

## ABSTRACT

Title of Dissertation:                   DECIPHERING HOW EGFR-GRB2-SOS1  
COMPLEX REGULATES KRAS4B  
ACTIVATION AND LEADS TO HIPPO  
SIGNALING THROUGH RASSF5

Tsung-Jen Liao, Doctor of Philosophy, 2021

Dissertation directed by:           Professor, Dr. David Fushman, Department of  
Chemistry and Biochemistry, University of  
Maryland, College Park  
Professor, Dr. Ruth, Nussinov, National Cancer  
Institute, National Institutes of Health

Ras is a small GTPase, which regulates cell proliferation and apoptosis. Its bifunctional switch is controlled by the nucleotide state: GTP-bound – switch on; GDP-bound – switch off. Among the three Ras isoforms, HRas, NRas, and KRas (with two splice variants of KRas4A and KRas4B), KRas4B is highly oncogenic, the most frequently mutated in lung, colorectal, and pancreatic cancers. However, Ras was thought to be “undruggable” due to the lack of effective pharmacological inhibitors over the past three decades. Most of the current focus has been directed at inhibiting the activation of Ras signaling. Ras proteins transduce signals between cell surface receptors and multiple intracellular signaling cascades. In response to epidermal growth factor receptor (EGFR) activation, growth factor receptor bound protein 2 (Grb2) establishes the connection between EGFR and Ras-specific nucleotide

exchange factor (RasGEF), son of sevenless 1 (SOS1). SOS1 activates Ras by exchanging GDP to GTP. In addition to Ras major effectors and pathways, e.g. MAPK and PI3K $\alpha$ /Akt, which are cell growth related, GTP-bound Ras associating with RASSF5 activates the Hippo pathway, which acts to suppress cell proliferation. In this serial study, we use NMR measurement and molecular dynamics (MD) simulation to investigate the interactions of Grb2–SOS1, SOS1–KRas4B, KRas4B–RASSF5, and the EGFR effects on the binding of Grb2–SOS1. Our findings successfully uncovered (1) a novel Grb2 binding site PKLPPKTYKREH on SOS1 and the most probable binding mode of Grb2–SOS1, (2) strong SOS1 peptide binders induce a closed conformation of Grb2 nSH3 domain but unchanged conformation of Grb2 cSH3 domain, (3) full length Grb2 performs high affinities for one-site SOS1 peptides, and the EGFR segment may facilitate the binding of Grb2 to the particular two-site SOS1 peptide, (4) KRas4B binding to SOS1 allosteric site induces the conformational changes of catalytic site and accelerate the KRas4B activation cycle, (5) the hypothesized mechanism that RASSF5 is a tumor suppressor in vivo but opposite in vitro, and (6) the dynamic mechanism of RASSF5 auto-inhibition. Our effort in elucidating the mechanism of Ras and Ras effectors results in 8 publications and offers a new venues for future therapeutic strategies.

DECIPHERING HOW EGFR-GRB2-SOS1 COMPLEX REGULATES KRAS4B  
ACTIVATION AND LEADS TO HIPPO SIGNALING THROUGH RASSF5

by

Tsung-Jen Liao

Dissertation submitted to the Faculty of the Graduate School of the  
University of Maryland, College Park, in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
2021

Advisory Committee:  
Professor David Fushman, Chair  
Professor Ruth Nussinov  
Professor Segei Sukharev  
Professor Garegin Papoian  
Assistant Professor Pratyush Tiwary

© Copyright by  
Tsung-Jen Liao  
2021

## Preface

*RAS* genes (*HRAS*, *NRAS*, and *KRAS*) were first identified in human cancer cells in 1980s. Expression of *RAS* gene activates MAPK and PI3K pathways, inducing cell proliferation, differentiation, and survival. Mutations of *RAS* gene usually render the continuous MAPK/PI3K signaling and result in cell overgrowth – cancer. The cancer database, catalogue of somatic mutations in cancer (COSMIC), indicates that oncogenic Ras mutants participate 20 to 25% of human malignant tumors. Moreover, KRas mutations are the most frequently observed isoform (22%) in cancer compared to HRas (8%) and NRas (3%). Targeting Ras has become one of the most significant task in cancer therapy.

While Ras has been considered undruggable over decades, recent progress in targeting oncogenic Ras invigorated the Ras research community. On May 28, 2021, the U.S. food and drug administration approved Lumakras (sotorasib) as the first targeted therapy for the genetic mutation KRas G12C, which aims to treat the non-small cell lung cancer. The direct inhibition of KRas mutant is a huge step forward and raising the hope of future therapeutics for Ras-induced cancer.

Oncogenic KRas involve multiple mutations such as G12C, G12V, G12D, G13C, G13D, A18D, Q61H, and K117N. These mutations perform various biological functions and contribute to different cancer diseases. Therefore, beside the Ras direct inhibition, more efforts turned to search the alternative targets from Ras effectors. In this study, we investigated the serial interaction of EGFR-Grb2, Grb2-SOS1, SOS1-KRas4B, and KRas4B-RASSF5 by using NMR experiments and MD simulation. Our

effort in deciphering the interaction mechanism of Ras upstream and downstream effectors provides an insight for future drugging strategies.

Our past work has yielded 8 publications, which are listed below and following the reverse-chronological order:

1. **T.J. Liao**, H. Jang, R. Nussinov, D. Fushman: *NMR data indicate Grb2 nSH3/cSH3 domains high affinity for SOS1 proline-rich domain motifs*. 2021. Drafted and will submit it soon.
2. **T.J. Liao**, H. Jang, D. Fushman, R. Nussinov: *SOS1 interacts with Grb2 through regions that induce closed nSH3 conformations*. J. Chem. Phys., 2020, 153(4), 045106.
3. **T.J. Liao**, H. Jang, R. Nussinov, D. Fushman: *High-affinity Interactions of the nSH3/cSH3 Domains of Grb2 with the C-terminal Proline-rich Domain of SOS1*. J. Am. Chem. Soc., 2020, 142 (7), 3401-3411.
4. R. Nussinov, M. Zhang, CJ. Tsai, **T.J. Liao**, D. Fushman, H. Jang: *Autoinhibition in Ras effectors Raf, PI3K $\alpha$ , and RASSF5: A comprehensive review underscoring the challenges in pharmacological intervention*. Biophysical reviews, 2018, 1-20.
5. **T.J. Liao**, H. Jang, D. Fushman, R. Nussinov: *Allosteric KRas4B can modulate SOS1 fast and slow activation cycles*. Biophysical Journal, 2018, 115 (4), 629-641.
6. **T.J. Liao**, H. Jang, CJ. Tsai, D. Fushman, R. Nussinov: *The dynamics mechanism of RASSF5 and MST kinase activation by Ras*. Physical Chemistry Chemical Physics, 2017, 19, 6470-6480.
7. R. Nussinov, H. Jang, CJ. Tsai, **T.J. Liao**, S. Li, D. Fushman & J. Zhang: *Intrinsic protein disorder in oncogenic KRas signaling*. Cellular and Molecular Life Science, 2017, 74(17):3245-3261.

8. **TJ. Liao**, CJ. Tsai, H. Jang, D. Fushman, R. Nussinoc: *RASSF5: An MST activator and tumor suppressor in vivo but opposite in vitro*. Current Opinion in Structural Biology, 2016, 41:217-224.

## Acknowledgements

Time is fleeting and is approaching to the end of being a PhD student. Studying in our biophysics program at University of Maryland is the most meaningful period of my life and the most treasurable memory of mine. In the past seven years, I received many helps from many people. Without any of you, I wouldn't have any chances to get through the tough time and survive.

I would like to thank my advisers Dr. Ruth Nussinov and Dr. David Fushman first. In the spring of 2015, Ruth was invited to give a talk in our biophysics weekly seminar, and David was the host. Before the end of the semester, I visited Ruth's group and told her I want to publish many papers and become a superstar in academia (no matter how many years have passed, I feel so embarrassed whenever I recall that scene). In the fall of 2015, I started my research in Ruth's group and was assigned the project of RASSF5. The first lesson that Ruth taught me is the importance of making a good presentation. In order to apply the fellowship of NCI-UMD partnership in the upcoming year, we were seeking a chance to collaborate with David. I could clearly recall the day before discussion. Ruth sat with me and modified the slides I prepared. She told me that "people in academia are very busy, but today, they are willing to listen to your talk. There is no reason not to prepare your presentation well. You have to cherish every chance of giving a talk and respect your audience." These words always remind me of making perfect slides for each presentation.

Unfortunately, in my first application for the fellowship, our proposal was not accepted. Ruth always encouraged me and said "Don't be frustrated. This is very common for all the scientists. We just need to do it again." In order to draft a better

proposal for the next application, we aimed to add some experimental data. Luckily, we received the plasmids of RASSF5 from the foreign groups. I started to prepare the protein samples for NMR experiments. In the meantime, Ruth and I participated many of the seminars held by NCI-UMD partnership and tried to build the connections with Dr. Wolfgang Losert (the head at UMD) and Dr. Daniel Larson (the head at NCI). Even though we finally won the fellowship in the following year, I think making a robust social connection might be more effective than making a good proposal. Anyway, my limited words are unable to express how grateful I am for everything Ruth did and everything I learned from her. Your profound insight and strong self-discipline deeply impact me. I wish I could stay and work longer for you.

I am very lucky having a great opportunity to learn NMR spectroscopy in David's lab. Even though about 95% of time I was struggling in the protein expression and purification, I still pretty enjoyed the rest 5% of time doing simple NMR measurements. NMR is a super powerful tool. I remember in the class, David said "NMR is the only experimental tool in the world to investigate the details of protein dynamics." I also clearly remember the day that David sat with me in the NMR room and showed me how to distinguish the signals of W193ε and W194ε of cSH3 and how we can label all the signals through NOESY and TOCSY. That was an unforgettable afternoon. I appreciate David that showed me the beauty and magic of NMR even though I didn't go far down the path.

I would like to thank Dr. Sergei Sukharev. In my first semester, I only received a half teaching assistantship from the physics department due to my poor language. As our biophysics director, Sergei was willing to offer me a half research assistantship to

compensate my tuition and stipend while I almost knew nothing about biochemistry at that time. One day, when I was going to pick up my cell stocks in the centrifuge, I found the rotor was broken in pieces. I was extremely terrified; however, Sergei didn't blame me but told me "This happens" and "We have money to fix it". Sergei is a very kind person, and he set up an example for me – we should always tolerate people's mistake. Because of this event, I was extremely careful in David's lab. There is no instrument that would be ruined with my hands ever.

I want to especially thank Dr. Hyunbum Jang, a true grand master of science and art. Hyunbum is a great teacher, a decent friend, and an erudite person. He built my understanding of the fundamental molecular dynamics and showed me the ways to investigate protein-protein interactions. The most important lesson I learned from Hyunbum is how to make beautiful figures. At the beginning of the research work in Ruth's group, one of the most common sentences I heard from him is "This doesn't look good.", and Ruth never stood by me. Sometimes I felt he wanted to train me into an artist instead of a scientist. However, high quality figures indeed promote the quality of a paper. I truly benefit from this skill in each of my following presentation and publication. Hyunbum's great painting skills also largely enhanced my aesthetic sense. I am very grateful for having Hyunbum's training and assistance during these years.

I want to thank the lab members in Ruth's and David's groups. Dr. Buyong Ma, Dr. Chung-Jung Tsai, and Dr. Mingzhen Zhang always gave me many useful suggestions to improve my research design and direction. My colleagues Andrew Boughton and Betsegaw Lemma gave me huge help and saved me from the endless biochemistry experiments. Shall every of you have a bright and successful future.

In the last, I want to thank my parents Shu-Hung Liao, Li-Chen Cheng, and my brother Tsung-Han Liao for your warmest supports. Leaving 8,000 miles from home did not separate us apart but tightly connected our hearts. Your voices always warmed me up in the countless nights and gave me energy to keep on going. Besides, I want to thank my beloved wife Jingwen Zhao for your care, company, and loving support. We have gone through the severe pandemic of 2020 together. This is a very special journey. I owe you a formal wedding as I promised.

I sincerely appreciate every person I met, every lesson I learned, and everything I experienced. They all mean a lot to me and ultimately shape me a scientist. I am proud of ever being a member of two top groups of the world. Thanks to all the help of all of you.

# Table of Contents

Preface.....	ii
Acknowledgements.....	v
Table of Contents.....	ix
Chapter 1: Introduction.....	1
Chapter 2: High-affinity interactions of the nSH3/cSH3 domains of Grb2 with the C-terminal proline-rich domain of SOS1.....	6
2.1 Introduction to chapter 2.....	6
2.2 Results.....	10
2.2.1 Candidate nSH3/cSH3 binding sites on SOS1 PR domain.....	10
2.2.2 The binding interface on nSH3/cSH3 domains.....	11
2.2.3 The nSH3 shows stronger affinity for PxxPxR motif.....	14
2.2.4 S4 and S10 display strong and comparable affinity for cSH3.....	17
2.2.5 Construction of Grb2-SOS1 complex models.....	21
2.3 Discussions.....	25
2.4 Material and Methods.....	29
2.4.1 Sample preparation and NMR measurements.....	29
2.4.2 Grb2-SOS1 modeling and simulation.....	30
Chapter 3: SOS1 interacts with Grb2 through regions that induce closed nSH3 conformations.....	32
3.1 Introduction to chapter 3.....	32
3.2 Results.....	35
3.2.1 Simulations of nSH3/cSH3–SOS1 complexes define the binding interfaces.....	35
3.2.2 The effective free energy validates the multiple binding sites on SOS1 PR domain.....	37
3.2.3 The nSH3 n-Src loop strengthens the interaction with SOS1, inducing closed nSH3 conformation.....	40
3.2.4 No conformational change of cSH3 upon binding to SOS1.....	43
3.3 Discussion.....	46
3.4 Methods.....	49
3.4.1 Generating initial configurations of nSH3/cSH3–SOS1 PR segments.....	49
3.4.2 Atomistic molecular dynamics simulations.....	49
3.4.3 Effective free energy calculations.....	50
Chapter 4: NMR data indicate Grb2 nSH3/cSH3 domains high affinity for SOS1 proline-rich domain motifs.....	52
4.1 Introduction to chapter 4.....	52
4.2 Results.....	55
4.2.1 The one-site and two-site segments of SOS1 PR domain.....	55
4.2.2 Grb2 nSH3 and cSH3 have high affinities for S1, S2, S3, S4, S5.....	57
4.2.3 EGFR segment has limited effect on Grb2 binding to S1, S2, S3, S4, S5.....	62
4.2.4 Grb2 interacting with S12, S23, and S45 exhibits different complex conformations.....	70
4.2.5 EGFR segment may strengthen the association of Grb2-S45.....	74

4.3 Discussion .....	78
4.4 Material and Methods .....	82
4.4.1 Protein expression and purification .....	82
4.4.2 NMR spectroscopy.....	82
4.4.3 Transverse <sup>15</sup> N spin-relaxation measurement .....	84
4.4.4 Computational modeling of EGFR–Grb2–SOS1 complex.....	85
Chapter 5: Allosteric KRas4B can modulate SOS1 fast and slow Ras activation cycle .....	86
5.1 Introduction to chapter 5 .....	86
5.2 Results.....	91
5.2.1 The C-terminal catalytic region of SOS1 exhibits conformational changes upon binding to KRas4B.....	91
5.2.2 Competing interactions of KRas4B at the SOS1 binding sites.....	97
5.2.3 Conformational changes of KRas4B at the SOS1 catalytic site during activation.....	100
5.2.4 Allosteric signaling from KRas4B at the allosteric site to that at catalytic site in SOS1 activation.....	103
5.3 Discussion .....	106
5.4 Material and Methods .....	110
5.4.1 Generating initial configurations of KRas4B-SOS1 complex.....	110
5.4.2 Atomistic molecular dynamics simulations .....	111
5.4.3 Binding free energy calculation .....	112
Chapter 6: RASSF5 is an MST activator and tumor suppressor in vivo but opposite in vitro .....	115
6.1 Introduction to chapter 6.....	115
6.2 Hypothesis.....	117
6.2.1 RASSF5 is an adaptor protein.....	117
6.2.2 How activated RASSF5 can act to activate MST1/2 and suppress cancer in vivo .....	122
6.2.3 The similar roles of MAPK and Hippo signaling in the cell cycle.....	125
6.3 Discussion .....	128
Chapter 7: The dynamic mechanism of RASSF5 and MST kinase activations by KRas4B .....	129
7.1 Introduction to chapter 7.....	129
7.2 Results.....	134
7.2.1 Strong interaction of the RASSF5-MST2 SARAH heterodimer compared to the SARAH homodimers.....	134
7.2.2 A kinked helical motif predominantly observed in the RASSF5 SARAH monomer .....	137
7.2.3 The self-associated RASSF5 conformation in the inactive state .....	140
7.2.4 RASSF5 in complex with GTP-bound KRas4B .....	147
7.3 Discussion .....	154
7.4 Materials and Methods.....	158
Chapter 8: Conclusion.....	162
Appendices.....	164
Appendix A.....	164

Appendix B .....	173
Appendix C .....	181
Appendix D .....	221
Appendix E .....	232
Bibliography .....	249

## Chapter 1: Introduction

Ras is a small GTPase, which switches between the active GTP-bound and inactive GDP-bound states and participates in the regulation of transduces multiple intracellular signaling cascades, regulating cell proliferation and survival. Ras is activated by the Ras-specific guanine nucleotide exchange factor (RasGEF) and deactivated by GTPase-activating protein (GAP). Mutations in Ras disrupt the GTP/GDP exchange cycle and lock Ras in the active GTP-bound state, resulting in the constantly downstream signaling and leading to malignant tumor. Oncogenic mutations of Ras proteins were reported approximately 24% in all cancers. Ras family includes three main isoforms: HRas, NRas, and KRas. KRas has two splice variants: KRas4A and KRas4B. Among the Ras family members, KRas4B is recognized the most frequently oncogenic Ras, which results in 86% pancreatic cancer, 41% colorectal cancer, and 32% lung cancer. Therefore, targeting oncogenic Ras is of particular importance in Ras-driven cancer. However, pharmaceutical approaches to directly target oncogenic Ras have not been successful. Alternatively, most efforts focus on inhibiting the Ras upstream and/or the two main downstream signaling pathways: mitogen-activated protein kinase (MAPK, Raf/MEK/ERK), and phosphatidylinositide 3-kinase (PI3K)/Akt/mTOR pathways. An additional Ras downstream signaling, Hippo pathway, has recently spurred increasing attention. MAPK signaling promotes cell proliferation whereas Hippo's abolishes it. The cross-talk between critical pathways such as MAPK and Hippo points to the opportunity to exploit it in the search for targets to Ras-induced cancers. Our aim is to reveal how KRas4B is activated by the upstream signaling from cell membrane and regulates the Hippo pathway. Accomplishing the aims of KRas4B activation will provide a broad view on tumor proliferation, helping drug discovery efforts.

Ras signaling initiates at the plasma membrane. When the transmembrane epidermal growth factor receptor (EGFR) receives the signal of EGF, EGFR dimerizes and undergoes its kinase domain self-phosphorylation. The phosphorylated EGFR segment is capable of grasping growth factor receptor-bound protein 2 (Grb2). Grb2 binds to the RasGEF, son of sevenless 1 (SOS1), and brings SOS1 to plasma membrane. SOS1 thus associates with the membrane-anchored Ras and activates Ras by exchanging GDP to GTP. Active Ras dimerizes and forms nanocluster on the membrane, transducing Ras signaling through MAPK and PI3K pathways. To fully activate MAPK pathway, the dimeric GTP-bound Ras binds to Raf through Ras binding domain (RBD) and promotes Raf dimerization. However, Ras association domain family 5 (RASSF5), which competes with Raf for the same Ras binding site via similar RBD, is a module of the Hippo pathway, which acts to suppress cell proliferation. In this study, we focus on the activation mechanism of KRas4B. By collaborating NMR experiments and molecular dynamics (MD) simulation, we elucidated the interactions of KRas4B upstream/downstream effectors: EGFR–Grb2–SOS1, Grb2–SOS1, SOS1–KRas4B, and KRas4B–RASSF5.

We first investigated the association of Grb2 with SOS1. Grb2 consists of an SH2 domain flanked by N- and C-terminal SH3 domains (nSH3/cSH3). Grb2 nSH3/cSH3 domains have strong binding affinity for SOS1 proline-rich (PR) domain that mediates the Grb2–SOS1 interaction. The nSH3/cSH3 domains have distinct preferred binding motifs: PxxPxR for nSH3 and PxxxRxxKP for cSH3 (x represents any natural amino acid). Several nSH3-binding motifs have been identified in the SOS1 PR domain but none specific for cSH3 binding. Even though both nSH3 and cSH3 exhibit the strongest binding to the SOS1 peptide PVPPPVPPIRRRP, this mutually-exclusive binding combined with other potential nSH3/cSH3 binding regions in SOS1 make understanding the Grb2–SOS1 interaction challenging. To identify the SOS1–cSH3 binding sites, we selected

seven potential binding segments in SOS1. The synthesized peptides were tested for their binding to nSH3/cSH3. Our NMR data reveal that PKLPPKTYKREH peptide has strong binding affinity for cSH3, but very weak for nSH3. The binding specificity suggests that the most likely Grb2–SOS1 binding mode is through nSH3–PVPPPVPPIRRRP and cSH3–PKLPPKTYKREH interactions, which is supported by replica-exchange simulations for the Grb2–SOS1 complex models. We propose that nSH3/cSH3 binding peptides, which effectively interrupt Grb2–SOS1 association, can serve as tumor suppressors.

To comprehensively explore the nSH3/cSH3 interaction with linked and truncated SOS1 PR domain, we constructed the computational models of nSH3/cSH3 with all probable PR segments and applied MD simulations. Our 248 microseconds ( $\mu$ s) simulations include 620 distinct trajectories, each 400 nanoseconds (ns). We construct the effective free energy landscape to validate the nSH3/cSH3 binding sites. The nSH3/cSH3–SOS1 peptide complex models indicate that strong peptide binders attract the flexible nSH3 n-Src loop, inducing a closed conformation of nSH3; by contrast, the cSH3 conformation remains unchanged.

While the detail interactions of isolated nSH3/cSH3 with SOS1 one-site peptides were elucidated, how full length Grb2 interacting with SOS1 PR remains unclear. Here, we showed the unexpected high affinities of full length Grb2 with the nSH3/cSH3 binding peptides of SOS1 and validates the proposed binding models of Grb2-SOS1 in chapter 2. Four nSH3 binding peptides and one cSH3 binding peptide were selected. The full length Grb2 associating with these one-site peptides does not perform the anticipated primary and secondary bindings for the nSH3 and cSH3. Instead, the affinities of nSH3/cSH3 binding to each of the one-site peptides have the same magnitude of  $K_D$  values, and both nSH3 and cSH3 domains of Grb2 perform much stronger binding for the one-site peptides than the isolated nSH3/cSH3. Moreover, the interactions of Grb2

with three two-site SOS1 segments resemble the pairwise affinities of isolated nSH3/cSH3–SOS1 peptides. Lastly, the phosphorylated EGFR segment has little influences on the affinities and interactions of Grb2 binding to the SOS1 one-site peptide but may strengthen association of Grb2 with the particular two-site peptide.

To investigate the complex interactions of SOS1-KRas4B, we perform MD simulations on twelve SOS1 systems, scrutinizing each step in two possible KRas4B activation cycles, fast and slow. To activate KRas4B at the CDC25 catalytic site, the allosteric site in the Ras exchanger motif (REM) domain of SOS1 needs to recruit a (nucleotide-bound) KRas4B molecule. Our simulations indicate that KRas4B-GTP interacts with the REM allosteric site more strongly than with the CDC25 catalytic site, consistent with its allosteric role in the GDP-to-GTP exchange. In the fast cycle, the allosteric KRas4B-GTP induces conformational change at the catalytic site. The conformational change facilitates loading KRas4B-GDP at the catalytic site and opening the KRas4B nucleotide-binding site for GDP release and GTP binding. GTP binding reduces the affinity of KRas4B-GTP to the CDC25 catalytic site, resulting in its release. By contrast, in the slow cycle, KRas4B-GDP binds at the allosteric REM site. The limited, altered conformational change that it induces prevents the exact alignments of Switch I and II of KRas4B. The increasing binding strength at both binding sites due to interactions of regions other than Switch I & II retards GDP release from the catalytic KRas4B, thus KRas4B activation. The accelerated activation cycle supports a positive feedback loop with allosteric signals communicating between the two Ras molecules, and is the predominant, native function of SOS.

Besides two Ras major signaling pathways MAPK and PI3K, Ras is involved in the activation of Hippo signaling. Here, we proposed the highly significant aim of unraveling the crosstalk between Ras/Raf-elicited MAPK and Ras/RASSF5-elicited Hippo signaling by deciphering the

enigmatic mechanism of RASSF5 tumor suppressor. Our mechanistic view explains how RASSF5 can activate MST1/2 and suppress cancer *in vivo*; but inhibits MST1/2 *in vitro*. Both activation and inhibition of MST1/2 can take place via SARAH heterodimerization. Our thesis is that RASSF5 acts as adaptor. *In vivo*, membrane-anchored Ras dimers (or nanoclusters) can promote SARAH domain hetero-dimerization, Raf-like MST1/2 kinase domain homo-dimerization and trans-autophosphorylation. *In vitro*, KRas binding also releases the RASSF5 SARAH stimulating MST1/2's SARAH heterodimerization; however, without membrane, no MST1/2 kinase domain homodimerization/trans-autophosphorylation.

How KRas4B activating the auto-inhibited RASSF5 was validated by using MD simulation. We modeled the RASSF5 Ras association (RA) domain with a covalently connected SARAH to elucidate the dynamic mechanism of how SARAH mediates between autoinhibition and Ras triggered-activation. Our results show that in inactive RASSF5 the RA domain retains SARAH, yielding a self-associated conformation in which SARAH is in a kinked  $\alpha$ -helical motif that increases the binding interface. When RASSF5 binds K-Ras4B-GTP, the equilibrium shifts toward SARAH's interacting with MST. Since the RA/SARAH affinity is relatively low, whereas that of the SARAH heterodimer is in the nM range, we suggest that RASSF5 exerts its tumor suppressor action through competition with other Ras effectors for Ras effector binding site, as well as coincidentally its recruitment to the membrane to help MST activation. Thus, SARAH plays a key role in RASSF5's tumor suppressing action by linking the two major pathways in tumor cell proliferation: Ras and the MAPK (tumor cell proliferation-promoting) pathway, and the Hippo (tumor cell proliferation-suppressing) pathway.

## Chapter 2: High-affinity interactions of the nSH3/cSH3 domains of Grb2 with the C-terminal proline-rich domain of SOS1

Chapter 2 was adapted from: TJ. Liao, H. Jang, R. Nussinov, D. Fushman: *High-affinity Interactions of the nSH3/cSH3 Domains of Grb2 with the C-terminal Proline-rich Domain of SOS1*. J. Am. Chem. Soc., 2020, 142 (7), 3401-3411.

### 2.1 Introduction to chapter 2

Growth factor receptor bound protein 2 (Grb2) is an adaptor protein that transmits downward growth factor signals. In response to epidermal growth factor receptor (EGFR) activation, Grb2 establishes a link between an active receptor tyrosine kinase (RTK) and Ras-specific guanine nucleotide exchange factor (GEF) such as Son of sevenless 1 (SOS1) at the plasma membrane [1-5]. SOS1 exchanges GDP by GTP, activating small GTPase Ras [6, 7]. Ras regulates cell proliferation through the mitogen-activated protein kinase (MAPK, Raf/MEK/ERK) and phosphatidylinositide 3-kinase (PI3K)/Akt/mTOR pathways [8-11]. Ras oncogenes frequently result in human somatic malignancies. However, therapeutic targeting of oncogenic Ras remains unsuccessful. Most efforts focus on blocking Ras regulators and effectors, such as SOS1, B-Raf, and PI3K $\alpha$  [12-14]. The Grb2–SOS1 interaction has attracted increasing attention in the clinics, and recent targeted therapy has shown that synthesized peptide inhibitors can interrupt the Grb2–SOS1 interaction, effectively preventing tumor development [15-20].

