reaction 1	**Reaction 2**
Template DNA (5-15 ng)	Template DNA (5-15 ng)
Oligonucleotide 1 (~10 µM)	Oligonucleotide 2 (~10 µM)
Phusion HF buffer 1X	Phusion HF buffer 1X
dNTPs (0.5 mM each)	dNTPs (0.5 mM each)
Phusion® High-Fidelity DNA Polymerase (0.5 U) (New England Biolabs)	Phusion® High-Fidelity DNA Polymerase (0.5 U) (New England Biolabs)
dH_2O/final volume 12.5 µL	dH_2O/final volume 12.5 µL

Table 4**:** First PCR reaction where each strand is amplified separately

Temperature (°C)	**Time**	**Step**
98	30 s	Initial denaturation
98	15 s	20
55	30 s	
72	4 min (30 s/kb)	
72	5 min	Final extension
4	∞	Hold

Table 5**:** Second PCR reaction, or "annealing step", where both strands are mixed and linked together

Temperature (°C)	**Time (s)**
95	30
90	30
80	30
70	60
65	60
60	60
55	60
50	60
40	60
4	∞

Liposome preparation

The phospholipids 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG), 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1,2-dioleoyl-sn-glycero-3-[(N-(5-amino-1-carboxypentyl)iminodiacetic acid)succinyl] (nickel salt) (18:1 DGS-NTA(Ni)) and 1,2-dipalmitoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (DPPG) were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL). Large unilamellar vesicles (LUVs) were prepared by extruding a multilamellar lipid suspensions in 10 mM sodium phosphate buffer, 10 mM NaCl, pH 7.4 (buffer B) through a 100 nm NucleoporeTrack-Etechmembrane filters (Whatman) using an Avanti Polar Lipids mini-extruder.

Synchrotron Radiation Circular Dichroism (SRCD)

SRCD experiments were performed in the UV-CD12 beamline of ANKA – KIT synchrotron (Karlsruhe, Germany). Protein concentration was ca. 18 μM in a 25 mM sodium phosphate, 50 mM NaCl, 1 mM β-mercaptoethanol, 10% glycerol, pH 7.4 buffer. Three scans were collected over the wavelength range from 280 to 180 nm with 0.5 nm bandwidth, 1.5 s dwell time in a 100 μm path length demountable quartz cell (Hellma Ltd, UK), at 20 °C; the baseline for each sample consisted of the buffer and all components present in the sample other than the protein. The results were processed using CDTools [121], including averaging, baseline subtraction and normalization to $\Delta\varepsilon$.

A dried film of CnGRASP was formed on the surface of a Suprasil quartz circular plate by the deposition of a protein solution (1 mg/mL, 20 μL) followed by evaporation of the solvent. SRCD spectra was collected over the wavelength range from 280 to 160 nm, with 0.5 nm step size, at 20 ºC, at four different rotation angles on the plate (increment by 90 deg). The final spectrum reported (mdeg units) is an average of all the scans, after the subtraction of the SRCD signal of the plate.

Circular Dichroism (CD)

Far-UV (190-260 nm) CD experiments were carried out in a Jasco J-815 CD spectrometer (JASCO Corporation, Japan) equipped with a Peltier temperature control and using a quartz cell with a path length of 1 mm. The experimental parameters were: scanning speed of 50 nm·min^{-1}, spectral bandwidth of 1 nm, response time of 0.5 s and temperature of 25 °C. The protein samples for the alcohol titration were based on a protein dilution in aqueous solution containing the appropriate amount of pure alcohol (methanol, ethanol or isopropanol), from a high concentrated protein stock solution (dilution of less than 5% for all experiments). The following dielectric constant values (ε) for the pure solvents were used [190]: 78.3 (water), 33.1 (methanol), 25.3 (ethanol), and 20.2 (isopropanol). The dielectric constant of the alcohol/water mixtures was estimated by using the relation $\varepsilon = f_{water}\varepsilon_{water} + f_{alcohol}\varepsilon_{alcohol}$ with the fractions (v/v) $f_{water} + f_{alcohol} = 1$. For all other experiments, the protein sample was previously diluted in buffer B to a final concentration of around 0.15 mg/mL. Stock solutions of Ficoll 70 (Sigma-Aldrich) were prepared in buffer B to a final concentration of 50%, w/v, i.e., 500 g/L. The solution was stirred by several hours until full solubilization. The software CDSSTR [133], with an appropriate database available at DICHROWEB [134], was used for spectral deconvolution whenever possible.

Steady-State Fluorescence Spectroscopy

Static fluorescence was monitored using a Hitachi F-7000 spectrofluorimeter equipped with polarized filters for anisotropy experiments and with a 150 W xenon arc lamp. The protein concentration was around 7 μM in buffer B or in the same water/alcohol mixture described for the CD experiments. The alcohol titration experiments monitoring the 1-anilino-8-naphthalenesulfonic acid (ANS – 250 μM) fluorescence were performed with an excitation wavelength of 360 nm and the emission spectrum monitored from 400 to 650 nm. Data analysis was performed based on the normalization of the wavelengths at maximum intensity (λ_{max}) of the

ANS emission fluorescence for each alcohol concentration used. The set of λ_{max} was normalized according to the function $\left(\frac{\lambda_{max,i}-\lambda_0}{\lambda_{MAX}-\lambda_0}\right)$, where λ_{MAX} and λ_0 are the highest and lowest λ_{max} of the dataset, respectively, and $\lambda_{max,i}$ is the measured λ_{max} for the i^{th} point of the alcohol titration. The normalized λ_{max} set was plotted as a function of the dielectric constant, calculated as described in the CD section above. For tryptophan fluorescence experiments, the selective tryptophan excitation wavelength was set at 295 nm and the emission spectrum was recorded from 310 to 450 nm. For the anisotropy measurements, tryptophan was selectively excited at 300 nm. Each mutant was directly diluted in a solution with the proper amount of urea, which was prepared with the working buffer solution and equilibrated overnight at 4 ^{o}C.

Time-Resolved Fluorescence Spectroscopy

The decays of the fluorescence intensity of ANS ([CnGRASP] = 0.3 mg/mL and [ANS] = 500 μM) were recorded in a picosecond laser spectrometer using the time-correlated single-photon counting technique. The excitation source was a mode-locked Ti:saphire laser (Tsunami 3950 pumped by Milennia X Spectra Physics) producing 1 ps FWHM pulses with 8.0 MHz pulse repetition rate (3980 Spectra Physics pulse picker). The laser wavelength was selected with second harmonic generators (LBO crystal, GWN- 23PL Spectra Physics) to yield 309 nm excitation pulses that were directed to an L-format Edinburgh FL900 spectrometer with a monochromator in the emission channel. Single photons were detected by a cooled Hamamatsu R3809U microchannel plate photomultiplier yielding an instrument response function of 100 ps. Data analyses were performed by using a nonlinear least-squares formalism inserted in a commercial package (Edinburgh Instruments). Data were analyzed according to the reduced-χ^2 values and the residual distribution. For tryptophan-monitored experiments, excitation was performed at 295 nm and the sample conditions were the same as those used in the static fluorescence experiments.

Results

CnGRASP is not significantly disturbed by variations in molecular crowding, metal ions or pH

It is well known that the concentration of macromolecules inside the cell can be as high as 400 mg/mL [191], which leads to pronounced effects related to excluded volume that can significantly influence protein structure and/or its function [192,193]. Some IDPs can go through multiple disorder-to-order transitions upon increasing molecular crowding, in a clearly functional-dependent manner [194,]. However, Szasz *et al* observed a tendency of disorder prevalence upon increasing molecular crowding for α-casein, MAP2c and $p21^{Cip1}$ by using Ficoll70 as a mimetic of the intracellular environment [194]. Therefore, the results so far in the literature do not point to any specific behavior of IDPs in crowded environments.

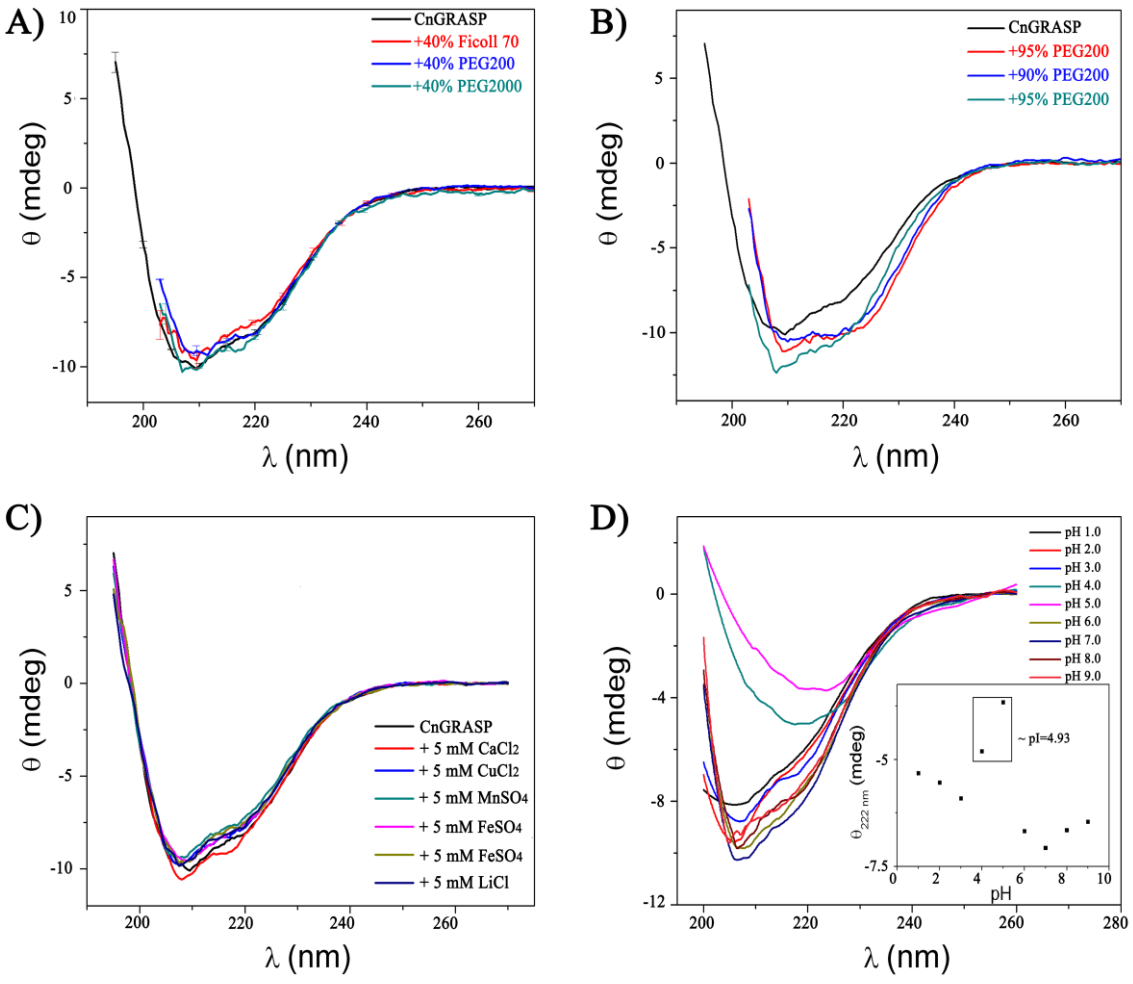


Figure 22: Probing the CnGRASP structural plasticity by CD and SRCD A) CD was used to monitor the impact of macromolecular crowding on the CnGRASP secondary structure using 40% Ficoll 70, PEG 200 and PEG 2000 as crowding mimetics. B) Effect of high PEG200 concentration in CnGRASP secondary structure monitored by CD. C) Different counter ions were chosen to investigate whether charge stabilization could induce any disorder-to-order transition on CnGRASP structure. D) The impact on CnGRASP secondary structure under different pH values was also investigated. The inset shows the ellipticity at 222 nm as a function of pH. All the CD spectra were acquired with the same protein concentration of 0.15 mg/mL.

Here, Ficoll70 and poly(ethylene glycol) with average molecular weights of 200 (PEG200) and 2000 (PEG2000) were used as molecular crowding agents to mimic the intracellular concentration of molecules. CD was used to monitor the structural changes of CnGRASP under crowded conditions. It was observed that CnGRASP is not affected by the increase in molecular crowding, even with mimetic concentrations reaching up to 40% (Figure 22A). Our results suggest that the secondary structure of CnGRASP is not significantly disturbed under the tested conditions, maintaining its original molten-globule features. This indicates that the existence of multiple disordered regions in CnGRASP is not an artefact of the ideal thermodynamic dilution from the purified sample, and that they are likely to be found even in the very crowded intracellular milieu.

However, significant structural rearrangements were observed for PEG200 concentrations higher than 40% v/v (Figure 22B). In all crystal structures of the GRASP domain (PDB ID: 3RLE, 4EDJ, 4KFW, 4KFV and 4REY) determined so far PEG was utilized in the crystallization assays, with cases (4DEJ, for instance) in which 30% PEG 4000, 10% PEG 20,000 and 10% glycerol were used. Our results cannot discriminate whether the structural rearrangements are due to changes in the entire GRASP structure or to alterations in the GRASP domain only. However, they do suggest that the overcrowded crystallization conditions could affect the equilibrium among coexisting conformations of GRASP, trapping the protein structure in one specific 3D arrangement, which does not necessarily represent the whole conformational ensemble of GRASP.

Different IDPs also experience structural rearrangements when metal ions are present in solution, a phenomenon that happens especially because of charge stabilization [91,195]. However, the presence of Fe^{+2}, Cu^{+2}, Mn^{+2}, Li^{+1} and Ca^{+2} in solution promotes no noticeable

influence on CnGRASP secondary structure content as monitored by CD (Figure 22C). Additionally, fluctuations in pH values have also been shown to modulate IDP structures [99,196]. Indeed, a pronounced decrease in the secondary structure content of CnGRASP is observed by increasing or decreasing pH values (Figure 22D), with the maximal structural content close to the pH used in the working buffer solution. However, since this behavior is not significantly different from those observed for other well-structured proteins [158,197], changes in ± 2 units from the neutral pH do not do not affect significantly the CnGRASP structure, even though this protein is membrane associated and the pH can be more acidic at the lipid/water interface [210]. The very broad CD spectra obtained at pHs 4 and 5 (Figure 22D) reflect protein aggregation, since these pH values are close to the theoretical pI of CnGRASP (~ 4.9).

CnGRASP does not interact directly with membrane models

Proteins belonging to the GRASP family are peripherally associated with membranes by a dual anchor [53,63,70,198]. Such association can occur via different mechanisms, which include post-translational protein lipidation (Cys-palmitoylation or G2-myristoylation), interaction with a membrane protein partner (GM130 for GRASP65 or Golgin45 for GRASP55) and even through an acetylated N-terminal amphipathic helix [50,70,198]. The exact mechanism used by CnGRASP to associate with membranes is still unknown, but this protein has a glycine at the position 2 of its sequence that allows it to be N-myristoylated. Here, we used large unilamellar vesicles comprised of pure POPC, POPG, DPPC, or DPPG phospholipids to test the hypothesis of a direct interaction between CnGRASP and membranes. Our SRCD results illustrated in Figure 23A show that the membrane models do not disturb the CnGRASP secondary structure. Besides, both steady-state and time-resolved fluorescence experiments did not indicate any change in the Trp microenvironment and in the fluorescence lifetime in the presence of the lipid vesicles used in this study (Figure 24B). Even though we found no evidence for direct protein-membrane interaction, we cannot exclude the possibility of a weak and superficial membrane association.

Therefore, N-myristoylation and possible interactions with a membrane protein partner might be the dual anchoring factors used for CnGRASP to be bound on the membrane surface.

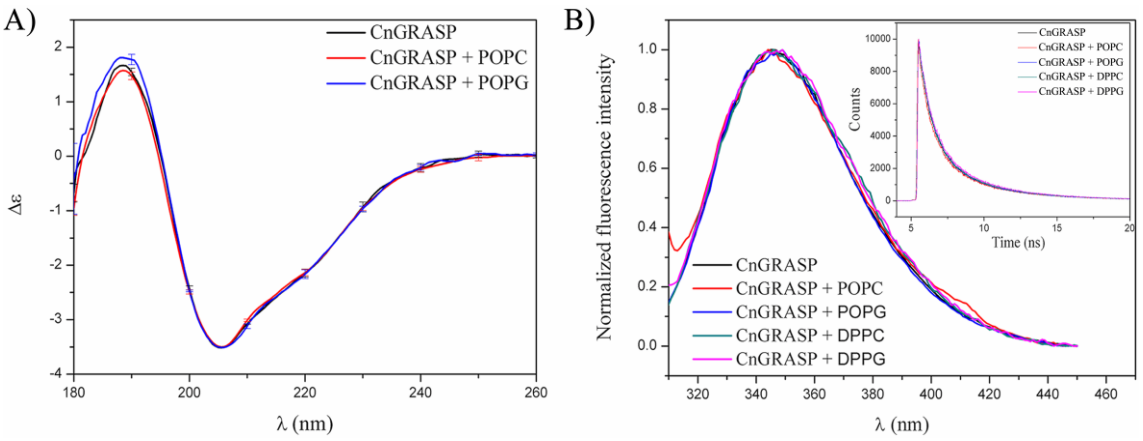


Figure 23: Probing the CnGRASP/membrane interaction using spectroscopic techniques The possible interaction of CnGRASP with model membranes of different compositions was monitored by (A) SRCD and (B) fluorescence techniques. A) SRCD spectra in the absence and in the presence of POPC and POPG LUVs. The results point to only minor (if any) changes in secondary structure content after incubation with LUVs. B) Steady-state fluorescence spectra of the CnGRASP Trp residues in the absence and in the presence of POPC, POPG, DPPC, and DPPG membranes. All experiments were performed at 25ºC, a temperature that DPPC and DPPG are both in the lamellar gel phase and POPC and POPG are in the liquid-crystalline phase. All the spectra were acquired with the same protein concentration of 0.15 mg/mL.

Phosphomimetic mutations do not induce disorder-to-order transitions of CnGRASP

Disorder predictors have shown that the intrinsic disorder of the GRASP SPR domain seems to be a common feature of the family [174] even though the amino acid sequence of such region is not conserved among members of the family [127,174]. It is well known that the SPR domain is the target of multiple kinases and that phosphorylation is responsible for breaking the homo-dimerization [68,69,63]. This property is explored by the cell in the processes of assembly and disassembly of the Golgi cisternaes. One convenient approach to study the effect of single site-directed phosphorylation is based on the use of phosphomimetic mutations [199,200]. In this approach, the specific target amino acid residue is mutated to a negatively charged amino acid (aspartic or glutamic acid), thus inducing a permanent phosphorylated-like protein. The putative regions in the SPR that could represent the possible targets for phosphorylation were predicted by using a specific server (http://www.cbs.dtu.dk/services/NetPhosYeast/, accessed in 2016) especially developed for prediction of phosphorylation sites in yeast [201]. *C. neoformans* is a yeast-like fungus, so we expect that this server will yield more accurate results than those

developed for general systems. The five S/T residues with the higher probability of being phosphorylated in the SPR domain (higher than 75%) were chosen and mutated to aspartic acid: T217D, S220D, S224D, S241D and S245D (Figure 24A). As shown in Figure 24B, the single mutants do not induce disorder-to-order transitions, since all substitutions led to no significant changes in the CD spectra compared to the CnGRASP one, with the only exception being S245D mutation that promotes a change in the CD line shape (Figure 24B). Furthermore, the aspartic acid substitutions do not promote any protein structural stabilization, since in all cases the urea-induced transition to the disorder state is slightly decreased (Figure 24C) compared to the native protein. The results suggest that phosphorylation induce neither disorder-to-order transitions nor changes in the overall protein structure stabilization, although it does play an important role in GRASP oligomerization [53,63].

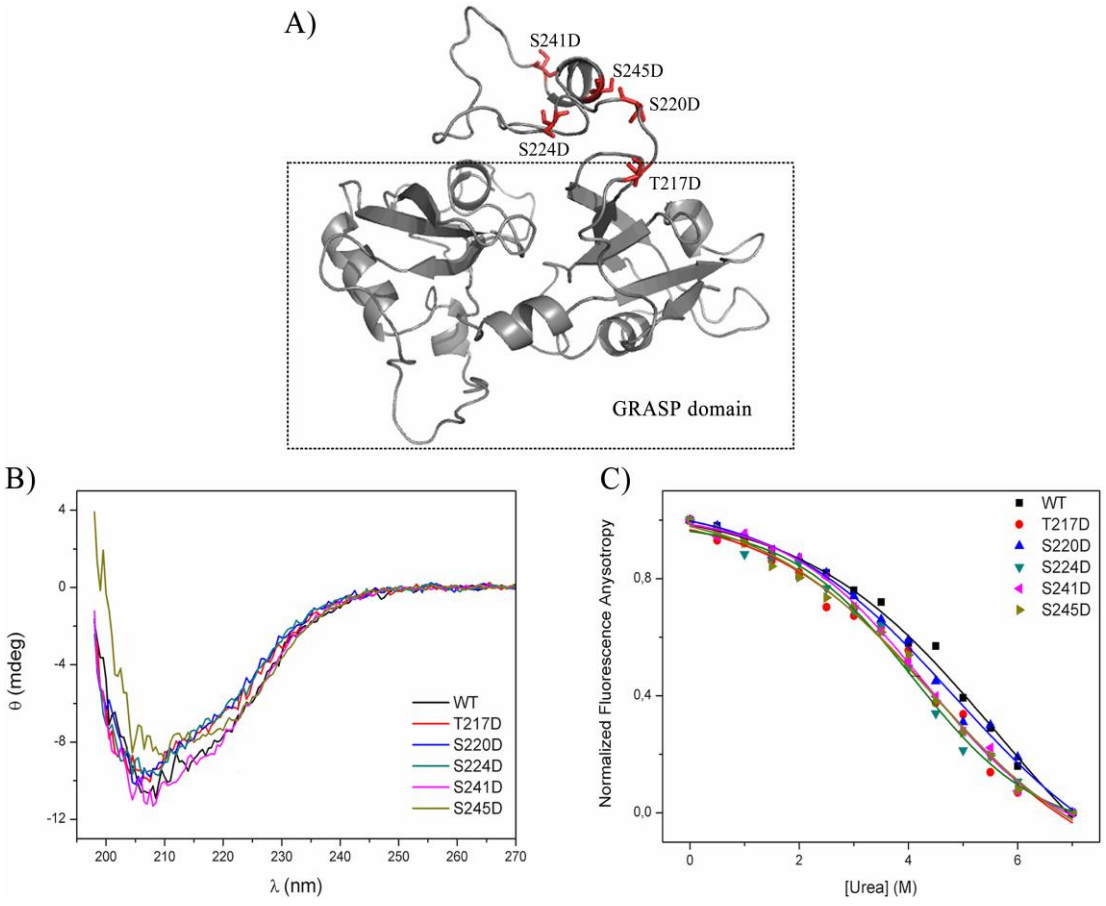


Figure 24: Monitoring the impact of site-directed phosphorylation on CnGRASP structure. Different phosphomimetic mutations in the SPR domain were constructed and checked for effects on CnGRASP secondary structure and overall stability. A) Molecular model of CnGRASP [21] showing the residues chosen for the phosphomimetic mutations (red). B) CD analyses of each mutant showing that none of them induce disorder-to-order transitions in CnGRASP. All the CD spectra were acquired with the same protein

concentration of 0.15 mg/mL. C) The changes in structural stability upon increasing urea concentration were monitored by fluorescence anisotropy and fitted using a sigmoidal curve for clarity.

Local dehydration induces multiple disorder-to-order transition in CnGRASP

Many of the protein/protein and protein/membrane interactions are entropically allowed due to the local dehydration induced on the protein surface to accommodate the structures of the interacting partners [202,203]. It has been shown that IDPs can go through a disorder-to-order transition after/before (induced fit/conformation selection model) interaction with a protein partner to strengthen the complex stability [204,205,206,207]. Moreover, it has been recently observed that both native globular proteins and IDPs have very different structural responses after total dehydration [208]. While in the former no significant structure rearrangement is observed, in the latter multiple disorder-to-order transitions are induced [208]. Therefore, we investigated whether protein dehydration could play a role in inducing a disorder-to-order transition in CnGRASP, a native molten globule-like protein that might behave as an intermediate between regular globular proteins and extended IDPs. As shown in Figure 25A, complete dehydration of CnGRASP leads to structural rearrangement, as inferred from its SRCD spectrum. Since an accurate normalization of the SRCD signal after water removal is not reliable, spectral changes can only be analysed by variations of the relative peak positions. The SRCD spectrum of CnGRASP after dehydration presents bands attributed to the characteristic α-helical transitions [130,131], but with shifts in the peak position from 205 to 208 nm (minimum) and from 190 to 195 nm (maximum). It is also noteworthy that the band centred around 222 nm is better defined after dehydration as compared to the CD spectrum of CnGRASP in solution (Figure 25A), leading to a decrease of almost 24% of the $\theta_{(208\ nm)}/\theta_{(222\ nm)}$ ratio (from ~1.7 to ~1.3). This result indicates that CnGRASP is folded in a richer α-helical structure upon dehydration. The same behavior is observed after the protein is treated with the fluorinated alcohol TFE and the sample monitored by SRCD (Figure 25B – Table 6). Spectral deconvolution yields an increase of the protein α-

helical content to 30% and 50% after treatment with 50% and 100% TFE, respectively, with a concomitant decrease of the unordered structure fraction from 39% to 30% and to 24%, respectively.