Grb2 (residues 1-217) contains three successive Src homology (SH) domains: N-terminal SH3 (nSH3, residues 1-58), SH2 (residues 60-152), and C-terminal SH3 (cSH3, residues 156-215) domains (Fig. 2-1A). X-ray crystal structure analyses illustrated that the adaptor protein forms an auto-inhibited homodimer [21]. Specific, high affinity interactions of the SH2 domain with the

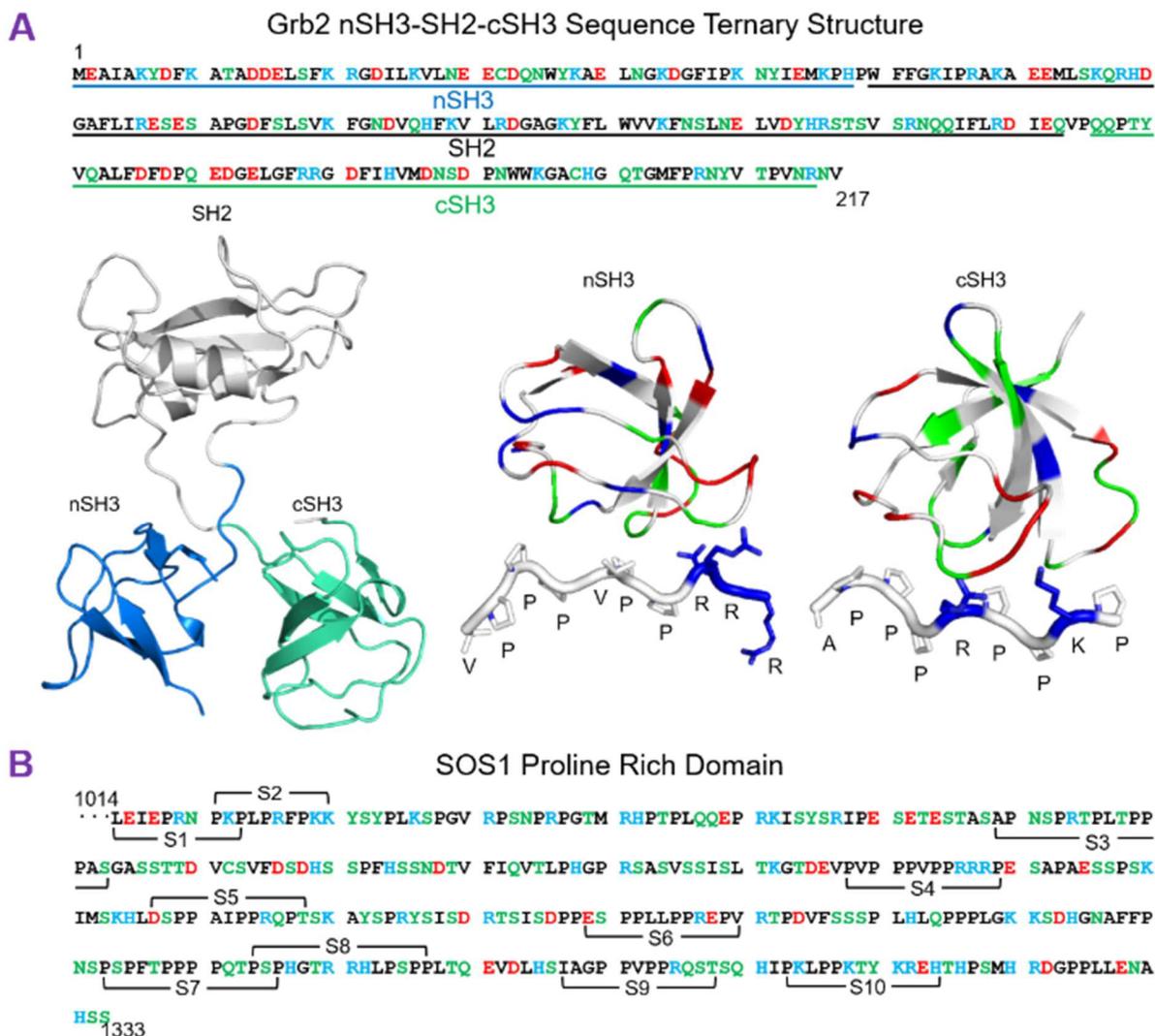
phosphorylated tyrosine (pY) containing motif pYxNx (x represents any natural amino acid) of the RTK [22, 23] suggest that Grb2 can release the autoinhibition upon binding to the pY-containing motif, liberating its “arms”, the nSH3/cSH3 domains. The released domains recruit SOS1 C-terminal proline-rich (PR) domain, relocating SOS1 to the plasma membrane [24]. GTP-bound Ras binding to the exposed SOS1 allosteric site at the Ras exchange motif (REM) leads to Ras GDP/GTP exchange at the catalytic site of CDC25 [25]. SOS1 activates Ras at the plasma membrane, which activates MAPK signaling in the cytoplasm. Thus, the Grb2–SOS1 interaction activating SOS1 is important for the Ras activation.

SH3-binding ligands typically contain the canonical PxxP motif that tends to fold into a polyproline type II (PPII) helix [26, 27]. Ligands binding to Grb2 nSH3 have the consensus PxxPxR motif with an arginine abutting the PxxP core. By contrast, cSH3-binding ligands do not have the PxxP motif. Instead, they contain the core sequence RxxK that stabilizes the cSH3 negatively charged pockets [28]. The cSH3 binding motif PxxxRxxKP, which contains a proline adjoining RxxK core, enhances the binding affinity to cSH3 even though the exact function of the proline has been unclear [29]. Grb2-associated-binding protein 2 (Gab2), with an APPPRPPKP motif with a proline at the last position, has strong binding affinity ( $K_D \sim 50 \mu\text{M}$ ) for cSH3 [30]. This motif contains a non-proline residue at the first position and folds into a PPII helix. Unlike the nSH3 binding motif PxxP, the cSH3 binding motif RxxK has not been only defined as a PPII helix; rather, upon binding cSH3, the PPVNRNLKP motif (RxxK) of Gab2 forms a  $3_{10}$  helix [30, 31], suggesting that the cSH3 binding motif has distinct binding modes, rendering the puzzle still unsolved.

Grb2 interaction with SOS1 PR domain is at least 100 fold stronger than the interactions of the individual nSH3/cSH3 domains with the truncated SOS1 PR peptides [32]. This suggests that Grb2-SOS1 association likely involves simultaneous binding of nSH3 and cSH3 domains to SOS1

PR domain. However, SOS1 PR domain lacks the cSH3 PxxxRxxKP binding motif. It was reported that cSH3 may interact with SOS1 PR domain through the nSH3 PxxPxR binding motif [33]. Sequence search for PxxPxR shows four distinct binding sites on SOS1: PVPPPVPPRRRP, DSPPAIPPRQPT, ESPPLLPPREPV, and IAGPPVPPRQST (Fig. 2-1B). Both nSH3/cSH3 have the strongest binding affinity for PVPPPVPPRRRP, but nSH3 outcompetes cSH3. Besides, cSH3 has extremely weak binding affinity for the rest of the segments ( $K_D > 1000 \mu\text{M}$ ). The mutually-exclusive nSH3/cSH3 binding to PVPPPVPPRRRP suggests that cSH3 may use an additional, still undiscovered, binding site.

To clarify how Grb2 interacts with SOS1, here we examine the binding affinity of nSH3/cSH3 for selected SOS1 PR peptides by NMR measurements. The analysis reveals a previously undiscovered strong cSH3 binding motif in SOS1. We suggest likely binding modes of Grb2–SOS1 based on the analysis of binding affinities. Since structural data of the Grb2–SOS1 complex are unavailable, to verify the concurrent nSH3/cSH3–SOS1 associations, we construct four most probable Grb2–SOS1 complex models and sample the conformational ensembles of complexes by using replica-exchange molecular dynamics (REMD) simulations. The sampling shows that concurrent nSH3/cSH3 binding is influenced by steric constraints. Clarifying how Grb2 nSH3/cSH3 domains associate with SOS1 may facilitate the development of peptide inhibitors to prevent Ras activation.



**Figure 2-1.** Grb2 and SOS1 PR domain. (A) Full-length human Grb2 (PDB: 1GRI) contains 217 residues and is composed of nSH3 (blue, residues 1-58), SH2 (white, residues 60-153), and cSH3 (green, residues 156-215) domains. The polar, nonpolar, positively charged, and negatively charged residues are colored green, black/white, blue, and red, respectively, in the amino acid sequence of Grb2 (top). Available NMR/crystal structures of nSH3/cSH3 domains with the binding partners are nSH3–VPPPVPVRRR (PDB: 1AZE) and cSH3–APPPRPPKP (PDB: 2W0Z). (B) Candidate nSH3/cSH3 binding sites on SOS1 are denoted as S1-S10 in the PR domain (residues 1014-1333) amino acid sequence, colored the same way as for Grb2.

## 2.2 Results

### 2.2.1 Candidate nSH3/cSH3 binding sites on SOS1 PR domain

In order to identify potential nSH3/cSH3 binding motifs, we analyzed the amino acid sequence of SOS1 PR and used NMR to confirm and characterize the binding interactions. The strong interaction between nSH3/cSH3 and SOS1 is due to electrostatics. Unlike the basic SH2 with isoelectric point, pI=9.11, nSH3/cSH3 are acidic with pI=4.96. The nSH3/cSH3 binding motifs always involve at least an arginine or lysine. For example, structural data show that nSH3 associates with SOS1 VPPPVPPIRRR and cSH3 with Gab2 APPPIRPPKP with high affinity. Four nSH3 binding peptides PVPPPVPPIRRR, DSPPAIPPIRQPT, ESPLLPPREPV, and IAGPPVPPIRQST were characterized. APNSPRTPLTPPIPAY S was selected due to its PxxPR, which is similar to nSH3 binding PxxPxR motif (with an extra tyrosine present between alanine and serine). To examine the cSH3 binding sites, we selected the three R/KxxK-embedding peptides LEIEPRNP KP, PKPLPRFPKK, and PKLPPKTYKREH. Since the continuous arginine are critical for nSH3/cSH3 binding, we also tested the peptide PSPHGTRRHLPSP. A recent study indicates Grb2 cSH3 binding to Vav nSH3 via the PPPPG motif [34, 35]. Thus, the PSPFTPPPIQTPSP sequence, which contains four successive Pro, was also selected. These SOS1 peptides are designated S1–S10 following the sequential residue numbers (Table 2-1).

**Table 2-1. Candidate of nSH3/cSH3 binding sites**

Peptides	Residues	Segments	Motifs
S1	1014-1023	LEIEP <b>R</b> NP <b>K</b> P	PRxPKP
S2	1021-1030	PKPLP <b>R</b> FP <b>K</b> K	PxPxPRxPKK
S3	1079-1093	A <b>P</b> NS <b>P</b> R <b>T</b> PLT <b>P</b> PPAYS	PxxPRxPxxP
S4	1148-1159	PV <b>P</b> P <b>P</b> V <b>P</b> PP <b>R</b> R <b>R</b> P	PxxPxR
S5	1177-1188	DSP <b>P</b> A <b>I</b> P <b>P</b> RQ <b>P</b> T	PxxPxR
S6	1209-1220	ESP <b>P</b> LL <b>P</b> P <b>R</b> EPV	PxxPxR
S7	1253-1266	P <b>S</b> P <b>F</b> T <b>P</b> PP <b>P</b> Q <b>T</b> P <b>S</b> P	PxxPPPPxxP
S8	1264-1277	PSPHG <b>T</b> <b>R</b> R <b>H</b> LP <b>S</b> P <b>P</b>	RR
S9	1287-1298	IAG <b>P</b> P <b>V</b> P <b>P</b> RQ <b>S</b> T	PxxPxR
S10	1303-1314	<b>P</b> K <b>L</b> P <b>P</b> K <b>T</b> Y <b>K</b> REH	PxxPxKxxKR

The various SOS1 PR peptides are selected by their unique motifs and denoted S1 to S10. The segments are aligned by the critical ‘R/K’ residues.

### 2.2.2 The binding interface on nSH3/cSH3 domains

Structural analysis uncovered the binding surface on nSH3/cSH3 domains (Fig. 2-2A and 2-3A). SH3 binding motifs usually form a PPII helix and require positively charged residues. PPII helix lacks internal hydrogen bonds, and the backbone adopts the SH3 hydrophobic pockets. The nSH3 hydrophobic pocket is composed of residues Tyr7, Phe9, Trp36, Phe47, Pro49, and Tyr52 [36]. The nSH3–VPPPVP**P**RRR peptide interaction shows that Asp15 and Asp33 of nSH3 form salt bridges with peptide’s arginine (Fig. 2-2A). Noticeably, Asp15 and Glu16 are surrounded by non-polar residues Phe9, Ala11, Ala13, Trp36, and Phe47. Thus, the VPPPVP**P**RRR arginine particularly target the isolated Asp15 and Glu16. The high specificity and strong interaction of nSH3 with VPPPVP**P**RRR results in high binding affinity. By contrast, the cSH3 hydrophobic pocket consists of residues Phe165, Phe167, Trp193, Phe205, Pro206, and Tyr209 [37]. For the cSH3–AP**P**RR**P**PKP complex, Glu171 and Glu174 of cSH3 form salt bridges with the arginine and lysine from the peptide (Fig. 2-3A) [38]. Unlike the isolated Asp15 and Glu16 of nSH3,

Glu171 and Glu174 together with Asp168, Gln170, and Asp172 form a large negatively charged cluster. The arginine and lysine of APPPRPPKP stabilize the interaction with cSH3.

Sequence alignment of nSH3 and cSH3 shows an agreement in the binding interface for the hydrophobic residues but not for the negatively charged residues (Fig. A1). The hydrophobic residues Tyr7, Phe9, Trp36, Phe47, Pro50, and Tyr52 of nSH3 are aligned with Phe165, Phe167, Trp193, Phe205, Pro206, and Tyr209 of cSH3. However, the critical residues Asp15 and Glu171 for the salt bridge interaction are not aligned, suggesting that nSH3 and cSH3 have distinct binding motifs.

The  $^1\text{H}$ - $^{15}\text{N}$  NMR spectra of free nSH3/cSH3 and the assigned amide signals are shown in Fig. A2 and A3. During the course of the titration, we observed strong signal attenuation in nSH3 residues Lys10, Lys26, and Asn35. Apo-nSH3 Lys10 and Lys26 exhibited slow exchange on the chemical shift time scale. When SOS1 peptides were titrated into nSH3 approximately to the 2:1 molar ratio, the signals of these residues exhibited a switch to the fast exchange regime. Conversely, Asn35 exhibited significant signal attenuation upon addition of the peptides.

**Table 2-2 Dissociation constants for nSH3/cSH3–SOS1 peptides interaction**

Peptides	This work				Ref.[33]	
	nSH3		cSH3		nSH3	cSH3
	$K_D$ ( $\mu\text{M}$ )	$\Delta\delta_{\text{max}}$ (ppm)	$K_D$ ( $\mu\text{M}$ )	$\Delta\delta_{\text{max}}$ (ppm)	$K_D$ ( $\mu\text{M}$ )	$K_D$ ( $\mu\text{M}$ )
S1	*730 $\pm$ 890	0.05	NBD	0.02	--	--
S2	250 $\pm$ 96	0.16	440 $\pm$ 146	0.12	--	--
S3	*280 $\pm$ 103	0.05	NBD	0.01	--	--
S4	37 $\pm$ 22	0.38	142 $\pm$ 18	0.39	39 $\pm$ 1	125 $\pm$ 13
S5	--	--	--	--	56 $\pm$ 5	1396 $\pm$ 87
S6	--	--	--	--	117 $\pm$ 2	1718 $\pm$ 33
S7	*580 $\pm$ 462	0.04	NBD	0.02	--	--
S8	208 $\pm$ 42	0.14	*400 $\pm$ 447	0.05	--	--
S9	--	--	--	--	82 $\pm$ 1	1318 $\pm$ 44
S10	660 $\pm$ 320	0.11	156 $\pm$ 33	0.46	--	--

The dissociation constants ( $K_D$ , in  $\mu\text{M}$  units) in this work (left columns) were obtained by globally fitting experimental NMR titration data for select residues to a single-site binding model.  $\Delta\delta_{\text{max}}$  is the largest CSP over all analyzed residues at the titration endpoint. The asterisks mark the  $K_D$  values determined from a limited number of titration curves (hence poor statistics and large standard deviations) due to the very small observed CSPs (maximal  $\Delta\delta < 0.05$  ppm), reflecting weak binding. In these cases, the mean  $K_D$  values from the individual fits for the same residues, are 1250  $\mu\text{M}$  (nSH3-S1), 270  $\mu\text{M}$  (nSH3-S3), 950  $\mu\text{M}$  (nSH3-S7), and 650  $\mu\text{M}$  (cSH3-S8). NBD stands for ‘no binding detected’. See further details in the Methods section. The  $K_D$  values in the right columns are published data from the previous work [33].

### 2.2.3 The nSH3 shows stronger affinity for PxxPxR motif

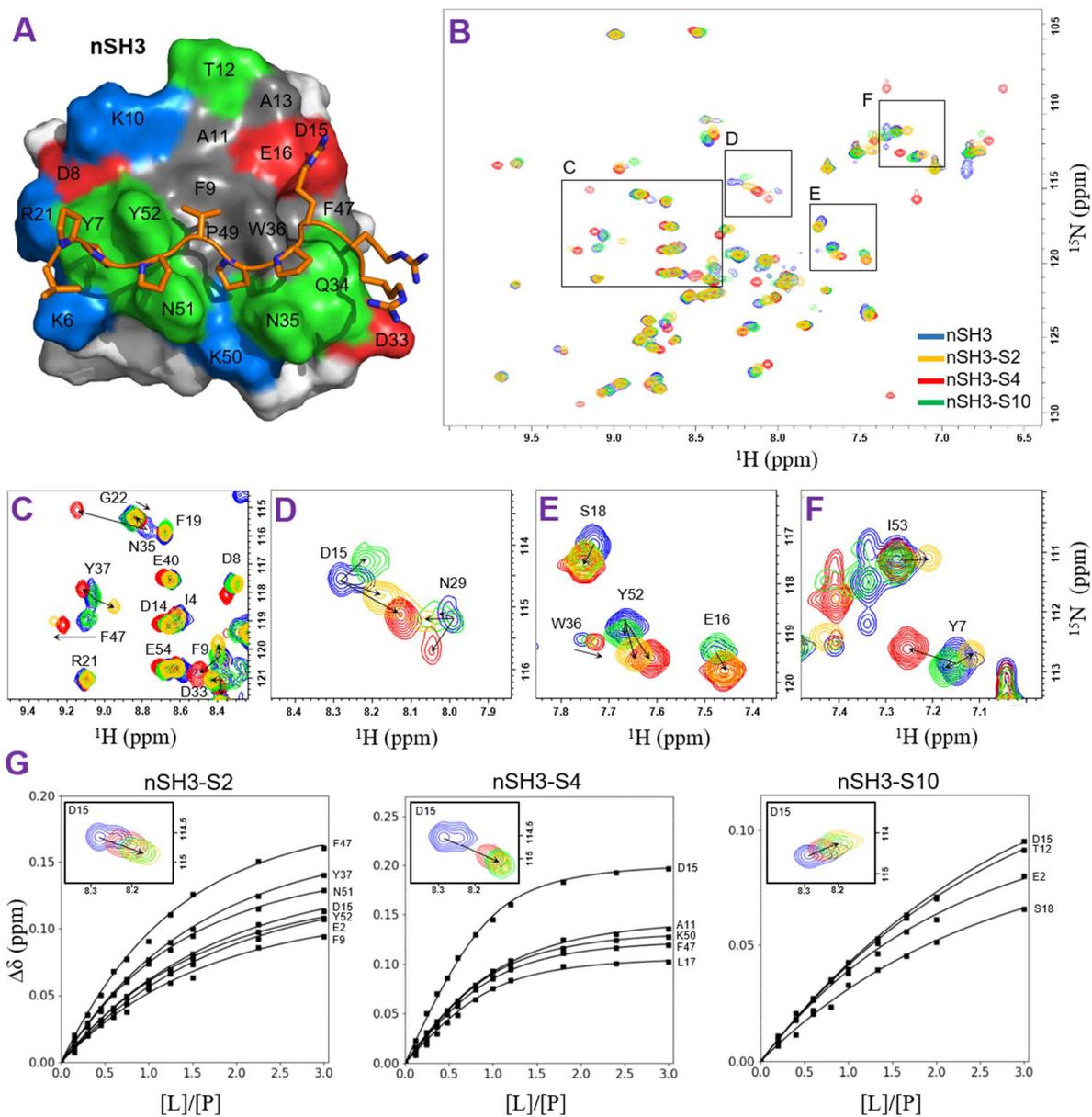
S1 (LEIEPRNPKP) displays weak affinity for nSH3 (Table 2-2). S1 misses nSH3 binding PxxPxR motif but involves two Glu followed by a PxxP motif embedding 'RxxK'. Small amide signal shifts were observed when the titrant reached the endpoint of titration (Fig. A1B and A5A). Similar to S1, S2 (PKPLPRFPKK) does not follow the PxxPxR motif. S2 has a 'PxPxP' motif followed by 'RxxK'. The 'PxPxP' motif less likely forms a regular PPII helix. Surprisingly, S2 showed fairly strong binding to nSH3, with  $K_D \sim 250 \mu\text{M}$  (Table 2-2) and induced large CSPs in Glu2, Phe9, Asp15, Asn35, Tyr37, Phe47, and Tyr52 (Fig. 2-2, and 2-4A), suggesting that nSH3 may alternatively bind to S2 through the PxxPxK motif. Thus, a substituted lysine for arginine reduces the binding affinity but retains the nSH3 binding motif.

S4 (PVPPPVPPIRRRP) peptide has the strongest affinity for nSH3 (Table 2-2). With several prolines, S4 forms a well-structured PPII helix, and the three successive arginine 'RRR' form salt bridges with Asp15 and Glu16 (Fig. 2-2A). Asp15 pairs the first arginine, and Asp33 interacts with the last arginine. The middle arginine has the least binding effect. In addition to Asp15, residues Tyr7, Phe9, Trp35, Phe47, and Tyr52 exhibited large CSPs likely due to the hydrophobic interactions (Fig. 2-2 and 2-4C). Similarly, S5 (DSPPAIPPRQPT), S6 (ESPPLLP PREPV), and S9 (IAGPPVPPRQST) peptides also follow the nSH3 binding motif PxxPxR, showing small  $K_D$  for nSH3 (Table A1) [33]. The  $K_D$  values of nSH3–S5/S6/S9 are 56, 117, and 82  $\mu\text{M}$ . Unlike the strong binding between nSH3 and S4, S5, S6, and S9, S3 (APNSPRTPLTPPPAYS) contains a critical arginine adjacent to multiple prolines but still shows weak binding to nSH3 judging by the overall small CSPs (Fig. A5C, Table 2-2). The amino acid sequence of S3 contains PxxPR motif instead of PxxPxR. A missing amino acid between PxxP and 'R' results in the mismatch to the

peptide binding pocket on nSH3. The weak affinity of nSH3–S3 suggests nSH3 does not favor the PxxPR motif (Fig. A2D and A5C).

Without any positively charged residues, S7 (PSPFTPPPPQTPSP) weakly interacts with nSH3 and induces small chemical shifts (Table 2-2, Fig. A2F and Fig. A5E), suggesting that to trigger the association the nSH3-binding motif requires a critical arginine or lysine. Since the positively charged residues are essential for interaction with nSH3, S8 (PSPHGTRRHLPSP) involving ‘RR’ may associate with nSH3; however, unlike the other SH3 binding candidates, S8 contains no prolines proximal to ‘RR’. Nonetheless, even though S8 does not follow the nSH3 binding PxxPxR motifs, it displays a fair affinity for nSH3 ( $K_D \sim 210 \mu\text{M}$ ) (Table 2-2). Interestingly, Asp15, Phe47 and Tyr52 still form strong interactions (inferred from large CSPs); however, Tyr7 and Phe9 show smaller CSPs. Instead, Glu2, Ile4 and Thr12 exhibit large CSPs (Fig. A2G and A5G), which suggests that the nSH3–S8 interaction may have an atypical binding interface, resulting in the binding affinity not as strong as of the typical nSH3 binding motif.

S10 (PKLPPKTYKREH) has weak affinity for nSH3 with  $K_D \sim 660 \mu\text{M}$  (Table 2-2). Even though this peptide contains a PxxPxK motif, which is similar to S2, the nSH3–S10 interaction is much weaker, and the overall signal shifts are smaller than for S2 (Fig. 2-2 and 2-4E). Except for Asp15, typical interface residues Tyr7, Phe9, Phe47, and Tyr52 exhibit relatively smaller CSPs. Instead, Glu2, Ile4, Thr12, and Ser18 have larger CSPs. Noticeably, S10 and S8 induce similar CSP patterns in nSH3, implying similar binding interface to nSH3–S8, but weaker binding affinity.



**Figure 2-2.** NMR spectra and titration curves of nSH3-S2/S4/S10 binding. (A) Solution structure of nSH3-VPPPVPPRRR complex (PDB: 1AZE) shows the typical nSH3-peptide binding interface. Residues colored green, gray, blue, and red represent the polar, nonpolar, positive, and negative side chains, respectively. Others are colored white. Binding peptide VPPPVPPRRR derived from SOS1 PR domain is shown in yellow, and the nitrogen atoms of backbones and sidechains are colored blue. The nSH3 residues Y7, F9, W36, F47, P49 and Y52 form hydrophobic

interactions with the peptide; D15, E16, and D33 form salt bridges. (B) Superposition of  $^1\text{H}$ - $^{15}\text{N}$  NMR spectra of nSH3 in the unbound state (blue) and upon saturation with 3-fold molar excess of S2 (yellow), S4 (red), and S10 (green). Signals belonging to typical interface residues are emphasized by rectangles and zoomed on in (C) F9, D33, and F47; (D) D15; (E) E16, W36, and Y52; (F) Y7. (G) Titration curves for selected residues of nSH3 ([L]) upon addition of S2, S4, or S10 ([P]) are plotted as a function of [L]/[P] ratio; the symbols depict experimental chemical shift perturbations, the curves are the results of fit to a single-site binding model. The insets show shows signal shifts during titration for residue D15: [L]/[P] = 0 (blue), 1 (red), 2 (green), and 3 (yellow). In  $K_D$  calculations, 14 residues (E2, F9, A11, D15, E16, L17, S18, Y37, A39, F47, K50, N51, Y52, and I53) were used for S2 binding, 15 residues (E2, K6, A11, A13, D15, E16, L17, S18, G22, D23, D33, Y37, F47, K50, and N51) for S4, and 7 residues (E2, Y7, T12, D15, S18, N51, Y52) for S10. The buffer condition and the protein concentrations are detailed in the Materials and Methods section.

#### 2.2.4 S4 and S10 display strong and comparable affinity for cSH3

S1 (LEIEPRNPKP) has very weak affinity for cSH3 (Table 2-2) and induces small CSPs (Fig. A3B and A5B). The observed CSPs are too small to determine the  $K_D$  value. Comparison of S1 sequence with the cSH3 binding PxxxRxxKP motif indicates that S1 lacks the first proline and has a glutamate instead. Even though cSH3 binding favors the RxxKP motif, the S1 glutamate may repulsively interact with Asp166, resulting in weak association. Similar to S1, S2 (PKPLPRFPKK) peptide also has the core 'RxxK' motif. However, the front proline is not at the right position. The 'PxPxP' motif is not favored to form a stable PPII helix. At the rear of S2 there are two consecutive lysine instead of 'KP'. Even though S2 barely follows the cSH3 binding PxxxRxxKP motif, it binds cSH3 with a  $K_D \sim 400 \mu\text{M}$ , much stronger than in case of S1 (Table 2-2). Interestingly, S2

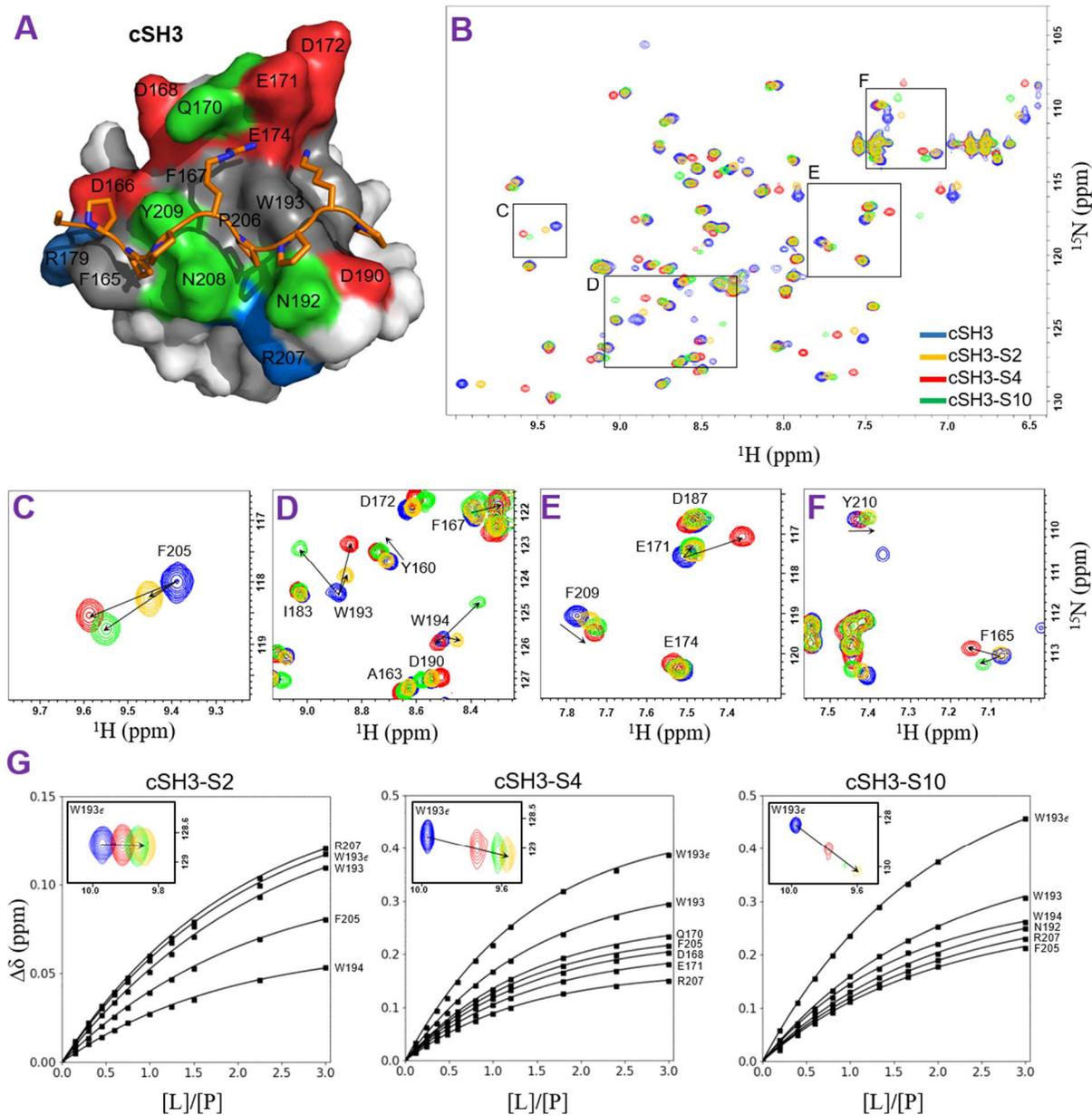
RxxK interaction with cSH3 induces small CSPs around Glu171, but the PxPxP region results in larger CSPs at Trp193 and Phe205 (Fig. 2-3 and 2-4B).

Of all the peptides, S4 (PVPPPVPPRRRP) displays the highest affinity for cSH3 with  $K_D \sim 138 \mu\text{M}$  (Table 2-2). Unlike the 'RxxK' motif interacting with Glu171 and Glu174, the 'RRR' motif strengthens the interaction with Glu171, Asp187, and Asp190 [33]. In addition, Trp193, Phe205, and Tyr209 also have large CSPs due to hydrophobic interactions (Fig. 2-3 and 2-4D). Our S4 results are consistent with a previous study [33]. However, without the 'RRR' motif, the single salt bridge [33] formed by cSH3's Glu171 with the arginine in S5 (DSPPAIPPRQPT), S6 (ESPPLLPPREPV), S9 (IAGPPVPPRQST), and possibly S3 (APNSPRTPLTPPPAYS), does not strengthen the cSH3 binding. Our NMR data indicate that cSH3-S3 interaction is very weak (Table 2-2), as inferred from the practically negligible signal shifts (Fig. A3D and A5D). This is consistent with the result of the previous study, which found that cSH3 also has very weak affinity for S5, S6, and S9 ( $K_D > 1000 \mu\text{M}$ ). Since 'RRR' is so important for nSH3/cSH3 binding, alanine scan was performed for S3, showing that PVPPPVPPRRAP has  $K_D = 72 \mu\text{M}$  for nSH3 and  $K_D = 318 \mu\text{M}$  for cSH3 [33]. Even though the alanine mutation decreases the binding affinity, the 'RR' still results in S3 binding to cSH3 stronger than S5, S6, or S9. Apart from S4, another 'RR' segment also exists in SOS1 PR domain, and thus S8 (PSPHGTRRHLPSP) was selected for the test.

S8 (PSPHGTRRHLPSP) contains an 'RR' motif but exhibits weak binding affinity for cSH3 (Table 2). Since the 'RR' motif lacks the proximal prolines, S8 cannot form a PPII helix and fit into the cSH3 hydrophobic pocket, resulting the overall small CSPs of cSH3 (Fig. A3G and A5H). The weak affinity suggests that the proximal prolines in 'RR' are important for cSH3 binding. By contrast, the S7 (PSPFTPPPPQTPSP) peptide contains multiple prolines and has the 'PPPP' motif. Unlike the strong binding between cSH3 and the 'PPPPG' motif of Vav nSH3 domain [34, 35], S7

induced only minor signal shifts in cSH3 (Fig. A3F and A5F), indicative of very weak binding (Table 2-2). Even though S7 contains PxxPxxPxxP motif, three continuous PxxP, most likely forming a robust PPII helix, without positively charged residues it is incapable of forming a stable complex with cSH3.

S10 (PKLPPKTYKREH) has strong binding affinity for cSH3 ( $K_D \sim 154 \mu\text{M}$ ), which is a novel interaction unreported previously. The affinity of cSH3–S10 is as strong as the cSH3–S4 interaction (Table 2-2). Noticeably, S10 peptide does not follow the cSH3 PxxxRxxKP binding motif. The first proline is not located in the right position. Like the RxxK motif, S10 contains a ‘KxxK’ motif which, however, is followed by an additional arginine instead of proline. For the KxxK motif, the substituted first lysine for arginine does not result in an expected low binding affinity of cSH3–S10. Surprisingly, S10 induces very large CSPs near Trp193, Phe205, and Tyr209 but relatively low CSPs around Glu171 (Fig. 2-3 and 2-4F), which suggests that cSH3–S10 complex is characterized by stronger hydrophobic interactions than the salt bridges.



**Figure 2-3.** NMR spectra and titration curves of cSH3–S2/S4/S10 binding. (A) Crystal structure of cSH3–Gab2 APPPRPPKP complex (PDB: 2W0Z) illustrates how cSH3 interacts with RxxK motif forming the binding interface. The coloring of residues follows that in figure 2-2. The critical cSH3 residues at the interface: F165, F167, W193, F205, P206, Y209 form hydrophobic interactions with the peptide; E171, E174, and D190 form salt bridges. (B) Superposition of  $^1\text{H}$ - $^{15}\text{N}$  NMR spectra of cSH3 in the unbound state (blue) and at the endpoint of titration with S2

(yellow), S4 (red), and S10 (green). Signals belonging to the interface residues are emphasized by rectangles and zoomed on in (C) F205, (D) F167, D190, W193, (E) E171, E174, F209, and (F) F165. (G) Titration curves for the selected residues of cSH3 ([P]) upon addition of S2, S4, or S10 ([L]) are plotted as a function of [L]/[P] ratio; the symbols depict experimental chemical shift perturbations, the curves are the results of fit to a single-site binding model. The insets show signal shifts during titration for W193 indole NH group (W193 $\epsilon$ ): [L]/[P] = 0 (blue), 1 (red), 2 (green), and 3 (yellow). In  $K_D$  calculation, 6 residues (D168, W193 $\epsilon$ , W194, F205, R207, and N208) were used for S2 binding, 16 residues (Y160, F165, D166, F167, D168, Q170, E171, D172, G173, W193, W193 $\epsilon$ , G196, M204, F205, R207, N208, and Y209) for S4, and 15 residues (Y160, V161, F167, D168, E171, D172, G196, A197, N192, W193, W193 $\epsilon$ , W194, M204, F205, R207, and Y209) for S10. The buffer condition and the protein concentrations are detailed in the Materials and Methods section.

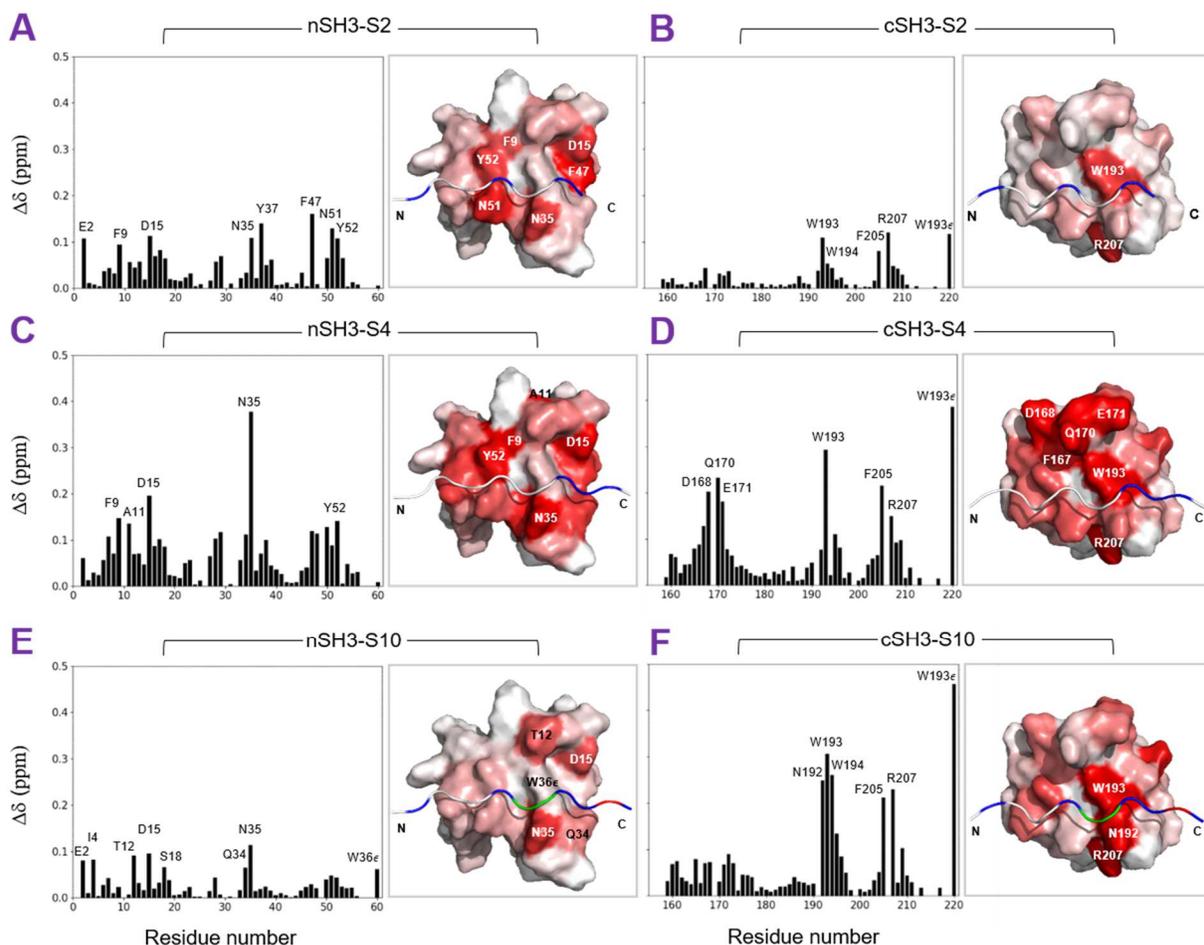
### 2.2.5 Construction of Grb2-SOS1 complex models

To probe the Grb2-SOS1 interaction, we constructed the models for Grb2-SOS1 complexes formed by concurrent binding of both nSH3 and cSH3 to SOS1, based on the (pairwise) affinities of nSH3/cSH3 binding to individual SOS1 peptides. The pairwise affinities suggest the most probable complex models might be nSH3-S4/cSH3-S10, followed by nSH3-S5/cSH3-S4, nSH3-S5/cSH3-S10, nSH3-S9/cSH3-S4, nSH3-S9/cSH3-S10, nSH3-S6/cSH3-S4, nSH3-S6/cSH3-S10, and etc. The four most probable Grb2-SOS1 binding modes are nSH3-S4/cSH3-S10 (mode 1), nSH3-S5/cSH3-S4 (mode 2), nSH3-S5/cSH3-S10 (mode 3), and nSH3-S9/cSH3-S4 (mode 4). The corresponding models were initially constructed using the available NMR/crystal structures and sampled by replica exchange simulation (Fig. A6 and Fig. 2-5). The contact frequencies between nSH3/cSH3 and SOS1 were analyzed over the conformational ensembles of each mode.

The high contact frequency of modes 1, 3, and 4 suggests a stable Grb2–SOS1 interaction; for mode 2, nSH3 loses the contact with S5 and tends to dissociate while the cSH3–S4 contact remains stable.

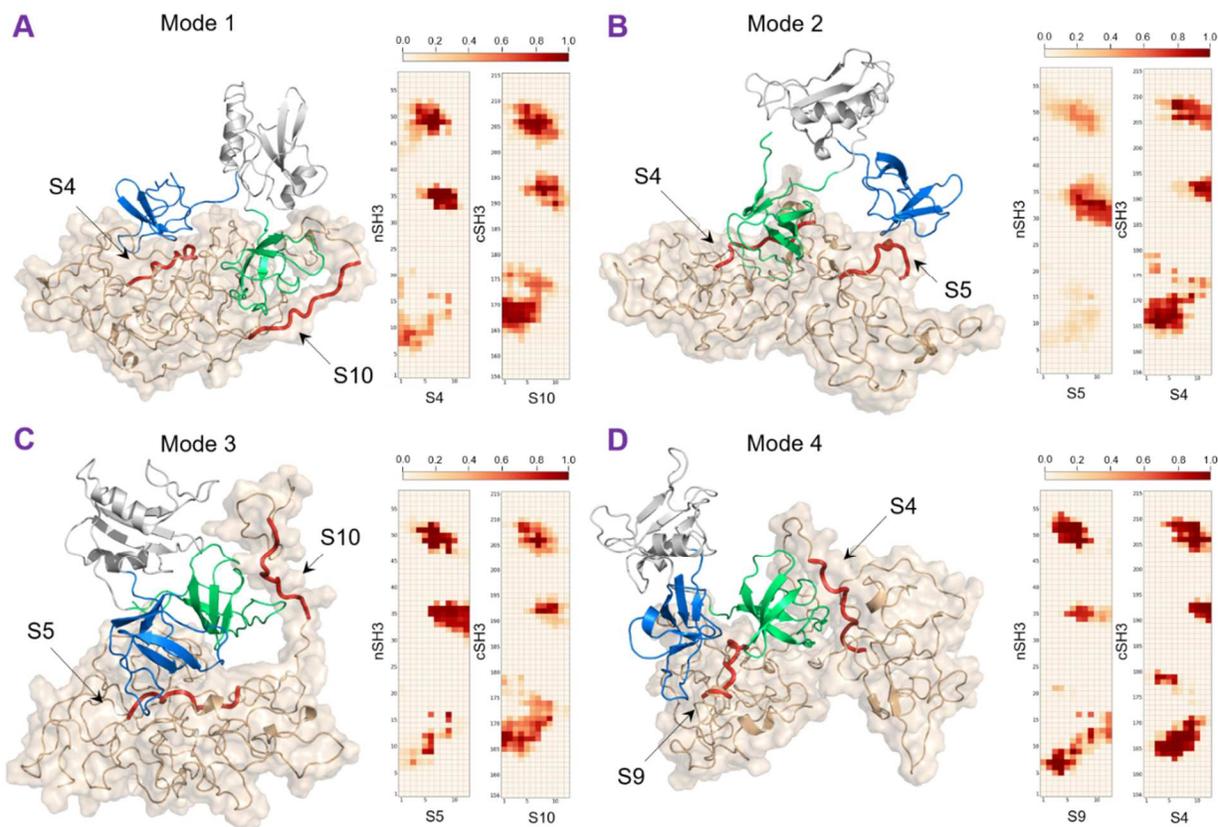
The linkers between two distinct nSH3/cSH3 binding sites may play a critical role in stabilizing Grb2–SOS1 interaction. Among the four binding modes, the linkers are 143 residues long for mode 1, 17 residues for mode 2, 114 residues for mode 3, and 127 residues for mode 4. The long linkers of modes 1, 3, and 4 allow the large structural fluctuation needed to relocate nSH3/cSH3. By contrast, the short linker of mode 2 restricts the structural flexibility between S4 and S5. While S4 contains ‘RRR’, S5 has only a single ‘R’, hence weaker nSH3 binding. In addition, linker fluctuations unstructured S5, which further destabilizes the nSH3–S5 association. Thus the nSH3–S5 interaction is more likely affected by the structural relaxation of the short linker. The orientation of nSH3 retains the salt bridge with S5 but loses hydrophobic interactions. The simulations suggest that the Grb2–SOS1 interaction may rely on modes 1, 3, and 4 with a long linker more than on the less stable mode 2 with the short linker.

The contact frequency shows that the nSH3–SOS1 and cSH3–SOS1 interactions are interdependent. For modes 1 and 3, cSH3 binds to S10 while nSH3 binds to S4 or S5, respectively; nSH3–S4 has higher contact frequency than nSH3–S5 does, resulting in closer cSH3–S10 contact. Similarly, in the case of modes 2 and 4, unlike nSH3–S9 complex that shows stronger interaction in mode 4, nSH3–S5 complex in mode 2 is weaker and loses contact, leading to the lower contact frequency of cSH3–S4. Thus, nSH3–SOS1 binding enhances cSH3–SOS1 interaction, which then strengthens the nSH3–SOS1 interaction. The mutual influence of nSH3–/cSH3–SOS1 interactions limits the probable binding modes of Grb2–SOS1.



**Figure 2-4.** NMR characterization of the peptide binding interface on nSH3 and cSH3. Amide chemical shift perturbations (CSPs,  $\Delta\delta$ ) at the titration endpoint are shown as a function of nSH3 or cSH3 residue number for the following interactions: (A) nSH3–S2, (B) cSH3–S2, (C) nSH3–S4, (D) cSH3–S4, (E) nSH3–S10, and (F) cSH3–S10. The CSPs for indole NH signals of W36 (nSH3) and W193 (cSH3) are shown as bars at positions 60 and 220, respectively, and marked as W36 $\epsilon$  and W193 $\epsilon$ . Structure images to the right of the plots show CSPs mapped onto surfaces of nSH3/cSH3 (PDB: 1GRI) in complexes with the corresponding modeled peptides, respectively. The surface coloring is based on the observed CSP values and ranges from white (0 ppm) to red (maximal  $\Delta\delta = 0.15$  ppm); residues with  $\Delta\delta$  greater than one standard deviation are indicated. The

bound peptides are colored based on non-polar (white), polar (green), negatively-charged (red), and positively-charged (blue) residues. N and C indicate the N and C termini of peptide.



**Figure 2-5.** The four most probable structural models of Grb2-SOS1 complex. Grb2-SOS1 interactions are suggested by the strong binding affinity of (A) nSH3-S4/cSH3-S10 (mode 1), (B) nSH3-S5/cHS3-S4 (mode 2), (C) nSH3-S5/cHS3-S10 (mode 3), and (D) nSH3-S9/cHS3-S4 (mode 4). The models of Grb2-SOS1 PR are obtained from replica exchange simulation for 100 ns. Grb2 nSH3, SH2, cSH3 and SOS1 PR domain are colored blue, white, green, and yellow, and the binding peptides (S4, S5, S9, and S10) are shown as red tubes. The heat maps of contact frequency show the interactions between nSH3/cSH3 residues (y-axis) and various SOS1 peptides residues (x-axis) during the simulation. The color of dark brown indicates highly frequent contacts. The three main interaction regions of nSH3/cSH3 are around residues 10, 35, and 50 for nSH3 and

170, 195, and 205 for cSH3. These maps show correlation between nSH3–SOS1 and cSH3–SOS1 interactions. For mode 2, the unfolded S5 tends to dissociate from nSH3, resulting in unstable Grb2–SOS1 association.

### 2.3 Discussions

Activation of Ras by hormone-activated EGFR is mediated by Grb2 adaptor protein, which recruits SOS1 to the plasma membrane. SOS1 exchanges GDP by GTP, activating Ras. Ras then activates its effectors, e.g. Raf, PI3K $\alpha$ , RASSF5 (NORE1A) and their respective signaling pathways, MAPK, PI3K/Akt/mTOR and Hippo [39, 40]. To obtain a mechanistic insight into how Grb2 recruits SOS1, we aim to elucidate their interaction which to date is still not entirely understood. Exploiting NMR experiments and replica exchange computer simulations we examined seven potential SH3 binding sites on SOS1: S1, S2, S3, S4, S7, S8, and S10. Including the previously tested S5, S6, and S9, the results suggest that nSH3 has stronger binding affinity for SOS1 than cSH3 does. SOS1 PR domain involves several occurrences of the nSH3 binding PxxPxR motif (S4, S5, S6, and S9) but lacks the cSH3 binding motif PxxxRxxKP. Even though cSH3 has weak binding affinity for the nSH3 binding motif, it displays strong binding to S4 (PVPPPVPPIRRR), which contains three successive arginine ('RRR'). However, the nSH3–S4 binding is still stronger than cSH3–S4. Therefore, nSH3–SOS1 association can outcompete cSH3–SOS1. Previous studies indicated that cSH3 has strong binding affinity for Gab1/2, which suggests that Grb2 is more likely to form a ternary structure with SOS1/2 and Gab1/2 where nSH3 binds to SOS1/2 and cSH3 associates with Gab1/2. SOS1/2 and Gab1/2 transduce MAPK and PI3K signaling separately [29, 41]. However, Grb2–SOS1/2–Gab1/2 ternary structure does not explain the high affinity of 1:1 Grb2–SOS1 interaction. We notice that the cSH3 core binding motif 'R/KxxK' exists at both N- and C-termini of SOS1 PR domain, which is different from the nSH3

binding motif PxxPxR that locates in the middle. The ‘R/KxxK’ motifs (S2 and S10) provide alternative binding sites for cSH3. We suggest that nSH3–SOS1 binding does not block the cSH3–SOS1 binding. The multiple alternative binding sites could allow simultaneous nSH3/cSH3 binding to SOS1 peptides in various binding modes.

Our NMR data (CSPs) point to distinct binding modes of nSH3/cSH3 with SOS1 peptides (Fig. 1-4, A4, and A5). For nSH3, S4 (PVPPPVPPRRP) follows PxxPxR, which forms a robust PPII helix and provides large positively charged residues, inducing high CSPs around Tyr7, Phe9, Asp15, Trp35, Tyr37, Phe47, and Tyr52. S2 (PKPLPRFPKK) shows weaker affinity albeit similarly high CSPs at the typical nSH3 binding interface. However, S10 (PKLPPKTYKREH) contains the PxxPxK motif, which resembles PxxPxR but induces relatively high CSPs at the nSH3 atypical binding interface Glu2, Ile4, Thr12, Asp15, and Ser18. The nSH3 atypical binding interface is also involved in nSH3 interaction with S8 (PSPHGTRRHLPSP). Lacking the PxxPxR motif, S1 (LEIEPRNPKP), S3 (APNSPRTPLTPPPAYS), and S6 (PSPFTPPPPQTPSP) induce very small CSPs indicative of low binding affinity. Therefore, we conclude that the nSH3 binding motif strictly follows PxxPxR. The R→K mutation results in an atypical binding interface of nSH3 and lower binding affinity. For cSH3, even though S4 region (PVPPPVPPRRRP) does not follow the cSH3 binding motif PxxxRxxKP, it induces large CSPs at the typical peptide-binding surface of cSH3 (Asp168, Gln170, Glu171, Trp193, Phe205, Arg209, and Tyr209) and displays strong binding affinity. By contrast, S1 (LEIEPRNPKP) contains the ‘RxxKP’ motif but results in weak binding likely due to the repulsion between negatively charged residues of S1 and cSH3. S2 (PKPLPRFPKK) also has the ‘RxxK’ motif and shows stronger affinity than S1. The ‘PxPxP’ region of S2 induces higher CSPs in Trp193, Phe205, and Arg209 than in Asp168 and Glu171. Unlike S2, S10 (PKLPPKTYKREH) involves the ‘PxxP’ motif followed by ‘KxxK’ and shows a

surprisingly strong binding to cSH3. S10 does not induce the same large CSPs at the typical binding surface on cSH3. The 'KxxK' motif forms weaker salt bridge interactions than the 'RxxK' does, reflected in significantly smaller induced CSPs at Asp168 and Glu171. However, large CSPs are observed for Asn192, Trp193, Trp194, Phe205, and Arg207. This implies that KxxK triggers cSH3–S10 association, in which the weaker salt bridges are compensated by hydrophobic interactions. Thus, S10 interaction with cSH3 involves a binding mode distinct from cSH3–S4 but exhibits comparable and strong affinity. Notably, S10 is specific to cSH3, not nSH3. Therefore, unlike the S4-dependent inhibitors blocking both nSH3/cSH3-SOS1 interaction, we suggest that a S10-based inhibitor may only interrupt cSH3–SOS1 association, suppressing PI3K signaling.

Taken together, the binding affinities of individual nSH3/cSH3–SOS1 peptides interactions elucidate how Grb2 may associate with the SOS1 PR domain by engaging nSH3/cSH3 concurrently. Among S1-S10, S4 demonstrates the strongest binding affinity for nSH3. The similar affinities of nSH3 for S4, S5, S6, and S9 suggest that nSH3 has four alternative binding sites on SOS1 PR domain. cSH3 shows the strongest binding to S4, which is an order of magnitude stronger than to S1, S3, S5, S6, S7, S8, or S9. Because of the exclusive cSH3 binding at S4, the previous study suggested that the interaction of Grb2–SOS1 requires cSH3 binding to S4 and nSH3 binding to one of the other alternative binding sites, most likely S5, which has the second strongest affinity. Here we identified S10 as another cSH3 binding site, comparable to S4 in affinity. Based on our results, cSH3 also has three alternative binding sites: S2, S4, and S10. While cSH3 binding to S4 is about two times stronger than to S2, the cSH3–S10 binding is roughly similar. We thus propose that there are multiple binding modes of Grb2–SOS1 interaction, and the most probable binding complex might be nSH3–S4/cSH3–S10.

To conclude, the nSH3–SOS1 interaction is stronger than cSH3–SOS1[42]. nSH3 has four similarly strong binding sites S4, S5, S6, and S9 in SOS1 PR. By contrast, cSH3 has three alternative binding sites S2, S4, and S10, and cSH3–S2 is weaker than cSH3–S4/S10 interaction. nSH3/cSH3 binding to the various SOS1 peptides takes place through distinct binding modes. The cSH3–S4 interaction is arginine-dependent, inducing large CSPs at residues involved in salt bridges and hydrophobic interactions. By contrast, the CSPs of cSH3–S10 are small for salt bridges but much larger for hydrophobic interactions. The positively charged residues from S10 trigger the association with cSH3, and the proline segment complements the interaction, showing comparable binding affinity to that of cSH3–S4. Thus, when both nSH3 and cSH3 are present, nSH3–S4 association is expected to outcompete cSH3–S4, resulting in the cSH3–S10 complex. Based on the binding affinities for the isolated nSH3 and cSH3, the four most probable binding modes of Grb2–SOS1 were constructed. The simulation suggests that the long linkers of modes 1, 3, and 4 relocate nSH3/cSH3 and stabilize Grb2–SOS1 interaction. The short linker of mode 2, however, exhibits spatial constraints and large fluctuations, resulting in nSH3–S5 dissociation. Even though nSH3 binds S5 strongly, the high affinity of individual nSH3/cSH3–SOS1 peptides may not favor Grb2–SOS1 association in mode 2. Our experimental and computational results suggest that the most probable binding mode, nSH3–S4/cSH3–S10, mediates Grb2–SOS1 association, thus potentially resolving the Grb2/SOS1 interaction enigma to better understand the mechanism of EGFR-induced Ras activation. Further experimental studies involving full-length Grb2 and SOS1 PR domain are required to test these predictions.

## 2.4 Material and Methods

### 2.4.1 Sample preparation and NMR measurements

Wild type human Grb2 nSH3/cSH3 plasmids and selected SOS1 PR peptides were commercially obtained from Biomatik Corporation. The plasmids encoding His<sub>6</sub>-tagged constructs were transformed into BL21 competent *E.coli* cells. The cells were grown in M9 media and induced with 1 mM IPTG at 16 °C overnight. The cells were then collected and disrupted by sonication in 50 mM Tris, 100 mM NaCl, 50 mM imidazole, pH 7.6 and NaN<sub>3</sub> 0.02%. The <sup>15</sup>N-labeled proteins were purified by Ni-NTA chromatography and eluted with 200 mM imidazole. The preliminarily purified proteins were buffer exchanged into 50 mM Tris, 100 mM NaCl, 5 mM DTT, pH 7.6 and NaN<sub>3</sub> 0.02%, followed by the size exclusion chromatography.