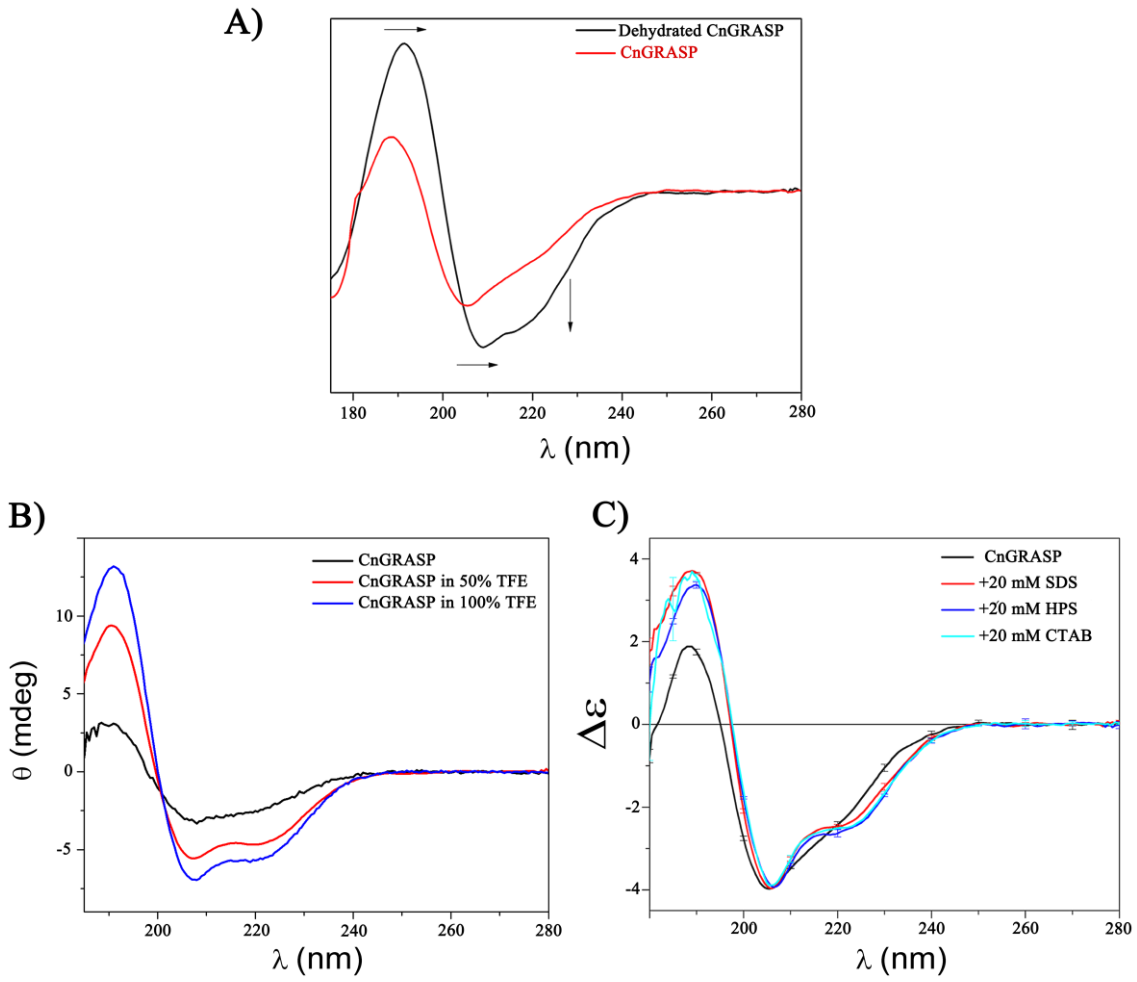


Figure 25: Different conditions were used to disturb protein structure and the corresponding changes were monitored by SRCD to check for disorder-to-order transition of CnGRASP. A) Effect of total water removal on CnGRASP structure. Black arrows indicate the shifts in the resonance peaks and the changes in signal lineshape after water removal. B) SRCD spectra of CnGRASP in the presence of different TFE concentrations. TFE is a well-known inducer of secondary structure in proteins, especially by making the intramolecular hydrogen bounds more favorable than those with the bulk water. C) Effect of micelles of different charge composition (positive CTAB, negative SDS and neutral HPS) on the CnGRASP SRCD spectra. All the CD spectra were acquired with the same protein concentration of 0.15 mg/mL.

Table 6**:** CD spectra deconvolution using the web-server DICHROWEB

	α-Helix	**β-Sheet**	**Turn**	**Unordered**	**RMSD**
Native Protein	0.21	0.25	0.14	0.39	0.01
+ HPS	0.28	0.23	0.15	0.37	0.014
+ CTAB	0.29	0.18	0.16	0.36	0.024
+ SDS	0.23	0.24	0.14	0.38	0.016
+ 50% TFE	0.45	0.13	0.13	0.3	0.007

+ 100% TFE	0.56	0.09	0.11	0.24	0.005

Detergent micelles were also used to probe the influence of local dehydration on the CnGRASP structure. The much less packed structure and much higher curvature of micelles, as compared to liposomes, could facilitate protein-micelle interaction and, thus, decrease the protein solvent accessible surface area, promoting local dehydration in the CnGRASP structure. Therefore, we used micelles of surfactants exhibiting different surface electric charges, such as the positively charged CTAB, the negatively charged SDS and the neutral HPS as potential molecular partners. Unlike liposomes, the presence of micelles led to remarkable changes in the CnGRASP SRCD spectra (Figure 25B), especially in the region close to 190 nm. All micelles increased the α-helix content of CnGRASP by approximately 8% and decreased the content of the remaining secondary structures (turn + β-sheet + other structures) by the same amount (Figure 25C). Interestingly, detergent micelles are able to induce conformational changes in CnGRASP in a charge-independent manner. These data suggest that micelles are able to promote disorder-to-order transitions on CnGRASP structure in a charge-independent manner supposedly by water removal from the protein surface upon micelle binding.

The results presented in this section suggest that CnGRASP might have multiple sites capable of undergoing disorder-to-order transitions after local dehydration, and this plasticity could potentially be used to regulate multiple functionalities *in vivo*.

The membrane electric field has a great impact on CnGRASP structure and dynamics

Structural studies focusing on the GRASP domain estimated its length as 6.5 nm [65,70] thus suggesting that its dimer could adopt an adequate conformation to fit within the estimated 11-nm space between the Golgi cisternae [209]. Because of this specific localization, it is expected that disturbances in the physicochemical parameters induced by biological membranes can have a great influence on GRASP structure and function. Membranes may act as denaturing

agents for some proteins in the cell [210,211,212] due mainly to changes in the dielectric constant (ε) and interfacial pH values, i.e, the "membrane field". Typically, the hydrophobic core of biological membranes, comprised by the lipid acyl chains, presents very low ε values (around 2-4), whereas the bulk solvent exhibits ε of ~ 80 [210,213]. Therefore, a dielectric gradient is observed at the membrane/water interface, which can be modelled by an exponentially increasing function from ε = 2-4 at the lipid bilayer to 80 at approximately 5-6 nm from the interface [213]. Because regular proteins are of this order of size, the ε gradient can have a great impact on protein structure close to membrane surface [210]. Variations in ε can partially denature well-structured proteins [214,215], increase the secondary structure content of many different IDPs [91,190,216,217] and induce, in some well-folded proteins, a transformation to a molten globule-like state in aqueous solution under mild denaturing conditions [210,218]. The latter phenomenon is hypothesized to be important for protein translocation across the membrane [219,220,221].

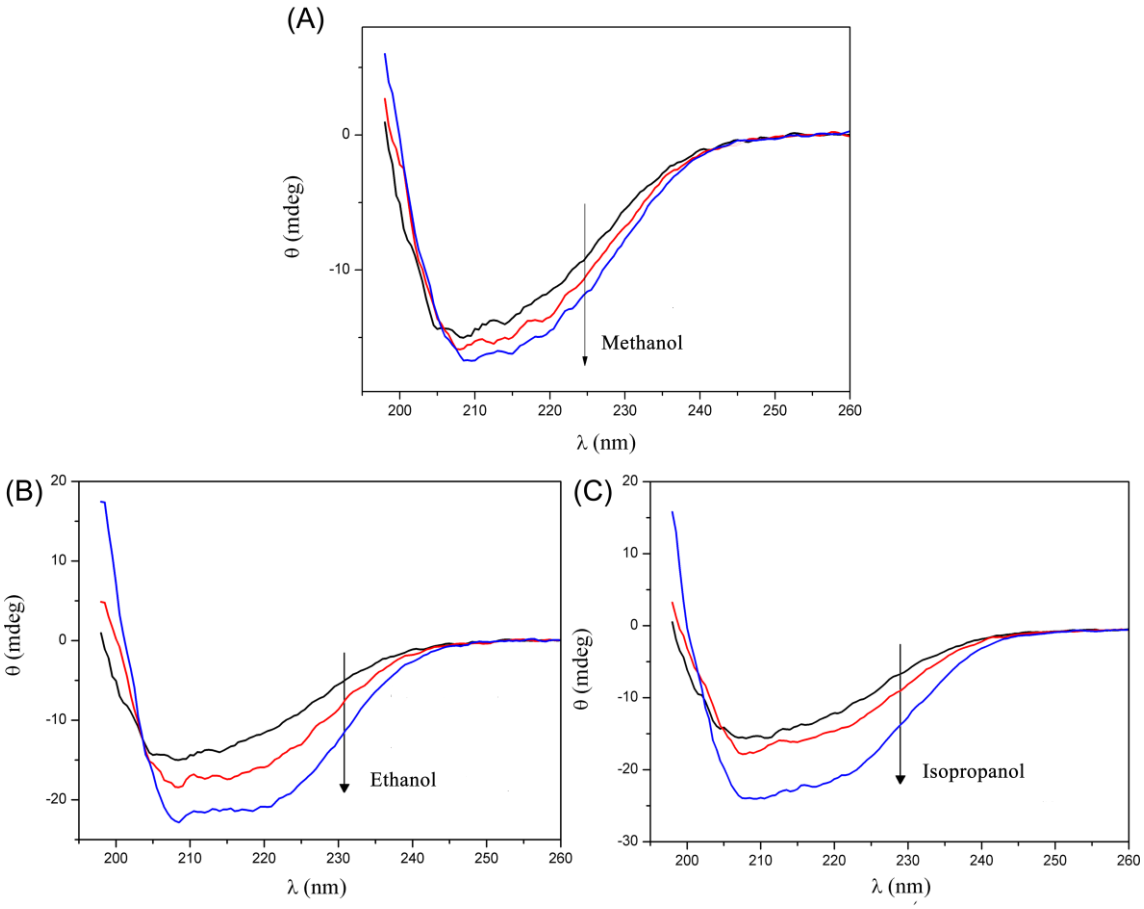


Figure 26: Disorder-to-order propensity on CnGRASP induced by alcohols. The effect of changes in the dielectric constant on CnGRASP structure was monitored by CD using different alcohols: A) methanol, B) ethanol and C) isopropanol. Only representative concentrations are shown to illustrate the changes in the CD spectra upon alcohol titration. All the CD spectra were acquired with the same protein concentration of 0.15 mg/mL.

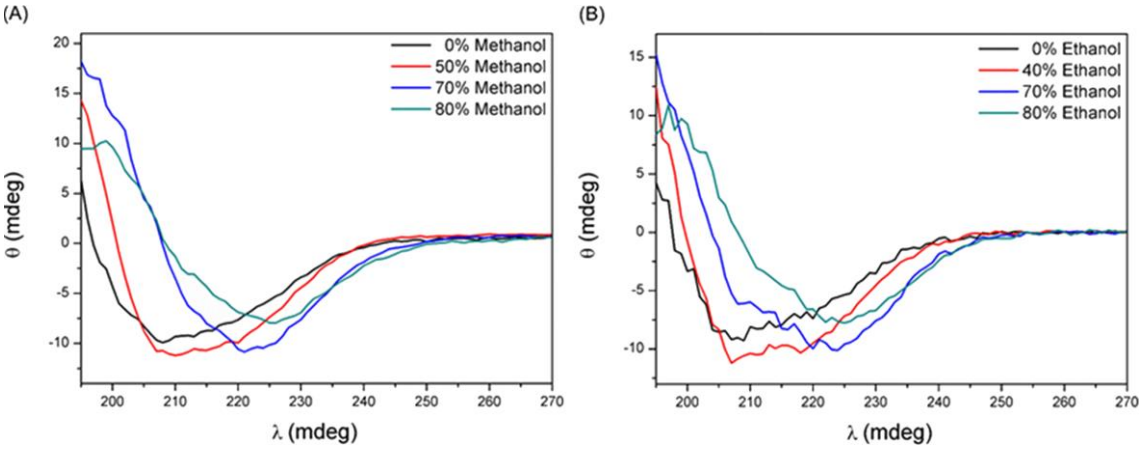


Figure 27: CnGRASP secondary structure response upon extreme alcohol concentrations. The effect of extreme alcohol concentrations on CnGRASP structure was monitored by CD using A) methanol, and B) ethanol. Only representative concentrations are shown to illustrate the changes in the CD spectra upon alcohol titration. All the CD spectra were acquired with the same protein concentration of 0.15 mg/mL.

Since we did not observe structural changes of CnGRASP in the presence of phospholipid vesicles presumably due to weak binding (see above), we decided to utilize alcohol/water mixtures to model the membrane field as ε changes in the medium. This is a widely used strategy in protein structural studies [212,222,223], especially in the case of IDPs [91,99]. In this work, we used methanol, ethanol and isopropanol as agents to change the dielectric constant of the medium and, thus, to separate the alterations promoted by ε in the protein structure from artefacts induced by nonspecific protein/alcohol interaction. The increase of alcohol concentration up to 60% v/v leads to significant changes in the CnGRASP CD spectra (Figure 26). Once these high concentrations are achieved, the protein starts to unfold upon further ethanol and methanol titration (Figure 27) and precipitates with isopropanol. Interestingly, the transition induced by the alcohol molecules, as monitored by the fraction of folded protein as a function of alcohol concentration, follows a two-state scheme characterized by half transition concentrations depending on the alcohol type (data not shown). However, the two-state transitions follow the same pattern if the native fraction is plotted as a function of the dielectric constant rather than the

alcohol concentration, i.e., the curves are alcohol independent (Figure 28A and Table 7). This is an indicative that the transition is induced by the decrease of the dielectric constant upon titration rather than by nonspecific alcohol/protein interaction. The half transition takes place for ε values between 55 and 60 (Figure 28A). Interestingly, if one considers that the dielectric constant at the water-membrane interface increases from 4-6, in the first water layer, to values around 78 at 25 nm away from the membrane interface [224], ε values in the order of 55-60 could be the average values experienced by CnGRASP, provided the protein is peripherally membrane associated.

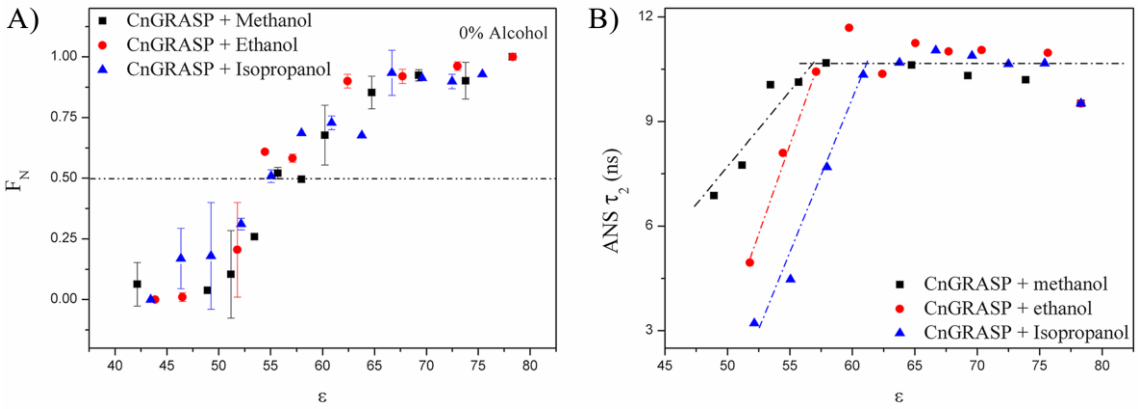


Figure 28: Changes in the dielectric constant affect CnGRASP structure and compaction. A) Fraction of the native conformation, calculated from the ellipticity at 222 nm, as a function of the dielectric constant. A two-state transition can be modeled with $\varepsilon_{1/2}$ ~ 55-60 irrespective of the alcohol type. B) Protein compaction as a function of the dielectric constant. The compaction was monitored by measuring the longer ANS fluorescence lifetime (τ_2), which is associated with the ANS molecules embedded into the protein structure.

Table 7: Estimated values of the dielectric constant in each alcohol/water mixture

Fraction	**Dielectric Constant**		
	Methanol	*Ethanol*	*Isopropanol*
0	78.3	78.3	78.3
10	73.78	73	72.49
20	69.26	67.7	66.68
30	64.74	62.4	60.87
40	60.22	57.1	55.06
45	57.96	54.45	52.16
50	55.7	51.8	49.25
55	53.44	49.15	46.35
60	51.18	46.5	43.44
65	48.92	43.86	40.54
70	46.66	41.2	37.63

In order to gain further insights into the protein conformation upon ε reduction, we performed time-resolved fluorescence experiments using ANS as an extrinsic probe. The fluorescence lifetime of ANS is not influenced by the presence of the alcohol [190,225]. The decay of the ANS fluorescence intensity in aqueous solution follows a monoexponencial behavior and occurs very rapidly (τ = 0.27 ns under our experimental conditions) [225]. In the presence of a protein, the decay is better described by a biexponencial function with two characteristic lifetimes [225], one directly related to the fraction of ANS molecules bound to the protein surface (τ_1), and the other ($\tau_2 > \tau_1$), related to protein-embedded ANS molecules [190,225]. Therefore, τ_2 gives us insights about protein compactness [190], which we associate with the fraction of CnGRASP that remains compact enough to allow ANS binding to its interior. For CnGRASP, τ_2 is somewhat constant for 55-60 < ε < 80 (Figure 28B), suggesting that some protein molecules maintain their compactness in that range, while also increasing their secondary structure content as observed by CD (for ε > 55-60). For ε < 55-60, the CnGRASP molecules able to accommodate ANS are disturbed in an alcohol-dependent manner (Figure 28B).

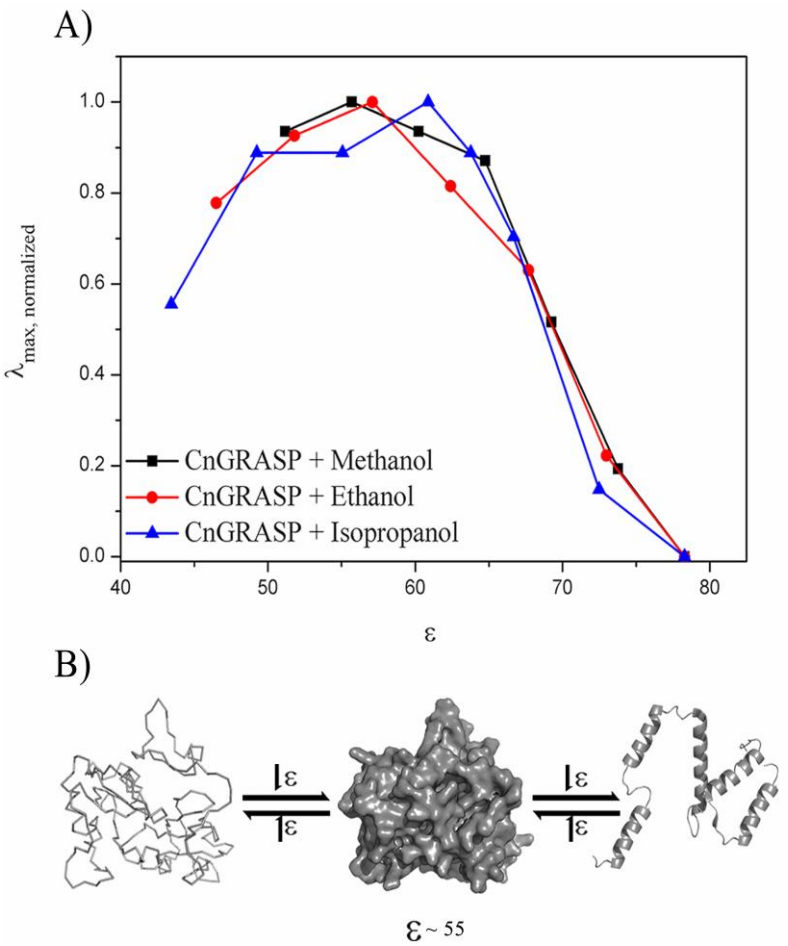


Figure 29: Changes in dielectric constant affect CnGRASP compaction. A) Plot of the normalized set of wavelength at maximum intensity ($\lambda_{max,norm}$) of the ANS emission fluorescence against the dielectric constant. Please refer to the materials and methods section for further information about the normalization process. B) A three-state model used to understand the possible intermediates observed in this study. The first structure is a molecular model of CnGRASP [23] in ribbon representing the native structure upon high values of ε (no alcohol, or far from the membrane field). Once the protein gets closer to the membrane and ε decreases to 55, an intermediate appears that seems to be more structured than the native state and to exhibit almost the same compaction. This intermediate state is illustrated by the molecular model of CnGRASP in a surface representation. Upon further ε reduction ($\varepsilon < 55$), a highly helical and extended conformation appears, which is a characteristic unfolded state of proteins in high alcohol concentrations. This conformation is figuratively illustrated here as an extended array of helices without any common pattern.

The general trend of keeping the protein compactness upon ε reduction is also observed in the ANS static fluorescence. A blueshift of the wavelength of maximum fluorescence intensity (λ_{max}) in the ANS fluorescence emission spectrum is an indicative of the increase in the hydrophobicity around the ANS vicinity, but multiple parameters may also influence [157]. The normalized set of λ_{max} values for ANS in the presence of CnGRASP, and upon alcohol titration, shows an umbrella-shaped curve with a plateau around 55 and a monotonic decrease for $\varepsilon < 55$

and ε > 55 (Figure 29A). Because the ANS static fluorescence is influenced by multiple factors such as the alcohol presence (unlike τ_2), changes in polarity and other parameters that vary in the titration course (especially the fractions of surface bound/embedded ANS), conclusions made from these assays are only possible by taking into account the previous experiments. The result suggests that a fraction of CnGRASP remains compact upon decreasing ε to values around 55-60. Furthermore, it is worth emphasizing that the induction of a molten globule state in well-folded proteins by alcohol titration has been shown to require a moderate acidic condition to be achieved [190,210,218]. Such a condition is paramount to stress the protein structure and make it more sensitive to the dielectric constant changes. However, our data suggest that the transition to a more structured state without loss of protein compaction takes place for CnGRASP without the need of an acidic condition or any other mild denaturant. These results indicate that CnGRASP is much more sensitive to changes in the medium polarity than well-folded proteins, which is likely the main factor that drives the disorder-to-order transitions in the unstructured regions of CnGRASP provided the protein is associated with membranes.

As discussed before, GRASPs in general have a double attachment to the membrane necessary for the correct orientation and trans-oligomerization. It has been proposed that the double attachment is achieved firstly by a lipidation at the N-terminus and secondly by a direct interaction with a Golgi integral membrane protein [63]. N-terminal lipidation anchors the protein to the membrane, therefore changes in the dielectric constant surrounding the protein are expected. Since information regarding the putative interaction of CnGRASP with membrane proteins is still elusive, we decided to mimic the N-terminal lipidation to strengthen our conclusions based on the dielectric gradient changes. To do so, we used a strategy previously adopted for a CnGRASP homologues, where a His-Tag was engineered at the protein's N-terminus to lock GRASP in model membranes doped with a Ni^{2+}-nitrilotriacetic acid-containing lipid [71]. Therefore, the specific interaction of the engineered His-Tag with 18:1 DGS-NTA(Ni) lipid embedded in lipid bilayers is responsible for the protein/membrane coupling. We used samples containing membranes with the same properties as those measured in Figure 23, i.e., large unilamellar vesicles of 100 nm, PC (phosphocholine) and PG (phospho-1'-rac-glycerol)

headgroups, and protein/lipid ratio of 1:100. The CD spectra of CnGRASP measured in those conditions show small differences in the presence of PC or PG vesicles, mainly a shift of the minimum from 205 to 208 nm, and changes in the spectrum intensity. A more dramatic change in the CD intensities happens in the presence of PC/PG LUVs (Figure 30). In this case, changes in intensity can be seen and may be related to an increase in the amount of ordered elements in the protein structure. This data suggest that the lipid bilayer can induce order in some CnGRASP disordered regions. Since this His-Tag-DGS-NTA(Ni) coupling strategy is a non-native way of attaching CnGRASP to the membrane, extra care should be taken when comparing the dielectric gradient felt by the protein in this case to the one found in a more native biological environment.

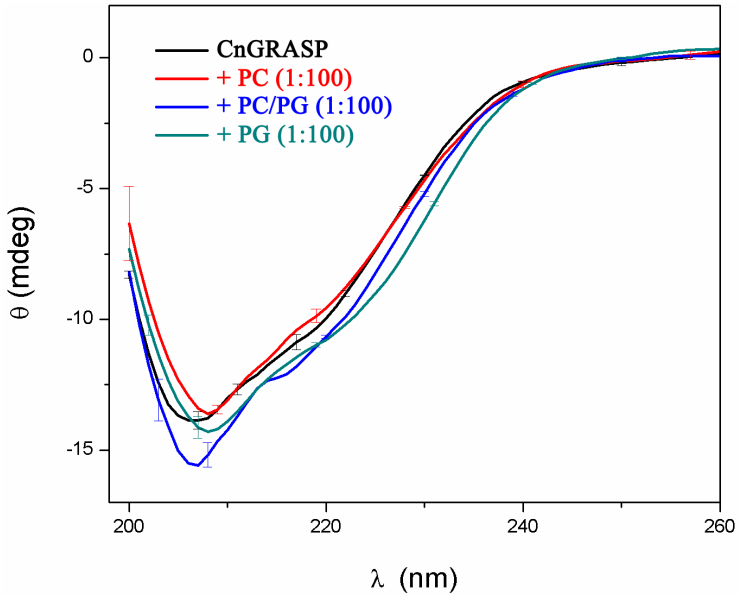


Figure 30: Attachment of CnGRASP in membrane model using an N-terminal anchoring. Interaction of His-tagged CnGRASP with LUVs of 100 nm with different headgroups and doped with 5 mol% of 18:1 DGS-NTA(Ni) (protein/lipid ratio of 1:100) monitored by CD.

Discussion

IDPs constitute a very broad class of macromolecules with high structural plasticity, especially because of its unique amino acid composition that is capable of driving atypical responses according to changes in their environment [156]. Even though many efforts have been made to understand this structural plasticity, most of the work has focused on extended IDPs [91,93,99]. It was previously suggested that CnGRASP is a molten globule protein, even in

physiological pH or in the absence of any mild denaturing condition [174]. This previous work focused on the structural behavior of CnGRASP in solution, but GRASPs form a high dynamic class of proteins with many of their functions being membrane associated [63,66,69,70]. Moreover, it is not clear yet how changes in the physicochemical parameters of the solution could affect CnGRASP structure. To address these issues, this work focuses on how GRASP responds to several mimetic conditions so as to unravel its structural plasticity.

One of the unique properties of the cell environment is its high macromolecule concentration that leads to pronounced excluded volume effects, nonspecific interactions and local dehydration [191]. Even though many IDPs present structural rearrangements and folding coupled with the increase of molecular crowding, the same behavior was not observed for CnGRASP. Our results indicate that the multiple disordered regions of CnGRASP are not an artifact of the diluted conditions used during the experiments, but rather a native property of this protein.