NMR experiments were performed at 298K on Bruker Avance III 600 MHz spectrometer equipped with a cryoprobe. All samples were prepared in 20 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>, 50 mM NaCl, 5 mM DTT, 10% D<sub>2</sub>O, and 0.02% NaN<sub>3</sub> at pH 7.6. The initial protein concentrations of nSH3 were 233, 196, 157, 196, 227, 122, and 122 μM in binding assays with S1, S2, S3, S4, S7, S8, and S10, respectively; similarly for cSH3, the concentrations were 160, 106, 151, 106, 81, 76, and 81 μM for binding to S1, S2, S3, S4, S7, S8, and S10. The stock solution concentrations of various SOS1 peptides were 10 times the corresponding initial nSH3/cSH3 concentrations. The data were processed in TopSpin 3.6 and analyzed using Sparky. NMR titrations were performed by adding unlabeled peptides to <sup>15</sup>N labeled proteins and monitored by <sup>1</sup>H-<sup>15</sup>N SOFAST-HMQC spectra recorded at each titration point. The signals of unbound nSH3 and cSH3 domains were assigned based on the published data [43, 44]. Starting protein concentrations were ~100-200 μM; the titration continued until the peptide/protein molar ratio of 3:1 was reached. Signal shifts were quantitated as amide chemical shift perturbations (CSP) calculated as follows:

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2}$$

, where  $\Delta\delta_H$  and  $\Delta\delta_N$  are signal shifts for the  $^1\text{H}$  and  $^{15}\text{N}$  resonances, respectively.

The titration curves were fitted to a single-site binding model using in-house Matlab program Kdfit [45]. The reported  $K_D$  values were determined by global fit for the selected residues, and the reported error was calculated as the standard deviation of the  $K_D$  values for the individual fits for the same residues. For the weak binders such as nSH3–S1/S3/S7 and cSH3–S8, the overall small magnitude of the observed CSPs (maximal  $\Delta\delta < 0.05$  ppm) and the limited number of titration curves precluded statistically reliable determination of the  $K_D$  values. These  $K_D$  were denoted by the asterisk in Table 2, and both the global-fit  $K_D$  and the mean  $K_D$  are reported. The cases when the  $K_D$  could not be determined due to negligibly small CSPs are denoted by ‘no binding detected’ (NBD).

#### 2.4.2 Grb2-SOS1 modeling and simulation

To construct Grb2–SOS1 complex, Grb2 structure was extracted from the Grb2 homodimer structure (PDB: 1GRI). However, full-length SOS1 PR structure is unavailable. The structure of SOS1 S4 peptide was obtained from the nSH3–SOS1 (VPPPVPPIRRR) complex (PDB: 1AZE) [36]. By contrast, the cSH3–SOS1 structure is unavailable. We used the backbone conformation of Gab2 peptide (APPPRPPKP) from the cSH3–Gab2 complex (PDB: 2W0Z) to construct the cSH3–SOS1 structure [30]. We built four Grb2–SOS1 complexes: nSH3–S4/cSH3–S10 (mode 1), nSH3–S5/cSH3–S4 (mode 2), nSH3–S5/cSH3–S10 (mode 3), and nSH3–S9/cSH3–S4 (mode 4) (Fig. S6). The side chains of Gab2 APPPRPPKP were mutated to the SOS1 peptides accordingly for the four complexes. The rest of the SOS1 disordered regions and Grb2 residues 28-33 were modeled using the Modeller server [46].

The Grb2–SOS1 structures were sampled by the replica-exchange molecular dynamics (REMD) simulation. The simulation was performed for 100 ns by using Generalized Born implicit solvent (GBIS) of NAMD2.12 [47] and the force field of CHARMM42 [48]. The temperature range was 300-320K and divided into 4 windows. The time step was 1 fs, and 1000 steps for the exchange attempt. During the simulation, the temperatures are exchanged between trajectories of the 4 windows. After the simulation, the trajectories were organized by combining the same-temperature frames into the same trajectories. The representative complex structures (Fig. 2-5) were selected from the window with temperature 300-305K. The analysis of contact frequency was performed by measuring the distances between  $\alpha$  carbons ( $C\alpha$ ) of the nSH3/cSH3–SOS1 interfaces. The contact was counted when the  $C\alpha$ – $C\alpha$  distance was smaller than 12 Å. The protocol follows our previous work [40].

## Chapter 3: SOS1 interacts with Grb2 through regions that induce closed nSH3 conformations

Chapter 3 was adapted from: TJ. Liao, H. Jang, D. Fushman, R. Nussinov: *SOS1 interacts with Grb2 through regions that induce closed nSH3 conformations*. J. Chem. Phys., 2020, 153(4), 045106.

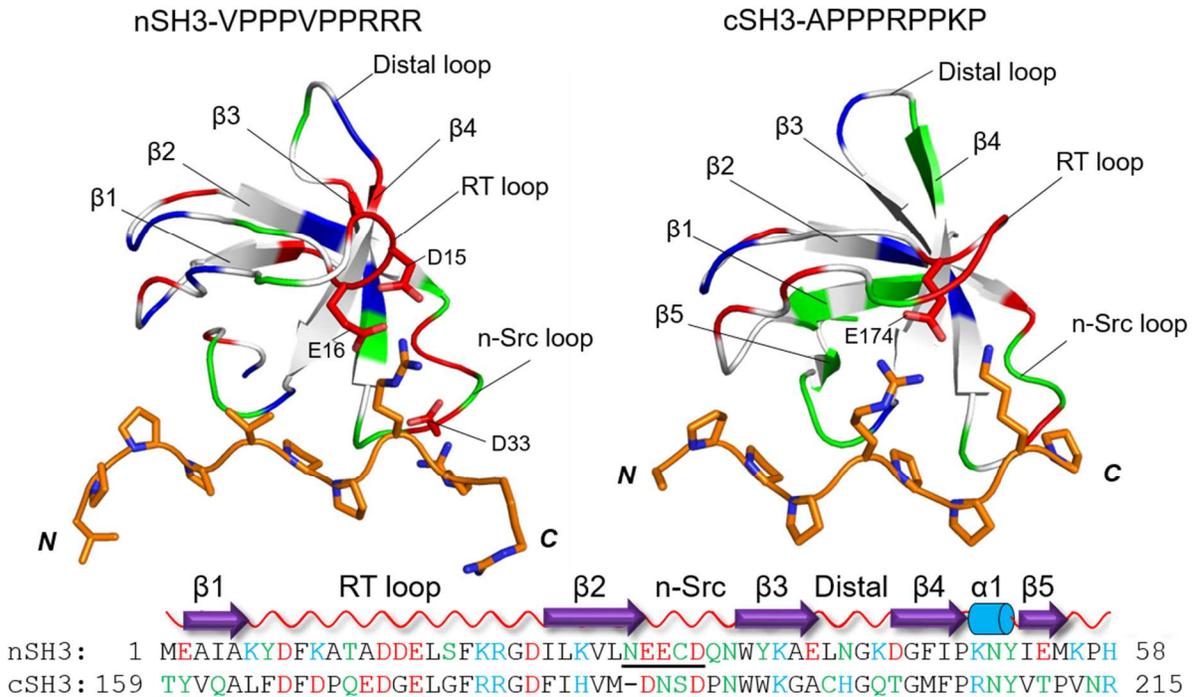
### 3.1 Introduction to chapter 3

In chapter 2, our experimental studies uncovered possible nSH3/cSH3 binding sites on the SOS1 PR domain [33, 49]. We and others tested ten PR segments, denoting them S1 to S10 (Fig. 2-1B). Sequence analysis indicated that S1, S2, and S10 contain the R/KxxK motif, and S4, S5, S6, and S9 involve the PxxPxR motif. The binding data for these ten segments indicated that nSH3 has binding affinity for S4, S5, S6, and S9 ( $K_D < 200 \mu\text{M}$ ), while cSH3 displays binding affinity for S4 and S10 ( $K_D < 200 \mu\text{M}$ ) [49]. The S4 peptide (PVPPPVPpRRRP) exhibits the highest affinity for both nSH3 ( $K_D \sim 40 \mu\text{M}$ ) and cSH3 ( $K_D \sim 140 \mu\text{M}$ ) due to a three consecutive arginine ('RRR') sequence [33]. The mutually exclusive nSH3/cSH3 binding to S4 suggested that cSH3 may use an alternative binding site on the PR region with the S10 segment [49]. SOS1 PR domain with deletion of S4, S5, S6, and S9 segments displays strong binding affinity for Grb2 *in vitro* [50]. This suggests that nSH3 and cSH3 can have multiple binding motifs of SOS1 PR domain beyond the PxxPxR motif of the S4, S5, S6, and S9 segments.

While Grb2 SH2 domain contains seven  $\beta$ -strands flanked by two  $\alpha$ -helices, nSH3/cSH3 conserve a  $\beta$ -barrel-like structure with five antiparallel  $\beta$ -strands,  $\beta_1$  to  $\beta_5$ , a short  $3_{10}$  helix,  $\alpha_1$ , as well as three loops including RT, n-Src, and distal loops (Fig. 3-1) [30, 36, 51]. The RT nomenclature reflects the observations that mutations of arginine and threonine residues within

this SH3 loop in Src tyrosine kinase are important for transforming activity. The prominent RT and n-Src loops as well as  $\beta$ 3,  $\beta$ 4, and  $\alpha$ 1 are involved in the association of the nSH3/cSH3 with the peptides. Even though nSH3/cSH3 have high sequence identity (41%), they prefer distinct ligand motifs, e.g. PxxPxR for nSH3 and PxxxRxxKP for cSH3 [29, 33, 52]. nSH3/cSH3-binding partners typically form a polyproline type II (PPII) helix, followed by one or more positively charged residues, arginine or lysine [26, 27]. When searching for the typical nSH3/cSH3 binding motifs on the SOS1 PR region, we identified four PxxPxR motifs for nSH3, but none of the PxxxRxxKP motifs for cSH3.

To comprehensively investigate the cryptic nSH3/cSH3 binding sites on SOS1 PR domain, we construct computational models of nSH3/cSH3 associating with all possible truncated SOS1 PR segments. Molecular dynamics (MD) simulations are applied to validate the stability of each complex. The effective free energy landscape of nSH3/cSH3–SOS1 reveals the most probable nSH3/cSH3 binding sites. The structural details show that the SOS1 PR segments regulate the nSH3 open/closed conformation, albeit retaining the cSH3 conformation. Importantly, the simulated complexes also provide a structural view of the unsolved nSH3/cSH3–SOS1 interactions. By providing the conformational details of nSH3/cSH3 interacting with the various SOS1 PR segments, our results may help the development of peptide inhibitors to block Grb2–SOS1 interaction, thus SOS1 recruitment to the membrane, Ras GDP/GTP nucleotide exchange and consequently activation.



**Figure 3-1.** Grb2 nSH3/cSH3 domains and the sequence of SOS1 proline-rich (PR) domain. The available NMR/crystal structures of nSH3 (residues 1-58) and cSH3 (residues 159-217) are nSH3–SOS1 VPPPVPPRRR (PDB: 1AZE) and cSH3–Gab2 APPPRPPKP (PDB: 2W0Z). Both nSH3/cSH3 contain five  $\beta$ -strands ( $\beta$ 1,  $\beta$ 2,  $\beta$ 3,  $\beta$ 4, and  $\beta$ 5) and 3 loop regions (RT, n-Src, and distal loops). Five modeled amino acids, NEECD, of nSH3 n-Src loop is indicated. The  $3_{10}$  helix,  $\alpha$ 1, is unfolded in the nSH3/cSH3 binding complexes. The polar, non-polar, positively charged, and negatively charged residues/atoms are colored by green, white, blue, and red, respectively. The backbone of nSH3/cSH3 binding ligands are indicated by orange tube and the side chains are shown as sticks. The critical residues, D15, E16, and D33 of nSH3 and E174 of cSH3, involved in the interaction with the peptides are marked.

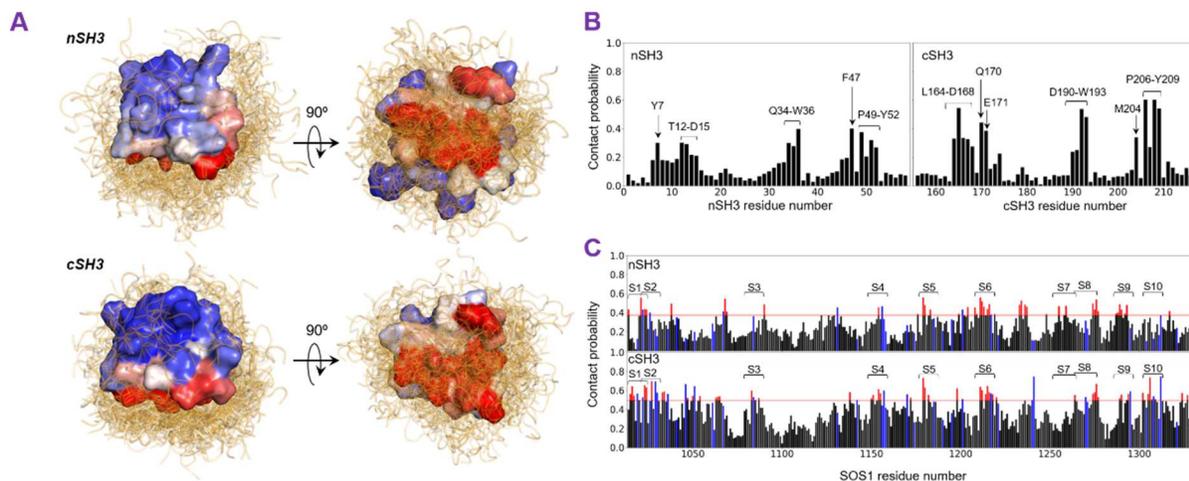
## 3.2 Results

### 3.2.1 Simulations of nSH3/cSH3–SOS1 complexes define the binding interfaces

During the MD simulations, we observed that conformations of nSH3/cSH3-favored SOS1 PR peptides fit into nSH3/cSH3 pockets and form stable nSH3/cSH3–SOS1 complexes. By contrast, unfavorable SOS1 peptides dissociated from nSH3/cSH3. Distributions of all simulated SOS1 PR peptides on the surfaces of nSH3/cSH3 clearly delineate highly populated binding interfaces (Fig. 3-2A). To define interface residues, we calculated contact probability between nSH3/cSH3 and SOS1 PR peptides (Fig. 3-2B). As indicated in the figure, residues of nSH3/cSH3 with contact probability greater than one standard deviation are defined as nSH3/cSH3 interface residues. The binding interfaces are mainly composed of three residue clusters: RT loop, n-Src loop to  $\beta$ 3, and  $\beta$ 4 to  $\alpha$ 1. For nSH3, Y7, T12, A13, D14, and D15 are in RT loop; Q34, N35, and W36 are located in n-Src loop and  $\beta$ 3; F47, P49, K50, N51, and Y52 are in  $\beta$ 4 and  $\alpha$ 1. Similarly, for cSH3, L164, F165, D166, F167, D168, Q170, and E171 are in the RT loop; D190, P191, N192, and W193 are in n-Src loop and  $\beta$ 3; M204, P206, R207, N208 and Y209 are in  $\beta$ 4 and  $\alpha$ 1. Noticeably, cSH3 has 16 interface residues, which is greater than the 13 interface residues of nSH3. Even though the interface residues involved in the cSH3–SOS1 interaction show higher contact probabilities than those in the nSH3–SOS1 interaction, the larger number of cSH3 interface residues suggests that cSH3 has less specificity for binding to SOS1 PR peptides.

To reveal the interface residues of SOS1 PR domain, we projected the nSH3/cSH3 interface residues onto the PR region. In the projection, we considered the contact probabilities between the SOS1 residues and the nSH3/cSH3 interface residues (Fig. 3-2C). The threshold (red horizontal line) was set as the average plus one standard deviation of the probability. The high contact probabilities suggest likely nSH3/cSH3 binding sites. On the probability distribution, we labeled

the experimentally defined SOS1 PR peptides, S1 to S10, based on our previous work [49]. Regions with high contact probability typically contain the PxxP motif followed by R/K, which facilitates the formation of stable nSH3/cSH3–SOS1 complex (Fig. B1). Noticeably, some regions in the PR domain also exhibit high contact probability despite partial binding of SOS1 peptide to nSH3/cSH3. These can be seen in the interactions of S1, S3, and S8 with nSH3/cSH3. In the absence of the non-polar interaction, a salt bridge interaction of R1019 (S1), R1084 (S3), and R1270/R1271 (S8) with D15 (nSH3) and E171 (cSH3), respectively, marginally prevents dissociation of the peptide (Fig. B2). However, without any R/K in S7, we observed complete dissociation of the peptide from nSH3/cSH3.



**Figure 3-2.** Molecular dynamics (MD) simulations elucidate the interface of nSH3/cSH3–SOS1.

(A) During the simulation, the SOS1 PR peptides can fully bind the nSH3/cSH3, only partially bind, and dissociate. Superimposing all 309 average structures of the nSH3/cSH3–SOS1 complex, indicate the most frequent contact regions. The surface of nSH3/cSH3 is colored from blue (0, low contact probability) to red (1, high contact probability). (B) Residues with one standard deviation above the average contact probability are viewed as surface residues and indicated for nSH3 and cSH3. (C) The contact probabilities of SOS1 PR residues are obtained from the interaction with

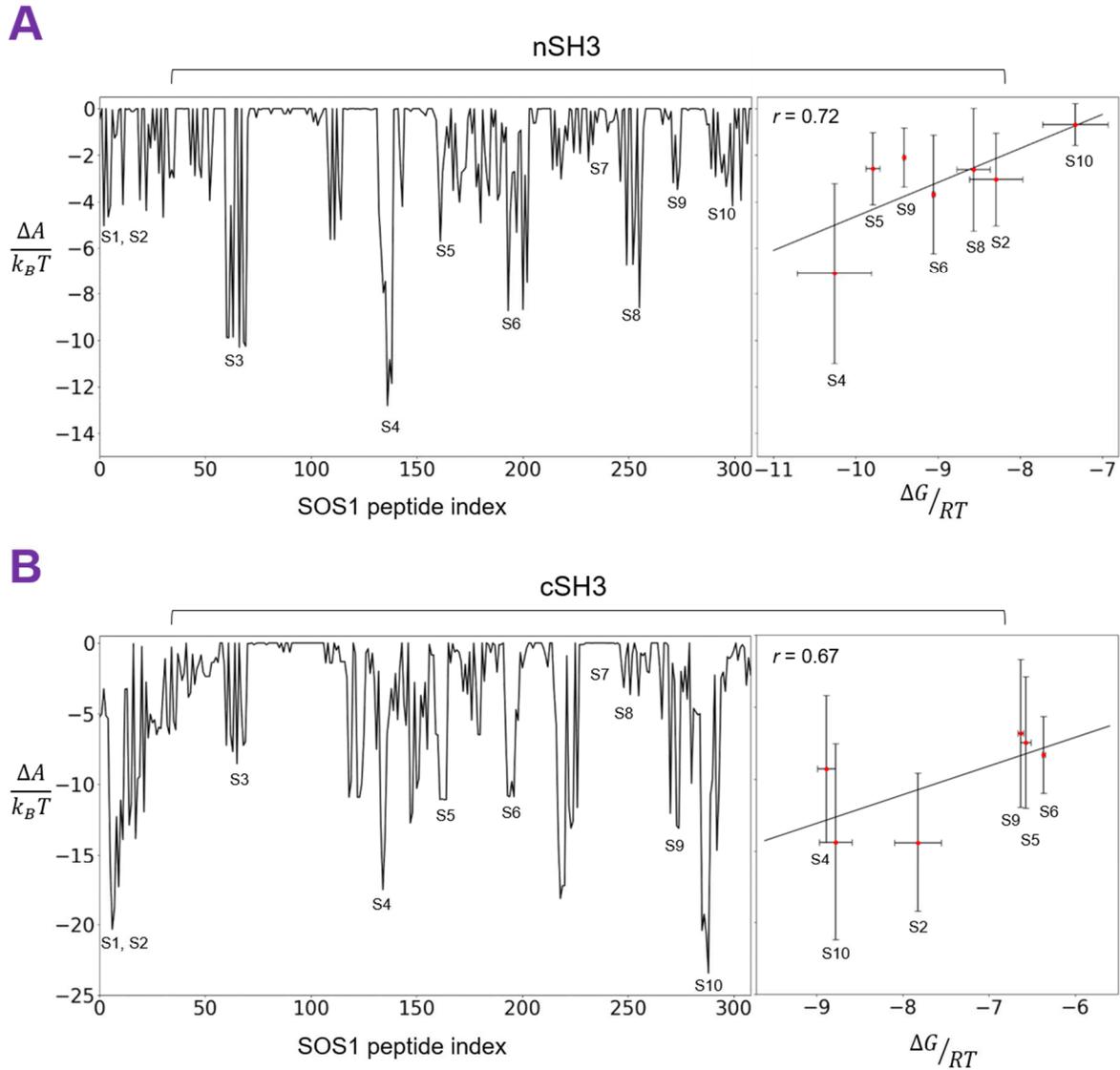
the surface residues of nSH3/cSH3. The threshold of the average contact probability plus one standard deviation is shown by the red horizontal line, and the residues above the threshold are colored blue (positively charged residues, arginine or lysine) and red (non-positively charged residues). Segments with existing experimental data are labeled S1-S10.

### 3.2.2 The effective free energy validates the multiple binding sites on SOS1 PR domain

Our simulations support the NMR data that indicated that the interactions of the SOS1 peptides with the nSH3/cSH3 domains of Grb2 are stable. Further, we observed that the peptides with sequences shifted by one or more residues from the experimental peptides adjust their nSH3/cSH3 binding sites, converging their conformations to the experimental peptides. For example, cSH3–S5<sup>-1,+1</sup> complex exhibits a final structure highly similar to cSH3–S5 (Fig. B3). Here,  $\pm n$  represents the residue number shifts in the N-terminal (-) or the C-terminal (+) direction. Similarly, cSH3–S6<sup>-1,+1</sup> converges to cSH3–S6, and cSH3–S9<sup>-3,+1</sup> converges to cSH3–S9 (Figs. B4 and B5).

The specific atom-atom interactions were observed in multiple nSH3/cSH3–SOS1 systems. Based on these observations, we calculated contact probabilities and derived the effective free energy. In the calculations, the contact probabilities of the same atom-atom interaction in different systems were converted into the partition function, which can be applied to the effective free energy landscape. For the nSH3–SOS1 interaction, S4 exhibits the most negative value of the effective free energy, followed by S3, S6, S8, S5, S2, S9, S10, S1, and S7 (Fig. 3-3A). The calculated effective free energy shows a good correlation with the experimental Gibbs free energy [49], extracted from the experimental  $K_D$  values, with the correlation coefficient of 0.72. By contrast, cSH3 yields distinct effective free energy landscape (Fig. 3-3B). Unlike nSH3, cSH3 exhibits the most negative value of the effective free energy with S10, followed by S2, S4, S9, S5, S6, S3, S1, S8, and S7, with the correlation coefficient of 0.67.

The calculated free energy predicts the stability of complex formation. However, some conformations yield a low free energy value with unstable complex formation. For example, S3 has the second negative value of the effective free energy, but exhibits partial binding, marginally attaching to nSH3 D15 through a single salt bridge with R1084 (Fig. B2A). This prompted us to employ the average structures and the root-mean-square fluctuation (RMSF) of each system in order to define probable binding modes in the nSH3/cSH3–SOS1 interaction. Stable nSH3/cSH3–SOS1 complexes require hydrophobic matching and multiple salt bridge formations between nSH3/cSH3 and SOS1. Based on this, we can designate the most stable complexes as nSH3–S2<sup>3</sup>/S4/S5/S6/S9/S10<sup>-3</sup> and cSH3–S2/S4/S5/S6/S9/S10 (Fig. B6).



**Figure 3-3.** The effective free energy landscape and the correlation with experimental affinities. The effective free energy is an approximation, which does not represent the real values of Gibbs free energy. (A) nSH3–SOS1 shows the most negative effective free energy for S4, followed by S3, S6, S8, S5, S2, S9, S10, S1, and S7. The calculated free energy has a correlation coefficient  $r = 0.72$  with the available experimental affinity data. (B) cSH3–SOS1 exhibits a distinct pattern of effective free energy landscape compared to nSH3–SOS1 with the most negative energy for S10,

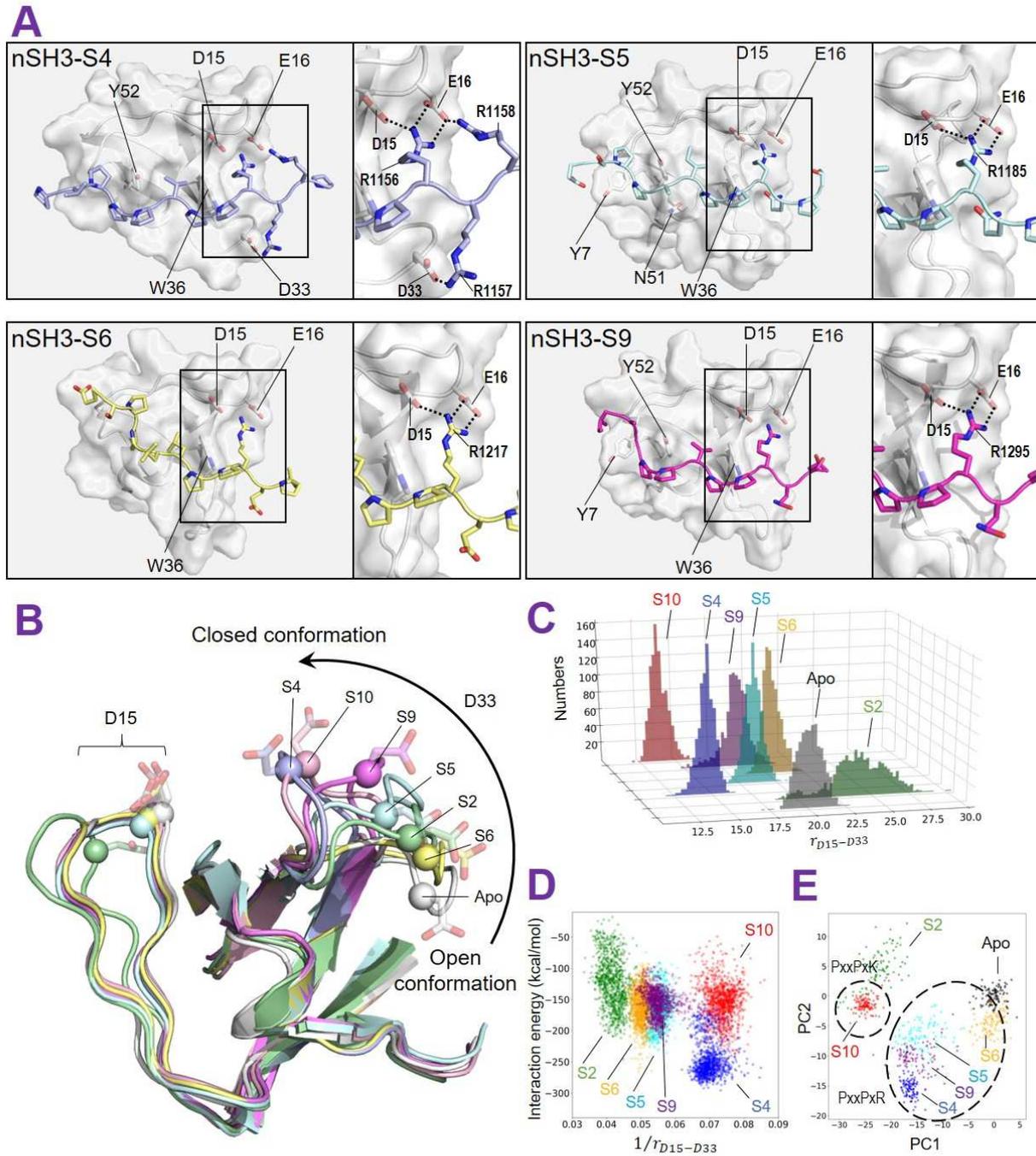
followed by S2, S4, S9, S5, S6, S3, S1, S8, and S7. The correlation coefficient to the experimental affinity is 0.67. Affinity data with poor statistics such as nSH3–S1/S3/S7 and cSH3–S8 are excluded, and cSH3 binding to S1, S3, S7 was not detected.

### 3.2.3 The nSH3 n-Src loop strengthens the interaction with SOS1, inducing closed nSH3 conformation

The nSH3 forms a stable complex with S2<sup>-3</sup>, S4, S5, S6, S9, and S10<sup>-3</sup>. S4, S5, S6, S9 contain the typical nSH3 binding PxxPxR motif. S2<sup>-3</sup> and S10<sup>-3</sup> have the PxxPxR and PxxPxK motifs, respectively. Within the PxxPxR motif, S4, S5, S6, and S9 use the arginine to form salt bridges with nSH3 D15 and D16. However, the interaction details of individual complexes vary (Fig. 3-4A). For example, only S5 and S9 fit well into the nSH3 hydrophobic pocket formed by Y7, W36, and Y52. S4 lacks the interaction for Y7 but forms an additional salt bridge with D33. S5 A1181 forms a hydrogen bond (H-bond) with nSH3 N51, and S6 loses the non-polar interaction with both Y7 and Y52. By contrast, S2<sup>-3</sup> only has the non-polar interaction with Y52 and establishes a salt bridge between K1029 and nSH3 D14. S10<sup>-3</sup> K1308 forms a salt bridge with nSH3 D15, and its PxxP motif fits into the nSH3 hydrophobic pocket formed by Y7, W36, and Y52 (Fig. B6). S10<sup>-3</sup> Y1310 forms an H-bond with nSH3 D33, resulting in n-Src loop approaching S10<sup>-3</sup>.

By superimposing the average structures of all nSH3–S2<sup>-3</sup>/S4/S5/S6/S9/S10<sup>-3</sup> complexes, we observed that nSH3 exhibits open/closed conformations regulated by the n-Src loop (Fig. 3-4B). The distance  $r$  between D15 of the RT loop and D33 of the n-Src loop can quantify the magnitude of nSH3 conformational changes. We found that  $r^{S2^*} > r^{apo} > r^{S6} > r^{S5} > r^{S9} > r^{S10^*} > r^{S4}$  (Fig. 3-4C), where S2\* and S10\* denote S2<sup>-3</sup> and S10<sup>-3</sup>, respectively. The reciprocal distance between D15 and D33 negatively correlates with the interaction energy of nSH3–SOS1, indicating that the closer conformation tends to have stronger interaction (Fig. 3-4D). Although nSH3–S10<sup>-3</sup> exhibits the

closed conformation, the interaction is reduced due to replacement of arginine with lysine. Principal component analysis (PCA) suggests that nSH3 has similar conformations when binding to S4, S5, and S9 (Fig. 3-4E). However, for S6 (ESPPLLPPREPV) with the PxxPxR motif, the glutamic acid following arginine induces a repulsive force with D33. As a result, the conformation of nSH3–S6 is more similar to that of apo-nSH3.



**Figure 3-4.** Structural details of nSH3–S4/S5/S6/S9 complexes and the nSH3 opened/closed conformations regulated by SOS1 PR peptides. (A) The average structures of nSH3 are colored white, and the peptides S4, S5, S6, and S9 are colored blue, cyan, yellow, and purple, respectively.

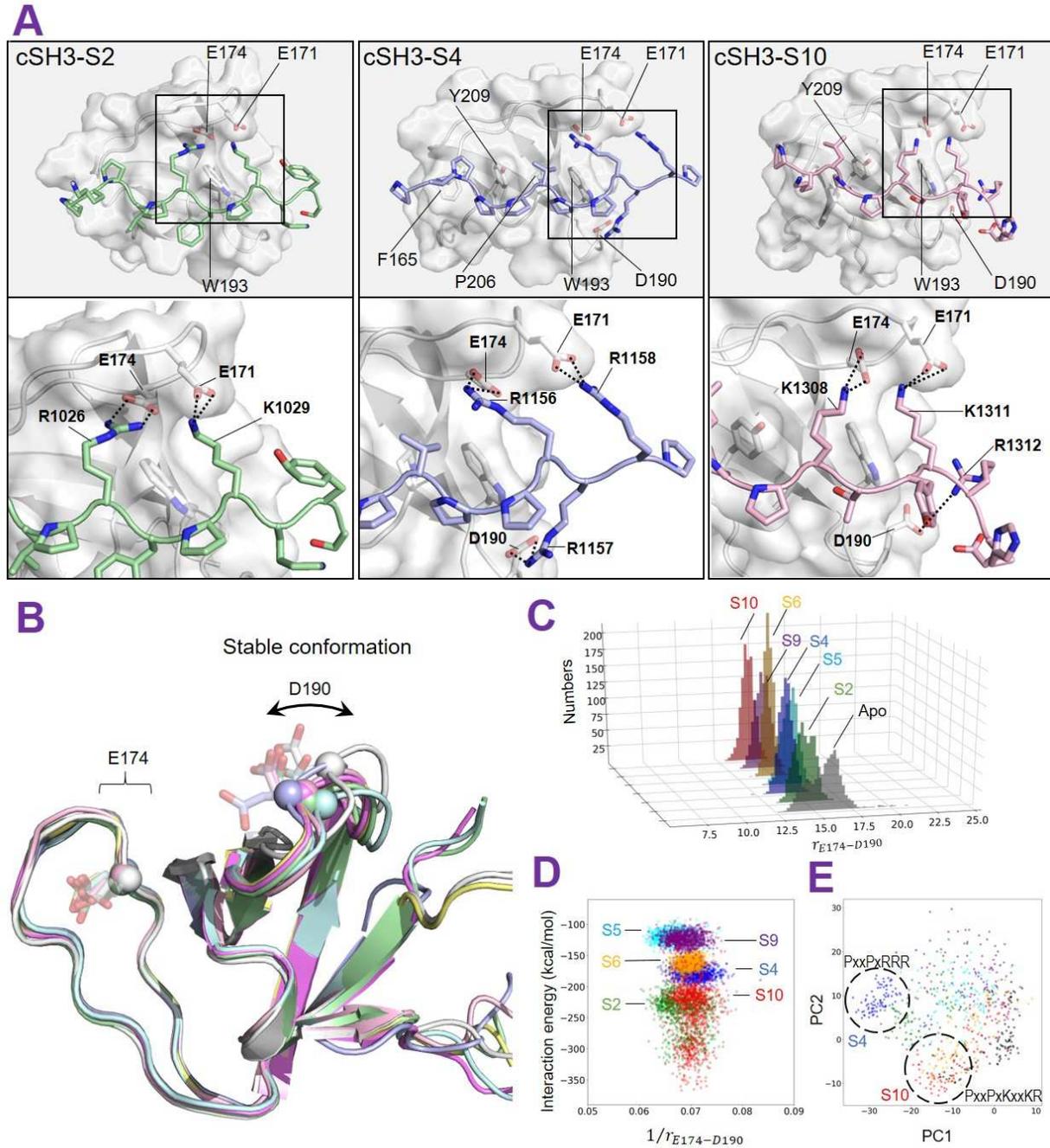
The nSH3 residues with contact probability greater than 50% for SOS1 peptides are indicated. The critical nSH3 D15 and E16 form salt bridges with S4 R1156, S4 R1158, S5 R1185, S6 R1217, and S9 R1295. The additional salt bridge of nSH3 D33 with S4 R1157 results in nSH3 having stronger affinity for S4 than S5, S6, and S9. (B) By overlaying the average structures, the nSH3 open/closed conformations modulated by the n-Src loop are observed. The average structures of nSH3 are colored green, blue, cyan, yellow, purple, and pink when binding to S2<sup>-3</sup>, S4, S5, S6, S9, and S10<sup>-3</sup>, respectively, and the apo-nSH3 is colored white. The color code applies to the following figures. While the apo-nSH3 has open conformation, the strongest binding of nSH3–S4 has a closed conformation. The critical residues D15 on RT loop and D33 on n-Src loop are indicated. (C) The histogram of distances  $r$  between D15 and D33 quantifies the magnitude of nSH3 open/closed conformations and shows that  $r^{S2^*} > r^{apo} > r^{S6} > r^{S5} > r^{S9} > r^{S10^*} > r^{S4}$ , where S2\* and S10\* denote S2<sup>-3</sup> and S10<sup>-3</sup>, respectively. (D) The reciprocal of distance  $r_{D15-D33}$  against interaction energy shows a negative correlation, suggesting that the closed conformation induces a more negative interaction energy. (D) Principal component analysis (PCA) for nSH3–S2\*/S4/S5/S6/S9/S10\* shows that nSH3 has similar conformations when binding to S4, S5, and S9. nSH3 binding to PxxPxR (S4, S5, S6, and S9) is separated from binding to PxxPxK (S2<sup>-3</sup> and S10<sup>-3</sup>). When bound to S6, nSH3 structure is closer to apo-nSH3.

### 3.2.4 No conformational change of cSH3 upon binding to SOS1

In the cSH3–SOS1 interaction, we observed that cSH3 stably forms a complex with S2, S4, S5, S6, S9, and S10, even though none of these contains the typical cSH3 PxxxRxxKP binding motif. Among them, S2, S4, and S10 form large number of salt bridges with cSH3 residues. The cSH3 E171 and E174 form salt bridges with R1026/K1029 (S2), R1156/R1158 (S4), and K1308/K1311 (S10) (Fig. 3-5A). For S5, S6, and S9, the arginine residue at the end of the sequence only interacts

with cSH3 E174 (Fig. B6). The interfaces between cSH3 and SOS1 PR segments vary. S4, S5, S6, and S9 fit well into the cSH3 hydrophobic pocket formed by F165, W193, and Y206. However, S2 lacks the interaction with F165 and Y209, and similarly, S10 loses the interaction with F165. Noticeably, S4 R1157 and S10 R1312 form additional salt bridges with D190 and strengthen the interaction with cSH3.

Superimposition of the average structures reveals that upon binding to S2, S4, S5, S6, S9, and S10, cSH3 retains the stable and consistent conformation (Fig. 3-5B). To quantitatively measure the conformational change in cSH3, we calculated the distance between E174 in the RT loop and D190 in the n-Src loop. Our calculations show highly similar values of the distance among all cSH3–SOS1 systems including apo-cSH3 (Fig. 3-5C). The distance between the RT and n-Src loops appears to be independent of the interaction energy between cSH3 and SOS1 PR segments (Fig. 3-5D). PCA fails to separate the different clusters due to the highly consistent conformation of cSH3 (Fig. 3-5E). Unlike nSH3 exhibiting open/closed conformations upon binding to ligand, cSH3 shows limited conformational changes regardless of ligands containing the PxxPxR or R/KxxK motifs.



**Figure 3-5.** The structural details of cSH3–S2/S4/S10 complexes and the cSH3 conformation. (A) The average structures of cSH3 are colored white, and the peptides S2, S4, and S10 are colored green, blue, and pink, respectively. The cSH3 residues with contact probability greater than 50% for SOS1 peptides are indicated. The cSH3 uses both E174 and E171 forming the two-to-two salt

bridges with S2 R1026 and K1029, S4 R1156 and R1158, and S10 K1308 and K1311. S4 R1157 and S10 R1312 could form additional salt bridges with cSH3 D190, enhancing the interaction of cSH3–S4 and cSH3–S10. (B) The superposition of average structures shows little difference among cSH3–S2/S4/S5/S6/S9/S10, suggesting that cSH3 tends to retain its conformation. The critical E174 of RT loop and D190 of n-Src loop are indicated. The average structures of cSH3 are colored green, blue, cyan, yellow, purple, and pink when binding to S2, S4, S5, S6, S9, and S10, respectively, and the apo-cSH3 is colored white. The color code applies to the following figures. (C) The distance between E174 and D190 is conserved among apo-cSH3 and all cSH3–SOS1 complexes. (D) The reciprocal of distance  $r_{E174-D190}$  is independent of the interaction energy. The interaction energy varies but  $r_{E174-D190}$  remains similar, and nSH3–S2/S10 have the most negative interaction energy. (E) Due to the highly consistent conformation of cSH3, the principal component analysis (PCA) hardly clusters different groups of the cSH3–SOS1 complex. The higher affinity of cSH3–S4 and cSH3–S10 could be found at the edge of the diagram.

### 3.3 Discussion

The effective free energy exhibits higher correlation to experimental affinity data for nSH3–SOS1 than cSH3–SOS1 interaction due to the better fitting nSH3–SOS1 interface. Through statistical analysis of the comprehensive simulations, we find that nSH3 has 13 interface residues (Y7, T12, A13, D14, D15, Q34, N35, W36, F47, P49, K50, N51, and Y52), and cSH3 has 16 interface residues (L164, F165, D166, F167, D168, Q170, E171, D190, P191, N192, W193, M204, P206, R207, N208, and Y209). The large number of the interface residues may result in less accurate energy estimation. Since the effective free energy is calculated by specific atom-atom interactions in the interface, the more interface residues cause more energy terms to be added. For example, when associating with S2, cSH3 has only 3 residues (E171, E174, and W193) with high

contact probability ( $> 0.5$ ). The additional energy terms deviate  $\Delta A^{cSH3-S2}/k_B T$  from the fitted curve, reducing the correlation coefficient. Further, some strong salt bridges are binding-site specific and are not reflected in the statistical analysis. For example, within the consensus PxxPxR motif, the arginine of S4, S5, S6, and S9 interacts with nSH3 E16 and cSH3 E174. However, the low contact frequency excludes nSH3 E16 and cSH3 E174 from the interface residues. Nevertheless, the non-empirical approach of the effective energy yields a fair correlation with the experimental affinity and could help explore the cryptic binding sites.

Structural convergence is only observed for the weak affinity interactions of cSH3–S5/S6/S9 ( $K_D > 1000 \mu\text{M}$ ). This implies that the energy landscape nearby S5, S6, and S9 may form a basin with shallow depth and large width. When binding to cSH3, the segments  $S5^{-1,+1}$ ,  $S6^{-1,+1}$ , and  $S9^{-3,+1}$  could shift the local energy minimum and attain structures that are highly similar to cSH3–S5/S6/S9 (Figs. B3, B4, and B5). By contrast, complexes with stronger affinity such as nSH3–S4/S5/S6/S9 and cSH3–S2/S4/S10 may have deep and narrow local energy minimum, and thus shifts by one or more residues of these SOS1 segments would result in partial association for nSH3/cSH3. S5, S6, and S9 are incapable of forming salt bridges with cSH3 E171, resulting in weak affinities. Similarly, S2 and S10 shows weaker affinity for nSH3 due to the lack of salt bridges with nSH3 D15. nSH3/cSH3 binding ligands form salt bridges with both nSH3 D15 and E16 and both cSH3 E171 and E174, and the additional interaction with nSH3 D33 and cSH3 D190 could further enhance the affinity.

Upon binding to various SOS1 PR segments, nSH3 forms open/closed conformations, while cSH3 preserves its conformation. The nSH3 open/closed conformations are modulated by the n-Src loop. The fact that nSH3 n-Src loop was not observed in the Grb2 homodimer crystal structure (PDB: 1GRI), but is present in the nSH3–VPPPVPPRRR complex (PDB: 1AZE) in solution,

suggests that the nSH3 n-Src loop is flexible and could be stabilized by the nSH3 binding ligands (Figs. 3-1A and B7). By contrast, the cSH3 n-Src loop does not regulate the closed conformation. However, the cSH3 n-Src loop participates in Grb2 homo-dimerization. Grb2 dimerizes by forming H-bonds between SH2 E87 and cSH3 Y160 as well as between cSH3 N188 and N214 (Fig. B7) [21]. Y160 and N214 are located in the cSH3 N- and C-termini, and N188 is in the n-Src loop. Notably, even though cSH3 has comparable strong affinity for S4 and S10, binding to S4 shows much higher fluctuation of N188 than binding to S10 (Fig. B8). The n-Src loops in nSH3 and cSH3 have different and critical roles: the nSH3 n-Src loop strengthens the interaction with nSH3 binding ligands and regulates the nSH3 open/closed conformation; the cSH3 n-Src loop participates in Grb2 homo-dimerization and could be partially affected by cSH3 binding ligands.

To conclude, here we carried out comprehensive simulations to investigate the interaction between Grb2 nSH3/cSH3 domains and SOS1 PR domain. We simulated the nSH3/cSH3 with continuous and truncated SOS1 PR segments and calculated the effective free energy based on the specific atom-atom interactions at the interface of nSH3/cSH3–SOS1 segments. The effective energy landscape together with the structural analysis reveal the cryptic nSH3/cSH3 binding sites on SOS1 and provide a fair correlation with the experimental affinity data. Notably, the high affinity interaction requires that nSH3/cSH3 binding ligands form salt bridges with nSH3 D15 and E16 as well as cSH3 E171 and E174. The additional interaction between nSH3 binding ligands with nSH3 D33 could further enhance the affinity and result in nSH3 closed conformation. Taken together, our results elucidate the structural mechanism for nSH3/cSH3 interaction with truncated SOS1 segments and help peptide design to inhibit Grb2 interactions with SOS1, thereby hindering SOS1 recruitment to the plasma membrane and blocking Ras activation. In particular, the multiple energetically favorable SOS1 segments identified here provide essential detailed data of the

multiple ways SOS1 can be recruited and can aid more promising designs of peptidimers for interrupting Grb2–SOS1 interaction, blocking Ras signaling pathway.

### 3.4 Methods

#### 3.4.1 Generating initial configurations of nSH3/cSH3–SOS1 PR segments

To comprehensively investigate the interactions between nSH3/cSH3 and SOS1 PR, nSH3/cSH3–SOS1 models were constructed by taking the entire sequence of the SOS1 PR domain. The nSH3/cSH3 structures were extracted from the crystal structure of Grb2 (PDB: 1GRI) [21], and the bound SOS1 peptides were modeled by using the backbone conformation of SOS1 VPPPVPPIRRR (PDB: 1AZE) [36] for nSH3 and Gab2 APPPRPPKP (PDB: 2W0Z) [30] for cSH3. The peptide structures were extended to 12 residues and sequentially mutated to SOS1 PR (residues 1014-1333). The missing Grb2 nSH3 n-Src loop (residues 28-33) and structure extension of SOS1 peptide were modeled by Modeller [46], and the mutations were accomplished with the CHARMM program [48]. In the modeled n-Src loop, the initial structure of nSH3 is in the open conformation. Sequentially mutating a 12-residue binding peptide from the entire SOS1 PR (320 residues) yielded 309 different peptides. Superimposing the extracted nSH3/cSH3 and the mutated peptides to the 1AZE and 2W0Z created 618 models for all possible nSH3/cSH3-SOS1 complexes. The apo-nSH3/cSH3 were also simulated. All complexes were solvated by TIP3P water model that constitutes the isometric unit cell box of  $64 \times 64 \times 64 \text{ \AA}^3$ . In addition, to neutralize the system,  $\text{Na}^+$  and  $\text{Cl}^-$  were added to satisfy a total ion concentration near 100 mM.

#### 3.4.2 Atomistic molecular dynamics simulations

To generate the set of starting points for a production-ready stage, we employed MD simulations using the updated CHARMM [48] all-atom additive force field [53]. Our simulations closely

followed the protocol described in our previous works [11, 25, 54-56]. A series of minimization and dynamics cycles using the steepest decent and adopted basis Newton-Raphson algorithms were performed for the solvents around the harmonically restrained protein backbone. The pre-equilibrium minimization and dynamics simulations under the same conditions were repeated until the solvent reached 310 K. At the final pre-equilibrium stage, the harmonic restraints on the backbone were gradually released with a force constant from  $k = 5$  to 0 (kcal/mol/Å<sup>2</sup>/atom). Each cycle with different  $k$  value was performed for 500,000 steps with the full particle mesh Ewald (PME) electrostatics calculation. Following the pre-equilibrium stages, each independent simulation was performed for 400 ns with the Langevin temperature control that maintains the constant temperature at 310 K and the Nosé-Hoover Langevin piston pressure control that sustains the pressure at 1 atm. With 620 independent systems (thus trajectories), and 400 ns each, the time scales of our simulations reached a total of 248 μs. For the production runs, the NAMD parallel computing code (version 2.12) [47] with the CHARMM [48] force field on a Biowulf cluster at the National Institutes of Health (Bethesda, MD) was employed. The average structures were taken for the last 200 ns to ensure equilibration of the system.

### 3.4.3 Effective free energy calculations

The calculation of effective free energy for each system is based on the atom-atom interaction between nSH3/cSH3 and SOS1 peptides. The same atom-atom interaction in different models has different interaction probabilities. For each model, the probability shows the number of contacts over the trajectory length, obtained over the last half of the trajectories. The discrete probabilities show unnormalized distribution. The total number of contacts,  $N$ , is the sum of the contacts of each model,

$$N = \sum_{i=0}^n N_i \quad (3-1)$$

, where,  $N_0$  is the biggest number of contacts and may represent the most stable state among the various models. We use the probability to obtain the partition function (dividing  $N$  by  $N_0$ ) and denote it as the effective partition function,  $Z^{eff}$ ,

$$Z^{eff} = \frac{N}{N_0} = \frac{1}{N_0} \sum_{i=0}^n N_i \quad (3-2)$$

In terms of  $Z^{eff}$ , the effective free energy,  $A^{eff}$ , can be deduced by

$$A^{eff} = -k_B T \ln Z^{eff} \quad (3-3)$$

, where  $k_B$  is the Boltzmann constant, and  $T$  is the simulation temperature. The total effective free energy for the  $i$ -th model is a sum of contributions from salt bridges, hydrogen bonds, and non-polar interactions,

$$A^{eff} = \sum_j (A_j^E + A_j^H + A_j^N) \quad (3-4)$$

, where  $j$  denotes the specific atom-atom interaction, and  $A^E$ ,  $A^H$ , and  $A^N$  represent the effective free energy of electrostatics, hydrogen bond, and non-polar interactions, respectively.

## Chapter 4: NMR data indicate Grb2 nSH3/cSH3 domains high affinity for SOS1 proline-rich domain motifs

### 4.1 Introduction to chapter 4

The Grb2–SOS1 interaction can be negatively regulated by the MAPK signaling cascade [57-60]. SOS1 activates Ras by exchanging GDP to GTP [25]. Active GTP-bound Ras proteins form nanoclusters in the plasma membrane which promote Raf kinase domain dimerization, leading to Raf activation [56]. Dimeric Raf kinase domains gather in large complexes, which includes mitogen-activated protein kinase 1/2 (MEK1/2) and extracellular signal-regulated kinase 1/2 (ERK1/2) dimers [61] and KSR. Negative feedback occurs when active ERK1/2 phosphorylates SOS1 residues, Ser1132, Ser1167, Ser1178, and Ser1193 in the PR domain, disrupting the association with Grb2 [59]. Ribosomal S6 kinase 2 (RSK2) known as the ERK1/2 effector also phosphorylates SOS1 residues, Ser1134 and Ser1161 in the PR domain, which create the 14-3-3 docking sites [62]. The association of 14-3-3 with the PR domain disrupt Grb2–SOS1 association and reduces the SOS1 plasma membrane localization [62]. Grb2–SOS1 regulates MAPK signaling, leading to ERK1/2 activation, which in turn inhibits Grb2–SOS1 binding [60]. Other SOS1 residues, Ser1078, Ser1082, Ser1210, Ser1229, and Ser1275 in the PR domain are also phosphorylated [63, 64], but the role of these phosphorylated serine, if there is one, are unclear [65].

Released from SOS1, Grb2 can bind to Gab2 and transmit PI3K signaling. However, although Grb2 binding to Gab2 promotes PI3K signaling, the negative feedback regulation is also subject to Grb2–Gab2 interaction [66]. Gab2 contains a highly

conserved N-terminal pleckstrin homology (PH) domain and an intrinsically disordered C-terminal tail [67]. In the central region, adjacent to the PH domain, Gab2 contains the PR domain with many PxxP motifs for binding to Grb2 SH3 domains. The C-terminal tail retains the phosphorylated tyrosine motifs, recruiting the SH2 domain of p85 of PI3K [68]. AKT phosphorylates Gab2 residues, Ser210 and Thr391 in the C-terminal tail [66]. Binding of 14-3-3 protein to the Gab2 phosphorylated Ser210 and Thr391 promotes the dissociation of Gab2 from Grb2 [66].

Grb2 binds to SOS1 with 1:1 stoichiometry, and the simultaneous binding of nSH3 and cSH3 domains to SOS1 PR domain shows high affinity of  $K_D \sim 0.3 \mu\text{M}$  [32]. By contrast, Grb2 nSH3 domain is not involved in the interaction with Gab1, binding of Grb2 to Gab1 exhibits 2:1 stoichiometry with the affinity of  $K_D \sim 7 \mu\text{M}$  [69]. Since Gab family members, Gab1 and Gab2 lack the nSH3 binding motifs but retain the highly conserved cSH3 binding sites [69], Grb2 may follow a similar binding mechanism to Gab1/2, using both nSH3 and cSH3 binding to SOS1 PR, whereas only cSH3 can associate with Gab1/2. Grb2 cSH3 domain becomes a “bridge”, which communicates between MAPK and PI3K pathways by modulating the interactions of Grb2–SOS1 and Grb2–Gab complexes. However, Grb2 binding to SOS1 exhibits approximately 20-fold higher affinity than to Gab1. Even though serine phosphorylation on SOS1 PR domain blocks part of nSH3/cSH3 binding sites, Grb2 cSH3 binding to SOS1 may still outcompete cSH3–Gab1 association. Therefore, why Grb2 cSH3 domain shifts the binding from SOS1 to Gab1/2, forming a SOS1–Grb2–Gab ternary structure, remains unclear.

Binding of the phosphorylated EGFR peptide, EpYINQSV, to Grb2 SH2 domain does not change the affinity of SH3 domains for the one-site SOS1 PR peptides [70, 71]. Regardless of the presence of EpYINQSV, SOS1 peptide PVPPPVPPIRRRPESA performs the same affinity of  $K_D \sim 22 \mu\text{M}$  for the intact Grb2 and the EpYINQSV–Grb2 complex [71]. However, how the one-site SOS1 PR peptides interact with the individual nSH3 and cSH3 domain are unknown. Moreover, even though the SH2 and nSH3/cSH3 bindings seem to be independent, association of EpYINQSV–Grb2 is affected by the interaction of Grb2 with the full length SOS1 [32]. EpYINQSV binding to Grb2–SOS1 ( $K_D \sim 0.3 \mu\text{M}$ ) is stronger than Grb2 alone ( $K_D \sim 0.7 \mu\text{M}$ ). The indirect interaction between EpYINQSV and SOS1 is mysterious.

Here, to clarify how Grb2 nSH3 and cSH3 domains recruit two separate PR segments, we examine the interaction of monomeric Grb2 with the truncated SOS1 PR segments by NMR measurements. This is the first time the individual nSH3 and cSH3 bindings of full length Grb2 are elucidated by using our binding model. Our affinity data reveal that both nSH3 and cSH3 have much higher affinity for the selected one-site SOS1 PR peptides than the isolated nSH3/cSH3. Grb2 binding to the two-site SOS1 peptides shows larger chemical shift perturbation (CSP) than to the one-site SOS1 peptides, suggesting that the interaction of cSH3 with nSH3 binding motif was promoted. While the binding of Grb2 nSH3/cSH3 to SOS1 peptides shows high cooperativity, the EGFR segment binding to Grb2 SH2 domain has zero cooperativity for Grb2 nSH3/cSH3–SOS1 peptides. Understanding the mechanism of full length Grb2 interacting with one-/two-site SOS1 peptides may facilitate the development of inhibitors to prevent MAPK signaling.

## 4.2 Results

### 4.2.1 The one-site and two-site segments of SOS1 PR domain

To examine the interaction of Grb2 with SOS1, we extracted five potential nSH3/cSH3 binding sites from the PR domain of SOS1, PVPPPVPPIRRRP (S1), DSPPAIPPRQPT (S2), ESPPLLPPREPV (S3), IAGPPVPPRQST (S4), and PKLPPKTYKREH (S5) (Table 4-1). The SOS1 peptides, S1-S4 were extended and/or inserted with a tyrosine in order to obtain accurate concentration by measuring the UV absorption. The modified S1 to S4 peptides were denoted S1\* to S4\* (Table 4-1). In S1, we inserted three additional residues, glutamate, serine, and a tyrosine between the serine and alanine, obtaining PVPPPVPPIRRRPESYA (S1\*). The S2 peptide was elongated by including five more wild-type residues, generating DSPPAIPPRQPTSKAYS (S2\*), which contains a tyrosine. Similarly, The S3 peptide was elongated by adding two residues tyrosine and threonine, yielding ESPPLLPPREPVYT (S3\*). In S4, a serine was added to the N-terminus and two residues tyrosine and serine were added to the C-terminus, generating SIAGPPVPPRQSTYS (S4\*). Recent data indicate that the typical SOS1 binding surfaces include Tyr7, Phe9, Asp15, Glu16, Trp36, Phe47, Pro49, and Tyr52 for the nSH3 domain, and Asp168, Gln170, Glu171, Glu174, Trp193, Phe205, Arg207, and Tyr209 for the cSH3 domain [49, 72]. For the modified peptides, we ensured that the additional residues including the tyrosine modifications are distal from the nSH3/cSH3 binding surfaces and should not affect the binding affinities. By considering the neighbor nSH3/cSH3 binding sites, the two-site SOS1 PR segments, S12, S23, and S45 were selected, which contain 48, 48, and 29 amino acids, respectively (Table 4-1). We

abandoned the two-site segment involving S3 and S4, since the peptide of the segment is 90 residues long, indicating that the synthesis of the longer peptide has the lower success rate. The S34 peptide contains two nSH3 binding PxxPxR motifs, which is similar to S23.