Under neutral and physiological pH, extended IDPs present a large number of uncompensated local charge that tends to destabilize the protein structure, which, along with the low content of hydrophobic amino acid residues, keep this class of proteins essentially unfolded [100]. Counter-ions can then cause a great impact on IDP structure by simple charge compensation. This is not the case for CnGRASP, where no changes in protein structure during metal ions titration were observed. Besides, CnGRASP shares the same characteristics observed for classical well-folded proteins upon pH changes. This result indicates that the unusual features observed for CnGRASP are not a consequence of uncompensated electrical charges along its amino acid sequence. We also observed that CnGRASP mutants intended to mimic site-directed phosphorylation in the SPR domain do not show disorder-to-order transitions, even though phosphorylation has proven important for breaking GRASP homo-dimerization and contributing to Golgi structural dynamics [52,53,57,60].

Thus far, most of the previous results do not differentiate CnGRASP from well structured/globular proteins, but this was expected based on the amino acid composition of CnGRASP, which is not similar to that observed in regular extended IDPs [174]. GRASPs are

highly regulated proteins with the important structural function of linking Golgi cisternae, and which are peripherally membrane associated in regular cell conditions or free in the cytosol during mitosis and apoptosis [63,198]. The microenvironment close to the membrane surface exhibits very different physicochemical properties compared to that in the bulk solution, especially regarding pH and the dielectric constant [210]. Regular well-folded proteins are not sensitive to small changes in the dielectric constant, and start to denature only when ε is abruptly decreased. Consequently, many proteins lose regular secondary structure upon membrane binding. This is the reason why the biological membrane is often called a denaturing agent of (soluble) proteins in the cell [210,211]. Interestingly, this effect is not observed for CnGRASP, which responds to changes in the dielectric constant in an opposite way: CnGRASP tends to increase the content of secondary structure upon ε reduction (in the range ~ 50-55 < ε < 78), as similarly observed for some extended IDPs and well-folded proteins under mild denaturing conditions. Furthermore, the degree of CnGRASP compactness does not seem to change for ε values in the range 50-55 < ε < 78. It is also worth mentioning that the "final" denatured states of CnGRASP induced by strong denaturants, such as urea and high temperature [174], or by mixtures of water and organic solvents [210,226] are different. While in the former the protein is simply unfolded, the latter induces a highly extended helical structure [226], which is formed for ε < 55 according to our fluorescence data. Therefore, upon ε reduction, CnGRASP seems to undergo two major conformational changes: 1) from the native state to a more structured intermediate state with similar compactness; and 2) from the compact, intermediate state to an extended, highly helical conformation (Figure 29B).

However, how does an ε value of 55 relate to the *in vivo* environment around any GRASP when it is attached to the membrane? Cherepanov *et al* proposed a model for the dielectric constant variation in the water/membrane interface at high ionic strength [213]. In their model, the dielectric function varies monotonically from the membrane surface, taken as z = 0, to the bulk solution as $\varepsilon_{(z)} = \varepsilon_{max}\left[1 + \left(\frac{\varepsilon_{max}}{\varepsilon_{min}} - 1\right)\exp\left(-\frac{z}{\lambda}\right)\right]^{-1}$. In this equation, ε_{min} = 4 represents the dielectric constant of the membrane layer, ε_{max} = 78 is the dielectric constant of the bulk

solution, and λ represents the characteristic length, which is taken as 1 nm for a solution with high ionic strength (0.1 M) [213]. With this equation, we estimate that, to experience an ε of 55, the protein should be around 3.8 nm from the membrane/water interface. Based on this result and assuming that each cisternae of the Golgi behaves independently, and that they are 11 nm apart from each other [209], we propose the model in Figure 31 for CnGRASP arrangement within the Golgi stacks. The CnGRASP dimension was estimated based on the molecular model previously proposed [174] and the orientation close to the membrane was assumed to be the same as the one determined for GRASP55 using neutron reflection [71]. We suggest that when positioned in between the Golgi stacks, CnGRASP likely experiences an ε value around 55 (Figure 31), and hence shows increased amount of ordered secondary structure due to the reduction of the dielectric constant, a propensity also observed upon dehydration, TFE treatment and micelle interaction.

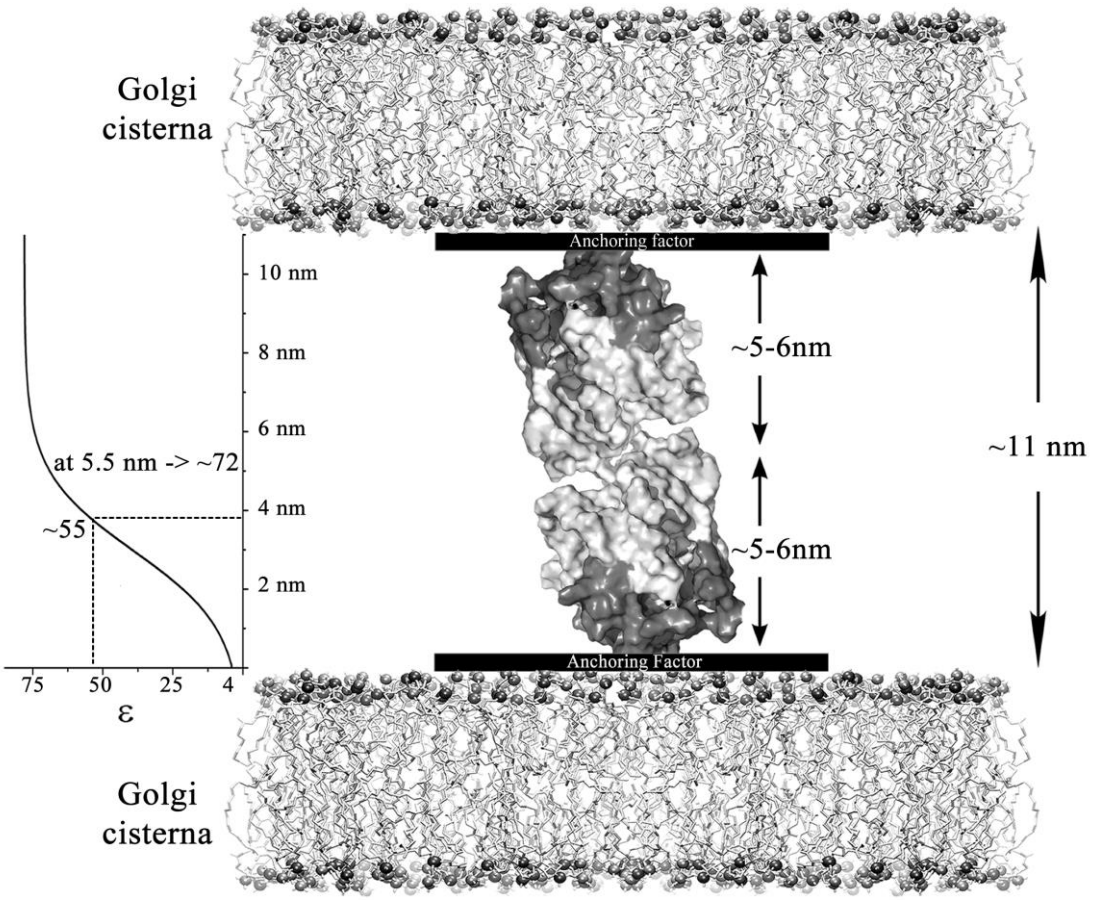


Figure 31: A model to account for the effect of the membrane field in GRASP organization. As a dimer and attached to the membrane, GRASPs feel the variation of the dielectric constant according to the graph in the figure. Because at half the distance between the cisternae (5.5 nm) ε is close to 72 (as compared to 78 from the bulk solvent), we considered that both membranes would induce an ε gradient independently of

each other so that only the variation for one membrane is presented. If we further consider the putative GRASP orientation in membranes [71] and the protein size estimated with molecular modelling [174], $\varepsilon \sim 55$ is a good approximation for the mean dielectric constant of the CnGRASP microenvironment.

The change in the dielectric constant can be thought as a consequence of the "membrane field" generated by the lipid bilayer on its surroundings. It is therefore tempting to speculate how the cell could use this structural plasticity for function. Once they are membrane associated, GRASPs are more structured, with less promiscuity during protein-protein interaction and with lower proteolysis sensitivity. It is well known that stable secondary structures are able to protect proteins from proteolytic activity [159]. Upon dimer dissociation due possibly to C-terminal phosphorylation or any other disturbance, each GRASP monomers separate apart (or go to the cytosol) and, therefore, a significant increase in the dielectric constant experienced by the protein is expected. When the membrane field turns weaker, the protein behaves more similarly to a disordered protein (since an increase of molecular crowding or phosphorylation do not induce ordering), with higher proteolysis sensitivity and higher promiscuity for protein-protein interaction. Our results provide new insights into GRASP dynamics, plasticity, and how the structural promiscuity of a molten globule protein can be manipulated in order to accomplish different cell functionalities.

Chapter 4 – Why would Nature give two PDZ domains to the “Golgi Reassembly and Stacking Protein"?

Abstract

The Golgi complex is an organelle responsible for receiving synthesized cargo from the endoplasmic reticulum for subsequent modifications, sorting and secretion. A family of proteins named Golgi Reassembly and Stacking Proteins (GRASP) is essential for the correct assembly and laterally tethering of the Golgi cisternae, a necessary structuration to keep this organelle working correctly. The GRASP structure is mainly composed of two regions: an N-terminal formed by two PDZ domains connected by a short loop and a non-conserved C-terminal region, rich in serine and proline residues. Even though there are now a few crystal structures solved for the N-terminal domain, it is surprising to notice that very limited information is available regarding the structural differences between the two PDZs in the N-terminal, which is the main functional region of this protein. It is hard to observe any particular difference between both PDZs by using the static crystal structures, but why would nature evolve to give two similar PDZs (and not just one) for GRASPs? We used solution NMR and synchrotron radiation circular dichroism to probe the structural differences between both PDZs. We clearly observed that both PDZs behave far from being equal, with the first PDZ being significantly more dynamic and unstable compared with PDZ2. We have also shown that there are significant more disordered regions in PDZ1, what gives new insights related to the PDZ1 promiscuity of interacting with much more protein partners than PDZ2

Introduction

Cryptococcosis is a fungal disease with a high mortality rate in immunosuppressed patients [107,108]. Although it is caused by the well-known pathogen *Cryptococcus neoformans*, the disease is very difficult to treat due to the lack of efficient drugs with low collateral effects [75]. The fungus has an external polysaccharide capsule, whose morphology is comparable to some bacterial systems [106], and it is responsible for fungal virulence. Glucuronoxylomannan (GXM), the main component of this external capsule, is synthesized in the intracellular milieu [80]. Data have shown that GXM can be secreted by conventional pathways, although blocking these pathways does not completely inhibit capsule synthesis [80,83]. Rodrigues and coworkers have shown that GXM can be secreted by non-conventional pathways with the poorly understood participation of a Golgi-related protein called Golgi Reassembly and Stacking Protein (GRASP) [84]. Fungal cells that had the GRASP gene silenced showed a reduction in the effective diameter of the capsule. [84]. This phenotype modification increased the rate of fungal encapsulation by macrophages and validated the GRASP protein as a possible target for medical research [84].

GRASP proteins were first reported as the main protein responsible for the Golgi architecture, tethering the cisternae laterally and stacking them, thus building the Golgi stacks [57]. This shape of this organelle is necessary for the pathways found in post-translational protein modification as well as for protein packing for delivery through the classical secretion pathway [112]. Although there is a clear role for GRASP in maintaining the Golgi shape, the absence of any clear homologue in the plant kingdom, whose Golgi shape is identical to other vertebrates, and the absence of the vertebrate Golgi shape in some lower eukaryotes, despite the presence of GRASP, suggest an additional role for GRASPs [64].

Even though there are now several reports showing the importance of GRASPs in eukaryotic cells, structural data is still incipient. Albeit many functions associated with GRASPs have been identified, their detailed molecular mechanisms are still unclear. The crystal structures of the N-terminal domain of human GRASP55 [64], *Rattus norvegicus* GRASP55 [65] and *Rattus*

norvegicus GRASP65 [65] have been only recently reported. The overall structure is composed of two structurally similar PDZ (named after the three first PDZ-proteins found: PSD-95, DLG and ZO-1) domains distinct from typical eukaryotic PDZ domains, which together compose the so-called GRASP domain (dGRASP) (Figure 32). This finding also suggests that a new sequence analysis criterion might reveal a host of unidentified eukaryotic PDZ domains [64]. A second region in the GRASP structure, usually larger than the dGRASP, consists of a non-evolutionary conserved domain that is rich in serine and proline (SPR domain) with regulatory function [63,127] and intrinsically disordered characteristics [174].

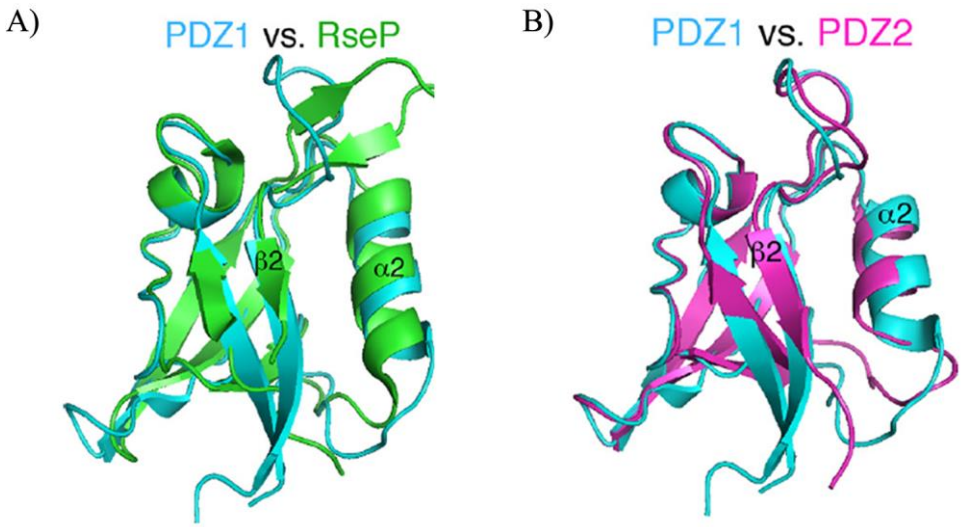


Figure 32: GRASP55 dGRASP domain crystal structure. A) Graphic representation of GRAS55 PDZ1 (blue) and RseP PDZ2 (green), a bacterial protein that also contains two circularly permuted tandem PDZ domains. GRASPs PDZs have a protozoan fold, distinct from typical eukaryotic PDZ domain. B) Structural data demonstrate that the dGRASP is composed of two structurally similar PDZ domains. Figure adapted from reference [64].

The structural data, together with several *in vivo* studies aiming at functional elucidation, suggest that GRASPs are involved in a large set of functions that likely includes a great number of interacting partners. However, the behavior of a full-length GRASP in solution still deserves more attention, since its biophysical properties in solution are still unknown.

In the previous chapters, we showed that CnGRASP presents features usually observed for a molten globule state, even in the absence of any mild denaturing conditions [174]. The molten globule state is considered as a collapsed IDP where the secondary structure content is similar to the native state but the hydrophobic core is not stable, so that the overall structure is collapsed but has a conformational fluctuation ("molten") in a characteristic timescale of μs-ms. The classification of CnGRASP as a member of the intrinsically disordered protein (IDP) family brings unexplored possibilities regarding the protein function *in vivo*, since IDPs are frequently observed as main components of the cellular signalling machinery [177], commonly functioning as central hubs [178], especially because of their unique structural plasticity [156]. Prion-forming sequences, which are especially enriched in asparagine, have been shown to promote molten globule-like structures, where amyloid-nucleating contacts can be made [179,180]. It is then becoming clear that the understanding of the structural behavior of molten globules in solution is of pharmacological interest.

Nevertheless, how the fungus, or any other GRASP-containing cell, could exploit this particular molten globule-feature remains elusive at this point. An appealing possibility is that different cell conditions could induce disorder-to-order transitions in GRASP allowing it to interact with multiple protein partners, thus taking different roles in cell processes. The presence of intrinsically disordered regions (IDRs) in GRASPs can also be used to understand how the protein is able to indiscriminately interact with a large number of proteins participating in secretion pathways. GRASP can directly interact with the C-terminal of the Transforming Growth Factor-a (TGF-a) and several members of the p24 family, adding an efficient retention of these proteins to the Golgi complex [56]. GRASP proteins are also directly related to the unconventional secretion of a large number of proteins, such as the soluble acyl-coenzyme A binding protein (ACBP) in *Dictyostelium* [37] and the starvation-induced secretion of ACBP in *Saccharomyces cerevisiae* and *Pichia pastoris* [39,40]. Furthermore, GRASP is involved in the Golgi bypass of αPS1 integrin during the stage 10B of *Drosophila* embryogenesis [43]. The GRASP-mediated unconventional secretion pathway can be used to rescue the cell surface

expression of the mutant ΔF508 in the cystic fibrosis transmembrane conductance regulator (CFTR) [44] the most commonly occurs mutation in the cystic fibrosis. However, this rescue only happens with the correct interaction between both proteins. GRASP is apparently also required for the exporti of non-protein molecules. In *C. neoformans*, deletion of the CnGRASP resulted in inefficient secretion of GXM, reduced capsule formation and attenuated virulence [84]. These data indicate that GRASP is also implicated in polysaccharide secretion via unconventional pathways and, consequently, fungal virulence. Interestingly, a large number of proteins seem to be associated with GRASPs by direct interaction with its PDZ1. If PDZ1 is the chosen PDZ to guide and to support the interactions, how do GRASPs identify PDZ1 if PDZ2 is so structurally similar? Moreover, it is still not clear how GRASPs act in the unconventional protein secretion, even though it is one of the main components of this very important cellular process, with many relevant biological and medical roles associated with it.

The main objective of the project that give rise to this chaptes was to unravel the differences between both PDZs of the GRASP domain in solution. Because the crystallographic structures failed to give a proper explanation of why both PDZs behave so differently *in vivo*, we decided to use high-field solution state nuclear magnetic resonance (NMR) and synchrotron radiation circular dichroism (SRCD) to probe differences in flexibility and stability. Our data show that both PDZs behave differently in solution, with most of the flexible and promiscuous regions located right in the binding pocket of PDZ1. PDZ2 has a much more well-behaved structure, suggesting that GRASPs are formed by two very asymmetric PDZs. Our data gives a convincing explanation of why PDZ1 is capable of mediating almost all the interaction with other protein partners while PDZ2 is restricted, and suggest a new possible role for GRASPs in unconventional protein secretion.

Methods

CnGRASP expression and purification

Unlabeled CnGRASP samples were expressed and purified as described previously [174]. ^{13}C and/or ^{15}N labeled CnGRASP and PDZs were produced as follow:

Preparation of stock solutions

Solution 1: $FeCl_2$ solution (100 ml)

8 ml concentrated HCl

5 g $FeCl_2 \cdot 4H_2O$ [198.8 g/mol]

184 mg $CaCl_2 \cdot 2H_2O$ [147.0 g/mol]

64 mg H_3BO_3 [61.83 g/mol]

40 mg $MnCl_2 \cdot 4H_2O$ [197.9 g/mol]

18 mg $CoCl_2 \cdot 6H_2O$ [237.9 g/mol]

4 mg $CuCl_2 \cdot 2H_2O$ [170.5 g/mol]

340 mg $ZnCl_2$ [136.3 g/mol]

605 mg $NaMoO_4 \cdot 2H_2O$ [241.95 g/mol]

Make up to 100 ml with H_2O. The solution is green and stirring for several hours is required before everything dissolves. Storage at room temperature (RT).

• Solution 2: Vitamin solution (1000 ml)

1.1 mg biotin stored at 4°C

1.1 mg folic acid* stored at RT

110 mg PABA (para-aminobenzoic acid) stored in 4°C

110 mg riboflavin* stored at RT

220 mg pantothenic acid stored in 4°C

220 mg pyridoxine HCl* stored at RT

220 mg thiamine HCl* stored at RT

220 mg niacinamide stored at RT

*Note that these vitamins are light sensitive.

Add 500 ml H_2O + 500 ml high purity ethanol, then filter sterilize. The solution will be bright yellow. Store at 4°C in a container designed for light-sensitive material or cover the container with aluminum foil.

• Solution 3: "SBM" solution

16.5 g KH_2PO_4

87.5 g K_2HPO_4

18.25 g NaCl

Preparation of 1 L of minimal media

- 940 ml H_2O

- 40 ml "SBM" solution

- 1 ml "S" solution

Mix and autoclave

- 2 ml "O" solution

- 1 ml vitamin solution

- 1 ml thiamine solution (1 mg/ml)

Mix and filter sterilize into the autoclaved growth medium of "SBM" and "S" solutions.

*Dissolve 1 g $^{15}NH_4Cl$ or $(^{15}NH_4)2SO_4$ into 5 ml H_2O

Filter sterilize into the growth medium

Unlabelled compounds can also be used

*Dissolve 2 - 4 g [U-^{13}C] glucose into 10 ml H_2O

Filter sterilize into the growth medium

Unlabelled compounds can also be used

The protocol used for expression and purification was the same as those used for the non-labeled samples. We observed that the rate of bacteria growth decreased by half in minimal media, compared with LB growth.

Circular Dichroism

CD measurements were performed using a Jasco J-815 Spectropolarimeter fitted with a computer-controlled Peltier temperature control unit. High-grade quartz cuvettes of 1 mm path length were used for all data collections. The temperature was settled to be 20℃. Further parameters were: data pitch 0.5 nm, D.I.T. of 1 sec, 1.00 nm bandwidth and a scanning speed of 50 nm/min. The sample preparation for the urea titration consisted of a protein dilution (less than 5% of the total final volume) directly in a solution containing 50 mM Sodium Phosphate, 10 mM NaCl and 1 mM β-MercaptoEthanol, pH 7.4 with the correspondent urea concentration.

Synchrotron Radiation Circular Dichroism (SRCD)

SRCD experiments were performed on the B23 Synchrotron Radiation CD beamline at the Diamond Light Source, Oxfordshire, UK. Protein concentration was in the range of 5-10 mg/mL and demountable quartz cells of 20 μm and 50 μm were mainly used. The parameters were the same as those used for the CD experiments.

High-Field Solution Nuclear Magnetic Resonance (NMR)

All NMR experiments were carried out using spectrometers operating at ^{1}H frequencies of 600 and 950 MHz. The spectrometers are equipped with Oxford Instruments magnets and home-built triple-resonance pulsed-field gradient probes. Data were processed using NMRPipe [227] or Bruker TopSpin 3.5 and spectra were analyzed using the CCPN software [228]. ^{1}H-^{15}N HSQC spectra of CnGRASP GRASP domain in 25 mM Hepes, 100 mM NaCl, 5 mM β-MercaptoEthanol, pH 7.0 (95% H_2O/5% D_2O) and in a urea concentration ranging from 0 to 9 M (1 M steps) were collected. The sample with 9 M urea was the one with the highest number of resonance peaks in the ^{1}H-^{15}N HSQC spectrum and, therefore, chosen for the further assignment step. CBCACONH, CBCANH, HNCO, HN(CA)CO, (H)CC(CO)NH, HBHA(CBCACO)NH, HSQC-TOCSY, HSQC-NOESY, HSQC-NOESY-HSQC were collected for an assignment that was posteriorly transferred for all the other HSQC spectra with lower urea concentration. The temperature was 20℃ for all the experiments and the protein concentration varied from 150 to 400 µM. NMR data collection of PDZ 1 and PDZ 2 followed the same protocols and buffers used with the dGRASP construction. Shigemi tubes were used for all data collections.

Results and discussions

15N-13C labelled-GRASP expression and purification

Labeled samples (with ^{15}N or $^{15}N/^{13}C$) were prepared using the protocol described in the materials and methods section. The protein yields using this protocol were comparable to the regular LB expression of non-labeled samples. We observed only a 2-fold decrease in the cell growth rate using minimum media. The purification protocol was the same used for the non-labeled samples. The success of the protein purification was checked after each batch using SDS-PAGE (Figure 33). The absence of other bands when an approximately 10 µg of protein is initially loaded in the gel was considered as the criterion of purity.

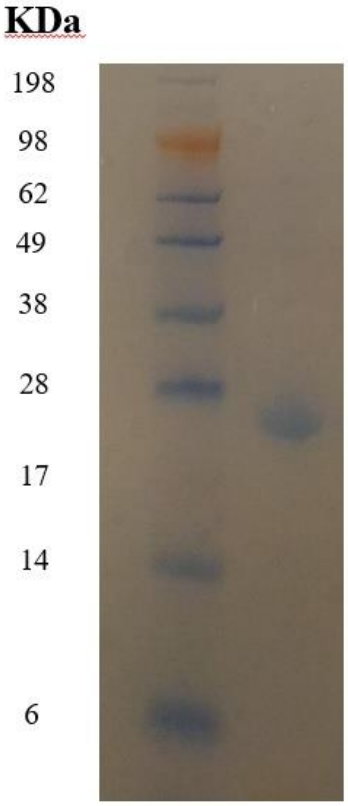


Figure 33: SDS-PAGE of final purified $^{15}N/^{13}C$ labeled dGRASP. Lane 1: molecular mass standards in kDa. Lane 2: purified dGRASP.

^{1}H-^{15}N HSQC analyses

As described in the introduction section, our main goal was to apply high-resolution solution NMR methods to describe the structural differences between both PDZs inside the GRASP domain of CnGRASP. NMR is still the most powerful technique available to describe structure and dynamics on an atomic resolution of stable (around 3-7 days at room temperature), low molecular mass proteins (<40 kDa in very favorable cases). We could have used three different strategies to obtain structural information for both PDZs: 1) Using the full-length protein; 2) working with our GRASP domain construction and 3) with the separated PDZ constructions. The idea of working with the full-length protein is the most obvious one at first sight. However, the full-length CnGRASP is not soluble at high concentrations and is highly sensitive to proteolysis at the C-terminus. The first limitation does not allow us to get 3D NMR data of reasonable quality and the second might generate some inhomogeneous peaks due to

different C-terminus. We then decided to start with our GRASP domain construction because it is considerably more soluble than the full-length protein and only carries a small portion of the SPR domain, without any detectable proteolysis during protein expression and purification. We could also, of course, have expressed the PDZs separately since the beginning but it was previously observed that PDZs in tandem might assist each other during folding [229,230]. For example, for the PDZ12 tandem of the GRIP1 PDZ1–3 cassette, the folding of PDZ1 strictly depends on the covalent attachment to PDZ2 [229]. In addition, PDZ5 of GRIP is completely unstructured in solution but when PDZ4 and PDZ5 are covalently connected, both PDZ domains become well folded and stable [230]. Since the behavior of each PDZ of CnGRASP when free in solution was not known, it seemed the logical choice to start the work with the GRASP domain.