Monomeric Grb2 was collected for examining the interaction with the SOS1 peptides. In the dimeric Grb2 Glu87<sup>SH2</sup> interacts with Tyr160<sup>cSH3</sup>, which blocks the binding site on cSH3. When SOS1 peptides bind to dimeric Grb2, the cSH3 domain in the dimer interface competes with the SOS1 interaction, suggesting that the measured affinity is smaller than the actual affinity for the cSH3–SOS1 interaction. To verify that the collected Grb2 was monomeric molecule, we measured the transverse ( $T_2$ ) <sup>15</sup>N spin-relaxation time for the sampled Grb2 and compared it with the EpYINQSV–Grb2 complex. The EpYINQSV peptide is a phosphorylated human EGFR segment (residue 1101-1107), which binds to Grb2 SH2 domain, retaining Grb2 as a monomer [22, 23]. The averaged  $T_2$  relaxation times for the sampled Grb2 and EpYINQSV–Grb2 complex are 52.4 and 53.2 ms, respectively. The similar values of  $T_2$  relaxation time suggest that the collected sample contains a Grb2 monomer (Fig. C1A).

Attenuation of NMR signals were observed in all our experiments (Table C1). Signal attenuation is usually resulted from the increased size of molecular complex and/or the local electronic environment changes of observed nucleus due to association/dissociation. The increased size of molecular complex generally caused the signal broadening across all residues. By contrast, residues close to binding surfaces switch between bound and unbound states. When the exchange rate is much slower than the resonance frequency difference in NMR measurement, the slow exchange rate

performs longer residence time in either bound or unbound state, resulting in severe signal broadening and disappearance. Since the molecular weight of Grb2 is much greater than the EpYINQSV and the SOS1 one-site peptide, the overall changes of signal intensity upon binding were not significant. However, due to the lack of binding model to describe the intrinsic interaction of Grb2 with two-site SOS1 peptides, the magnitude of signal attenuation was used to evaluate the association of Grb2–S12, Grb2–S23, and Grb2–S45.

**Table 4-1.** The selected nSH3/cSH3 binding peptides of SOS1

S1	PVPPPVPPRRRP
S2	DSPPAIPPRQPT
S3	ESPPLLPPREPV
S4	IAGPPVPPRQST
S5	PKLPPKTYKREH
S1*	PVPPPVPPRRRPESYA
S2*	DSPPAIPPRQPTSKAYS
S3*	ESPPLLPPREPVYT
S4*	SIAGPPVPPRQSTYS
S12	<u>EVPVPPPVPPRRRPESAPAESSPSKIMSKHLD</u> DSPPAIPPRQPTSKAYS
S23	DSPPAIPPRQPTSKAYS <u>SPRYSISDRTSISDPPE</u> ESPPLLPPREPVRTPD
S45	SIAGPPVPPRQSTS <u>QH</u> IPKLPPKTYKREH

SOS1 peptides, S1-S5 were identified as the nSH3/cSH3 binding sites. The modified S1\*-S4\* were introduced a tyrosine for accurate concentration measurements. The segments with two proximal nSH3/cSH3 binding sites S12, S23, and S45 were selected.

#### 4.2.2 Grb2 nSH3 and cSH3 have high affinities for S1, S2, S3, S4, S5

S1\* (PVPPPVPPRRRPESYA) associates with the nSH3 and cSH3 domains of full length Grb2 with  $K_D = 3$  and  $15 \mu\text{M}$ , respectively, which are about 10 fold stronger than the S1 interacting with the isolated nSH3 ( $K_D = 39 \mu\text{M}$ ) and cSH3 ( $K_D = 125 \mu\text{M}$ )

(Table 4-2). However, it is unclear how the affinities of full length Grb2 binding to S1\* are enhanced. Grb2 residues with CSPs greater than the average, summing up to two standard deviations were pointed out (Fig. 4-1). The nSH3 residues Lys10, Ala11, Asp15, Tyr37, Tyr52, and cSH3 residues Asp168, Asp171, Trp193, Trp193ε, Phe205, and Arg207 show large CSPs (Fig. 4-1A). The epsilon denotes the indole NH group of Tryptophan. Signal attenuations of Lys38 and Lys50 were observed in the NMR spectra. The corresponding CSPs were mapped to the nSH3/cSH3 domains. Noticeably, while Grb2 cSH3 interacting with S1\* has lower affinity than Grb2 nSH3–S1\*, cSH3 residues on the binding surface Asp168, Asp171, Trp193, Trp193ε, and Phe205 show larger CSPs than the nSH3 residues.

S2\* (DSPPAIPPRQPTSKAYS) also shows high affinities for Grb2 nSH3 ( $K_D = 3 \mu\text{M}$ ) and cSH3 ( $K_D = 7 \mu\text{M}$ ) (Table 4-2). S2 was previously reported to strongly bind to the isolated nSH3 ( $K_D = 56 \mu\text{M}$ ), but weakly to the isolated cSH3 ( $K_D = 1396 \mu\text{M}$ ) (Table 4-2). Grb2 nSH3 exhibits 20 folds stronger binding for S2\* than the isolated nSH3 binding to S2. The nSH3 residues Lys38, Phe47, and Lys50 exhibit signal attenuation. Grb2 cSH3 has roughly 200 folds higher affinities for S2\* than the interaction of isolated cSH3 with S2. Residues with large CSPs are nSH3 Tyr7, Asp8, Lys10, Ala11, Ala13, Asp15, Tyr37, Tyr52, and cSH3 Trp193ε (Fig. 4-2B). Even though most nSH3 residues exhibit larger CSPs than cSH3, the high affinity of nSH3–S2\* interaction is comparable to the cSH3–S2\*.

S3\* (ESPPLLPPREPVYT) contains two glutamic acids. The second, which follows the critical arginine, interferes with the formation of salt bridges between the arginine and the nSH3 Asp15 and Glu16 as well as the cSH3 Glu171 and Glu174 [72].

Therefore, S3 has the weakest binding affinities for the isolated nSH3 ( $K_D = 117 \mu\text{M}$ ) and cSH3 ( $K_D = 1718 \mu\text{M}$ ) among S1 to S4. However, similar to Grb2–S2\* interaction, S3\* has high affinities for Grb2 nSH3 ( $K_D = 14 \mu\text{M}$ ) and cSH3 ( $K_D = 9 \mu\text{M}$ ) (Table 4-2). Residues Tyr37, Lys38, Phe47 and Lys50 show signal attenuation. While the affinity of cSH3–S3\* is slightly higher than nSH3–S3\*, cSH3 residues have overall small CSPs. Residues with large CSPs are nSH3 Tyr7, Asp8, Ala11, Ala13, Asp15, Glu16, Ser18, Val27, Asn51, and Tyr52 (Fig. 4-1C).

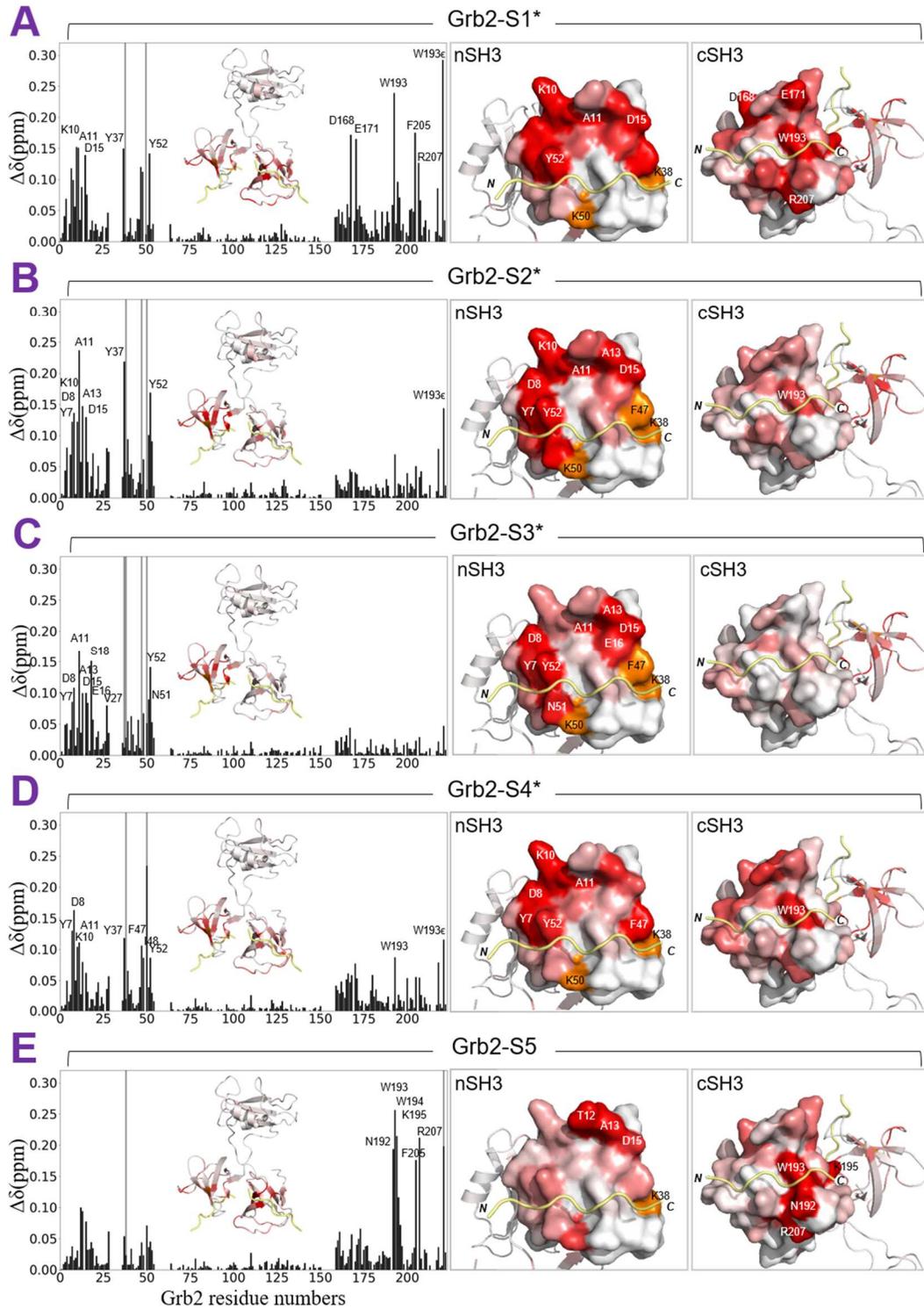
The affinities of S4\* (SIAGPPVPPRQSTYS) to Grb2 are  $K_D = 9$  and  $13 \mu\text{M}$  for Grb2 nSH3 and cSH3 domains, respectively, while S4 affinities are  $K_D = 82$  and  $1318 \mu\text{M}$  for the isolated nSH3 and cSH3, respectively (Table 4-2). Similarly, within the full length Grb2, the interactions of nSH3 and cSH3 with S4\* are much stronger than the isolated nSH3/cSH3–S4. The nSH3 residues Tyr7, Asp8, Lys10, Ala11, Tyr37, Phe47, Ile48, Tyr52, and cSH3 Trp193 and Trp193 $\epsilon$  show larger CSPs (Fig. 4-1D). Only Lys38 and Lys50 undergo signal attenuation.

S5 (PKLPPKTYKREH) weakly binds to the isolated nSH3 ( $K_D = 620 \mu\text{M}$ ) but shows higher affinity for the isolated cSH3 ( $K_D = 150 \mu\text{M}$ ). This suggests that S5 is a cSH3-specific binding site on SOS1. S5 interacting with full length Grb2 exhibits  $K_D = 21$  and  $23 \mu\text{M}$  for nSH3 and cSH3, respectively (Table 4-2). The nSH3 Trp36 $\epsilon$  and Lys38 show signal attenuation. The cSH3 residues Asn192, Trp193, Trp194, Lys195, Phe205, and Arg207 exhibit large CSPs (Fig. 4-1E). Interestingly, even though the affinities of Grb2 interacting with S5 are also enhanced, they are still smaller than the Grb2 interacting with S1\*, S2\*, S3\*, and S4\*.

**Table 4-2.** Binding affinities of SOS1 peptides with Grb2 and EpYINQSV– Grb2

SOS1	Grb2		EpYINQSV–Grb2		SOS1	Isolated	
	nSH3	cSH3	nSH3	cSH3		nSH3	cSH3
S1*	3 ± 2	15 ± 13	4 ± 6	9 ± 7	S1	39 ± 1	125 ± 13
S2*	2 ± 3	7 ± 4	2 ± 2	7 ± 6	S2	56 ± 5	1396 ± 87
S3*	14 ± 11	9 ± 11	25 ± 13	10 ± 15	S3	117 ± 2	1718 ± 33
S4*	9 ± 14	13 ± 11	4 ± 4	13 ± 15	S4	82 ± 1	1318 ± 44
S5	21 ± 18	23 ± 18	18 ± 19	26 ± 13	S5	620 ± 240	150 ± 30

The reported dissociation constants ( $K_D$  in unit  $\mu\text{M}$ ) were obtained by fitting NMR titration curves for the selected residues, followed by averaging the individual  $K_D$  values (left column). The errors were calculated as the standard deviation for the same residues. The  $K_D$  values in right column are published data from the previous work [33, 49].



**Figure 4-1.** NMR characterization of the SOS1 peptides interacting with full length Grb2. Amide chemical shift perturbation (CSP,  $\Delta\delta$ ) at the titration end point are shown

as function of Grb2 residue numbers for (A) Grb2–S1\*, (B) Grb2–S2\*, (C) Grb2–S3\*, (D) Grb2–S4\*, and (E) Grb2–S5. CSPs of indole NH signals of Trp36, Trp60, Trp121, Trp193, and Trp194 are denoted as Trp36 $\epsilon$ , Trp60 $\epsilon$ , Trp121 $\epsilon$ , Trp193 $\epsilon$ , and Trp194 $\epsilon$  and located at the residue number 218, 219, 220, and 221. The epsilon denotes the molar extinction coefficient of Tryptophan. Signals with attenuation during titration are shown as grey bars (Table C1). Residues with  $\Delta\delta$  greater than two standard deviation are indicated. CSPs are mapped onto the surfaces of Grb2 in complexes with the corresponding modeled peptides (yellow). Surface coloring based on the observed CSP is colored from white (0 ppm) to red (maximal  $\Delta\delta = 0.1$  ppm). Residues with signal attenuation are colored orange. Structure images to the right show the enlarged binding surfaces of nSH3 and cSH3 in detail.

#### 4.2.3 EGFR segment has limited effect on Grb2 binding to S1, S2, S3, S4, S5

To investigate how EGFR affects Grb2 interaction with the SOS1 PR peptides S1\*, S2\*, S3\*, S4\*, and S5, the phosphorylated EGFR segment, EpYINQSV, was first added to the Grb2. EpYINQSV associates with Grb2 SH2 domain and has the least interactions with nSH3/cSH3. The SH2 binding pocket is formed by Ile65, Ala68, Leu84, Arg86, Phe108, Leu111, Trp121, and Val122 [73-75]. The binding of EpYINQSV to SH2 domain induces large CSPs at Ile65, Val99, Gly102, Phe108, Leu111, and Trp121 $\epsilon$  (Fig. C1B). The signals of Lys109, Trp121, and Val122 are attenuated. By contrast, EpYINQSV induces small CSPs in the nSH3/cSH3 regions, and similarly, nSH3/cSH3 binding peptides S1\*, S2\*, S3\*, S4\*, and S5 also induce small CSPs on SH2 region (Fig. 4-2). The small CSPs indicates the specific bindings of SH2 and nSH3/cSH3 with their binding partners.

For EpYINQSV–Grb2 interacting with S1\*,  $K_D = 4$  and  $9 \mu\text{M}$  for the nSH3 and cSH3 domains, respectively (Table 4-2). The nSH3 residues Lys10, Ala11, Asp15, Tyr52, and cSH3 residues Asp168, Glu171, Trp193, Phe205, and Arg207 show large CSPs (Fig. 4-2A). By subtracting the CSPs of Grb2–S1\* from the CSPs of EpYINQSV–Grb2–S1\*, most of the residues show small changes of CSP ( $\Delta\Delta\delta$ ) (Fig. 4-3A). Residues with  $|\Delta\Delta\delta| > 0.05$  ppm such as Ala13, Trp36 $\epsilon$ , Met186, and Tyr209 are pointed out. The signal attenuation was evaluated by calculating the intensity ratio of titration end point to the free Grb2. Residues Trp36 $\epsilon$ , Lys38, Phe47, and Lys50 show severe signal attenuation ( $I/I_0 = 0$ ) while Lys10, Asp15, Tyr37, Phe125, Asp133, and Met186 exhibit high intensity ratio ( $I/I_0 > 1.5$ ) (Fig. 4-3A).

Similarly, S2\* exhibits high affinities for nSH3 ( $K_D = 2 \mu\text{M}$ ) and cSH3 ( $K_D = 7 \mu\text{M}$ ) of the EpYINQSV–Grb2 complex (Table 4-2). The residues Asp8, Ala11, Ala13, Asp15, Leu28, Phe37, Tyr52, and Trp193 $\epsilon$  show large CSPs (Fig. 4-2B). Four residues with large  $\Delta\Delta\delta$  are Tyr7, Arg21, Trp36 $\epsilon$ , Cys198, and Tyr209 (Fig. 4-3B). Signals of Lys38, Phe47, and Lys50 disappeared during titration. Signal intensity ratios of Ala3, Lys10, Thr12, Asp15, Glu16, Tyr37, Phe125, Asp133, Thr159, and Tyr160 were largely increased in the presence of EpYINQSV (Fig. 4-3B).

By contrast, S3\* interaction with EpYINQSV–Grb2 shows weaker affinity for nSH3 ( $K_D = 25 \mu\text{M}$ ) than cSH3 ( $K_D = 10 \mu\text{M}$ ) (Table 3-2). However, cSH3 residues show small CSPs, and nSH3 residues Asp8, Ala11, Asp15, Glu16, Ser18, Lys50, Asn51, and Tyr52 exhibit large CSPs (Fig. 4-2C). The overall  $\Delta\Delta\delta$  are within  $\pm 0.05$  ppm (Fig. 4-3C). Asp15, Tyr37, Lys38, Phe47, and Lys50 show signal attenuation. The increased

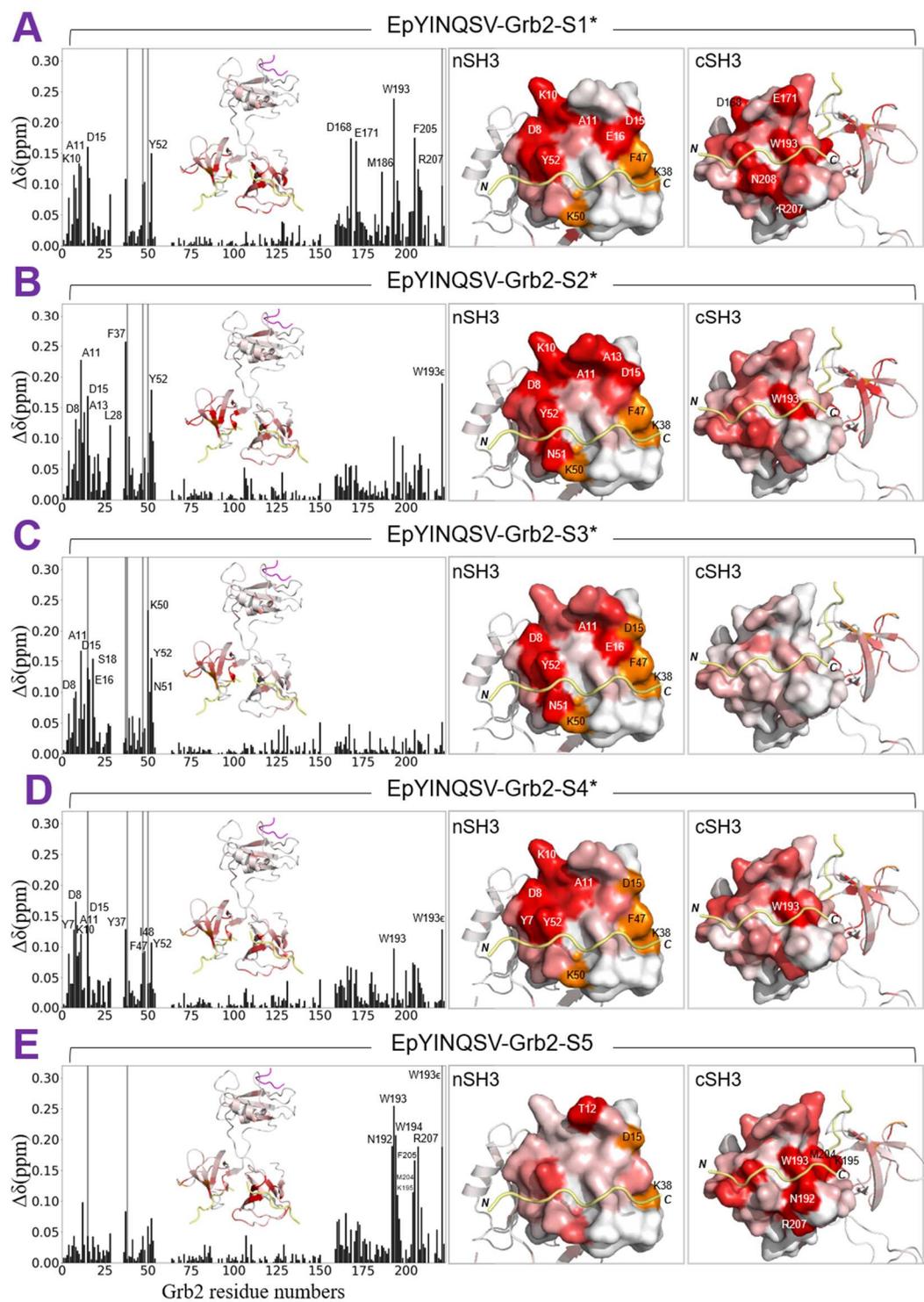
signal intensity ratio of Ala3, Lys15, Thr12, F125, D133, Tyr160, Met186, and Asn208 were observed (Fig. 4-3C).

The affinity of S4\* associating with EpYINQSV–Grb2 are  $K_D = 4$  and  $13 \mu\text{M}$  for nSH3 and cSH3 domains, respectively (Table 4-2). The nSH3 residues Tyr7, Asp8, Lys10, Ala11, Asp15, Tyr37, Phe47, Ile48, and Tyr52 as well as cSH3 Trp193 and Trp193 $\epsilon$  show large CSPs (Fig. 4-2D). Only residues Asp15, Trp36 $\epsilon$ , and Met204 have  $|\Delta\Delta\delta| > 0.05$  ppm (Fig. 4-3D). Similarly, residues Asp15, Lys38, Phe47, and Lys50 perform signal attenuation. While Lys10, Thr12, Glu16, Lys124, Phe125, and Asp133 have  $I/I_0 > 1.5$ , the overall signal intensity is reduced in the presence of EpYINQSV (Fig. 4-4).

S5 binding to EpYINQSV–Grb2 has  $K_D = 18$  and  $26 \mu\text{M}$  for nSH3 and cSH3 domains, respectively (Table 4-2). The cSH3 domain shows weaker affinity but larger CSPs on the residues Asn192, Trp193, Trp193 $\epsilon$ , Trp194, Lys195, Met204, Phe205, and Arg207 (Fig. 4-2E). Residues with  $|\Delta\Delta\delta| > 0.05$  ppm are Ala13, Phe165, Met204, and Tyr209 (Fig. 4-3E). Three residues Asp15, Trp36 $\epsilon$ , and Lys38 were vanished upon binding, and only Lys10, Lys124, and Phe125 have larger signal intensity ratio (Fig. 4-3E).

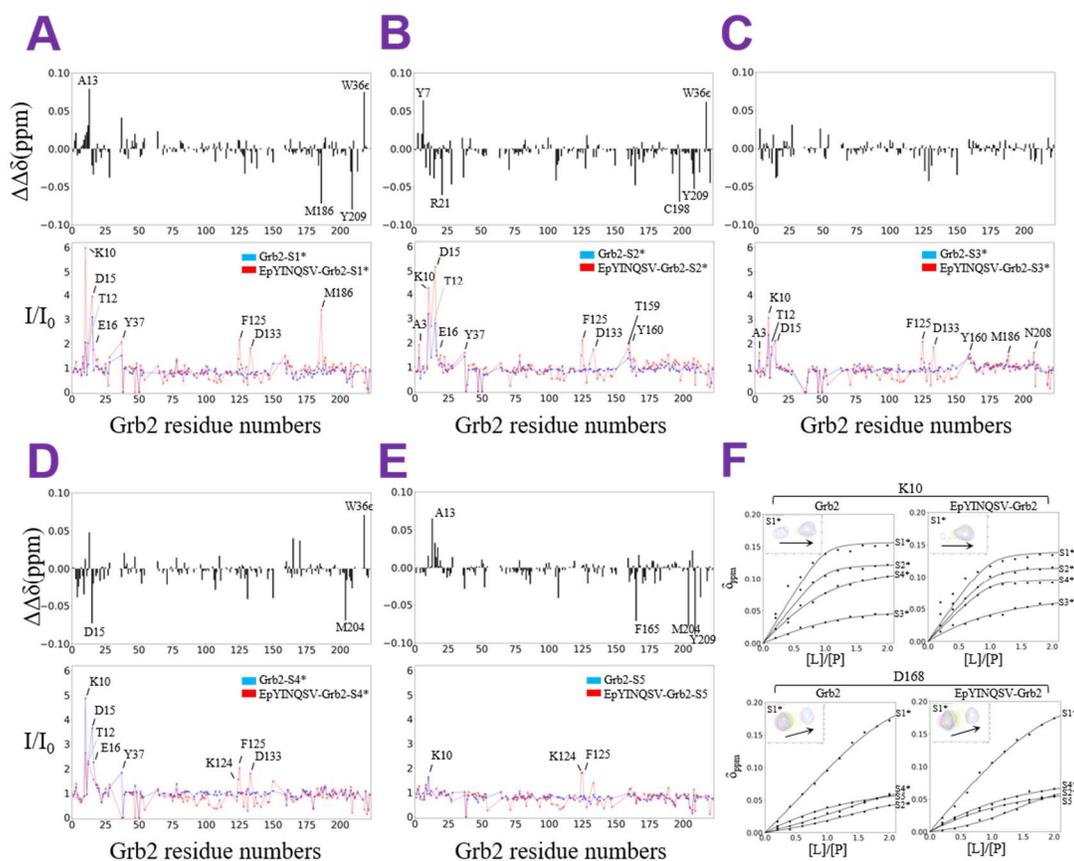
Binding of Grb2 with the one-site SOS1 peptides perform the signal intensity ratio  $I/I_0 \sim 1.0 \pm 0.2$  for nSH3, SH2, and cSH3 domains (Table 4-3). While the deviation is small, some residues such as Lys10 and Asp15 at or near the nSH3 binding sites exhibit large signal intensity ratio upon Grb2 binding to EpYINQSV. Therefore, we selected Lys10 in nSH3 and Asp168 in cSH3 to examine the relation between affinity and intensity ratio. Sequence alignment shows that Lys10 and Asp168 are the equivalent

residues in nSH3 and cSH3 domains. These two residues are also close to the nSH3/cSH3 binding sites. Noticeably, free Grb2 binding to the one-site SOS1 peptides increases signal intensity of Lys10 (Table 4-4). Even though the small CSPs of Lys10 induced by Grb2-S5 leads to the unavailable affinity, the signal intensity is still increased by 60%. Signal intensities of Lys10 are strengthened by the interaction of EpYINQSV-Grb2 with S1\*, S2\*, and S3; however, the intensities are decreased when EpYINQSV-Grb2 binds to S4\*. By contrast, Asp168 shows decreased intensity upon Grb2 binding to S1\*, S2\*, S4 and S5. The presence of EpYINQSV does not show significant changes on the signal intensity of Asp168. For Grb2-S3, the weak binding keeps the signal intensity ratio of Asp168 equal to 1. The titration curves of Lys10 and Asp168 were fitted to our binding model (Fig. 4-3F). Most  $K_D$  values are inversely proportional to the signal intensity ratios (Table 4-4).



**Figure 4-2.** NMR characterization of the SOS1 peptides interacting with the complex EpYINQSV–Grb2. Amide chemical shift perturbation (CSP,  $\Delta\delta$ ) at the titration end

point are shown as function of Grb2 residue numbers for (A) EpYINQSV–Grb2–S1\*, (B) EpYINQSV–Grb2–S2\*, (C) EpYINQSV–Grb2–S3\*, (D) EpYINQSV–Grb2–S4\*, and (E) EpYINQSV–Grb2–S5. CSPs of indole NH signals of Trp36, Trp60, Trp121, Trp193, and Trp194 are denoted as Trp36 $\epsilon$ , Trp60 $\epsilon$ , Trp121 $\epsilon$ , Trp193 $\epsilon$ , and Trp194 $\epsilon$  and located at the residue number 218, 219, 220, and 221. Signals with attenuation during titration are shown as grey bars. Residues with  $\Delta\delta$  greater than two standard deviation are indicated. CSPs are mapped onto the surfaces of Grb2 in complexes with the corresponding modeled peptides (yellow). Surface coloring based on the observed CSP is colored from white (0 ppm) to red (maximal  $\Delta\delta = 0.1$  ppm). Residues with signal attenuation are colored orange. Structure images to the right show the enlarged binding surfaces of nSH3 and cSH3 in detail.