CnGRASP GRASP domain (called dGRASP from now on) was first expressed and purified in minimal media supplemented with $^{15}NH_4Cl$ and a ^{1}H-^{15}N HSQC spectrum was collected. All the ^{15}N-edited experiments (with the exception of HSQC-TOCSY and HSQC-NOESY) were collected at both 600 MHz and 950 MHz but only the data from 950 MHz will be presented here. The dGRASP HSQC spectrum is very surprising since a total of 200 resonance points would be expected from the primary sequence but ~20 picks are observed (Figure 34). This behavior was observed for data collected at pH 6.5, 7 and 7.4, indicating that the rapid exchange of amine protons cannot explain this phenomenon alone (data not shown). CnGRASP is not stable on pH 6 and bellow, as shown in previous reports. We decided to work with the solution at pH 7.0 because dGRASP is not very stable at pH 6.5 with precipitations clearly appearing after a couple of hours. We also decided to keep the temperature below 25^{o}C for similar reasons. Another special feature observed in the dGRASP HSQC spectrum is that, for the small number of observable peaks, the proton dispersion is very low. From the most downfield to the most upfield peaks, a separation of less than 0.7 ppm is observed in the proton dimension, a strong indication that these peaks might be located inside disordered, highly dynamic regions.

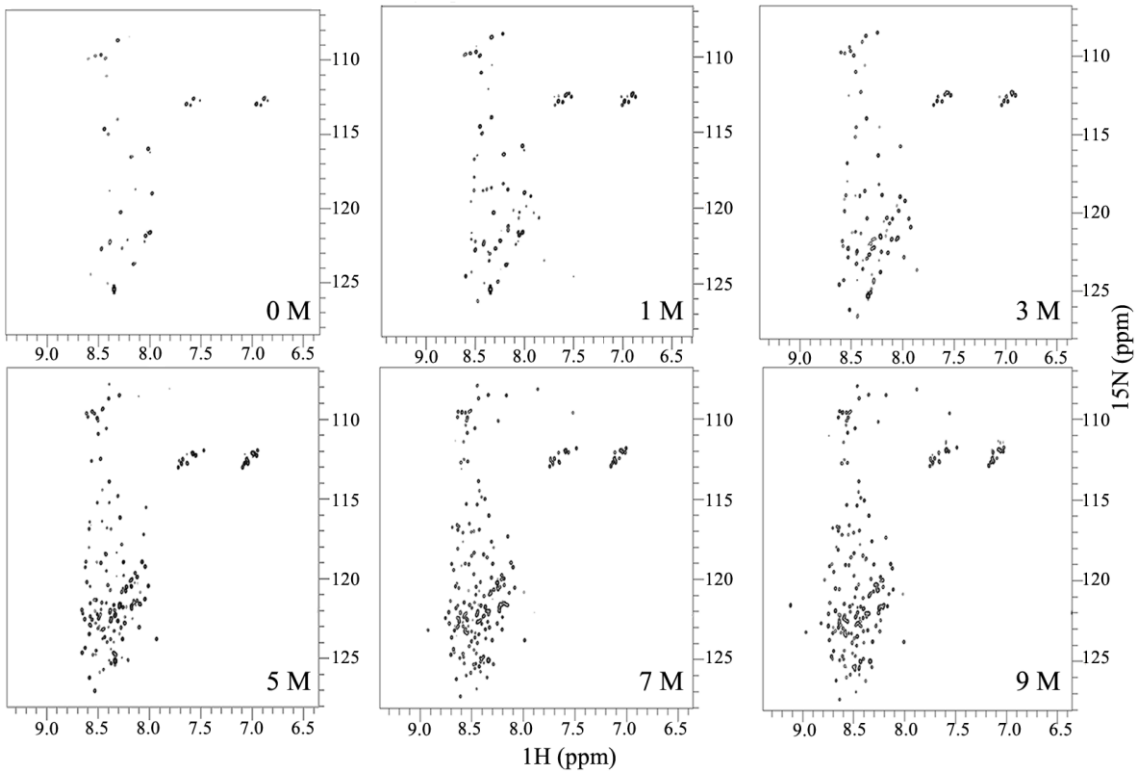


Figure 34: dGRASP urea titration monitored by ^{1}H-^{15}N HSQC at 950 MHz. The sample was titrated from 0 to 9 M in a 1 M step but just some representative spectra are presented.

We suggested in a previously published work that the full-length CnGRASP has a number of structural features that clearly resemble those observed for molten globule states of proteins [174]. If the dGRASP construction still shares the same property, we could expect that the structure is fluctuating in a µs-ms time scale, which is considered too slow for the HSQC time resolution. This explains why there are very few peaks in the dGRASP HSQC spectrum and suggests that the ones still visible come from dynamic/disordered regions, also explaining the poor proton dispersion.

If dGRASP is conformationally fluctuating in a time scale that is considered too slow on the NMR timescale, the peaks are "invisible" only because they are too broad to be detected. A good strategy to make these peaks "visible" again is to disrupt the structure so that we increase the local dynamics or decrease the number of "conformational states" that the protein visits. We decided to use urea because we observed previously that the temperature can promote aggregation. This

strategy was previously used to study the molten globule states of alfa lactoglobulin and p53 [162,163], however we did not induce a molten globule state in CnGRASP using any mild denaturating condition. This is a natural property of this protein and indicates that the number of tertiary contacts along its structure are inefficient to keep a stable and less dynamic folding.

Urea was titrated from 0 to 9 M in a 1 M steps and the ^{1}H-^{15}N HSQC spectra are showed in figure 34. It is clear that the resonance peaks start to appear during the titration indicating that that corresponding region is disrupted in that particular condition. It is interesting to observe that we have two moderately cooperative appearances of peaks: one around 3 M and another one at ~6 M, suggesting that there are two different regions of the protein that unfolds independently, at least. Some peaks only become visible at 9 M urea. Even though the spectra were collected at 950 MHz, the central region is very crowded, making it impossible to count the total number of resonances unambiguously. For this reason, we decided to start the resonance assignment based on the 9 M urea concentration and to later transfer the assignment to all the other nine HSQC spectra. It is also worth noticing that we optimized the HSQC spectrum of dGRASP by choosing a sweep width of 11.4 ppm in the proton dimension (centered in the water position) and 22.5 ppm in the nitrogen dimension (centered ~118 ppm). With these conditions, we naturally folded the tryptophan side chain resonances, a price that can be considered reasonable if we consider that the protein is unfolded (therefore, the tryptophan side chains resonances are not very useful) and we optimized the spectrum in the central region. All the subsequent 3D experiments were based on these HQSC conditions.

NMR assignment

Our initial assignment strategy consisted of using only ^{15}N-edited experiments, including HSQC-NOESY, HSQC-NOESY-HSQC, and HSQC-TOCSY, the so-called sequential assignment [231,232]. This procedure is very challenging since it relies only on the amide resonances and their NOEs, and is consequently limited to proteins smaller than 15 kDa. This method was previously used successfully for a molten globule structure unfolded by urea [233]

and we therefore decided to use this method, before going to higher field and to the double labelling of the protein (which is considerably more expensive).

Our assignment strategy can be divided in 2 steps. Firstly, the spin system identification and, for this, we used the HSQC-TOCSY (Figure 35A). The HSQC-TOCSY provides through-bond spin system information that can be even more reliable when working with an unfolded protein, especially using long TOCSY mixing times (50 ms in our case). The method is particularly useful for identifying alanine, valine, serine, and threonine but in all other cases, we cannot make a definitive identification (Figure 35B). Since the protein is also unfolded, we can roughly identify some amino acids, with similar properties, based only on the resonance position in the ^{1}H-^{15}N HSQC. This strategy works reasonably well to identify glycine, serine+threonine, and alanine+leucine.

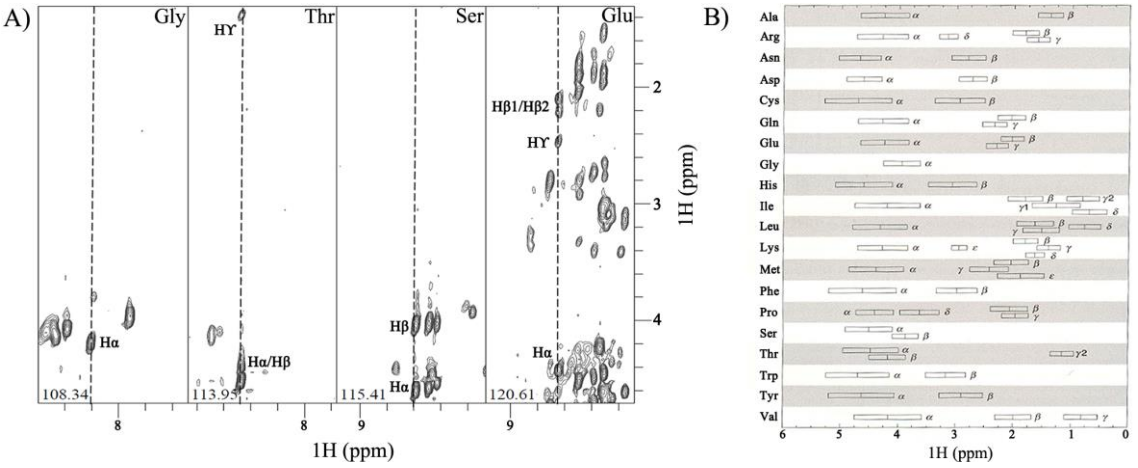
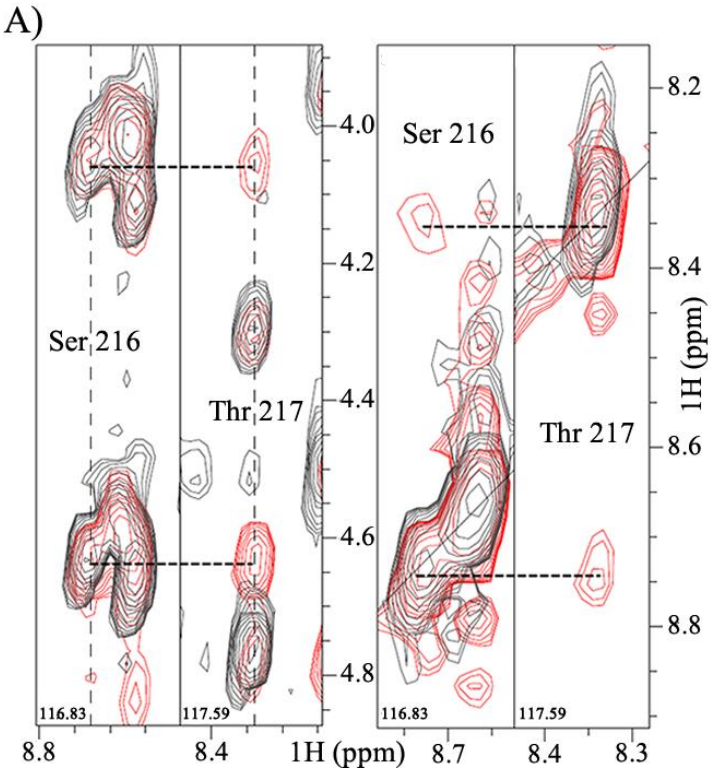


Figure 35: The spin system identification strategy for the ^{15}N-labelled samples. A) A set of ^{1}H-^{15}N HSQC-TOCSY planes is presented for some representative amino acids (Gly, Thr, Ser, and Glu). B) The set of proton resonances expected for each amino acid based in unfolded samples. The figure in B is adapted from [234].

The second step is to determine the sequential connectivity. For this, we started using the HSQC-NOESY coupled with the HSQC-TOCSY (Figure 36A). Wüthrich and coworkers showed that, for all sterically allowed values of φ, ψ and χ_1, at least one of the distances between HN, Hα,

and Hβ of residue i-1 and HN of residue i is short enough to give rise to an observable NOE effect [234,235]. Therefore, we identify intraresidue NOEs in the NOESY–HSQC strip plot by direct comparison with the TOCSY–HSQC strips; any peaks present in both the NOESY and TOCSY spectra are likely to be intraresidue connectivities. The remaining peaks come from the inter-residue connectivity identified by analyses of strips in the HSQC-TOCSY spectra (Figure 36A). Because the method relies on the through-space distance-dependent NOEs only, the analyze is complicated by NOEs to residues which are separated in the primary sequence but close enough in the tertiary fold. Of course, this imposes a severe limitation for folded protein but we could expect that this problem would be less important for an unfolded protein. The poor proton dispersion in our proton spectra, with the natural consequence of almost all the resonances of each particular amino acid falling in the same resonance regions, impose a significant challenge for the dGRASP system. We found that for our unfolded protein, the HN_{i-1}-HN_i NOEs are the most useful NOEs during the sequence assignment and we therefore also collected a HSQC-NOESY-HSQC spectrum (Figure 36B).



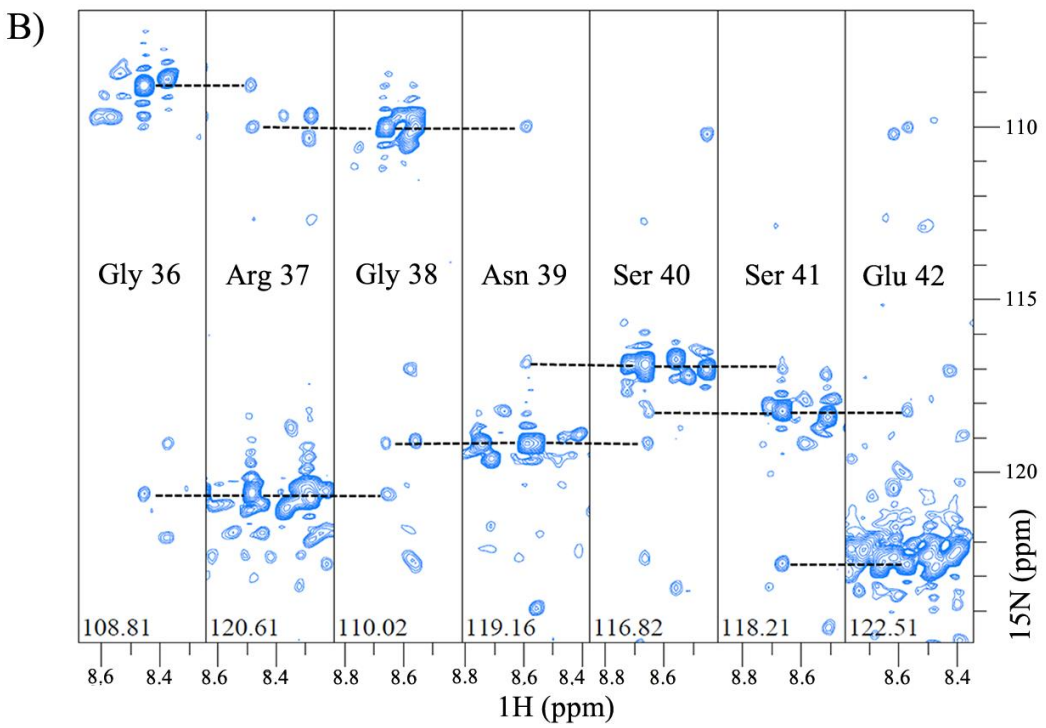


Figure 36: Sequential assignment using 15N-labelled dGRASP in 9M urea. A) HSQC-TOCSY and HSQC-NOESY used to probe the coupling between the Ser216 and Thr 217. B) HQSC-NOESY-HSQC plates showing coupling between amino acids 36 to 42.

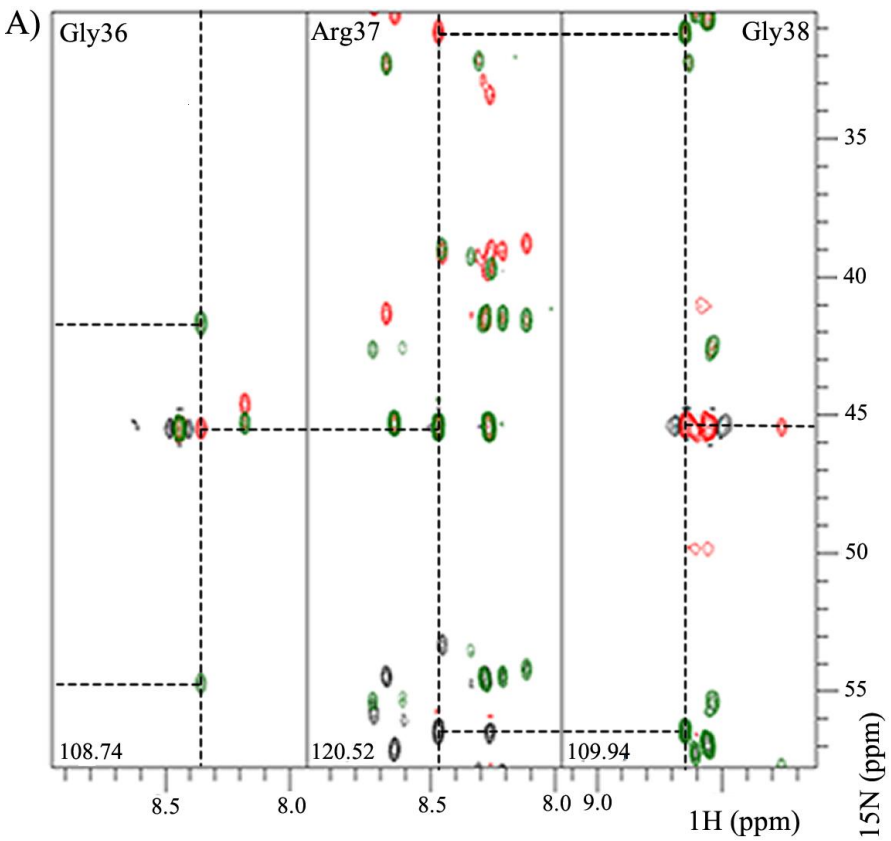
This data is very useful in identifying coupling between HNs that differ significantly in their ^{15}N resonances. Because of all the inherent limitation of NOE-based methods, we decided to be confident of correctly assigning each coupling unambiguously by identify at least three NOEs connectivities. Using these criteria, it was possible to assign approximately 40% of the HN resonances.

With the assigned resonances, we concluded that it would be interesting to complete the assignment by using the ^{1}H-^{15}N-^{13}C experiments at higher magnetic field (950 MHz, which is in 2017 still the highest magnetic field solution-NMR machine in the United Kingdom).

The sequential assignment methodologies revolutionized the field of protein NMR back in the 1980s, a time where all the analyses were based on proton experiments that greatly limits the complexity of applicable systems. As discussed previously, to rely only on through-space NOE effects is the most problematic limitation of the sequential method. In the early 90s,

techniques were developed for double labeling with ^{13}C and ^{15}N, which led to the development of an alternative assignment approach based solely on through-bond scalar couplings, greatly increasing the size of proteins that can be studied by NMR. Nowadays, the use of triple resonance techniques along with perdeuteration and transverse relaxation optimized spectroscopy (TROSY) methods has increased further the molecular mass limit for full protein assignment to beyond 40 kDa [236].

The dGRASP+9 M urea resonances assignments were based on two pairs of ^{1}H-^{15}N-^{13}C 3D experiments for sequential linking: CBCA(CO)NH + CBCANH (Figure 37A) and HN(CA)CO + HNCO (Figure 37B) (along with HSQC-NOESY-HSQC in Figure 36). In addition to HSQC-TOCSY, we used (H)CC(CO)NH and HBHA(CBCACO)NH (data not shown) for spin system identification. Because the HBHA(CBCACO)NH gives the H_β and H_α resonances of the previous residue, it can also be used with the HSQC-TOCSY for sequential connection.



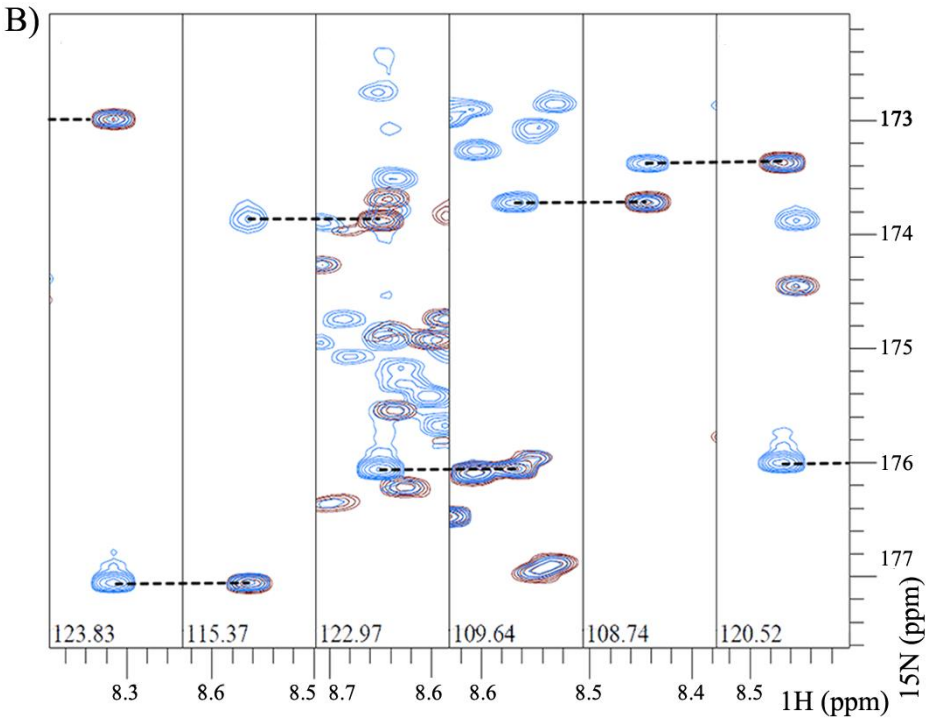


Figure 37: The Triple Resonance Assignment Method applied to 1H-15N-13C dGRASP in 9 M urea at 20℃ using a 950 MHz. A) CBCANH (green) and CBCA(CO)NH (black and red for positive and negative resonance pics, respectively) pairs. B) HNCO (purple) and HN(CA)CO (blue) pairs

The main reference used in the initial learning about assignments of protein resonances was the book chapter written by Christina Redfield (Biochemistry-University of Oxford) in "Protein NMR: Modern Techniques and Biomedical Applications" [234]. The website by Victoria A. Higman (http://www.protein-nmr.org.uk/ accessed in 2016/2017) includes some practical advices and tutorial for NMR data analyses using the CCPN routines. In short lines, because the resonance positions agree very well with those predicted for random coils, the most useful pair was the CBCA(CO)NH+CBCANH (Figure 37A) experiments that are not of the 'out-and-back' variety (starting in HN and coming back to the HN in a series of INEPT sequences). Instead, these 'straight-through' sequences start with a $^{1}H–^{13}C$ HSQC element in which the ^{13}C chemical shift is encoded in the direct dimension. Spin system information can also be estimated from the chemical shifts of Cα and Cβ, and is especially useful for identifying glycines (because the Cα signal is the only one that is 180 degrees out of phase), alanine, serine/threonine and valine. Besides, because it is not based on the ^{1}H-^{15}N HSQC of residue i-1, this residue can be identified

even if it is a proline. Using this pair of experiments, long stretches of unambiguously assigned residues were deduced (Figure 37A).

Unfortunately, a large number of overlaps were still present in the Cα and Cβ resonances. However, because of the unique relaxation properties of unfolded proteins, the HNCO+HN(CA)CO par was of high quality, allowing to solve most of the ambiguities from the CBCA analyses (Figure 37B). Some residues could not be unambiguously assigned using only the ^{13}C resonances and the ambiguity could only be solved using the HBHA(CBCACO)NH + HSQC-TOCSY pair (data not shown). We also collected the HSQC-NOESY-HSQC again at 950 MHz and this experiment was essential as a guide through the multiple strips. In order to make the assignment process very reliable, the residues were assigned only when all the pairs indicated that there was a couple.

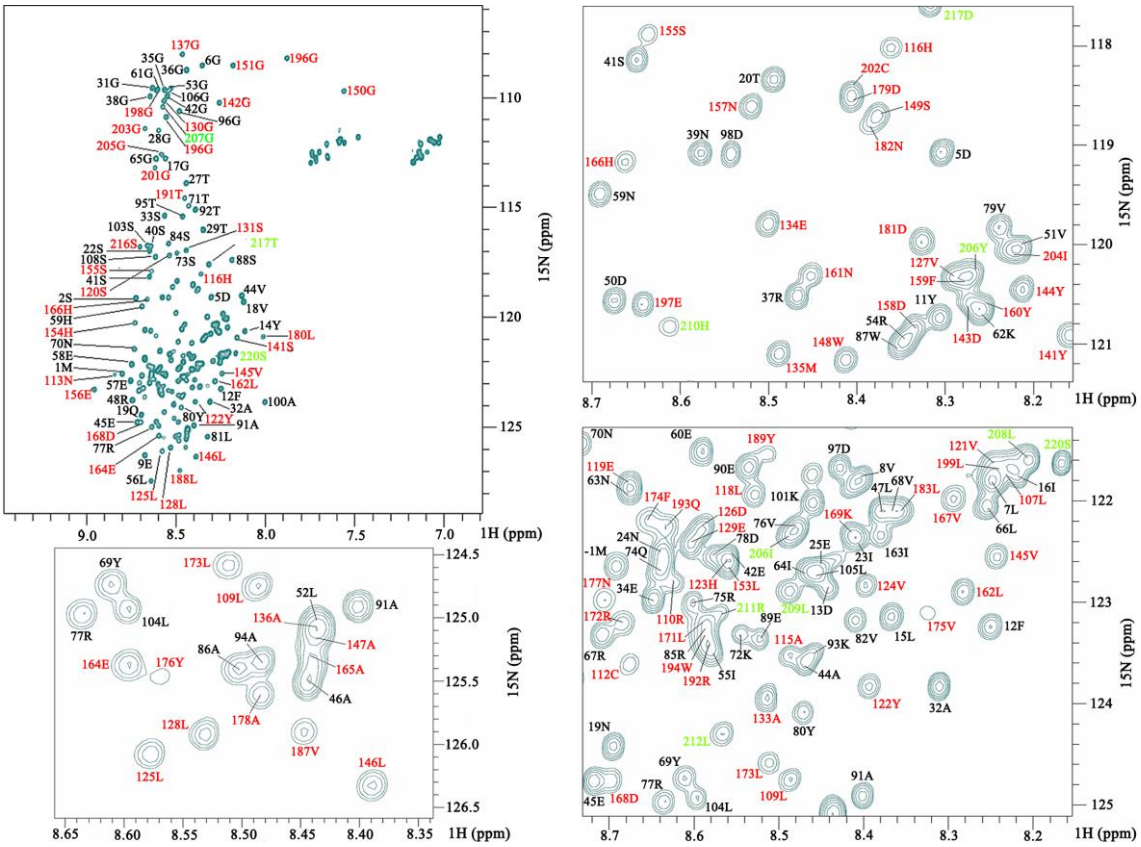


Figure 38: ^{1}H-^{15}N HSQC spectrum of dGRASP in 9 M urea showing the HN resonances that were successfully assigned. Resonances from PDZ1, PDZ2 and SPR region were colored in black, red and green, respectively. Some regions were amplified for better visualization.