**Figure 4-3.** The influence of EpYINQSV on the interaction of Grb2 with one-site SOS1 peptides. CSP difference ( $\Delta\Delta\delta$ ) and signal intensity ratio ( $I/I_0$ ) between the interactions of Grb2 and EpYINQSV–Grb2 with (A) S1\*, (B) S2\*, (C) S3\*, (D) S4\*, and (E) S5. Negative  $\Delta\Delta\delta$  values suggest the residues have greater impact by EpYINQSV. Residues with  $|\Delta\Delta\delta| > 0.05$  ppm are indicated. The individual signal intensity ratio were calculated by the signal intensity in the titration endpoint over the corresponding signal intensity in unbound Grb2. The titration endpoint of ligand-to-protein ratio is 2:1. The titrated Grb2 and EpYINQSV–Grb2 are colored blue and red respectively. Residues with signal intensity ratios  $I/I_0 > 1.5$  are pointed out. (F) The fitted titration curves were shown for the selected K10 and D168 in the association of Grb2 and EpYINQSV–Grb2 with the one-site SOS1 peptides. The ligand-to-protein

ratio ranges from 0 to 2 with interval 0.2. For the binding of S1\*, signal shifts with the same contour level follow the arrow direction showing the ligand-to-protein ratio = 0 (blue), 0.2 (red), 0.4 (green), 1.0 (yellow), and 2.0 (blue).

**Table 4-3.** Averaged signal intensity ratios of Grb2 nSH3, SH2, and cSH3 domains

Average signal intensity	nSH3	SH2	cSH3
$\langle I_i^{Grb2-S1^*} / I_i^{Grb2} \rangle$	1.0	0.8	0.8
$\langle I_i^{Grb2-S2^*} / I_i^{Grb2} \rangle$	1.0	0.9	0.9
$\langle I_i^{Grb2-S3^*} / I_i^{Grb2} \rangle$	0.9	0.9	1.0
$\langle I_i^{Grb2-S4^*} / I_i^{Grb2} \rangle$	1.2	1.0	0.9
$\langle I_i^{Grb2-S5} / I_i^{Grb2} \rangle$	0.9	0.8	0.8
$\langle I_i^{Grb2-S12} / I_i^{Grb2} \rangle$	0.7	0.8	0.7
$\langle I_i^{Grb2-S23} / I_i^{Grb2} \rangle$	0.5	0.6	0.6
$\langle I_i^{Grb2-S45} / I_i^{Grb2} \rangle$	0.5	0.6	0.4
$\langle I_i^{EpYINQSV-Grb2} / I_i^{Grb2} \rangle$	0.9	0.9	1.0
$\langle I_i^{EpYINQSV-Grb2-S1^*} / I_i^{Grb2} \rangle$	1.2	0.8	1.0
$\langle I_i^{EpYINQSV-Grb2-S2^*} / I_i^{Grb2} \rangle$	1.2	0.9	1.0
$\langle I_i^{EpYINQSV-Grb2-S3^*} / I_i^{Grb2} \rangle$	0.9	0.8	1.0
$\langle I_i^{EpYINQSV-Grb2-S4^*} / I_i^{Grb2} \rangle$	0.9	0.9	0.9
$\langle I_i^{EpYINQSV-Grb2-S5} / I_i^{Grb2} \rangle$	0.9	0.8	0.8
$\langle I_i^{EpYINQSV-Grb2-S12} / I_i^{Grb2} \rangle$	0.7	0.7	0.6
$\langle I_i^{EpYINQSV-Grb2-S23} / I_i^{Grb2} \rangle$	0.6	0.7	0.6
$\langle I_i^{EpYINQSV-Grb2-S45} / I_i^{Grb2} \rangle$	0.4	0.5	0.4

**Table 4-4.** Dissociation constants and signal intensity ratios of K10 and D168 in the bindings of Grb2 and EpYINQSV–Grb2 with the one-site SOS1 peptides

SOS1 peptide	Grb2				EpYINQSV–Grb2			
	K10		D168		K10		D168	
	$K_D$	$I/I_0$	$K_D$	$I/I_0$	$K_D$	$I/I_0$	$K_D$	$I/I_0$
S1*	1.2	2.1	9.5	0.5	2.1	6.0	8.7	0.7
S2*	0.9	3.2	1.2	0.8	0.1	4.3	12.3	0.8
S3*	3.3	2.4	-	1.0	40.1	3.1	-	1.0
S4*	11.1	4.9	1.1	0.8	1.8	2.6	31.2	0.7
S5	-	1.6	29.7	0.7	-	1.6	30.2	0.7

#### 4.2.4 Grb2 interacting with S12, S23, and S45 exhibits different complex conformations

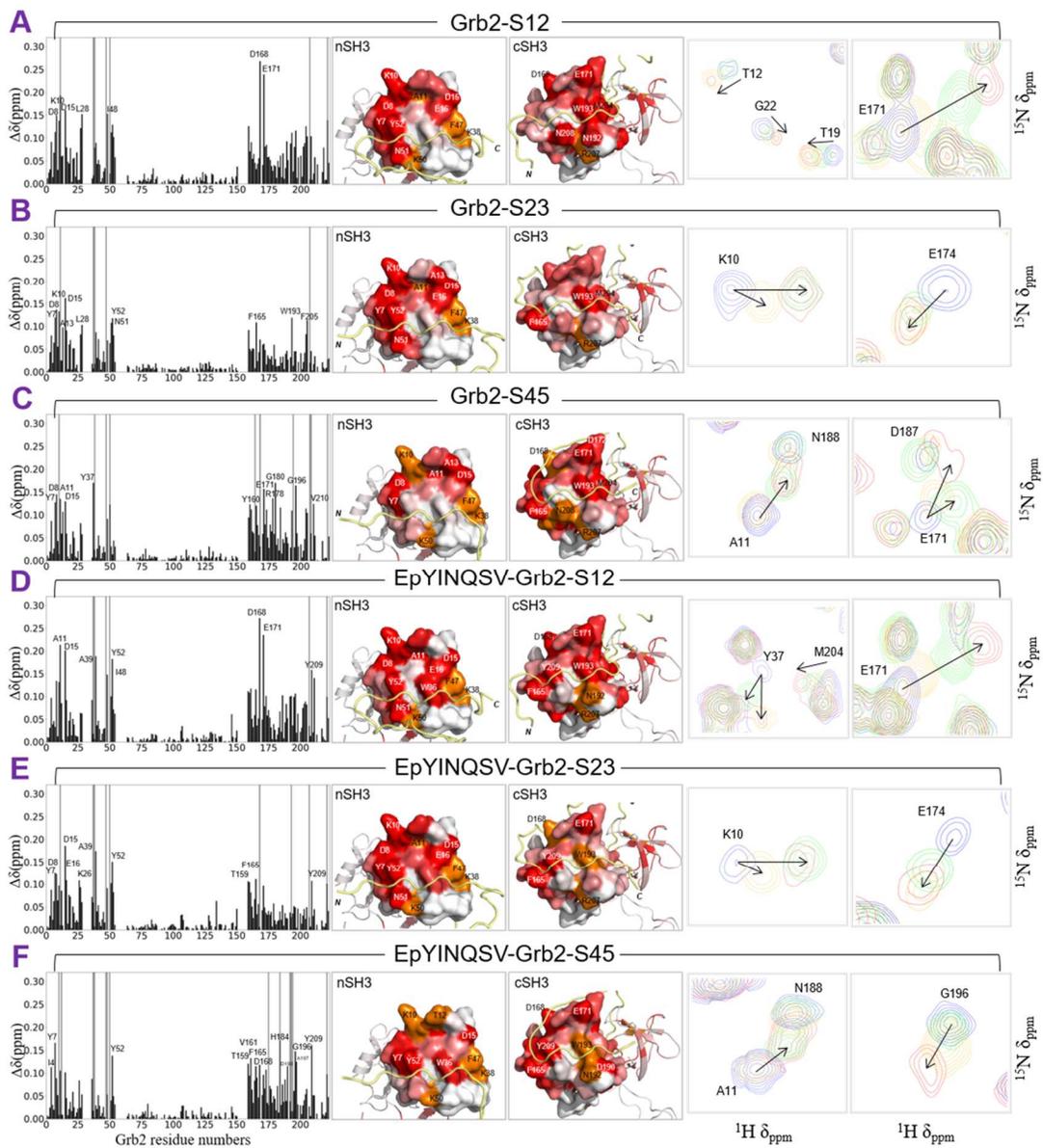
Grb2 associates with the wild type SOS1 PR segment S12 (EVPVPPPVP~~RRRRP~~-ESAPAESSPSKIMSKHLDSPPAIPPRQPTSKAYS) through the nSH3-S2 and cSH3-S1 interactions. While both S1 and S2 bind the nSH3 binding PxxPxR motif, the additional arginine residues in S1 enhance the binding affinity for cSH3 [57]. Even though the isolated nSH3 binding to S1 is about 3-fold stronger than that of isolated cSH3, the pairwise affinities suggest that one of the most probable binding modes of Grb2-SOS1 is through the nSH3-S2 and cSH3-S1 interactions [49]. To investigate how Grb2 uses the two binding domains (nSH3 and cSH3) to interact with the two binding sites (S1 and S2) of SOS1, the S12 peptide was added to the full length Grb2 with 2:1 stoichiometry. The nSH3 residues Asp8, Lys10, Asp15, Leu28, Ile48, and the cSH3 residues Asp168 and Glu171 show large CSPs (Fig. 4-4A). Ala11, Trp36, Tyr37, Lys38, Phe47, Lys50, and Arg207 show signal attenuation. By comparing the <sup>1</sup>H-<sup>15</sup>N NMR spectra of Grb2-S12 with Grb2-S1 and Grb2-S2, the selected nSH3 residues Thr12, Thr19, and Gly22 show the signals overlapping with the signals of Grb2-S2, while the signal of cSH3 Glu171 shifts closely to the signal of Grb2-S1. The similar signal shifts suggest the amides might experience similar changes of electronic environments. Therefore, the interaction of Grb2-S12 is more likely through nSH3-S2 and cSH3-S1 (Fig. 4-5A).

S23 (DSPPAIPPRQPTSKAYSSPRYSISDRTSISDPPESPPLLPPREPVRTPD) contains two nSH3 binding sites. Unlike the S12 the nSH3/cSH3 preferred binding site S1, both S2 and S3 strongly bind to the isolated nSH3 but weakly bind to the isolated

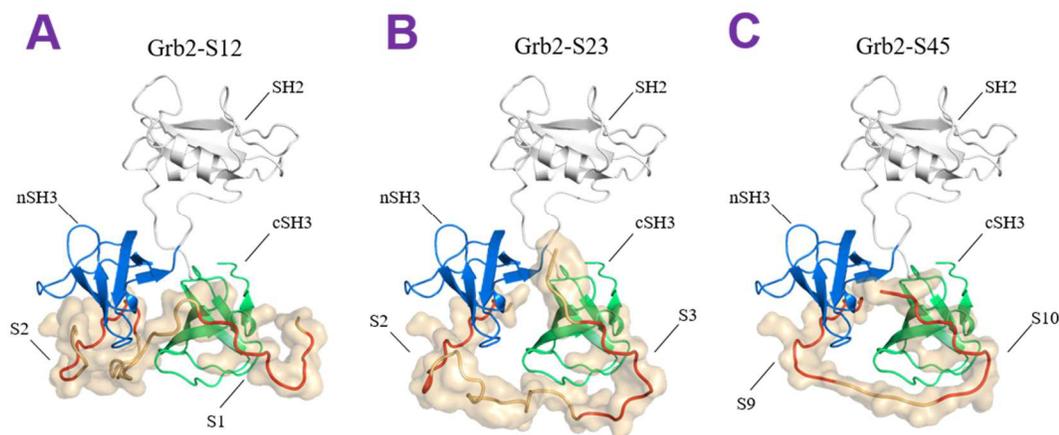
cSH3 ( $K_D > 1000 \mu\text{M}$ ). To verify whether Grb2 binding to S23 is via nSH3–S2/cSH3–S3 or nSH3–S3/cSH3–S2, the S23 peptide was added to Grb2 with the 2:1 molar ratio. Grb2 residues Tyr7, Asp8, Lys10, Ala13, Asp15, Leu28, Asn51, Tyr52, Phe165, Trp193, and Phe205 show large CSPs (Fig. 4-4B). Residues with signal attenuation are Ala11, Trp36 $\epsilon$ , Tyr37, Lys38, Phe47, and Arg207. Similarly, by comparing the  $^1\text{H}$ - $^{15}\text{N}$  NMR spectra of Grb2–S23 with Grb2–S2 and Grb2–S3, the selected nSH3 residues Lys10 and cSH3 residue Glu174 have the signals superimposing with the corresponding signals of nSH3–S2 and cSH3–S3, respectively. This suggests Grb2 associating with S23 is through nSH3–S2 and cSH3–S3 (Fig. 4-5B). Interestingly, upon Grb2 binding to S23, nSH3, SH2, and cSH3 domains show reduced signal intensities, which are about half of the intensities of free Grb2 (Table 4-3). Noticeably, even though SH2 domain is not involved in nSH3/cSH3 bindings, Grb2–S23 may have indirect interactions to the SH2 domain.

S45 (SIAGPPVPPRQSTSQH $\epsilon$ PKLPPKTYKREH) contains an nSH3 binding site S4 and a cSH3 binding site S5 with a short linker, which has merely 4 amino acids. S45 binding to Grb2 induces the large CSPs on the residues Tyr7, Asp8, Ala11, Asp15, Tyr37, Tyr160, Glu171, Arg178, Gly180, Gly196, and Val210 (Fig. 4-4C). Lys10, Trp36 $\epsilon$ , Lys38, Phe47, Lys50, Leu164, Asp168, Trp194, Arg207, and Asn208 show severe signal attenuation. By superimposing the  $^1\text{H}$ - $^{15}\text{N}$  NMR spectra of Grb2–S45, Grb2–S4, and Grb2–S5, the selected signal of nSH3 Ala11 overlaps the signal of Grb2–S4, and cSH3 Glu171 shifts to the same direction of the signal of Grb2–S5. Noticeably, S4 and S5 have the motifs PxxPxR and PxxPxKxxK, respectively. Without the same consensus motif, binding to S4 and S5 results in that the amide signals of Grb2 nSH3

and cSH3 shift to the different directions, and therefore, it is easier to identify the complex interactions of Grb2 with S45. As expected, Grb2 binding to S45 is via nSH3–S4 and cSH3–S5 (Fig. 3-5C). Similar to Grb2–S23, Grb2–S45 also largely reduces the signal intensities of nSH3, SH2 and, cSH3 domains (Table 4-3). The S5-bound cSH3 domain only have 40% signal intensity of the unbound state. This suggests the binding of Grb2–S45 may be stronger than Grb2–S12 and Grb2–S23.



**Figure 4-4.** Structural characterization of full length Grb2 with SOS1 two-site segments. CSPs ( $\Delta\delta$ ) at SOS1/Grb2 molar ratio equal to 2 are shown as function of Grb2 residue numbers for (A) Grb2–S12, (B) Grb2–S23, (C) Grb2–S45, (D) EpYINQSV–Grb2–S12, (E) EpYINQSV–Grb2–S23, and (F) EpYINQSV–Grb2–S45. Residues with  $\Delta\delta$  greater than two standard deviation are indicated, and the signal attenuation are shown by grey bars. The nSH3 and cSH3 structures are colored from white (0 ppm) to red (maximal  $\Delta\delta = 0.1$  ppm) based on the corresponding CSP values. Residues with signal attenuation are colored orange, and the modeled SOS1 segments are colored yellow. The selected signals in Grb2 (blue) are superimposed with (A) Grb2–S1\* (green), Grb2–S2\* (yellow), and Grb2–S12 (red); (B) Grb2–S2\* (green), Grb2–S3\* (yellow), and Grb2–S23 (red); (C) Grb2–S4\* (green), Grb2–S5 (yellow), and Grb2–S45 (red); (D) EpYINQSV–Grb2–S1\* (green), EpYINQSV–Grb2–S2\* (yellow), and EpYINQSV–Grb2–S12 (red); (E) EpYINQSV–Grb2–S2\* (green), EpYINQSV–Grb2–S3\* (yellow), and EpYINQSV–Grb2–S23 (red); (F) EpYINQSV–Grb2–S4\* (green), EpYINQSV–Grb2–S5 (yellow), and EpYINQSV–Grb2–S45 (red). The signal shifts suggest Grb2 associating with S12, S23, and S45 are via nSH3–S2/cSH3–S1, nSH3–S2/cSH3–S3, and nSH3–S4/cSH3–S5, respectively.



**Figure 4-5.** Computational modelling of Grb2 with the two-site SOS1 peptides. (A) Grb2 associates with S12 through nSH3–S2/cSH3–S1. (B) Grb2 binds to S23 via nSH3–S2/cSH3–S3, and (C) Grb2–S45 is through nSH3–S4/cSH3–S5. The nSH3, SH2, and cSH3 domains are colored white, blue, and green respectively. SOS1 peptides are shown yellow, and the corresponding nSH3/cSH3 binding sites are colored red.

#### 4.2.5 EGFR segment may strengthen the association of Grb2-S45

Association of EpYINQSV with Grb2 does not change the pairwise interactions of Grb2–S12, Grb2–S23, and Grb2–S45. For EpYINQSV–Grb2–S12, Grb2 residues Ala11, Asp15, Ala39, Ile48, Tyr52, Asp168, Glu171, and Tyr209 show large CSPs, and Trp36 $\epsilon$ , and Tyr37, Lys38, Phe47, Lys50, and Arg207 undergo severe signal attenuation (Fig. 4-4D). By comparing the  $^1\text{H}$ - $^{15}\text{N}$  NMR spectra of EpYINQSV–Grb2–S12, Grb2–S1\*, and Grb2–S2\*, signals of nSH3 residues Tyr37 and Met204 overlap the signals in Grb2–S2\*, and cSH3 residue Glu174 have signal shift close to the signal in Grb2–S1\*. EpYINQSV induces large changes of CSPs to the interaction of Grb2 with two-site SOS1 peptides. By subtracting the CSPs of Grb2–S12 from the CSPs of EpYINQSV–Grb2–S12, residues Tyr7, Leu28, Trp36, Trp36 $\epsilon$ , Ala39, Phe165,

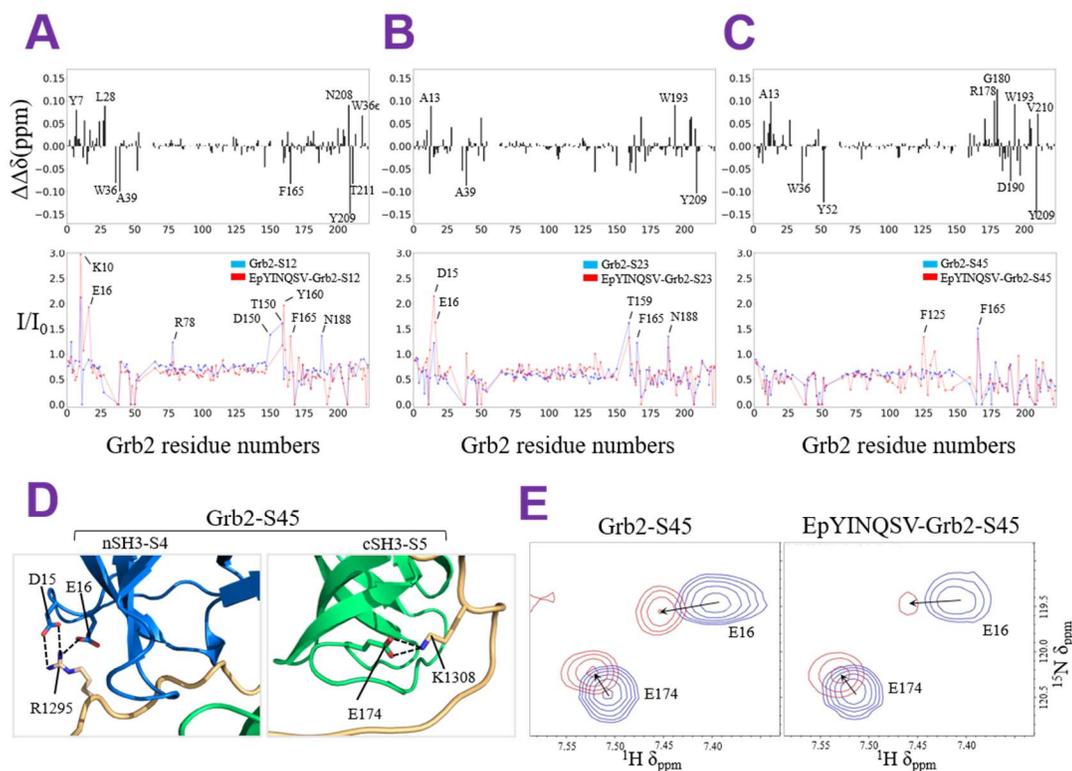
Asn208, Tyr209, Thr211 show  $|\Delta\Delta\delta| > 0.07$  ppm (Fig. 4-6A). The signal intensity ratio of EpYINQSV–Grb2–S12 is approximate to Grb2–S12 and smaller than Grb2–S1 and Grb2–S2 (Table 4-3). Lys10, Glu16, Tyr160, and Phe165 show signal intensity ratios  $I/I_0 > 1.2$  (Fig. 4-6A). This suggests that simultaneous bindings of nSH3/cSH3 to S12 are stronger than the one-site peptide S1 and S2. However, EpYINQSV may have little effect on the affinities of nSH3/cSH3. Overlaid signals will be ignored.

EpYINQSV–Grb2 associates with S23 via nSH3–S2 and cSH3–S3. Grb2 residues with large CSPs are Tyr7, Asp8, Asp15, Glu16, Lys26, Ala39, Thr159, Phe165, and Tyr209 (Fig. 4-4E). Signal attenuation results in that residues Ala11, Trp36 $\epsilon$ , Tyr37, Lys38, Phe47, Lys50, Asp168, Trp193, and Arg207 are incapable to be tracked. The nSH3 residue Lys10 and cSH3 residue Glu174 superimpose the corresponding signals in Grb2–S2 and Grb2–S3 respectively, suggesting the interaction of nSH3–S2 and cSH3–S3. By comparing the spectra of Grb2–S23 and EpYINQSV–Grb2–S23, residues with  $|\Delta\Delta\delta| > 0.07$  ppm are Ala13, Ala39, Trp193 and Tyr209 (Fig. 4-6B). Only Asp15, Glu16, and Thr159 exhibited signal intensity ratios  $I/I_0 > 1.2$ . Since signal intensity of Grb2 is largely reduced upon binding to S23, the presence of EpYINQSV, however, slightly increase the intensity ratios of nSH3 and SH2 domains, which suggests the reduced affinity of Grb2–S23 (Table 4-3).

EpYINQSV–Grb2 binding to S45 shows the interaction of nSH3–S4 and cSH3–S5. The nSH3 residues Ile4, Tyr7, Tyr52, and cSH3 residues Thr159, Val161, Phe165, Asp168, Asp190, Gly196, Ala197, and Tyr209 perform large CSPs (Fig. 4-4F). Lys10, Thr12, Trp36 $\epsilon$ , Tyr37, Lys38, Phe47, Ile48, Lys50, Leu175, His184, Asn192, Trp193, Trp194 exhibit severe signal attenuation. Signal of Ala11 overlaps the signal in Grb2–

S4, and G196 is closer to the signal in Grb2–S5, showing the interaction of EpYINQSV–Grb2 with S45 follows nSH3–S4 and cSH3–S5 as anticipated. Ala13, Trp36, Tyr52, Arg178, Gly180, Asp190, Trp193, Tyr209, and Val210 have  $|\Delta\Delta\delta| > 0.07$  ppm between the spectra of Grb2–S45 and EpYINQSV–Grb2 (Fig. 4-6C). Phe125 and Phe165 show larger signal intensity ratios. Surprisingly, binding of EpYINQSV–Grb2 to S45 shows slightly small signal intensity ratio of nSH3 and SH2 domains than Grb2–S45 (Table 4-3), suggesting EpYINQSV may enhance the affinity of Grb2–S45.

EpYINQSV may facilitate the association of Grb2–S45 due to the reduced signal intensities of Grb2. Grb2–SOS1 interaction particularly relies on the formation of salt bridges for nSH3/cSH3 bindings [33, 72]. For example, in the binding of Grb2–S45, nSH3 Glu16 uses Asp15 and Glu16 to form salt bridges with SOS1 Arg1295, and cSH3 Glu174 interacts with SOS1 Lys1308 (Fig. 4-6D). Interestingly, the signal intensity ratios of Glu16 and Glu174 in EpYINQSV–Grb2–S12 and EpYINQSV–Grb2–S23 are larger than Grb2–S12 and Grb2–S23 (Table 4-5). However, EpYINQSV–Grb2 binding to S45 shows slightly decreased signal intensity ratios of Glu16 and Glu174 in Grb2–S45 (Fig. 4-6E). The differences of intensity ratio suggest that EpYINQSV may weaken the interactions of Grb2–S12 and Grb2–S23 but strengthen Grb2–S45.



**Figure 4-6.** Differences in changes of CSP and signal intensity ratios between the interactions of Grb2 and EpYINQSV–Grb2 with S12, S23, and S45. The CSP differences ( $\Delta\Delta\delta$ ) were calculated by subtracting CSPs of (A) Grb2–S12 to EpYINQSV–Grb2–S12, (B) Grb2–S23 to EpYINQSV–Grb2–S23, and (C) Grb2–S45 to EpYINQSV–Grb2–S45. Negative  $\Delta\Delta\delta$  values suggest the residues have greater impact by EpYINQSV. Residues with  $|\Delta\Delta\delta| > 0.07$  are indicated. NMR signal intensity ratios ( $I/I_0$ ) were obtained by calculating the residues' intensity of titration endpoint ( $I$ ) over the intensity of unbound Grb2 ( $I_0$ ). The titration endpoint of ligand-to-protein ratio is 2:1. The titrated Grb2 and EpYINQSV–Grb2 complex are colored blue and red respectively. Signal intensity ratios with  $I/I_0 > 1.2$  were indicated. (G) For Grb2–S45, the critical nSH3 residue D15 and E16 form salt bridges with S4 R1295, and cSH3 E174 interacts with S5 K1308. (F) Signal intensities of E16 and E174 are reduced in

the titration endpoints of to Grb2–S45 and EpYINQSV–Grb2–S45. Signals are shown the same contour level and colored red for S45-bound Grb2/ EpYINQSV–Grb2 and blue for the unbound Grb2.

**Table 4-5.** Maximal CSPs and signal intensity ratios of E16 and E17 in the bindings of Grb2 and EpYINQSV–Grb2 with the two-site SOS1 peptides

SOS1 peptide	Grb2				EpYINQSV–Grb2			
	E16		E174		E16		E174	
	$\Delta\delta$	I/I <sub>0</sub>						
S12	0.08	0.89	0.06	0.78	0.09	1.93	0.06	0.79
S23	0.09	0.58	0.04	0.67	0.11	1.63	0.03	0.74
S45	0.06	0.67	0.05	0.60	0.05	0.51	0.04	0.52

#### 4.3 Discussion

Grb2–SOS1 interaction is challenging due to the multiple nSH3/cSH3 binding sites and the numerous serine phosphorylation on the SOS1 PR domain. Even though the nSH3 binding sites S1, S2, S3, S4 and the cSH3-favored binding sites S1 and S5 were identified, the association of Grb2 nSH3/cSH3 with these five binding sites result in multiple possible binding modes of Grb2–SOS1 such as nSH3–S1/cSH3–S5, nSH3–S2/cSH3–S1, nSH3–S2/cSH3–S5, nSH3–S4/cSH3–S1, and etc. [49]. However, the Grb2–SOS1 interactions can be interfered by serine phosphorylation on SOS1 PR domain. Noticeably, Ser1161 closely follows S1, and Ser1178 and Ser1210 locate on S2 and S3. The phosphorylated Ser1161, Ser1178, and Ser1210 may result in the unfavorable interactions of Grb2 nSH3/cSH3 with S1, S2 and S3. The released nSH3/cSH3 may result in nSH3 searching for alternative binding sites on SOS1 and

increase the probability of cSH3 associating with Gab1/2. How Grb2 nSH3 and cSH3 domains recruit two individual PR motifs become of particular importance.