Using this assignment strategy, it was possible to unambiguously assign 96% of the total non-proline HN resonances (Table 8 and Figure 38). If we divide the analyses in three parts, separating the PDZs and the SPR regions, then 100% of the PDZ1 HNs resonances were successfully assigned. The protein is expressed with a SUMO-tag that is removed by the ULP1 protease, leaving at the N-terminus non-native amino acids ($G_{-4}SHM_{-1}$) where we were capable of assigne the M^{-1} residue. Unlike the PDZ1, the HSQC resonances attributed to the PDZ2 were not well resolved, and 93% of the resonances were unambiguously assigned. It was not possible to identify the residues 111V, 117A, 138L, 139V, 184R, 185E, and 186V. For the small portion of the SPR domain (20 amino acids located in the protein C-terminus), all the amino acids were assigned with only one exception (R_{214}). This would naturally be a very challenging residue to identify since it is flanked by two prolines ($P_{213}R_{214}P_{215}$).

Table 8: dGRASP backbone resonances assignment statistics.

HN	**96% (non-proline aminoacids)**
CO	**95%**
Cα	**94,5%**
Cβ	**94,4%**

The quality of the assignment can be assessed using the position of the HN resonances since the protein is unfolded. Because most, if not all, the H-bonds are formed with water molecules and the protein is not collapsed, we expect the resonances to be mostly influenced by the previous amino acid and by the side chain. As we can see in Figure 38, all the glycines are

located in the upfield region of the spectra, with serines/threonines below and alanines/leucines being mostly located in the downfield region. Most of the clusters are also formed by amino acids of the same type (Figure 38).

Because we collected the CBCACONH, CACANH, HNCO, and HN(CA)CO, it was possible to extend the assignment also to the Cα, Cβ and CO residues (Table 8). A total of 95% of the carbonyl atoms were assigned and approximately 95% of Cα and Cβ (Table 8). This statistic also includes the prolines since it is possible to identify those resonances if the following residue is not a proline.

The HBHA(CBCACO)NH and HSQC-TOCSY also allow the assignment of protons from the backbone and side chain. This work is still in progress.

^{1}H-^{15}N HQSC urea titration analyses

Once we had reached a satisfactory HN assignment, we could move forward with the studies involving the PDZs stability. As discussed previously, we collected 10 dGRASP HSQC spectra ranging from 0 to 9 M urea concentration (1 M steps). Since we assigned the 9 M urea spectrum, the assignment transfer was done by spectral superposition and by following the shifts during the titration (Figure 39A and B). Of course, peaks already present in the 0 M spectrum, were also present in all the other spectra (accompanied by the displacements in chemical shift position). The chemical shifts are expected to vary during the titration not only because of the increase in the urea concentration (that can change medium viscosity and dielectric constant) but also because of the natural disruption in the protein structure. We used the water peak position (4.71 at 20℃) to reference all the 2D and 3D spectra but this position is expected to vary in a urea-dependent manner. Therefore, we did not make any of the analyses relying on how is the chemical shift variation. Based on a logical approach (such as the one represented in Figure 39), all the assignments from the spectrum collected in 9 M were transferred to the others. This approach has

previously been used successfully with other non-native molten globule structures [161,162,163,233].

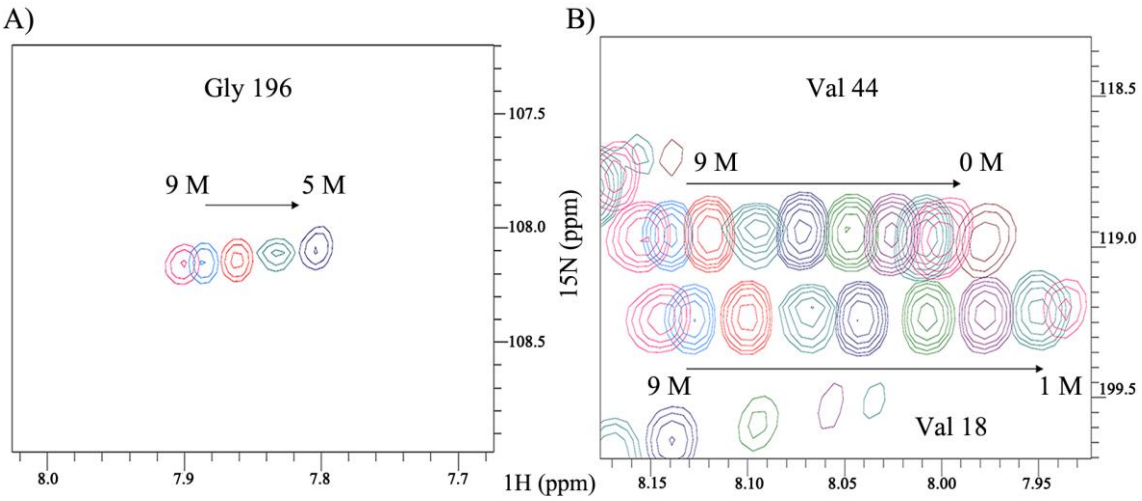


Figure 39: Urea titration monitored by NMR. The line broadening typically found in molten globule structures is the responsible for the "line disappearing" in the very beginning of the titration. The appearance of each and any resonance line was assumed to be due to the local unfolding of that region. A) Gly 196 and B) Val 44 and Val 18.

Not all the peaks were present at the very beginning of the urea titration and started to appear as the urea concentration was raised, which is a common feature of the molten globule state. The gradual appearance of the peaks is attributed to the unfolding of the particular region responsible for the peaks (Figure 39). All the analyses from now on rely on this assumption: If the resonance peak for a non-proline residue X was only observed in the ^{1}H-^{15}N HSQC at Y M urea and not on any other lower concentration, so this is the urea concentration where this residue is now considered to be in an unfolded structure. If we monitor peak by peak all those that could be tracked unambiguously during the titration, and plot the urea concentration where each peak appeared during the titration as a function of the amino acid residue index, we end up with the graphic in Figure 40.

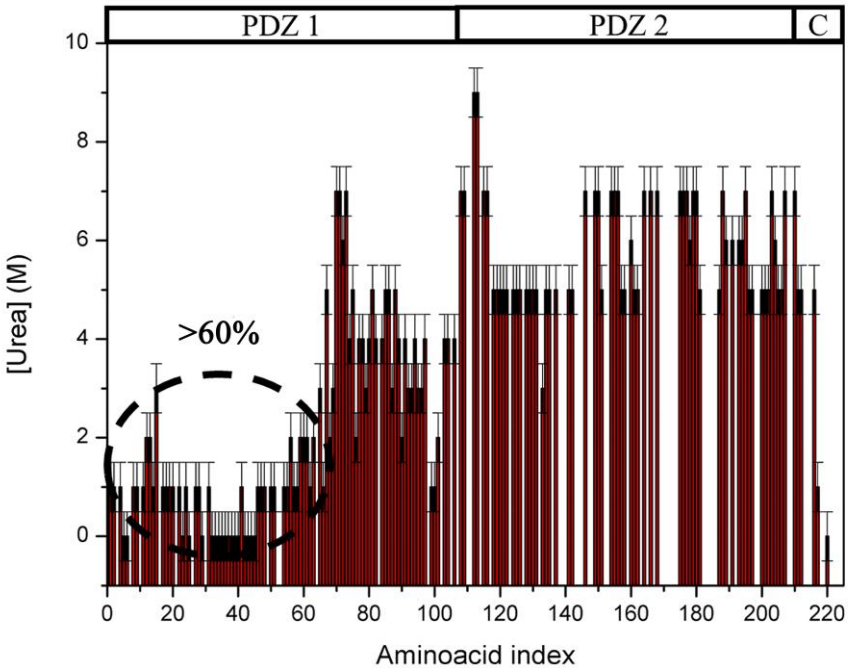


Figure 40: Urea titration by NMR. The graphic shows, for each assigned amino acid in the primary sequence, the urea concentration where the resonance line start to be visible in the ^{1}H-^{15}N HSQC. The low stability region inside the PDZ1 was highlighted in the graphic.

Figure 40 clearly shows that the PDZs behave very differently from each other in solution. This feature cannot be inferred from the available crystal structures, which actually point that they are very similar to each other. Of course we still do not have enough data to conclude what is really going on during the titration and a difference in oligomerization degree or long range interactions could also be favoring these differences observed. We can also see that all the resonance peaks already present at 0 M urea come from the PDZ 1. Because all these peaks are located in very sharp chemical shift dispersion close to 8 ppm in the proton dimension, we can safely suggest that those peaks are located in disordered or very flexible regions. In general, ~60% of the PDZ1 resonances appear at a urea concentration below 2 M, and with a total average concentration of 2 M. The most stable part of this domain seems to be the region between the amino acids 70 and 95 with a urea stability of around 4 M. This result suggests that the PDZ1 unfolding is weakly cooperative and we expect at least two main unfold transitions.

The opposite behavior was observed for PDZ2. The majority of the resonance peaks appeared at 5-7 M urea, a much sharper transition. The average urea concentration for PDZ2 unfolding is 5.7 M, suggesting that the GRASP domain is formed by much more stable PDZ2 and a dual PDZ1 formed mainly by a very unstable region (>60%). Interestingly, the most unstable region of PDZ1 is close to the predicted binding pocket for molecular partners suggested by the molecular model (Figure 41), indicating that this region is not only more unstable, but more flexible as well. The data so far did not discard a contribution coming from how oligomers have been formed and lost in our GRASP domain during titration. How GRASP dimers are formed, as discussed before, is still an open issue. This problem might be solved if the PDZs are studied separated, as will be discussed latter in this chapter.

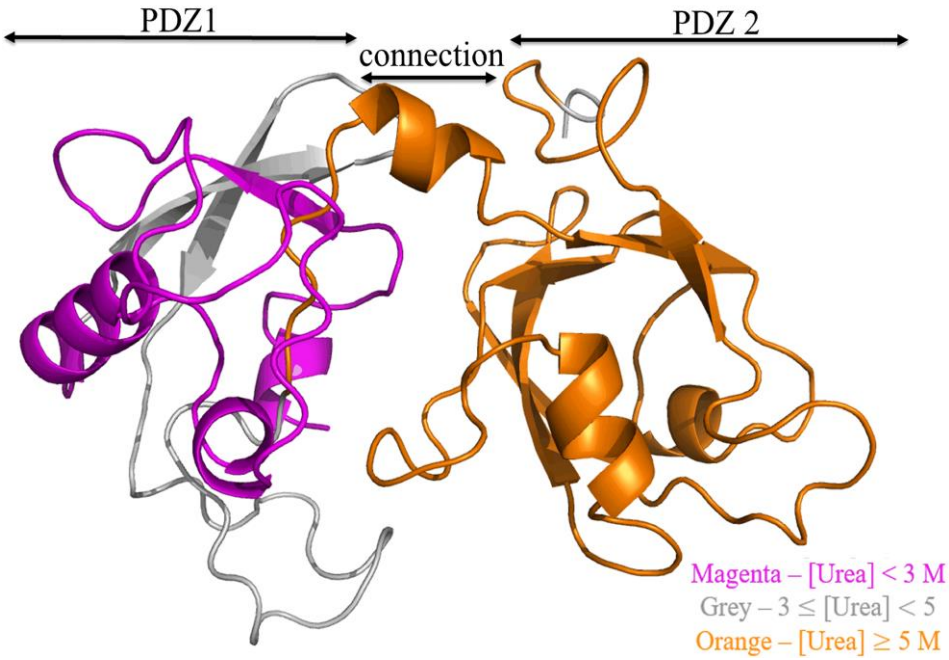


Figure 41: dGRASP molecular model (based on the structure of GRASP55 – PDB I.D. 3RLE) highlighted with different colors according to three different groups of urea stability. In magenta regions with the stability of less than 3 M, in grey between 3 M and 5 M and orange with higher than 5 M.

Because we wanted to futher understand the urea titration monitored by NMR, we decided to study the impact of the urea in the dGRASP secondary structure monitored by CD. As it can

be observed in Figure 42, the unfolding is remarkably non-cooperative, with the transition to the unfold state being linear Because the intensity remarkably decrease at 7M urea, we can confidently affirm that this is an unfolding rather than just a loss in protein oligomerization. This is the same behavior observed previously for the full-length protein [Figure 14, 174]. The great power of NMR is that it gives high-resolution, short-range structural information, which cannot be obtained in the same quantitys by any other experimental technique.

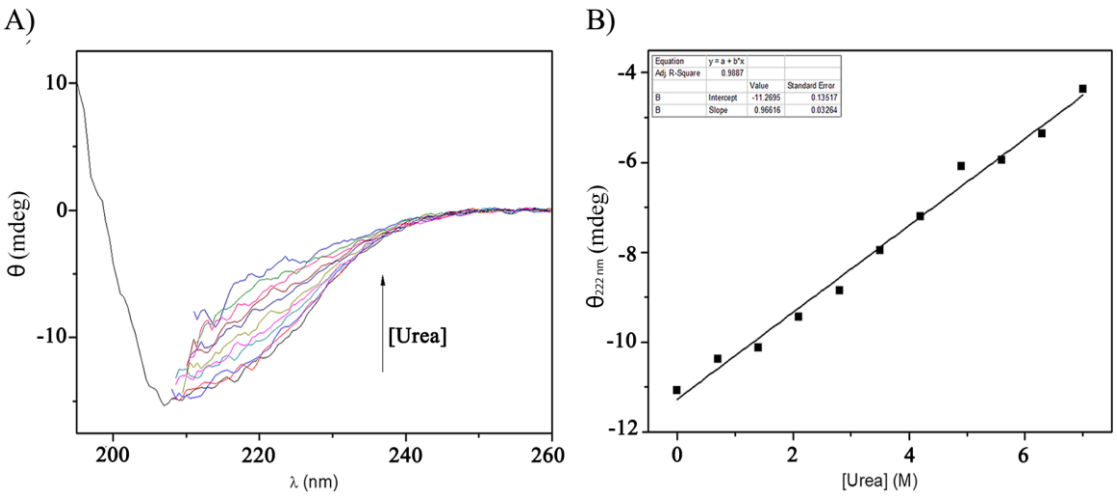


Figure 42: dGRASP urea titration monitored by CD. A) Far-UV CD curves for each step of the titration (0.7 M urea from 0 to 7 M). B) CD intensity at 222 nm plotted as a function of the urea concentration. The dataset was fitted to a linear function. Individual samples were prepared for each urea concentration at a fix protein concentration of 0.15 mg/mL.

Breaking the GRASP domain in two

We decided to initially work with the GRASP domain because it is considerably more stable than the full-length protein, but also because we were neither sure about where to cut the PDZs apart nor where they were going to fold at all without their counterpart. Since our main goal is to unravel all the structural features of both PDZs in solution, we decided to study the PDZs separately.

Based mainly on the PDZ prediction using the domain predictor Pfam (http://pfam.xfam.org, accessed in 2015), on the disordered prediction previously published by us [174] and on the information we obtained from the NMR experiments, we decided that the best positions to "cut"

the PDZs would be 1-115 for our "PDZ1" and 116-220 for the "PDZ2" construction (Figure 43A). We amplified these domains using the full-length CnGRASP DNA and standard cloning-sub cloning PCR strategies with the subcloning in pETSUMO (data not shown). DNA sequencing (data not shown) was used to check the integrity of the final constructs.

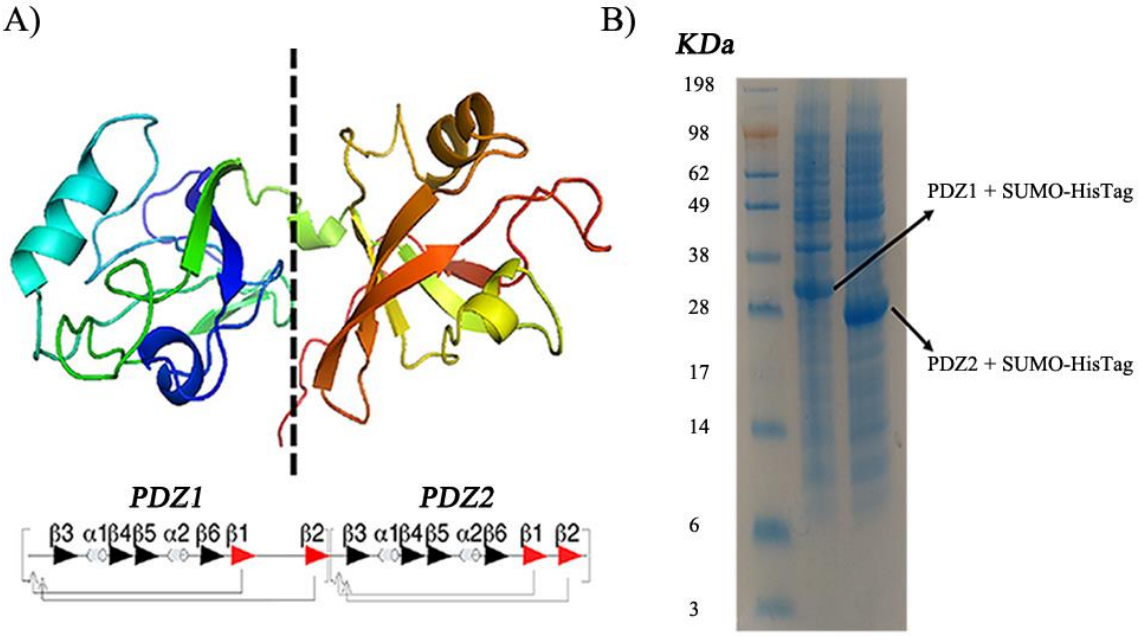


Figure 43: Splitting the PDZs apart. A) Crystal structure of GRASP 55 (PDB I.D. 3RLE) to ilustrate the region used to separate the PDZs. Figure adapted from reference [64]. B) SDS-PAGE of the expression analyses of the individual PDZs. Lane 1: molecular mass standard; Lane 2: PDZ1 + SUMO-HisTag and Lane 3: PDZ2 + SUMO-HisTag.

Both PDZs were expressed and purified using the same protocol used for the dGRASP. Both PDZs are expressed at very high yields (Figure 43B) given approximately 5-10 mg per liter of culture, and gave very sharp elution profile in the size-exclusion chromatography (Figure 44), suggesting high purity and monodispersity profile. Using globular proteins as elution standards, we estimated a molecular mass of 60 and 23 kDa for PDZ 1 and 2, respectively (Figure 44 and Table 9). This result suggests that PDZ1 is more likely a tetramer, while PDZ2 is a dimer in solution (like the dGRASP and the full-length protein), but we could also detect a small fraction of PDZ1 that elutes at the same position of PDZ2, which suggests a dimer conformation is also

present (Figure 44). Exchange between tetramer and dimer is observed if the tetramer fraction is isolated and reapplied to the SEC column. Why PDZ1 behaves as a tetramer is not known, but we might assume that this is related to the amount of disordered that this domain has along its structure bridging non-physiological oligomers. Because disordered proteins can have abnormal hydrodynamic properties, we also used chemical cross-linking and SDS-PAGE to validate the SEC results (data not shown) and the same properties were observed. We suggest that the PDZ1 tetramer is formed by non-specific dimerization of the functional relevant dimer and the full-length GRASP dimerization might be mediated by a PDZ1-PDZ1 and PDZ2-PDZ2 interaction.

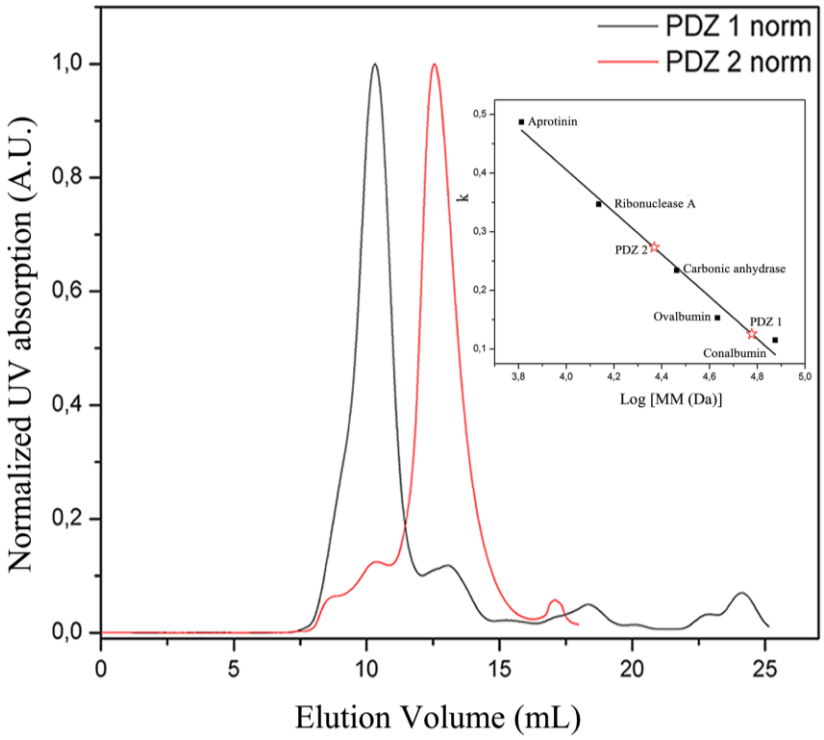


Figure 44: Normalized SEC elution profile of purified PDZ1 and PDZ2. The linear dependence of Rh as function of the partition coefficient obtained from SEC is shown in the inset. Standard proteins were obtained from Gel Filtration Calibration Kit LMW/HMW (GE Healthcare).

Table 9: Molecular masses of PDZ1 and 2 predicted based on the primary sequence and calculated from the hydrodynamic properties measured using SEC.

	Molecular mass (kDa)	Experimental MM (kDa)	Ratio
PDZ 1	12,3	60	4.87
PDZ 2	11,6	23	1.98

CD and SRCD analyses

Before moving further with the NMR experiments, we started a characterization of the PDZs using CD and SRCD. SRCD spectra of both PDZs show a very peculiar behavior at first (Figure 45) because the intensities are remarkably different (after normalization), with PDZ2 being much more intense in all the regions analyzed. It is important to emphasize that the concentrations were measured in two different machines, including one right before starting the measurements. Moreover, the results are reproducible (two different sample preparations). A precise concentration measurement is critical for the correct CD normalization and deconvolution.

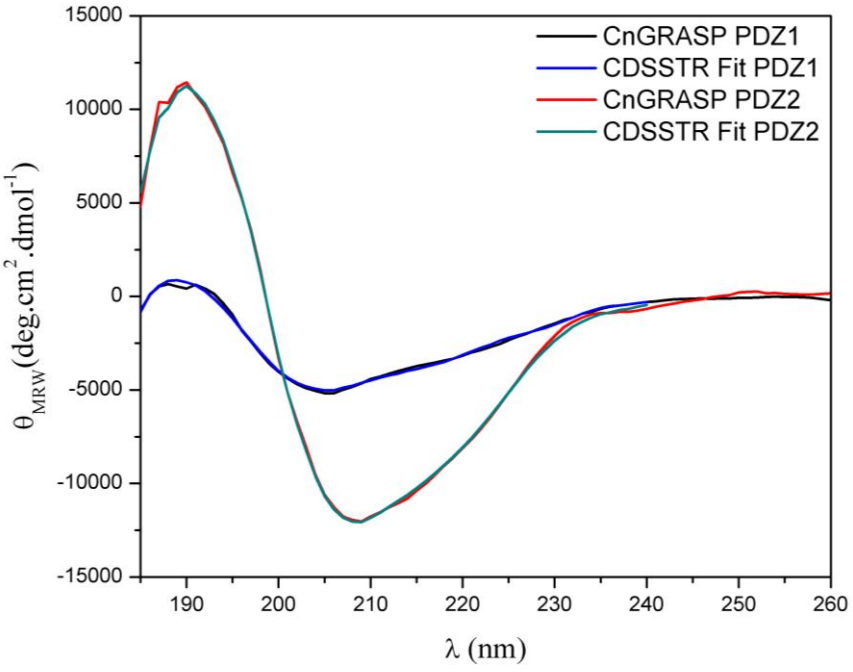


Figure 45: SRCD data of PDZ1 (black) and PDZ2 (red) together with the deconvolution fit (blue and green, respectively) obtained using the software CDSSTR and an appropriated database inside the online server Dichroweb.

Table 10: SRCD spectral deconvolution.

	Helix	**Strand**	**Turns**	**Unord.**	**NRMSD**
PDZ1	5%	37%	22%	34%	0.040
PDZ2	18%	30%	22%	30%	0.027

The signal deconvolution indicates that both PDZs have comparable amounts of unordered and regions classified as turns (Table 10). We were expecting a high amount of unordered regions in PDZ1 based on the disordered prediction reported previously [174]. The 30% of unordered structures found in PDZ2 is also consistent considering the total amount of this kind of structure observed previously for the full-length protein and considering the value for PDZ1 [174], but it does not agree with the disorder predictions. We observed that most of the peaks found in the dGRASP ^{1}H-^{15}N HSQC in the absence of urea comes from the PDZ1, and because they show low proton dispersion, we attributed it to disordered regions. It is not known why there are no peaks from PDZ2 if the number of disordered regions is comparable, but if we consider that the PDZ2 is conformational fluctuating in a µs-ms time scale, it might be experiencing several disordered-to-order fluctuations along its structure. We proposed previously in chapter 3 that the full-length GRASP has multiple sites capable of going through disorder-to-order transitions depending on the cell conditions. Thus, this might happen in a conformation selection fashion. It still needs further investigation, but of course, this could explain our hypothesis.

The CD deconvolution also shows that the PDZs differ themselves in the amount of predicted helix and strands (Table 10). PDZ1 possesses a lower helical content (approximately 3 times less

than PDZ2) and a little more strands. Because the 3D structure of CnGRASP is not available thus far, we can only try to correlate this amount with the crystal structures of the homologs. Only structures of the GRASP domain of GRASP55 and GRASP65 are currently available [64,65] and our deconvolution data show that the amount of regular ordered elements (helix and strands) in our model is greater than those predicted from the structures. Based on the data in chapter 3 showing that PEG, local dehydration and changes in the dielectric constant can remarkably disturb GRASP structure, we believe over reliance on the crystal structures (especially with conditions rich in PEGs) when interpreting biophysics data should be avoided.

The near-UV (250-350 nm) CD spectrum is a useful way of monitoring certain aspects of protein tertiary structure [130]. At these wavelengths, the chromophores are the aromatic amino acids (bands centered in the 250-270 nm, 270-290 nm, and 280-300 nm intervals are attributable to phenylalanine, tyrosine, and tryptophan, respectively) and disulfide bonds (broad weak signals throughout the near-UV spectrum), and the CD signals they produce are sensitive to the overall tertiary structure of the protein [130,131]. If a protein retains secondary structure but no well-defined 3D structure (e.g. IDPs or "molten-globule" structures), the signals in the near-UV region will be nearly zero [237]. On the other hand, the presence of significant near-UV signals is a good indication that the protein is folded into a well-defined structure [99,130].

PDZ1 (PDZ2) has 4 (6) tyrosines, 4 (3) phenylalanines, and 1 (3) tryptophans. Therefore, both PDZs have comparable amounts of aromatic amino acids that are reasonably distributed along the primary structure and that can be used as good probes for near-UV CD. The PDZ2 near-UV CD spectrum has multiple peaks in the region analyzed, suggesting that it has a more well-defined tertiary structure (Figure 46). The same pattern is not observed for the PDZ1, where only a bump is present in the tryptophan region, probably due to its single one in the sequence which might be in a more collapsed region. The near-UV CD analyses suggest that the isolated PDZ2 has a better-defined tertiary structure when compared to its partner PDZ1.

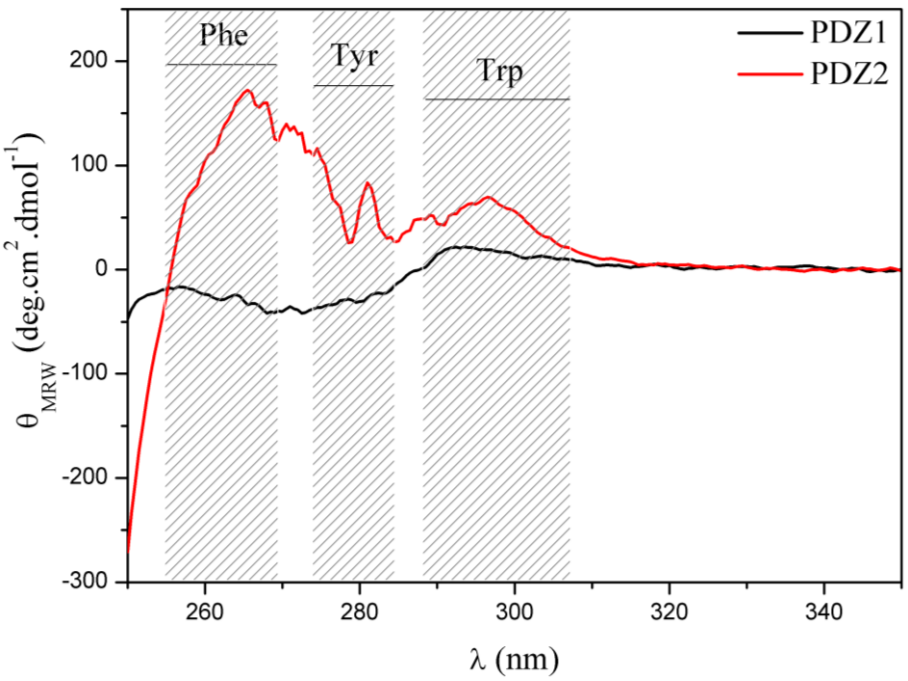


Figure 46: Near-UV CD of PDZ1 (black) and PDZ2 (red) with the regions of Phe, Tyr and Trp resonances highlighted in grey.

This analysis has revealed some important differences between PDZ1 and PDZ2 that cannot be determined from the crystallographic structures. The first one is a difference in structure stability as probed by NMR. Would a similar difference is stability be observed by CD using the separate PDZs domains? We showed before that the dGRASP unfolds in a very non-cooperative transition during urea titration. For the separate PDZs, the data now is consistent with the NMR results. Firstly, PDZ1 starts to unfold very early during the titration, around 1 M urea, and around 60% of the total population is already in the unfolded state at 2-3 M urea (Figure 47). This is consistent with the NMR analyses, where a significant part of the structure is unfolded right in the beginning of the urea titration. It is important to emphasize that both techniques provide somewhat different pieces of information. NMR allows us to analyse local denaturation and, in the CD experiment, we are looking for differences in population between folded and unfolded proteins. Still, they both agree that the PDZ1 is a less stable structure and with a less cooperative unfolding.

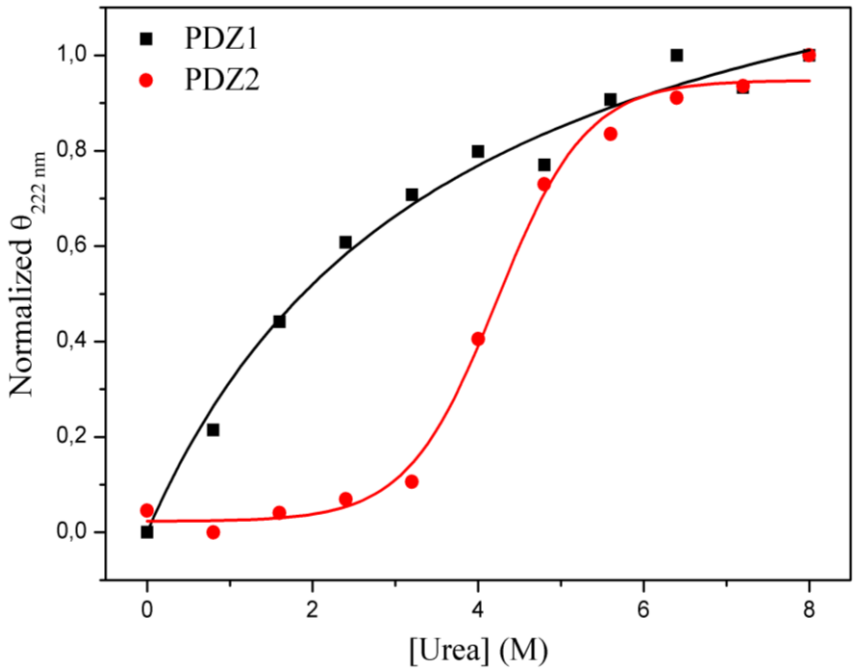


Figure 47: Urea titration monitored by the intensity at 222 nm using far-UV CD of PDZ1 (black) and PDZ2 (red) as a function of the urea concentration. The PDZ1 and PDZ2 data sets were fitted with a hyperbolic and a sigmoidal function, respectively.

A very different situation is observed for PDZ2, which shows a more cooperative transition to the unfolded state with a concentration of urea in the half transition of 4.2 M, consistent with the near-UV CD, suggesting that, even though the PDZs have comparable amounts of regular secondary structures and disordered regions, the PDZ2 has a better-defined tertiary structure. The latter feature is probably responsible for the highly cooperative transition observed in the titration by CD, and is consistent with the NMR data.

^{1}H-NMR and ^{1}H-^{15}N HSQC studies

^{1}H-NMR is probably the easiest, fastest and cheapest experiment to be collected using solution-NMR. First, because no isotopic labeling is required and second because protons are highly abundant (especially in organic materials). ^{1}H spectra of proteins are rarely informative because peaks are too crowded and cannot be assigned. By focusing first on the amide resonance regions, we can at least start to see some significant differences between the PDZs (Figure 48).

PDZ2 has multiple peaks spread over the region 6-10.5 ppm, with most of the lines being very sharp. This is a strong indication that PDZ2 behaves like a regular ordered protein and, because there is not usually a large number of effects that can compromise the nitrogen resonance dimension, PDZ2 might be a good protein for structure determination by NMR. We can even resolve the three tryptophan side-chain resonances (~10 ppm) even though two of them resonate very closely.

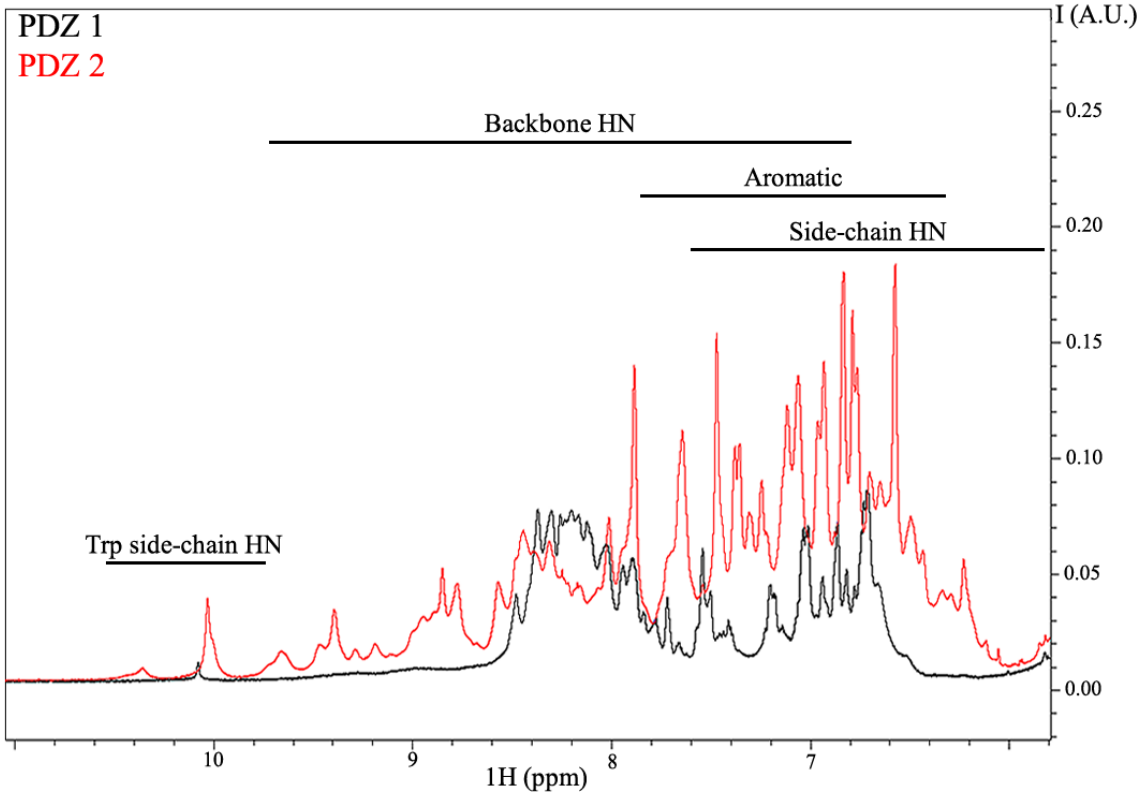


Figure 48: ^{1}H NMR of PDZ1 (black) and PDZ2 (red) with some resonance regions highlighted in the figure for illustration.

A remarkably different behavior is instead observed for PDZ1 (Figure 48). The amine proton resonances are distributed over a narrower region (~2 ppm) and the line superposition is much more severe (leading to apparent broadening). Even though PDZ1 has a poorer dispersion than PDZ2, this dispersion is still greater than those observed in extended intrinsically disordered proteins [238]. Of course, this was expected since neither the CnGRASP domain nor the dGRASP is an extended IDP.

However, the PDZ2-^{1}H NMR data does not agree with that observed previously for the dGRASP. It is interesting to notice that we are observing a very non-intuitive phenomenon where a protein domain seems to be better behaved when isolated than when it is part of the full-length protein. The opposite is more commonly observed, even for PDZs in other systems [229,230]. Most of these failed tentatives are usually discarded without even being submitted for publication because this is not an unexpected thing to happen. We have not found similar data in the literature for other systems thus far.

The PDZ1 and PDZ2 ^{1}H-^{15}N HSQC agrees with the ^{1}H-NMR analyses (Figure 49) showing a very narrow and a better dispersion of peaks in the proton dimension for the PDZ1 and PDZ2. Just a few number of resonance lines are observed for PDZ1 but this can be due to the tetramer conformation coupled with a number of disordered regions located throughout the structure (an increase of the effective size coupled with a decrease of the overall tumbling that ultimately leads to worse relaxation properties). However, for the few peaks present, the match with those previously observed for the dGRASP is excellent (Figure 49), showing that these regions are very disordered and have the same properties whether it has the PDZ2 attached to it or not. According to our urea titration data, all the peaks observed in the dGRASP HSQC in the absence of urea came from the PDZ1. PDZ2, on the other hand, behaves in a completely different way, with a proton dispersion, which looks like those observed and expected for well-structured proteins (Figure 49). Because PDZ2 is a dimer of ~20 kDa, the size of this molecule cannot be used to explain the absence of half of the expected peaks. This result suggests that the PDZ2 still presents the molten globule-like behavior observed in the full-length structure.

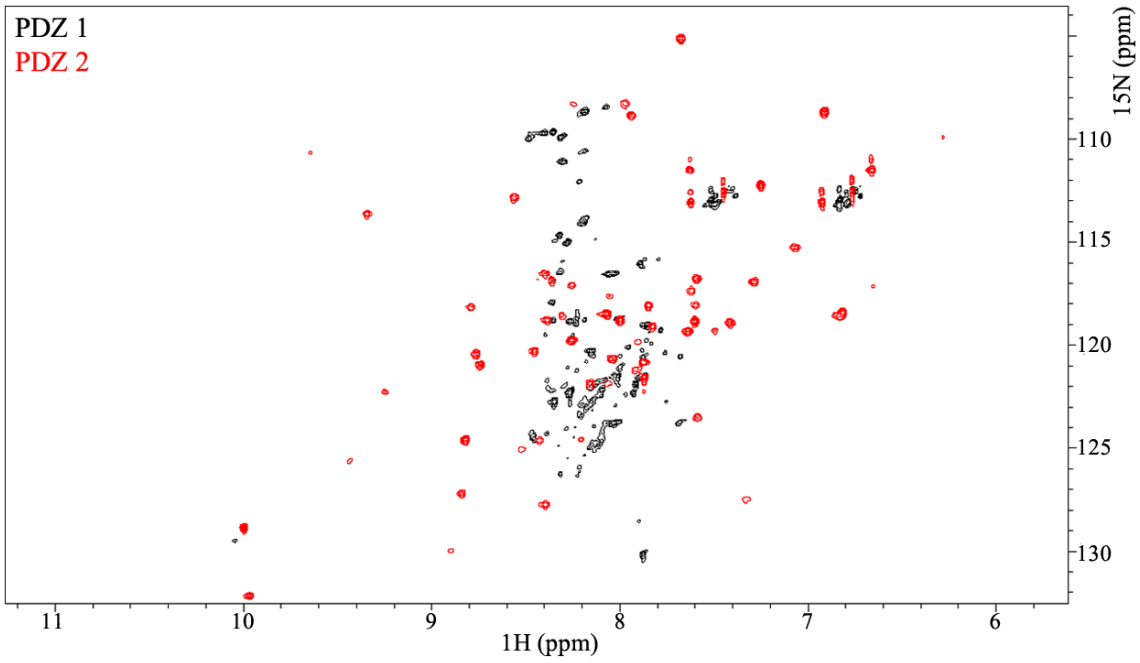


Figure 49: ^{1}H-^{15}N HSQC NMR of PDZ1 (black) and PDZ2 (red) showing the amine resonance region.

Urea titration monitored using ^{1}H-^{15}N HSQC

We showed before that upon increasing the urea concentration, the dGRASP construct unfolds non-cooperatively with the PDZ2 being more stable and with PDZ1 having at least two regions with different stability. We now repeated the analysis with the separated PDZs to validate the previous observations. PDZ1 shows the same pattern observed for dGRASP with fewer peaks at the very beginning of the titration but with a rapid appearance of the remaining lines when the urea concentration is increased (Figure 50). Most of the peaks absent in the spectrum of the native structure appeared at urea concentration around 1-2 M, the same pattern observed in dGRASP (Figure 34 and 40). This unfolding transition is associated with the part of the domain predicted to be the binding for molecular partners, and it is interesting to note that its ulfonding behaviour appers to be unaffected by the presence or absence of PDZ2. The remaining lines appeared around 3-4 M of urea, suggesting that the most stable region of PDZ1 in the dGRASP construction decreases its stability in the absence of PDZ2 (from ~5 to 3-4 M urea).

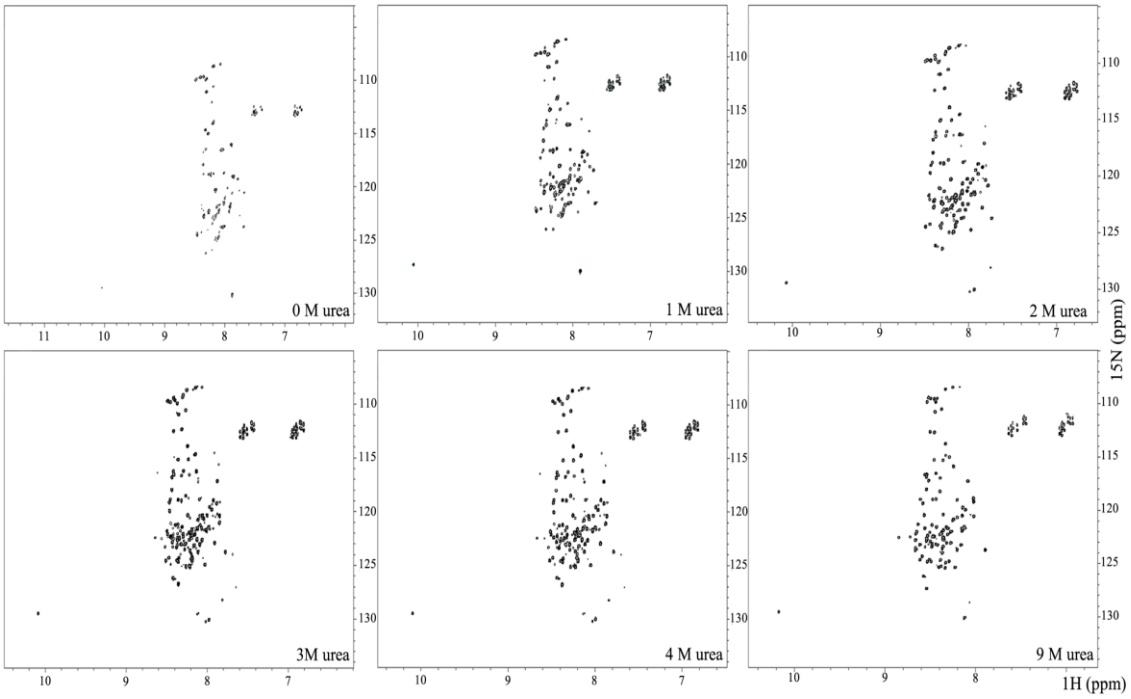


Figure 50: Isolated PDZ1 urea unfolds experiment monitored by 1H-15N HSQC in 950 MHz at 20℃.

Interestingly, if we superimpose the PDZ1 spectrum to the dGRASP one, both conditions without urea, it is possible to notice that all the resonance lines that were present for the dGRASP in the beginning of the titration correspond to PDZ1 residues (Figure 51). As we discussed before, this suggest that these regions are within disordered/flexible structures, and supports the transfer of the assignments from the 9 M urea condition.

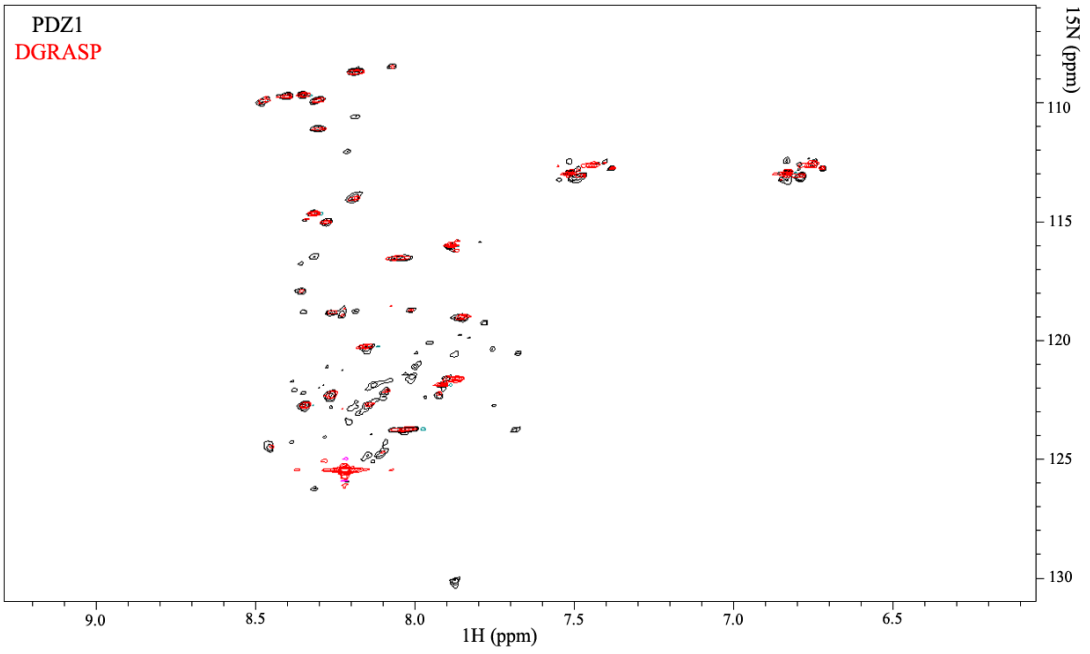


Figure 51: ^{1}H-^{15}N HSQC spectra of PDZ1 (red) and dGRASP (black) in the absence of urea.

For PDZ2 the profile is the same expected for regular well-folded proteins, with the native structure showing a reasonable spread in the proton dimension and, when the urea concentration reaches a specific value, the lines move to a very narrow region close to the ^{1}H resonance of 8 ppm (Figure 52). It is possible to see that the transition between these two states is highly cooperative and occurs for urea concentration around 3-4 M. The cooperative pattern is the same observed previously for the dGRASP but the overall stability was significantly reduced (in at least 1 M urea). This result shows that the PDZ2 is even more stable when it is connected to PDZ1 in tandem.

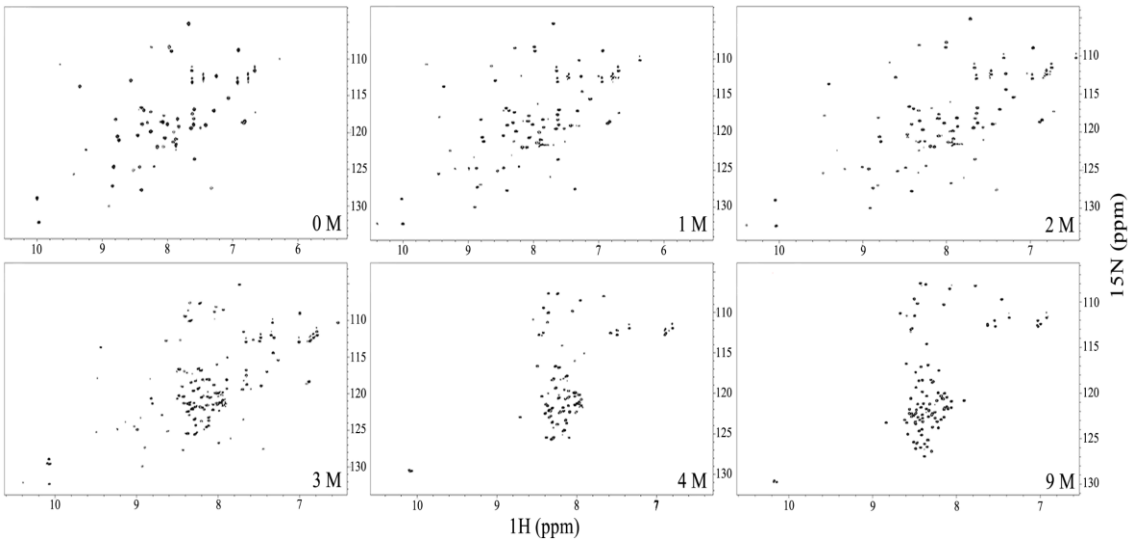


Figure 52: Isolated PDZ2 urea unfolds experiment monitored by 1H-15N HSQC in 950 MHz at 20℃.

The truth behind PDZ2

When we analyze all the data together it is possible to conclude that GRASP has two very asymmetric PDZs in terms of flexibility, compaction, and stability (even though they have similar amounts of regular secondary structure) and that this asymmetry is more promoted when the domains are together then when they are isolated. Looking at the ^{1}H-^{15}N HSQC of PDZ1 we can

find a direct correlation with the dGRASP HSQC, showing few peaks at the very beginning of the titration and with the other lines suddenly appearing when the urea concentration is increased. Besides, all the peaks observed in the dGRASP spectrum came from PDZ1 according to the urea titration analyses and in direct comparison with the PDZ1 spectrum. The results for PDZ1 agree with what was observed for dGRASP, whereas, for PDZ2, the opposite was observed. We can clearly see that PDZ2 is a much more stable and well-behaved structure. The near-UV CD suggests a more compact fold and the dispersion in the ^{1}H-^{15}N HSQC shows a regular well-folded structure pattern. The unfolding clearly agrees with the dGRASP data. However, why do we not see this greater dispersion in the ^{1}H-^{15}N HSQC dGRASP spectrum? There are a few hypotheses to be tested but one explanation could be just an increase of the molten properties of PDZ2 when PDZ1 is in tandem. The well-packed structure is still present but the conformational fluctuation is larger than when PDZ2 is alone and along with the decrease in relaxation properties expected with the larger molecular mass of dGRASP, we just do not observe the lines because they become too broad. Still, if this structure is true, we might recover some of the lines if we significantly increase our signal-to-noise ratio and that is what we did using ^{1}H NMR (Figure 53).

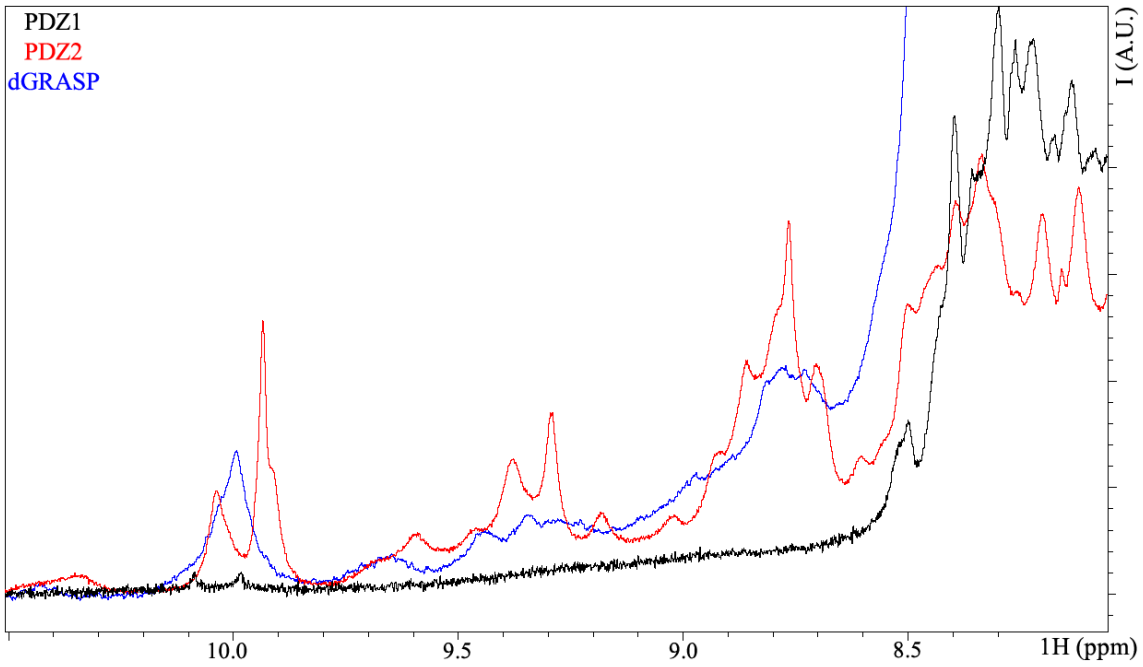


Figure 53: ^{1}H NMR of PDZ 1 (black), PDZ2 (red) and dGRASP (black) highlighting the region of resonance between 11 and 8 ppm. The scales are not the same and were selected to facilitate the analyze.

By collecting a great number of scans on a proton spectrum (1000), we could resolve some lines in the region around 9 ppm that have a clear correspondence with the PDZ2 spectrum, but not with the PDZ1. Furthermore, looking at the region around 0.5 ppm, which is where the methyl groups resonate, we start to also resolve some well-defined lines, a characteristic of well-folded proteins, with direct correspondence to PDZ2 but not with PDZ1 (Figure 54). The data strength the latter assumption that PDZ2 has a better-behaved structure than PDZ1 and this is not an artifact of the isolated structures.

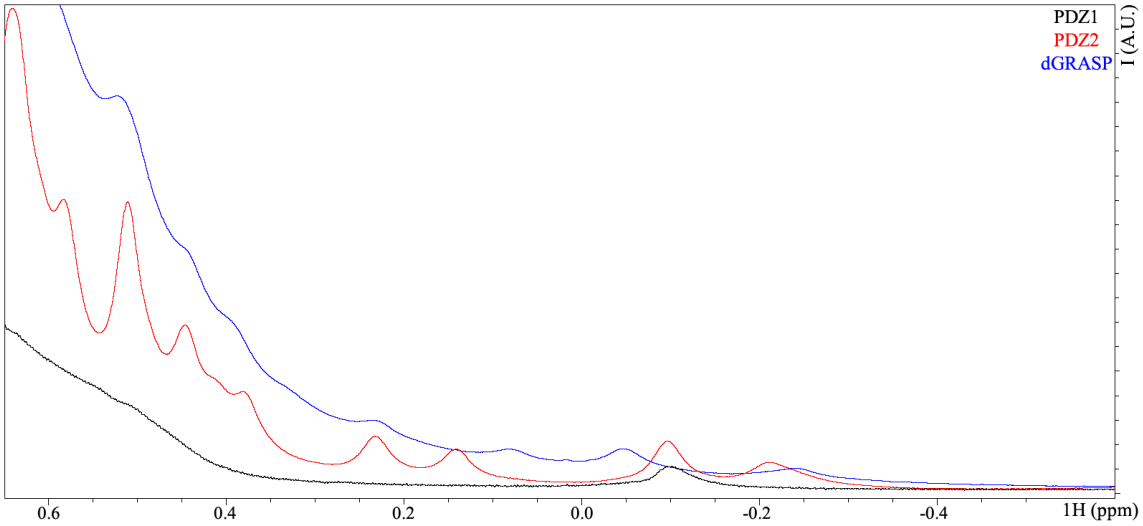


Figure 54: ^{1}H NMR of PDZ 1 (black), PDZ2 (red) and dGRASP (black) highlighting the region of resonance between 0.6 and -0.5 ppm. The scales are not the same and were selected to facilitate the analyze.

Conclusions

Our goal when starting the present project was to understand, if there was any, the differences between the two PDZs of CnGRASP. Based on numerous cell biology reports, it was clear that PDZ1 and PDZ2 were not similar, and the role played by GRASP in UPS might be directed related to the ability of this protein to interact with the target partner somehow using PDZ1 preferentially. Previous structural biology results suggested that PDZ1 and PDZ2 were structurally very similar,

with both resembling the PDZ fold commonly found in prokaryotic systems (GRASP are exclusive of the eukaryotic system though). We previously described the overall structural behavior of a full-length GRASP in solution and we observed that this protein has a molten globule-like structure as inferred by numerous spectroscopic techniques. Our data suggested that the GRASP structure might be much more complex in solution and that dynamics could be a possible explanation for the PDZ asymmetry. NMR is still the most powerful technique available for structural biology to describe the structure and dynamics of a protein in solution and at a relevant temperature. The only requirement for NMR is a uniform 13C/15N (and sometimes 2H) labeling, which is also the most non-invasive labeling protocol, especially compared with the use of bulky labels combined with site directed mutagenesis often used in fluorescence and electron paramagnetic resonance spectroscopy. Molten globules undergo conformational fluctuations at the μs/ms timescale, something that is very challenging to probe using the standard NMR techniques designed for regular well-folded proteins.

In conclusion, our data suggest that GRASPs are formed by two asymmetric PDZs domains, with PDZ1 being more unstable and flexible while PDZ2 is much more stable and structurally well behaved. Many of the unstable regions found in PDZ1 are in the predicted binding pocket, suggesting a structural promiscuity inside this domain that correlates with the functional promiscuity of interacting with multiple protein partners. In contrast, PDZ2 is more structured, and this correlates with the smaller number of associated proteins capable of anchoring with GRASPs using this domain. The consensus in the community that both PDZs have a similar structure, even though their function is remarkably different. However, it is worth keeping in mind that this conclusion came from static crystal structure obtained in conditions that we have previously shown that are capable inducing disorder-to-order transitions in GRASP structure. This chapter is the first work suggesting that in solution the structure and dynamics of the PDZs are not similar.

However, the issue with protein oligomerization and how can this interfere in our analyses must be explored in details. This happens specially because we cannot assume, for sure, that some

of the observed phenomenon are concentration-dependent. We observed that PDZ2 behaves as a stable dimer, while PDZ1 is more promiscuous and the tetramer is the most populated oligomer, but with observed dimers and hexamers as shown by chemical crosslinking assays. This promiscuity can be explained by our NMR data and gives more information on the problem regarding the dimerization of full-length GRASPs *in vivo*. How GRASPs form a dimer structure is still unclear, since biochemical data suggest a direct PDZ1/PDZ2 interaction [127] but with structural data suggesting PDZ1/PDZ1 and even SPR/PDZ1 binding [65]. We observed that, for CnGRASP, each PDZ is capable of forming homodimers. Experiments to look for cross interaction, and to quantify it, are currently being planned.

Chapter 5: General conclusions and future perspectives

As it was discussed throughout the text, GRASPs constitute a very broad and dynamic family of proteins with many important cell functionalities, especially regarding the Golgi complex organization and UPS. However and despite its relevance in the cell, structural information was limited to the N-terminal of GRASPs and when it comes to understanding its structural dynamics in solution, the amount of data in the literature was non existent. This thesis started with the main goal of giving, for the very first time, a structural characterization of a full-length GRASP. Using several biophysical and biochemical techniques, we were capable to show that our GRASP model has all the unique features previously associated with an intermediate of the protein-unfolding pathway called molten globule. These data inserted GRASP into the collapsed intrinsically disordered protein family. Besides, we could also observe that this protein is formed by a very dynamic structure and its C-terminal domain has no sequence conservation along the family, but the intrinsically disorder pattern is highly conserved.

Being a collapsed IDP means that this protein can have several unique responses to changes in the physical-chemical parameters of the medium since most of the well-established concepts applied to well-folded proteins cannot be directly applied to them. Moreover, all the characterization of molten globules reported in the literature relies on artificial molten globules induced by mild denaturation conditions, which can easily bias most studies. In chapter 3, we presented a very broad study aiming at understanding the molten globule structural plasticity and how the cell could modulate this structure promiscuity to achieve a desired function. We observe that a molten globule behaves as an intermediate between well-folded proteins and extended IDPs. The response to changes in counterions, molecular crowding and pH is similar to well-folded proteins, but the opposite happens when it comes to local dehydration and changes in the dielectric constant. Since GRASPs act in the Golgi keeping the cisternae “sandwich”, their environment is unique. We observed that the cell could play with GRASP structural promiscuity, especially its sensitivity to changes in the membrane field, by transferring GRASP from the cisternae

attachment to a free form in the cytosol. This is a relevant finding because we expect that, when attached to the Golgi, GRASPs cannot be very promiscuous to protein-protein interaction in order to keep the cisternae organization, whereas when involved in UPS it could. It was suggested that, before playing a role in UPS, GRASPs are first monomerized and then transfer to the cytosol. Our model for GRASPs when in the cytosol and when in the membrane-bounded state agree well with all the previous findings.

Interestingly, several reports have shown that GRASPs mediate interaction with protein partners (including the dimerization process) by interacting with the C-terminal (or internal peptide) of a protein partner, and this interaction is mainly mediated by the PDZ1. Besides an interaction of PDZ2 with GM130/Golgi45, no other partner was associated with this domain. It is a well-established concept in the community that both PDZs are structurally similar based on the available crystal structures, but how could they act so differently in solution if their structures are so very much alike? Of course, differences in the local amino acid composition within the binding pocket could explain these findings, but based on our pioneering studies on GRASP structure, we decided to take a closer look at both PDZs in solution. We observed that GRASPs are formed by a more unstable and flexible PDZ1 and much more stable and structurally well-behaved PDZ2. Many of the unstable regions found in PDZ1 are in the predicted binding pocket, suggesting a structural promiscuity inside this domain that correlates with the functional promiscuity of interacting with multiple protein partners. In conclusion, this thesis gave unique contributions related to the GRASP structure, which can remodel future GRASP studies in the community and add one more feature to this important family of proteins.

There are still many issues to be addressed in the future. It is a consensus already that our GRASP behaves as a molten globule and studies with other GRASPs in the group are now showing that this is not an exclusive finding of CnGRASP. However, if this feature prevails *in cell* is something that still needs to be proved. For this case, *in vivo* cell-NMR is the most powerful technique and has the power to validate also the model presented in chapter 3. To have a closer view on how the membrane field affects the GRASP structure, a work involving less artificial

membrane models would be interesting. However, it is still necessary to double attach the protein to the membrane and this need to be performed in a less artificial way, as well.

Another interesting and necessary aspect is to validate the proposed model of action of GRASPs exploring the asymmetry between PDZs. We observed that PDZ1 is much more stable and flexible than PDZ2 and this could be used during interaction with other proteins since disordered regions are famous for being promiscuous. If this is true, PDZ1 should become more rigid and well behaved after the interaction with a protein partner, a matter that still needs to be addressed in the near future.

References

1) Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular Biology of the Cell, 4th edition. New York, 2002

2) Palade G. 1975. Intracellular aspects of the process of protein synthesis. Science 189: 867.

3) Grieve AG, Rabouille C. Golgi bypass: skirting around the heart of classical secretion. Cold Spring Harb Perspect Biol. 2011 Apr 1;3(4). pii: a005298.

4) "George E. Palade - Facts". Nobelprize.org. Nobel Media AB 2014. Web. 11 Oct 2017. <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1974/palade-facts.html>

5) "Albert Claude - Facts". Nobelprize.org. Nobel Media AB 2014. Web. 11 Oct 2017. <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1974/claude-facts.html>

6) “Christian de Duve - Facts". Nobelprize.org. Nobel Media AB 2014. Web. 11 Oct 2017. <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1974/duve-facts.html>

7) Simon SM, Blobel G. A protein-conducting channel in the endoplasmic reticulum. Cell. 1991 May 3;65(3):371-80.

8) "Günter Blobel - Facts". Nobelprize.org. Nobel Media AB 2014. Web. 11 Oct 2017. <http://www.nobelprize.org/nobel_prizes/medicine/laureates/1999/blobel-facts.html>

9) Barlowe C, Orci L, Yeung T, Hosobuchi M, Hamamoto S, Salama N, Rexach MF, Ravazzola M, Amherdt M, Schekman R. COPII: a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell. 1994 Jun 17;77(6):895-907

10) Bonifacino JS, Glick BS. The mechanisms of vesicle budding and fusion. Cell. 2004 Jan 23;116(2):153-66

11) Lee MC, Miller EA, Goldberg J, Orci L, Schekman R. Bi-directional protein transport between the ER and Golgi. Annu Rev Cell Dev Biol. 2004;20:87-123.

12) "The Nobel Prize in Physiology or Medicine 2013". Nobelprize.org. Nobel Media AB 2014. Web. 11 Oct 2017. <http://www.nobelprize.org/nobel_prizes/medicine/laureates/2013/>

13) Akopian D, Shen K, Zhang X, Shan SO. Signal recognition particle: an essential protein-targeting machine. Annu Rev Biochem. 2013;82:693-721.

14) Rothman JE. Mechanisms of intracellular protein transport. Nature. 1994;372: 55–63.

15) Nagai K, Oubridge C, Kuglstatter A, Menichelli E, Isel C, Jovine L. Structure, function and evolution of the signal recognition particle. EMBO J. 2003;22: 3479–3485

16) Lingappa VR. Mechanism of protein translocation across the endoplasmic reticulum membrane. Annu Rev Cell Biol. 1986;2:499-516.

17) Walter P, Johnson AE. Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. Annu Rev Cell Biol. 1994;10:87-119.

18) Aebi M. N-linked protein glycosylation in the ER. Biochim Biophys Acta. 2013 Nov;1833(11):2430-7.

19) Budnik A, Stephens DJ. ER exit sites--localization and control of COPII vesicle formation. FEBS Lett. 2009 Dec 3;583(23):3796-803.

20) Stanley P. Golgi glycosylation. Cold Spring Harb Perspect Biol. 2011 Apr 1;3(4). pii: a005199..

21) Roth J. Protein N-glycosylation along the secretory pathway: relationship to organelle topography and function, protein quality control, and cell interactions. Chem Rev. 2002;102: 285–303.

22) Malsam J, Söllner TH. Organization of SNAREs within the Golgi stack. Cold Spring Harb Perspect Biol. 2011 Oct 1;3(10):a005249.

23) Dascher C, Matteson J, Balch WE. Syntaxin 5 regulates endoplasmic reticulum to Golgi transport. J Biol Chem, 1994;269: 29363–29366.
24) Rowe T, Dascher C, Bannykh S, Plutner H, Balch WE. Role of vesicle-associated syntaxin 5 in the assembly of pre-Golgi intermediates. Science, 1998;279: 696–700.
25) Schotman H, Karhinen L, Rabouille C. Integrins mediate their unconventional, mechanical-stress-induced secretion via RhoA and PINCH in Drosophila. J Cell Sci, 2009;122: 2662–2672.
26) Rabouille C. Pathways of Unconventional Protein Secretion. Trends Cell Biol. 2017 Mar;27(3):230-240.
27) Zhang M, Schekman R. Unconventional secretion, unconventional solutions. Science. 2013 May 3;340(6132):559-61.
28) Prudovsky I, Kumar TK, Sterling S, Neivandt D. (2013) Protein–phospholipid interactions in nonclassical protein secretion:problem and methods of study. Int. J.Mol.Sci. 14, 3734–3772
29) La Venuta G, Zeitler M, Steringer JP, Müller HM, Nickel W. The Startling Properties of Fibroblast Growth Factor 2: How to Exit Mammalian Cells without a Signal Peptide at Hand. J Biol Chem. 2015 Nov 6;290(45):27015-20.
30) Popoff V, Adolf F, Brügger B, Wieland F. COPI budding within the Golgi stack. Cold Spring Harb Perspect Biol. 2011 Nov 1;3(11):a005231.
31) Lippincott-Schwartz J, Yuan L, Tipper C, Amherdt M, Orci L, KlausnerRD. 1991. Brefeldin A's effects on endosomes, lysosomes, and the TGN suggest a general mechanism for regulating organelle structure and membrane traffic. Cell 67: 601–616.
32) Dupont N, Jiang S, Pilli M, Ornatowski W, Bhattacharya D, Deretic V. Autophagy-based unconventional secretory pathway for extracellular delivery of IL-1beta. EMBO J. 2011 30: 4701–4711
33) Zemskov EA, Mikhailenko I, Hsia RC, Zaritskaya L, Belkin AM. Unconventional secretion of tissue transglutaminase involves phospholipid-dependent delivery into recycling endosomes. PLoS One 6: 19414
34) Hughes RC. Secretion of the galectin family of mammalian carbohydrate-binding proteins. Biochim Biophys Acta. 1999;1473: 172–185
35) Glebov K, Schu¨tze S, Walter J. Functional relevance of a novel SlyX motif in non-conventional secretion of insulin-degrading enzyme. J Biol Chem. 2011;286: 22711–22715
36) Malhotra V. Unconventional protein secretion: an evolving mechanism. EMBO J. 2013;32: 1660–1664
37) Kinseth MA, Anjard C, Fuller D, Guizzunti G, Loomis WF, Malhotra V. The Golgi-associated protein GRASP is required for unconventional protein secretion during development. Cell. 2007 Aug 10;130(3):524-34
38) Anjard C, van Bemmelen M, Véron M, Reymond CD. A new spore differentiation factor (SDF) secreted by Dictyostelium cells is phosphorylated by the cAMP dependent protein kinase. Differentiation. 1997 Oct;62(1):43-9.
39) Duran JM, Anjard C, Stefan C, Loomis WF, Malhotra V. Unconventional secretion of Acb1 is mediated by autophagosomes. J Cell Biol. 2010 Feb 22;188(4):527-36.
40) Manjithaya R, Anjard C, Loomis WF, Subramani S. Unconventional secretion of Pichia pastoris Acb1 is dependent on GRASP protein, peroxisomal functions, and autophagosome formation. J Cell Biol. 2010 Feb 22;188(4):537-46.
41) Monteleone M, Stow JL, Schroder K. Mechanisms of unconventional secretion of IL-1 family cytokines. Cytokine. 2015 Aug;74(2):213-8.
42) Rabouille C, Malhotra V, Nickel W. Diversity in unconventional protein secretion. J Cell Sci. 2012 Nov 15;125(Pt 22):5251-5.

43) Schotman H, Karhinen L, Rabouille C. dGRASP-mediated noncanonical integrin secretion is required for Drosophila epithelial remodeling. Dev Cell. 2008 Feb;14(2):171-82.
44) Gee HY, Noh SH, Tang BL, Kim KH, Lee MG. Rescue of ΔF508-CFTR trafficking via a GRASP-dependent unconventional secretion pathway. Cell. 2011 Sep 2;146(5):746-60
45) Yankaskas JR, Marshall BC, Sufian B, Simon RH, Rodman D. Cystic fibrosis adult care: consensus conference report. Chest. 2004; 125:1S-39S.
46) Ward CL, Omura S, Kopito RR. Degradation of CFTR by the ubiquitin-proteasome pathway. Cell. 1995 Oct 6;83(1):121-7.
47) Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000.
48) Barr FA, Puype M, Vandekerckhove J, Warren G. GRASP65, a protein involved in the stacking of Golgi cisternae. Cell. 1997 Oct 17;91(2):253-62
49) Shorter J, Watson R, Giannakou ME, Clarke M, Warren G, Barr FA. GRASP55, a second mammalian GRASP protein involved in the stacking of Golgi cisternae in a cell-free system. EMBO J. 1999 Sep 15;18(18):4949-60
50) Behnia R, Barr FA, Flanagan JJ, Barlowe C, Munro S. The yeast orthologue of GRASP65 forms a complex with a coiled-coil protein that contributes to ER to Golgi traffic. J Cell Biol 2007;176:255–261.
51) Kondylis V, Spoorendonk KM, Rabouille C. dGRASP localization and function in the early exocytic pathway in Drosophila S2 cells. Mol Biol Cell 2005;16:4061–4072.
52) Sutterlin C, Polishchuk R, Pecot M, Malhotra V. The Golgi-associated protein GRASP65 regulates spindle dynamics and is essential for cell division. Mol Biol Cell 2005;16:3211–3222
53) Jarvela T, Linstedt AD. Golgi GRASPs: moonlighting membrane tethers. Cell Health and Cytoskeleton 2012:4 37–47
54) D'Angelo G, Prencipe L, Iodice L, et al. GRASP65 and GRASP55 sequentially promote the transport of C-terminal valine-bearing cargos to and through the Golgi complex. J Biol Chem. 2009;284(50): 34849–34860.
55) Kuo A, Zhong C, Lane W, Derynck R. Transmembrane transforming growth factor-α tethers to the PDZ domain-containing, Golgi membraneassociated protein p59/GRASP55. EMBO J. 2000;19(23):6427–6439.
56) Barr FA, Preisinger C, Kopajtich R, Körner R. Golgi matrix proteins interact with p24 cargo receptors and aid their efficient retention in the Golgi apparatus. J Cell Biol. 2001 Dec 10;155(6):885-91
57) Xiang Y, Wang Y. GRASP55 and GRASP65 play complementary and essential roles in Golgi cisternal stacking. J Cell Biol 2010;188:237–251
58) Bekier ME 2nd, Wang L, Li J, Huang H, Tang D, Zhang X, Wang Y. Knockout of the Golgi stacking proteins GRASP55 and GRASP65 impairs Golgi structure and function. Mol Biol Cell. 2017 Oct 15;28(21):2833-2842.
59) Lane JD, Lucocq J, Pryde J, Barr FA, Woodman PG, Allan VJ, Lowe M. Caspase-mediated cleavage of the stacking protein GRASP65 is required for Golgi fragmentation during apoptosis. J Cell Biol. 2002 Feb 4;156(3):495-509
60) Joshi G, Chi Y, Huang Z, Wang Y. Aβ-induced Golgi fragmentation in Alzheimer's disease enhances Aβ production. Proc Natl Acad Sci U S A. 2014 Apr 1;111(13):E1230-9.
61) Hiyoshi M, Suzu S, Yoshidomi Y, Hassan R, Harada H, Sakashita N, Akari H, Motoyoshi K, Okada S. Interaction between Hck and HIV-1 Nef negatively regulates cell surface expression of M-CSF receptor. Blood. 2008 Jan 1;111(1):243-50.
62) Hiyoshi M, Takahashi-Makise N, Yoshidomi Y, Chutiwitoonchai N, Chihara T, Okada M, Nakamura N, Okada S, Suzu S. HIV-1 Nef perturbs the function, structure, and signaling of the Golgi through the Src kinase Hck. J Cell Physiol. 2012 Mar;227(3):1090-7.

63) Vinke FP, Grieve AG, Rabouille C. The multiple facets of the Golgi reassembly stacking proteins. Biochem J. 2011 Feb 1;433(3):423-33
64) Truschel ST, Sengupta D, Foote A, Heroux A, Macbeth MR, Linstedt AD (2011). Structure of the membrane-tethering GRASP domain reveals a unique PDZ ligand interaction that mediates Golgi biogenesis. J Biol Chem. 10;286(23):20125-9
65) Feng Y, Yu W, Li X, Lin S, Zhou Y, Hu J, Liu X. Structural insight into Golgi membrane stacking by GRASP65 and GRASP55 proteins. J Biol Chem. 2013 Sep 27;288(39):28418-27.
66) Hu F, Shi X, Li B, Huang X, Morelli X, Shi N. Structural basis for the interaction between the Golgi reassembly-stacking protein GRASP65 and the Golgi matrix protein GM130. J Biol Chem. 2015 Oct 30;290(44):26373-82
67) Zhao J, Li B, Huang X, Morelli X, Shi N. Structural Basis for the Interaction between Golgi Reassembly-stacking Protein GRASP55 and Golgin45. J Biol Chem. 2017 Feb 17;292(7):2956-2965.
68) Sengupta D, Linstedt AD. Mitotic inhibition of GRASP65 organelle tethering involves Polo-like kinase 1 (PLK1) phosphorylation proximate to an internal PDZ ligand. J Biol Chem. 2010 Dec 17;285(51):39994-40003.
69) Truschel ST1, Zhang M, Bachert C, Macbeth MR, Linstedt AD. Allosteric regulation of GRASP protein-dependent Golgi membrane tethering by mitotic phosphorylation. J Biol Chem. 2012 Jun 8;287(24):19870-5
70) Bachert C, Linstedt AD. Dual anchoring of the GRASP membrane tether promotes trans pairing. J Biol Chem. 2010 May 21;285(21):16294-301
71) Heinrich F, Nanda H, Goh HZ, Bachert C, Lösche M, Linstedt AD. Myristoylation restricts orientation of the GRASP domain on membranes and promotes membrane tethering. J Biol Chem. 2014 Apr 4;289(14):9683-91
72) Sepkowitz KA. "AIDS—the first 20 years". N Engl J Med. 2001 Jun; 344(23): 1764–72.
73) https://www.cdc.gov/hiv/basics/livingwithhiv/opportunisticinfections.html accessed in 2017
74) Park BJ. Estimation of the current global burden of cryptococcal meningitis among persons living with HIV/AIDS. AIDS 2009;23 (4):525-30.
75) Barron MA, Madinger NE (November 18, 2008). "Opportunistic Fungal Infections, Part 3: Cryptococcosis, Histoplasmosis, Coccidioidomycosis, and Emerging Mould Infections". Infections in Medicine.
76) Gavriilidis G, et al. Cryptococcal Meningitis (CM): Review of Induction Treatment Guidance in Resource-Limited Settings (RLS). ICASA December 2011, Addis Ababa. (TUAB0505)
77) Rapid advice: diagnosis, prevention and management of cryptococcal disease in HIV-infected adults, adolescents and children. WHO Library Cataloguing
78) Kronstad JW, Hu G, Jung WH. An encapsulation of iron homeostasis and virulence in Cryptococcus neoformans. Trends Microbiol. 2013 Sep;21(9):457-65.
79) Feldmesser M, Kress Y, Casadevall A. Dynamic changes in the morphology of Cryptococcus neoformans during murine pulmonary infection. Microbiology. 2001 Aug;147(Pt 8):2355-65.
80) Yoneda A, Doering TL. A eukaryotic capsular polysaccharide is synthesized intracellularly and secreted via exocytosis. Mol Biol Cell. 2006 Dec;17(12):5131-40.
81) Zaragoza O, Rodrigues ML, De Jesus M, Frases S, Dadachova E, Casadevall A. The capsule of the fungal pathogen Cryptococcus neoformans. Adv Appl Microbiol. 2009;68:133-216.
82) Hu G, Steen BR, Lian T, Sham AP, Tam N, Tangen KL, Kronstad JW. Transcriptional regulation by protein kinase A in Cryptococcus neoformans. PLoS Pathog. 2007 Mar;3(3):e42.
83) Panepinto J, Komperda K, Frases S, Park YD, Djordjevic JT, Casadevall A, Williamson PR. Sec6-dependent sorting of fungal extracellular exosomes and laccase of Cryptococcus neoformans. Mol Microbiol. 2009 Mar;71(5):1165-76.
84) Kmetzsch L, Joffe LS, Staats CC, de Oliveira DL, Fonseca FL, Cordero RJ, Casadevall A, Nimrichter L, Schrank A, Vainstein MH, Rodrigues ML (2011). Role for Golgi reassembly and stacking protein (GRASP) in polysaccharide secretion and fungal virulence. Mol Microbiol. 2011 Jul;81(1):206-18.

85) Xu X, Zhao Y, Kirkman E and Lin X. Secreted Acb1 Contributes to the Yeast-to-Hypha Transition in Cryptococcus neoformans. Appl Environ Microbiol. 2015 Dec 4;82(4):1069-79
86) Lin X, Patel S, Litvintseva AP, Floyd A, Mitchell TG, Heitman J. Diploids in the Cryptococcus neoformans serotype A population homozygous for the alpha mating type originate via unisexual mating. PLoS Pathog. 2009 Jan;5(1):e1000283.
87) http://en.wikipedia.org/wiki/Cambrian_explosion accessed in 2017
88) http://en.wikipedia.org/wiki/Cambrian_explosion accessed in 2017
89) http://en.wikipedia.org/wiki/Cambrian_explosion accessed in 2017
90) Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol. 2005 Mar;6(3):197-208.
91) Uversky VN. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. Protein J. 2009 Oct;28(7-8):305-25
92) Peng Z, Xue B, Kurgan L, Uversky VN. Resilience of death: intrinsic disorder in proteins involved in the programmed cell death. Cell Death Differ. 2013 Sep;20(9):1257-67.
93) Uversky VN. Intrinsically disordered proteins from A to Z. Int J Biochem Cell Biol. 2011 Aug;43(8):1090-103.
94) DeForte S, Uversky VN. Order, Disorder, and Everything in Between. Molecules. 2016 Aug 19;21(8). pii: E1090.
95) Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature. 2001 May 3;411(6833):41-2.
96) Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. Trends Biochem Sci. 2008 Jan;33(1):2-8. Epub 2007 Nov 28
97) Fuxreiter M, Tompa P, Simon I, Uversky VN, Hansen JC, Asturias FJ. Malleable machines take shape in eukaryotic transcriptional regulation. Nat Chem Biol. 2008 Dec;4(12):728-37
98) Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model. 2001;19(1):26-59.
99) Tompa P. Structure and function of intrinsically disordered proteins. CRC Press, 2009. Boca Raton, FL
100) Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins. 2000 Nov 15;41(3):415-27.
101) Farquhar, M. G. and Palade, G. E. (1981). The Golgi apparatus (complex)-(1954–1981)-from artifact to center stage. J Cell Biol. 91, 77s–103s.
102) Xiang Y, Zhang X, Nix DB, Katoh T, Aoki K, Tiemeyer M, Wang Y. Regulation of protein glycosylation and sorting by the Golgi matrix proteins GRASP55/65. Nat Commun. 2013;4:1659
103) Lin CY, Madsen ML, Yarm FR, Jang YJ, Liu X, Erikson RL. Peripheral Golgi protein GRASP65 is a target of mitotic polo-like kinase (Plk) and Cdc2. Proc Natl Acad Sci U S A. 2000 Nov 7;97(23):12589-94.
104) Preisinger C, Körner R, Wind M, Lehmann WD, Kopajtich R, Barr FA. Plk1 docking to GRASP65 phosphorylated by Cdk1 suggests a mechanism for Golgi checkpoint signalling. EMBO J. 2005 Feb 23;24(4):753-65.
105) Tang D, Yuan H, Wang Y. The role of GRASP65 in Golgi cisternal stacking and cell cycle progression. Traffic. 2010 Jun;11(6):827-42
106) Bose I, Reese AJ, Ory JJ, Janbon G, Doering TL. A yeast under cover: the capsule of Cryptococcus neoformans. Eukaryot Cell. 2003 Aug;2(4):655-63.
107) Lessells RJ, Mutevedzi PC, Heller T, Newell ML. Poor long-term outcomes for cryptococcal meningitis in rural South Africa. S Afr Med J. 2011 Apr;101(4):251-2.
108) WHO. Revised Global Burden of Disease (GBD) 2002 Estimates. http://www.who.int/healthinfo/bodgbd2002revised/en/index.html

109) Denning DW, Bromley MJ. Infectious Disease. How to bolster the antifungal pipeline. Science. 2015 Mar 27;347(6229):1414-6.
110) Global Action Fund for Fungal Infections, 2015, http://www.gaffi.org
111) Kelley LA, Sternberg MJ. Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc. 2009;4(3):363-71.
112) Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010 Apr;5(4):725-38.
113) Zhang J, Liang Y, Zhang Y. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling.Structure. 2011 Dec 7;19(12):1784-95.
114) Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins. 2005;61 Suppl 7:176-82.
115) Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK. Predicting intrinsic disorder from amino acid sequence. Proteins. 2003;53 Suppl 6:566-72.
116) Yang ZR, Thomson R, McNeil P, Esnouf RM. RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics. 2005 Aug 15;21(16):3369-76.
117) Simossis VA, Heringa J. Integrating protein secondary structure prediction and multiple sequence alignment. Curr Protein Pept Sci. 2004 Aug;5(4):249-66.
118) Rost B, Yachdav G, Liu J. The PredictProtein server. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W321-6.
119) Cole C, Barber JD, Barton GJ. The Jpred 3 secondary structure prediction server. Nucleic Acids Res. 2008 Jul 1;36(Web Server issue):W197-201.
120) Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res. 2005 Jul 1;33(Web Server issue):W72-6.
121) Lees JG, Smith BR, Wien F, Miles AJ, Wallace BA. CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. Anal Biochem. 2004 Sep 15;332(2):285-9
122) Moroz OV, Dodson GG, Wilson KS, Lukanidin E, Bronstein IB. Multiple structural states of S100A12: A key to its functional diversity. Microsc Res Tech. 2003 Apr 15;60(6):581-92.
123) Kellerman G, Vicentin F, Tamura E, Rocha M, Tolentino H, Barbosa A, Craievich A, Torriani I. The small-angle x-ray scattering beamline at the Brazilian synchrotron light laboratory. J App Crystallogr. 1997; 30, 880-883
124) Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HD, Konarev PV, Svergun DI. New developments in the ATSAS program package for small-angle scattering data analysis. J Appl Crystallogr. 2012 Mar 15;45(Pt 2):342-350
125) Svergun, DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr. 1992;25, 495-503
126) Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI. PRIMUS - a Windows-PC based system for small-angle scattering data analysis. J Appl Cryst. 2003;36, 1277-1282
127) Wang Y, Satoh A, Warren G. Mapping the functional domains of the Golgi stacking factor GRASP65. J Biol Chem. 2005 Feb 11;280(6):4921-8
128) Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. J Mol Graph Model. 2001;19(1):26-59.
129) Huang Y, Liu Z. Do intrinsically disordered proteins possess high specificity in protein-protein interactions? Chemistry. 2013 Apr 2;19(14):4462-7.
130) Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. Biochim Biophys Acta. 2005 Aug 10;1751(2):119-39.
131) Woody R, Berova N, Nakanishi K. Circular dichroism: Principles and applications, VCH Publishers, 1994. New York
132) Sreerama N, Woody RW. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem. 2000 Dec 15;287(2):252-60.

133) Sreerama N, Woody RW. Computation and analysis of protein circular dichroism spectra. Methods Enzymol. 2004;383:318-51
134) Whitmore L, Wallace BA. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. Nucleic Acids Res. 2004 Jul 1;32(Web Server issue):W668-73
135) Barr FA, Nakamura N, Warren G. Mapping the interaction between GRASP65 and GM130, components of a protein complex involved in the stacking of Golgi cisternae. EMBO J. 1998 Jun 15;17(12):3258-68.
136) Wang Y, Seemann J, Pypaert M, Shorter J, Warren G. A direct role for GRASP65 as a mitotically regulated Golgi stacking factor. EMBO J. 2003 Jul 1;22(13):3279-90.
137) Erickson HP. Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy. Biol Proced Online. 2009 May 15;11:32-51.
138) Uversky VN. Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 2002 Apr;11(4):739-56.
139) Bychkova VE, Semisotnov GV, Balobanov VA, Finkelstein AV. The Molten Globule Concept: 45 Years Later. Biochemistry (Mosc). 2018 Jan;83(Suppl 1):S33-S47.
140) Siegel LM, Monty KJ. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. Biochim Biophys Acta. 1966 Feb 7;112(2):346-62.
141) Lassing I, Hillberg L, Höglund AS, Karlsson R, Schutt C, Lindberg U. Tropomyosin is a tetramer under physiological salt conditions. Cytoskeleton (Hoboken). 2010 Sep;67(9):599-607.
142) Murayama Y, Uhlmann F1. Biochemical reconstitution of topological DNA binding by the cohesin ring. Nature. 2014 Jan 16;505(7483):367-71.
143) Zasadzińska E, Barnhart-Dailey MC, Kuich PH, Foltz DR. Dimerization of the CENP-A assembly factor HJURP is required for centromeric nucleosome deposition. EMBO J. 2013 Jul 31;32(15):2113-24.
144) Pierce BD, Toptygin D, Wendland B. Pan1 is an intrinsically disordered protein with homotypic interactions. Proteins. 2013 Nov;81(11):1944-63.
145) Baker ES, Luckner SR, Krause KL, Lambden PR, Clarke IN, Ward VK. Inherent structural disorder and dimerisation of murine norovirus NS1-2 protein. PLoS One. 2012;7(2):e30534.
146) Narang P, Bhushan K, Bose S, Jayaram B. A computational pathway for bracketing native-like structures fo small alpha helical globular proteins. Phys Chem Chem Phys. 2005 Jun 7;7(11):2364-75.
147) Gast K, Damaschun H, Misselwitz R, Müller-Frohne M, Zirwer D, Damaschun G. Compactness of protein molten globules: temperature-induced structural changes of the apomyoglobin folding intermediate. Eur Biophys J. 1994;23(4):297-305.
148) Receveur-Bréchot V, Bourhis JM, Uversky VN, Canard B, Longhi S. Assessing protein disorder and induced folding. Proteins. 2006 Jan 1;62(1):24-45.
149) Rambo RP1, Tainer JA. Super-resolution in solution X-ray scattering and its applications to structural systems biology. Annu Rev Biophys. 2013;42:415-41.
150) Glatter O, Kratky O. Small Angle X-Ray Scattering. Academic Press, 1982. London.
151) Uversky VN. What does it mean to be natively unfolded? Eur J Biochem. 2002 Jan;269(1):2-12.
152) Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers. 2011 Aug;95(8):559-71.
153) Rambo RP, Tainer JA. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. Biopolymers. 2011 Aug;95(8):559-71.
154) Lakowicz JR. Principles of fluorescence spectroscopy. 3rd. New York: Springer. 2006. 26, 954 p. ISBN 9780387312781.
155) Baldwin RL, Rose GD. Molten globules, entropy-driven conformational change and protein folding. Curr Opin Struct Biol. 2013 Feb;23(1):4-10.
156) Uversky VN. Unusual biophysics of intrinsically disordered proteins. Biochim Biophys Acta. 2013 May;1834(5):932-51

157) Gasymov OK, Glasgow BJ. ANS fluorescence: potential to augment the identification of the external binding sites of proteins. Biochim Biophys Acta. 2007 Mar;1774(3):403-11
158) Garcia AF, Garcia W, Nonato MC, Araújo AP. Structural stability and reversible unfolding of recombinant porcine S100A12. Biophys Chem. 2008 May;134(3):246-53
159) Fontana A, Zambonin M, Polverino de Laureto P, De Filippis V, Clementi A, Scaramella E. Probing the conformational state of apomyoglobin by limited proteolysis. J Mol Biol. 1997 Feb 21;266(2):223-30
160) Rawlings ND, Barrett AJ. Families of serine peptidases. Methods Enzymol. 1994;244:19-61.
161) Redfield C. Using nuclear magnetic resonance spectroscopy to study molten globule states of proteins. Methods. 2004 Sep;34(1):121-32.
162) Park SJ, Borin BN, Martinez-Yamout MA, Dyson HJ. The client protein p53 adopts a molten globule-like state in the presence of Hsp90. Nat Struct Mol Biol. 2011 May;18(5):537-41.
163) Schulman BA, Kim PS, Dobson CM, Redfield C. A residue-specific NMR view of the non-cooperative unfolding of a molten globule. Nat Struct Biol. 1997 Aug;4(8):630-4.
164) Kwan AH1, Mobli M, Gooley PR, King GF, Mackay JP. Macromolecular NMR spectroscopy for the non-spectroscopist. FEBS J. 2011 Mar;278(5):687-703.
165) Wang Y, Shortle D. A dynamic bundle of four adjacent hydrophobic segments in the denatured state of staphylococcal nuclease. Protein Sci. 1996 Sep;5(9):1898-906.
166) Giuliani F, Grieve A, Rabouille C. Unconventional secretion: a stress on GRASP. Curr Opin Cell Biol. 2011 Aug;23(4):498-504.
167) Tang D, Zhang X, Huang S, Yuan H, Li J, Wang Y. Mena-GRASP65 interaction couples actin polymerization to Golgi ribbon linking. Mol Biol Cell. 2016 Jan 1;27(1):137-52.
168) Mironov AA, Pavelka M. The Golgi Apparatus: State of the art 110 years after Camillo Golgi's discovery, Springer-Verlag Wien, 1° edition, 2008
169) Klumperman J. Architecture of the mammalian Golgi, Cold Spring Harb Perspect Biol. 3:a005181
170) Day KJ, Staehelin LA, Glick BS. A three-stage model of Golgi structure and function. Histochem Cell Biol. 2013 Sep;140(3):239-49
171) Banfield DK. Mechanisms of protein retention in the Golgi. Cold Spring Harb Perspect Biol. 2011 Aug 1;3(8):a005264.
172) Jarvela T, Linstedt AD. Isoform-specific tethering links the Golgi ribbon to maintain compartmentalization. Mol Biol Cell. 2014 Jan;25(1):133-44
173) Veenendaal T, Jarvela T, Grieve AG, van Es JH, Linstedt AD, Rabouille C. GRASP65 controls the cis Golgi integrity in vivo. Biol Open. 2014 May 2;3(6):431-43
174) Mendes LF, Garcia AF, Kumagai PS, de Morais FR, Melo FA, Kmetzsch L, Vainstein MH, Rodrigues ML, Costa-Filho AJ. New structural insights into Golgi Reassembly and Stacking Protein (GRASP) in solution. Sci Rep. 2016 Jul 20;6:29976
175) Puthenveedu MA, Bachert C, Puri S, Lanni F, Linstedt AD. GM130 and GRASP65-dependent lateral cisternal fusion allows uniform Golgi-enzyme distribution. Nat Cell Biol. 2006 Mar;8(3):238-48
176) Short B, Preisinger C, Körner R, Kopajtich R, Byron O, Barr FA. A GRASP55-rab2 effector complex linking Golgi structure to membrane traffic. J Cell Biol. 2001 Dec 10;155(6):877-83
177) Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015 Jan;16(1):18-29
178) Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. The roles of intrinsic disorder in protein interaction networks. FEBS J. 2005 Oct;272(20):5129-48
179) Alberti S, Halfmann R, King O, Kapila A, Lindquist S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell. 2009 Apr 3;137(1):146-58
180) Mukhopadhyay S, Krishnan R, Lemke EA, Lindquist S, Deniz AA. A natively unfolded yeast prion monomer adopts an ensemble of collapsed and rapidly fluctuating structures. Proc Natl Acad Sci U S A. 2007 Feb 20;104(8):2649-54
181) Dobson CM. The Amyloid Phenomenon and Its Links with Human Disease. Cold Spring Harb Perspect Biol. 2017 Jun 1;9(6)

182) Mendoza-Espinosa P, García-González V, Moreno A, Castillo R, Mas-Oliva J. Disorder-to-order conformational transitions in protein structure and its relationship to disease. Mol Cell Biochem. 2009 Oct;330(1-2):105-20
183) Skora L, Becker S, Zweckstetter M. Molten globule precursor states are conformationally correlated to amyloid fibrils of human beta-2-microglobulin. J Am Chem Soc. 2010 Jul 14;132(27):9223-5
184) Kisilevsky R. Amyloids: Tombstones or triggers? Nature Medicine. 2000; 6, 633 - 634
185) Uversky VN. Amyloidogenesis of natively unfolded proteins. Curr Alzheimer Res. 2008 Jun;5(3):260-87
186) Prajapati RS, Indu S, Varadarajan R. Identification and thermodynamic characterization of molten globule states of periplasmic binding proteins. Biochemistry. 2007 Sep 11;46(36):10339-52
187) Vamvaca K, Vögeli B, Kast P, Pervushin K, Hilvert D. An enzymatic molten globule: efficient coupling of folding and catalysis. Proc Natl Acad Sci U S A. 2004 Aug 31;101(35):12860-4
188) Ramboarina S, Redfield C. Probing the effect of temperature on the backbone dynamics of the human alpha-lactalbumin molten globule. J Am Chem Soc. 2008 Nov 19;130(46):15318-26
189) Papworth C, Bauer JC, Braman J, Wright DA. Site-directed mutagenesis in one day with >80 % efficiency. Strategies. 1996;9:3–4
190) Uversky VN, Narizhneva NV, Kirschstein SO, Winter S, Löber G. Conformational transitions provoked by organic solvents in beta-lactoglobulin: can a molten globule like intermediate be induced by the decrease in dielectric constant? Fold Des. 1997;2(3):163-72
191) Ellis RJ. Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci. 2001 Oct;26(10):597-604
192) Soranno A, Koenig I, Borgia MB, Hofmann H, Zosel F, Nettels D, Schuler B. Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. Proc Natl Acad Sci U S A. 2014 Apr 1;111(13):4874-9
193) Cheung MS, Klimov D, Thirumalai D. Molecular crowding enhances native state stability and refolding rates of globular proteins. Proc Natl Acad Sci U S A. 2005 Mar 29;102(13):4753-8
194) Szasz CS, Alexa A, Toth K, Rakacs M, Langowski J, Tompa P. Protein disorder prevails under crowded conditions. Biochemistry. 2011 Jul 5;50(26):5834-44
195) Uversky VN, Li J, Fink AL. Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. A possible molecular NK between Parkinson's disease and heavy metal exposure. J Biol Chem. 2001 Nov 23;276(47):44284-96
196) Uversky VN, Li J, Fink AL. Trimethylamine-N-oxide-induced folding of alpha-synuclein. FEBS Lett. 2001 Nov 30;509(1):31-5
197) Zeraik AE, Staykova M, Fontes MG, Nemuraitė I, Quinlan R, Araújo AP, DeMarco R. Biophysical dissection of schistosome septins: Insights into oligomerization and membrane binding. Biochimie. 2016 Dec;131:96-105
198) Rabouille C, Linstedt AD. GRASP: A Multitasking Tether. Front Cell Dev Biol. 2016 Jan 26;4:1
199) Wagner LE 2nd, Li WH, Joseph SK, Yule DI. Functional consequences of phosphomimetic mutations at key cAMP-dependent protein kinase phosphorylation sites in the type 1 inositol 1,4,5-trisphosphate receptor. J Biol Chem. 2004 Oct 29;279(44):46242-52
200) Fraser JA, Worrall EG, Lin Y, Landre V, Pettersson S, Blackburn E, Walkinshaw M, Muller P, Vojtesek B, Ball K, Hupp TR. Phosphomimetic mutation of the N-terminal lid of MDM2 enhances the polyubiquitination of p53 through stimulation of E2-ubiquitin thioester hydrolysis. J Mol Biol. 2015 Apr 24;427(8):1728-47
201) Ingrell CR, Miller ML, Jensen ON, Blom N. NetPhosYeast: prediction of protein phosphorylation sites in yeast. Bioinformatics. 2007 Apr 1;23(7):895-7
202) Fernández A, Scheraga HA. Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. Proc Natl Acad Sci U S A. 2003 Jan 7;100(1):113-8
203) Fernández A, Scott R. Dehydron: a structurally encoded signal for protein interaction. Biophys J. 2003 Sep;85(3):1914-28

204) Koshland DE. Application of a Theory of Enzyme Specificity to Protein Synthesis. Proc Natl Acad Sci U S A. 1958 Feb;44(2):98-104
205) Changeux JP, Edelstein S. Conformational selection or induced fit? 50 years of debate resolved. F1000 Biol. 2011; Rep. 3, 19
206) Ramírez J, Recht R, Charbonnier S, Ennifar E, Atkinson RA, Travé G, Nominé Y, Kieffer B. Disorder-to-order transition of MAGI-1 PDZ1 C-terminal extension upon peptide binding: thermodynamic and dynamic insights. Biochemistry. 2015 Feb 17;54(6):1327-37
207) Belle V, Rouger S, Costanzo S, Liquière E, Strancar J, Guigliarelli B, Fournel A, Longhi S. Mapping alpha-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. Proteins. 2008 Dec;73(4):973-88
208) Yoneda JS, Miles AJ, Araujo AP, Wallace BA. Differential dehydration effects on globular proteins and intrinsically disordered proteins during film formation. Protein Sci. 2017 Apr;26(4):718-726
209) Cluett EB, Brown WJ. Adhesion of Golgi cisternae by proteinaceous interactions: intercisternal bridges as putative adhesive structures. J Cell Sci. 1992 Nov;103 (Pt 3):773-84
210) Bychkova VE, Basova LV, Balobanov VA. How membrane surface affects protein structure. Biochemistry (Mosc). 2014 Dec;79(13):1483-514
211) Endo T, Schatz G. Latent membrane perturbation activity of a mitochondrial precursor protein is exposed by unfolding. EMBO J. 1988 Apr;7(4):1153-8
212) Ptitsyn OB, Bychkova VE, Uversky VN. Kinetic and equilibrium folding intermediates. Philos Trans R Soc Lond B Biol Sci. 1995; 348(1323):35-41
213) Cherepanov DA, Feniouk BA, Junge W, Mulkidjanian AY. Low dielectric permittivity of water at the membrane interface: effect on the energy coupling mechanism in biological membranes. Biophys J. 2003 Aug;85(2):1307-16.
214) Chao H, Martin GG, Russell WK, Waghela SD, Russell DH, Schroeder F, Kier AB. Membrane charge and curvature determine interaction with acyl-CoA binding protein (ACBP) and fatty acyl-CoA targeting. Biochemistry. 2002 Aug 20;41(33):10540-53
215) Eggers DK, Valentine JS. Crowding and hydration effects on protein conformation: a study with sol-gel encapsulated proteins. J Mol Biol. 2001 Dec 7;314(4):911-22
216) Kumagai PS, DeMarco R, Lopes JL. Advantages of synchrotron radiation circular dichroism spectroscopy to study intrinsically disordered proteins. Eur Biophys J. 2017 Mar 3
217) Munishkina LA, Phelan C, Uversky VN, Fink AL. Conformational behavior and aggregation of alpha-synuclein in organic solvents: modeling the effects of membranes. Biochemistry. 2003 Mar 11;42(9):2720-30
218) Prasanna Kumari NK, Jagannadham MV. Deciphering the molecular structure of cryptolepain in organic solvents. Biochimie. 2012 Feb;94(2):310-7
219) Song M, Shao H, Mujeeb A, James TL, Miller WL. Molten-globule structure and membrane binding of the N-terminal protease-resistant domain (63-193) of the steroidogenic acute regulatory protein (StAR). Biochem J. 2001 May 15;356(Pt 1):151-8
220) Bychkova VE, Pain RH, Ptitsyn OB. The 'molten globule' state is involved in the translocation of proteins across membranes? FEBS Lett. 1988 Oct 10;238(2):231-4
221) Faudry E, Job V, Dessen A, Attree I, Forge V. Type III secretion system translocator has a molten globule conformation both in its free and chaperone-bound forms. FEBS J. 2007 Jul;274(14):3601-10
222) Bychkova VE, Dujsekina AE, Klenin SI, Tiktopulo EI, Uversky VN, Ptitsyn OB. Molten globule-like state of cytochrome c under conditions simulating those near the membrane surface. Biochemistry. 1996 May 14;35(19):6058-63
223) Lopes JL, Yoneda JS, Martins JM, DeMarco R, Jameson DM, Castro AM, Bossolan NR, Wallace BA, Araujo AP. Environmental Factors Modulating the Stability and Enzymatic Activity of the Petrotoga mobilis Esterase (PmEst). PLoS One. 2016 Jun 28;11(6):e0158146
224) Teschke O, Ceotto G, de Souza EF. Interfacial water dielectric-permittivity-profile measurements using atomic force microscopy. Phys Rev E Stat Nonlin Soft Matter Phys. 2001 Jul;64(1 Pt 1):011605

225) Uversky VN, Winter S, Löber G. Use of fluorescence decay times of 8-ANS-protein complexes to study the conformational transitions in proteins which unfold through the molten globule state. Biophys Chem. 1996 Jun 11;60(3):79-88
226) Alexandrescu AT, Ng YL, Dobson CM. Characterization of a trifluoroethanol-induced partially folded state of alpha-lactalbumin. J Mol Biol. 1994 Jan 14;235(2):587-99
227) Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. 6(3):277-93
228) Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins. Jun 1;59(4):687-96.
229) Long J, Wei Z, Feng W, Yu C, Zhao YX, Zhang M (2008). Supramodular nature of GRIP1 revealed by the structure of its PDZ12 tandem in complex with the carboxyl tail of Fras1. J Mol Biol. Feb 1;375(5):1457-68
230) Zhang Q, Fan JS, Zhang M. (2001). Interdomain chaperoning between PSD-95, Dlg, and Zo-1 (PDZ) domains of glutamate receptor-interacting proteins. J Biol Chem. Nov 16;276(46):43216-20
231) Marion D, Driscoll PC, Kay LE, Wingfield PT, Bax A, Gronenborn AM, Clore GM (1989) Overcoming the overlap problem in the assignment of H-1-NMR spectra of larger proteins by use of 3-dimensional heteronuclear H-1-N-15 hartmann-hahn multiple quantum coherence and nuclear overhauser multiple quantum coherence spectroscopy—application to interleukin-1-Beta. Biochemistry 28:6150–6156
232) Messerle BA, Wider G, Otting G, Weber C, Wüthrich K (1989) Solvent suppression using a spin lock in 2D and 3D NMR-spectroscopy with H2O solutions. J Magn Reson 85:608–613
233) Greene LH, Wijesinha-Bettoni R, Redfield C (2006). Characterization of the molten globule of human serum retinol-binding protein using NMR spectroscopy. Biochemistry. Aug 8;45(31):9475-84.
234) Berliner L. Protein NMR Modern Techniques and Biomedical Applications. Biological Magnetic Resonance. 2015. Volume 32, Springer.
235) Wüthrich K, Wider G, Wagner G, Braun W (1982) Sequential resonance assignments as a basis for determination of spatial protein structures by high-resolution proton nuclear magnetic-resonance. J Mol Biol 155:311–319
236) Pervushin K, Riek R, Wider G, Wüthrich K (1997). Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci U S A 94:12366–12371
237) Vassilenko KS, Uversky VN (2002). Native-like secondary structure of molten globules. Biochim Biophys Acta. Jan 31;1594(1):168-77.
238) Felli IC, Pierattelli R. Intrinsically Disordered Proteins Studied by NMR Spectroscopy. Advances in Experimental Medicine and Biology. 870