The five potential nSH3/cSH3 binding peptides S1\*, S2\*, S3\*, S4\*, and S5 were selected to examine the interactions with Grb2. Surprisingly, as a full length Grb2, nSH3 and cSH3 have much higher affinities for these truncated SOS1 peptides than the isolated nSH3/cSH3. Besides, S1\*, S2\*, S3\*, S4\*, and S5 are either nSH3- or cSH3-favored binding motifs. However, none of these five peptides shows the expected primary and secondary bindings for the nSH3/cSH3 of Grb2. Instead, both nSH3 and cSH3 domains of Grb2 always show strong bindings for S1\*, S2\*, S3\*, S4\*, and S5 with the same magnitude of  $K_D$  values. However, it remains puzzling that Grb2 nSH3/cSH3 binding to the one-site SOS1 peptides shows similar high affinities.

The phosphorylated EGFR segment, EpYINQSV, has limited effects on Grb2 binding to the one-site SOS1 peptides. The complex EpYINQSV–Grb2 association with S1\*, S2\*, S3\*, S4\*, and S5 shows affinities that are similar to Grb2 binding to these SOS1 one-site peptides. The differences in affinity are within the calculated error; the differences of CSPs and the deviation of signal intensity ratios are mostly small. Noticeably, some particular residues on the binding surface such as Lys10 and Asp15 perform higher intensity ratio in the presence of EpYINQSV, and the change of intensity ratios are inversely proportional to the  $K_D$  values.

Grb2 binding to the two-site SOS1 peptides S12, S23, and S45 indicates distinct complex conformations and induces larger CSPs than binding to the one-site peptides. Grb2 binds to S12 through the association of nSH3–S2 and cSH3–S1 even though the isolated nSH3–S1 binding is stronger than the isolated cSH3–S1. The actual binding

modes of Grb2–SOS1 complex are based on the pairwise affinities of isolated nSH3/cSH3 with SOS1 peptides. Grb2 binds to S23 and S45 through nSH3–S2/cSH3–S3 and nSH3–S4/cSH3–S5, respectively. Noticeably, for the Grb2–S23 complex, the CSPs of cSH3–S3 are much larger than the CSPs of cSH3–S3\* of the Grb2–S3\* complex. This suggests that nSH3 binding to S2 facilitates the recruitment of cSH3 for S3 even though cSH3 does not favor the nSH3 binding site. The simultaneous nSH3/cSH3 bindings lead to the strong association and show the highly decreased signal intensities of nSH3 and cSH3 domains. Interestingly, signal intensity of SH2 domain is also decreased while SH2 is not involved in nSH3/cSH3 bindings. The two-site SOS1 peptides binding to nSH3/cSH3 may provide an indirect link between nSH3/cSH3 and SH2 domain.

EpYINQSV–Grb2 binding to S12, S23, and S45 follows the same complex conformation of Grb2 interacting with these two-site SOS1 peptides. However, the affinity of Grb2–S12/S23/S45 may be affected by EpYINQSV. Since Glu16 and Glu174 dominate the nSH3 and cSH3 bindings respectively, binding of EpYINQSV results in the increased signal intensity ratios of these two critical residues in Grb2–S12 and Grb2–S23. By contrast, for EpYINQSV–Grb2–S45, Glu16 and Glu174 show slightly decreased signal intensity ratios. This suggests that EpYINQSV binding to SH2 domain is not independent of the simultaneous bindings of nSH3/cSH3 and may facilitate the specific two-site peptide such as S45.

Taken together, we investigated the interaction of full length Grb2 with the selected SOS1 one-site peptides S1\*, S2\*, S3\*, S4\*, S5, and the two-site peptides S12, S23, and S45. Even though previous data suggested that nSH3 favors S1\*, S2\*, S3\*, S4\*,

and cSH3 prefers S5\*, the full length Grb2 associating with these one-site peptides does not perform the anticipated primary and secondary bindings for the nSH3 and cSH3. Instead, the affinities of nSH3/cSH3 binding to each of the one-site peptides have the same magnitude of  $K_D$  values, and both nSH3 and cSH3 domains of Grb2 perform much stronger binding for the one-site peptides than the isolated nSH3/cSH3. Moreover, Grb2 binding to the two-site peptides S12, S23, S45 follows pairwise association of nSH3–S2/cSH3–S1, nSH3–S2/cSH3–S3, and nSH3–S4/cSH3–S5 respectively. Even though S23 lacks any cSH3 binding sites, cSH3–S3 interaction could be largely strengthened by the nSH3–S2 association. Lastly, the phosphorylated EGFR segment, EpYINQSV has little influences on the affinities and CSPs of Grb2 binding to the SOS1 one-site peptide but may strengthen association of Grb2 with the particular two-site peptide S45.

To conclude, strong association of Grb2–SOS1 in the early MAPK signaling will be subsequently inhibited by the ERK1/2 and RSK2 induced serine phosphorylation on SOS1 PR domain. Grb2 released from SOS1 capably recruits Gab1/2, forming SOS1–Grb2–Gab1/2. Our findings suggest the high cooperativity of nSH3/cSH3 bindings may strengthen the complex of SOS1–Grb2–Gab1/2. Targeting the individual nSH3/cSH3 could be more difficult than that using a two-site peptide inhibitor simultaneously blocks nSH3 and cSH3. In the subsequent MAPK signaling, while S1, S2, and S3 are probably blocked, S4 and S5 remain unaltered and vacant. Grb2–S45 becomes the most probable binding mode of Grb2–SOS1 complex. Our results suggest the peptide inhibitor that prevents Grb2–S45 interaction could be a potential clinical therapy in the subsequent MAPK signaling phase.

## 4.4 Material and Methods

### 4.4.1 Protein expression and purification

The plasmid of wild type human Grb2, the synthesized EGFR peptide EpYINQSV, and the synthesized SOS1 peptides S1\*, S2\*, S3\*, S4\*, S5, S12, S23, and S45 were purchased from Biomatik Cooperation. The Grb2 plasmid is encoded by the His6-tag and transferred into the BL21 competent *E. coli* cells. The cells were grown in M9 media and induced with 1 mM IPTG at 16 °C overnight. The cells were then collected and disrupted by sonication in 20 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>, 50 mM NaCl, pH 7.4, NaN<sub>3</sub> 0.02%, and 10 mM imidazole. The cell debris was removed by the ultracentrifuge, and the supernatant was collected and purified by Ni-NTA chromatography, followed by the elution of 20 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>, 50 mM NaCl, pH 7.4, NaN<sub>3</sub> 0.02%, and 200 mM imidazole. The preliminary purified proteins were buffer exchanged to 20 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>, 50 mM NaCl, pH 7.4, NaN<sub>3</sub> 0.02%, and 5 mM DTT, followed by the size exclusion chromatography.

### 4.4.2 NMR spectroscopy

NMR experiments were performed at 298 K on Bruker Avance III 600 MHz spectrometer equipped with a cryoprobe. All samples were prepared in 20 mM NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub>, 50 mM NaCl, 10% D<sub>2</sub>O, and 0.02% NaN<sub>3</sub> at pH 7.4. The initial protein concentrations of <sup>15</sup>N-labeled Grb2 were 80, 66, 66, 67, 67 60, 60, 60 μM in association with S1\*, S2\*, S3\*, S4\*, S5, S12, S23, and S45, respectively. Samples for Grb2–S12 were added 3 mM TCEP, and others included 5 mM DTT. EpYINQSV–Grb2 complex was formed by adding the unlabeled EpYINQSV to the <sup>15</sup>N-labeled

Grb2 with 1:1 stoichiometry. The initial concentrations of EpYINQSV–Grb2 complex were 92, 75, 75, 75, 75, 60, 60, 60  $\mu$ M for binding to S1\*, S2\*, S3\*, S4\*, S5, S12, S23, and S45. Samples for EpYINQSV–Grb2–S1\*/S2\*/S3\*/S4\*/S5 and EpYINQSV–Grb2–S12/S23/S45 were added 3 mM TCEP and 5 mM DTT respectively. The concentrations of all SOS1 peptides were prepared 10 folds the corresponding initial concentration of Grb2 and EpYINQSV–Grb2 complex. The signals were assigned by the reported spectra of Grb2 (Figure C5) [76]. The titration was proceeded by adding the unlabeled SOS1 peptides to the  $^{15}$ N-labeled proteins and monitored by  $^1$ H- $^{15}$ N SOFAST-HMQC spectra at each titration point (Figure C3, C4, and C5). The titrations ended when the molar ratio of peptide and protein reaches to 2. NMR data were processed using TopSpin3.6 and analyzed with Sparky [77]. The chemical shift perturbations (CSPs) for each residues were calculated as follows:

$$\Delta\delta = \sqrt{(\Delta\delta_H)^2 + (\Delta\delta_N/5)^2}$$

, where  $\Delta\delta_H$  and  $\Delta\delta_N$  correspond to chemical shift difference for  $^1$ H and  $^{15}$ N resonances.

The dissociation constant  $K_D$  was calculated by fitting the CSP values of titration points. The titration data were analyzed using the model assuming that the two nSH3 and cSH3 domains of Grb2 have independent non-equilibrium affinities for binding to the SOS1 peptides S1\*, S2\*, S3\*, S4\*, and S5. These peptides were partitioned between nSH3 and cSH3 at each titration points according to the individual affinity. Our two-site binding model illustrate the direct binding on the first binding site, which reflects the indirect binding on the second binding site. For example, the program fits the CSP of nSH3 and yields the  $K_D$  values for both nSH3 and cSH3 domains. The

putative  $K_D$  for cSH3 are not subjected to any specific cSH3 residues but merely reflected the indirect binding by nSH3. Similarly, the fits of cSH3 also perform the putative  $K_D$  for nSH3. The model was discussed in our previous publication [78], and the analysis was performed using our in-house Matlab program  $K_D\_fit\_2s$  [78]. Residues with maximal  $\Delta\delta > 0.03$  ppm were selected to fit, and the fitted curves with  $\chi^2 > 0.05$  were ignored. The individual  $K_D$  values and the fitted titration curves were shown in the supplementary (Table C3 to C6 and Figure C6 to C15). The reported  $K_D$  were determined by averaging the individual  $K_D$  values, and the reported error was calculated as the standard deviation of the individual  $K_D$  values for the same residues.

Signal intensity ratios were calculated by using the individual residue intensity in the titration endpoint over the corresponding residue intensity in the free Grb2. If multiple signals overlapped in the spectra of titration endpoint, the superimposed signals would be ignored. Signal intensity ratio of each residue shown in the supplementary (Table C7 to C23).

#### 4.4.3 Transverse $^{15}\text{N}$ spin-relaxation measurement

$^{15}\text{N}$  Carr–Purcell–Meiboom–Gill (CPMG) experiment was performed to determine the transverse ( $T_2$ ) relaxation of Grb2 and EpYINQSV–Grb2 complex with two relaxation delays 8 and 56 ms. The standard pulse sequence was described previously [79]. The concentrations of  $^{15}\text{N}$ -labeled Grb2 and EpYINQSV–Grb2 were 96 and 121  $\mu\text{M}$ .  $T_2$  was calculated by the following equation:

$$T_2 = -\frac{\Delta\tau}{\ln \frac{I}{I_0}}$$

, where  $\Delta\tau$  is the difference of the two relaxation delays (48 ms), and  $I/I_0$  represents to the ratio of signal intensity between the two delays. The intensity data were obtained by Sparky [77]. The  $T_2$  values and the intensity ratios of each residue are shown in Figure C1, Table C2, and Table C15.

#### 4.4.4 Computational modeling of EGFR–Grb2–SOS1 complex

To construct the EGFR–Grb2–SOS1 complex, Grb2 was extracted from the dimeric Grb2 structure (PDB: 1GRI) [21]. The EGFR segment EpYINQSV was modeled by the using the peptide KPFpYVNVEF, which also contains the SH2 binding pYxNx motif and forms the complex with SH2 domain (PDB: 1BMB) [23]. The KPFpYVNVEF was sequentially replaced its residues for EpYINQSV, and the additional N-terminal residues lysine and proline were removed. The structure of full length SOS1 PR domain is unavailable. The mere NMR structural data of SOS1 peptide is VPPPVPPIRRR, which associates with nSH3 (PDB: 1AZE) [36]. By contrast, the structure of cSH3–SOS1 complex is also unavailable. We used the backbone conformation of the Gab2 peptide APPPRPPKP, which forms the cSH3–Gab2 complex (PDB: 2W0Z) [30]. Peptides VPPPVPPIRRRP and APPPRPPKP were used to construct S12, S23, and S45. The side chain residues were mutated accordingly for these three two-site peptides. The linkers between the nSH3 and cSH3 binding sites were constructed by the CHARMM program [48] and modeled by Modeller [46].

## Chapter 5: Allosteric KRas4B can modulate SOS1 fast and slow Ras activation cycle

Chapter 5 was adapted from: TJ. Liao, H. Jang, D. Fushman, R. Nussinov: *Allosteric KRas4B can modulate SOS1 fast and slow activation cycles*. *Biophysical Journal*, 2018, 115 (4), 629-641.

### 5.1 Introduction to chapter 5

Small GTPase Ras proteins anchor to the plasma membrane through their C-terminal hypervariable regions (HVRs) [4]. Active GTP-bound Ras binds and activates downstream effectors, such as Raf kinase, phosphatidylinositide 3-kinase (PI3K), NORE1A (RASSF5) [39] and Ral guanine nucleotide dissociation stimulator (RalGDS). Among the three, HRas, NRas, and KRas isoforms, KRas is the most frequently mutated in *RAS*-driven cancers [80, 81]. Wild type KRas gene has two splicing protein isoform products, KRas4A and KRas4B; thus, mutations in KRas will be observed in both KRas4A and KRas4B. Mutant KRas4B is the most abundant in *RAS*-driven cancers. KRas isoforms have 98% sequence identity at the G-domain (residue 1-166) but differ in their C-terminal hypervariable region (HVR) (residue 167-189). The HVRs of both KRas4A and KRas4B are highly positively charged and avidly interact with the anionic membrane; however, that of KRas4B is more so than KRas4A's. KRas4A has two states: in state 1 the HVR is only farnesylated, but in state 2 it is also palmitoylated. The KRas4B HVR is only farnesylated. Anchored, prenylated HVR mediates assembly of active Ras molecules into nanoclusters [82, 83], promoting signaling through the mitogen-activated protein kinase (MAPK, Raf/MEK/ERK)

pathway, leading to cell proliferation [9, 10]. Son of sevenless 1 (SOS1) is a guanine nucleotide exchange factor (GEF) that specifically activates Ras proteins by exchange of GDP to GTP [6, 7]. SOS1 is a large multidomain protein with atomic mass of ~153 kDa [84, 85]. It consists of the membrane interacting N-terminal and Ras activating C-terminal regions. The N-terminal region contains histone-like fold (HF, residues 1–198), Dbl-homology (DH, residues 200–390), and pleckstrin-homology (PH, residues 444–548) domains. The C-terminal catalytic region is composed of Ras exchanger motif (REM, residues 567–741) and CDC25 (residues 780–1019) domains, followed by a C-terminal SH3 binding motif tail (residues 1020–1333). The N-terminal regulatory domains act in SOS1 recruitment to the plasma membrane [86], which is mediated by an EGFR phosphorylated tyrosine motif [87]. The C-terminal catalytic region contains two Ras binding sites, one locates at the REM and the other at the CDC25 domain, corresponding to the allosteric and catalytic Ras binding sites, respectively (Fig. 5-1).

The solved Ras–SOS1–Ras ternary complex [88, 89] revealed that the Switch I (residues 30–38) and II (residues 60–76) regions of the two Ras proteins are involved in both allosteric and catalytic sites SOS1 binding. Based on the ternary complex, the mechanism of Ras activation postulated that Ras binding to the REM allosteric site is crucial for activating GDP-bound Ras at the CDC25 catalytic site (hereafter, Ras interacting with SOS at the allosteric and catalytic sites refers to as allosteric and catalytic Ras, respectively). GTP-bound Ras at the allosteric site is known to facilitate  $\text{GDP} \rightarrow \text{GTP}$  exchange at the catalytic site [90-92]. The allosteric Ras-GTP elicits conformational changes of the tandem REM-CDC25 domains of SOS1, leading to

displacement of the helix-hairpin motif formed by  $\alpha F$ - $\beta A$ - $\beta B$ - $\alpha G$  at the CDC25 catalytic site [89, 93, 94] (domain structures shown in Fig. 5-1). The large movement of the helix-hairpin motif causes the  $\alpha F$  helix to sterically interfere in the Switch I region of Ras-GDP at the CDC25 catalytic site, exposing the nucleotide-binding site of Ras. The Switch I open conformation of Ras-GDP facilitates GDP's exit from the nucleotide-binding site, thus permitting loading cytosolic GTP. The crystal structures of the ternary complex depict nucleotide-free Ras at the CDC25 catalytic site with widely opened Switch I, implicating a snapshot conformation caught in the midst of an exchange event. Fully active SOS1 promotes hetero-nucleotide exchanges rather than homo-nucleotide exchanges [92, 95] because the cytosolic cellular concentration of GTP is ten times higher than GDP [96]. Upon exchange by GTP, Switch I shifts to the closed state, completing Ras activation. The weakened interaction with SOS of the activated Ras, culminates in Ras release. Thus, accelerated Ras activation requires Ras-GTP at the allosteric SOS1 site. This is a fast activation cycle via a positive feedback loop [91, 92].

SOS1 was crystallized in complex with HRas [88, 89]. Structural data relating to the highly oncogenic KRas are currently unavailable, and details of SOS1 conformational changes in KRas activation are missing. Here, we studied the wild type KRas4B interacting with SOS1. Since the HVR is long, disordered, and not involved in the Ras-SOS interactions, it is not considered in our work. Since the G-domains of the KRas isoforms are almost identical, our observations can be applicable to both KRas4A and KRas4B. Using all-atom molecular dynamics (MD) simulations, we examine the mechanism of KRas4B activation to clarify how the allosteric site in the REM domain

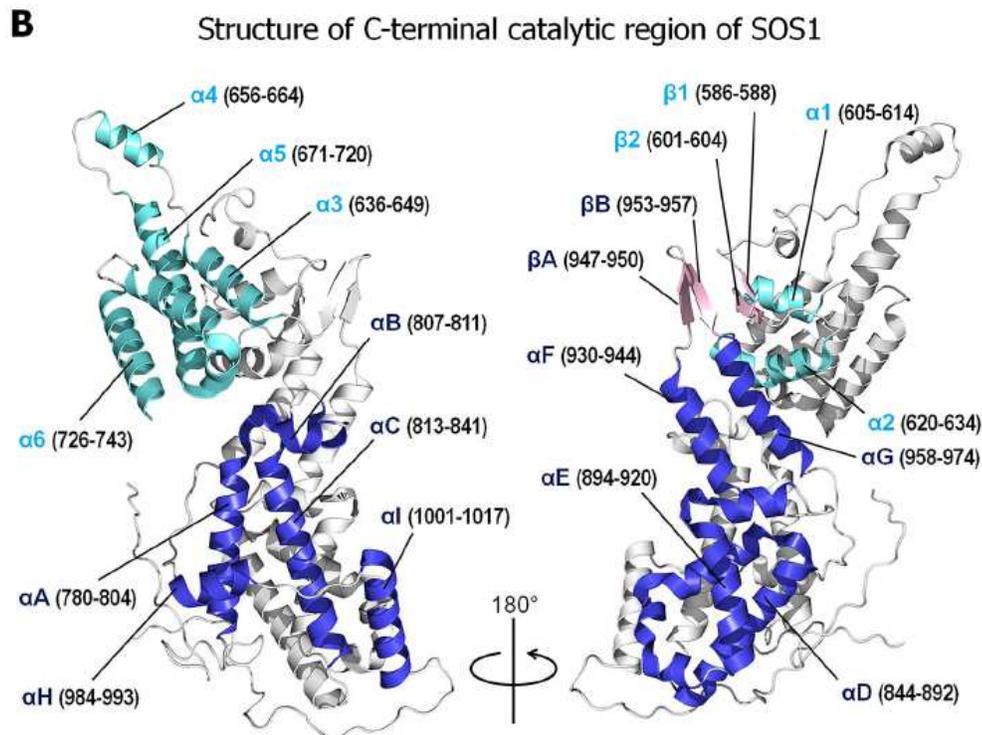
selectively accommodates GDP- and GTP-bound KRas4B. For comprehensive analysis, we constructed SOS1 systems with conformational ensembles representing the sequence of steps in Ras activation cycles. We investigated SOS1 systems including the SOS1–KRas4B dimers and the KRas4B–SOS1–KRas4B ternary complexes modeled with GDP- and GTP-bound, and nucleotide-free KRas4B. Using complementary techniques of conformational analysis, free energy calculations, and allosteric pathway elucidation, we show that allosteric KRas4B-GTP supports the fast cycle with positive feedback activation of SOS1, which mediates accelerated activation. By contrast, allosteric KRas4B-GDP impedes SOS1 in Ras activation, and regulates a limited, slow activation cycle. The GTP-bound Ras at the REM allosteric site promotes allosteric signals, which propagate through the REM-CDC25 tandem domains, shifting the SOS1 landscape, with the resulting conformational changes stimulating KRas4B activation. Clarifying how the KRas4B–SOS–KRas4B ternary complex influences SOS1 catalytic action may help development of small molecule drugs to inhibit allosteric activation [14, 88, 97-100].

**A** Sequence of C-terminal catalytic region of SOS1

```

..... 566
..... QMRLP SADVYRFAEP DSEENIIFEE NMQPKAGIPI 600
IKAGTVIKLI ERLTYHMYAD PNFVRTFLT YRSFCKPQEL LSLIIEERFEI 650
PEPEPTEADR IAIENGDOPL SAELKRFRKE YIQPVQLRVL NVCRHWVEHH 700
FYDFERDAYL LQRMEEFIGT VRGKAMKKWV ESITKIIQRK KIARDNGPGH 750
NITFQSSPPT VEWHRIRPGH IETFDLLTLH PIEIARQLTL LESDLYRAVQ 800
PSELVGSVWT KEDKEINSPN LLKMIRHTTN LTLWFEKIV ETENLEERVA 850
VVSRIIEILQ VFQELNNFNG VLEVVSAMNS SPVYRLDHTF EQIPSRQKKI 900
LEEAHELSED HYKKYLAKLR SINPPCVPF GIYLTNLIK EEGNPEVLKR 950
HGKELINFSK RRVVAEITGE IQOYONQPYC LRVESDIKRF FENLNPMGNS 1000
MEKEFTDYLF NKSLEIEPRN PKPLPFRPKK YSYPLKSPGV RPSNPR...
..................................................................... 1046

```



**Figure 5-1.** SOS1 sequence and structure. (A) The amino-acid sequence of the C-terminal catalytic region of SOS1. In the sequence, hydrophobic, polar/glycine, positively charged, and negatively charged residues are colored black, green, blue, and red, respectively. The cyan and blue underlines highlight the REM (residues 567–741) and CDC25 (residues 780–1019) domains of SOS1. (B) A crystal structure of the C-terminal catalytic region of SOS1 (PDB code: 4NYJ). The REM and CDC25 domains are colored cyan and blue, respectively.

## 5.2 Results

### 5.2.1 The C-terminal catalytic region of SOS1 exhibits conformational changes upon binding to KRas4B

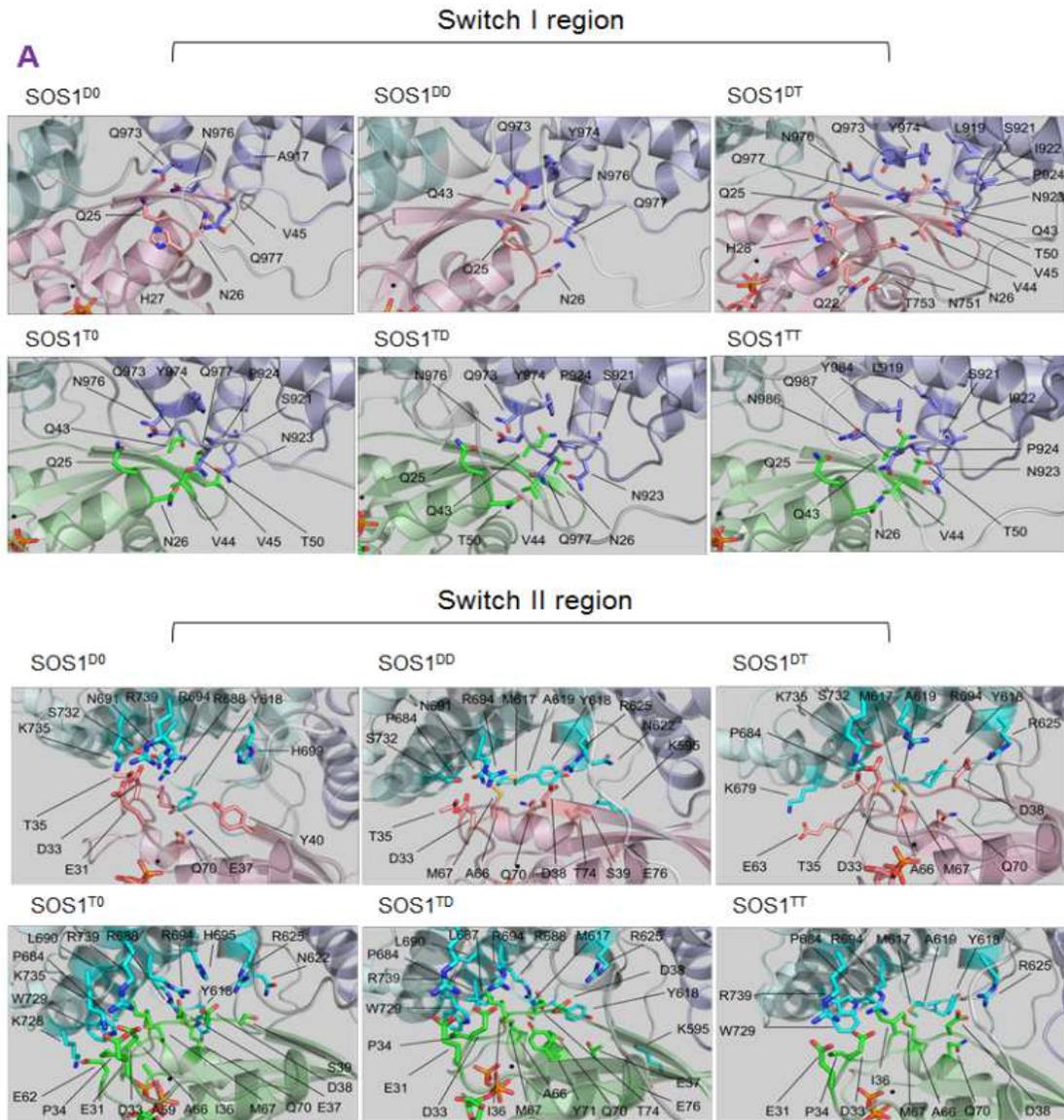
To decipher how SOS1 conformations with different Ras-binding modes delineate the activation cycle, we performed all-atom MD simulations on SOS1 N-terminal truncated, C-terminal catalytic region (hereafter referred to as SOS1) in complex with KRas4B in solution. The simulations were performed for SOS1 systems including a Ras-free apo-SOS1 monomer, four KRas4B–SOS1 dimers, and seven KRas4B–SOS1–KRas4B ternary complexes. To abbreviate the system notations throughout the text, we introduce two superscripted letters on SOS1. For instance, SOS1<sup>TD</sup> denotes GTP- and GDP-bound KRas4B interacting with SOS1 at the REM allosteric and CDC25 catalytic sites, respectively. Here, “T” and “D” denote GTP and GDP, with the former and latter superscripts corresponding to Ras interacting sites at REM and CDC25, respectively. Thus, for an apo-SOS1 monomer, we have SOS1<sup>00</sup>, where “0” denotes Ras-free. For the dimeric systems, we have SOS1<sup>D0</sup>, SOS1<sup>T0</sup>, SOS1<sup>0D</sup>, and SOS1<sup>0T</sup>. For the ternary systems, we have SOS1<sup>DD</sup>, SOS1<sup>DT</sup>, SOS1<sup>TD</sup>, and SOS1<sup>TT</sup>. To express ternary systems in the exchange event, we introduce SOS1<sup>TD\*</sup>, SOS1<sup>TR</sup>, and SOS1<sup>TT\*</sup>, where “R” denotes nucleotide-free Ras, and “\*” highlights Ras with Switch I open conformation as in the crystal structure. A total of twelve SOS1 systems were constructed (Fig. D1). During the simulations, we observed that SOS1 exhibits significant conformational adjustments, especially for the REM domain, depending on the KRas4B binding modes with respect to both allosteric and catalytic sites (Fig. D2). The inter-domain interface between REM and CDC25 can be divided into two regions: region 1 is formed by

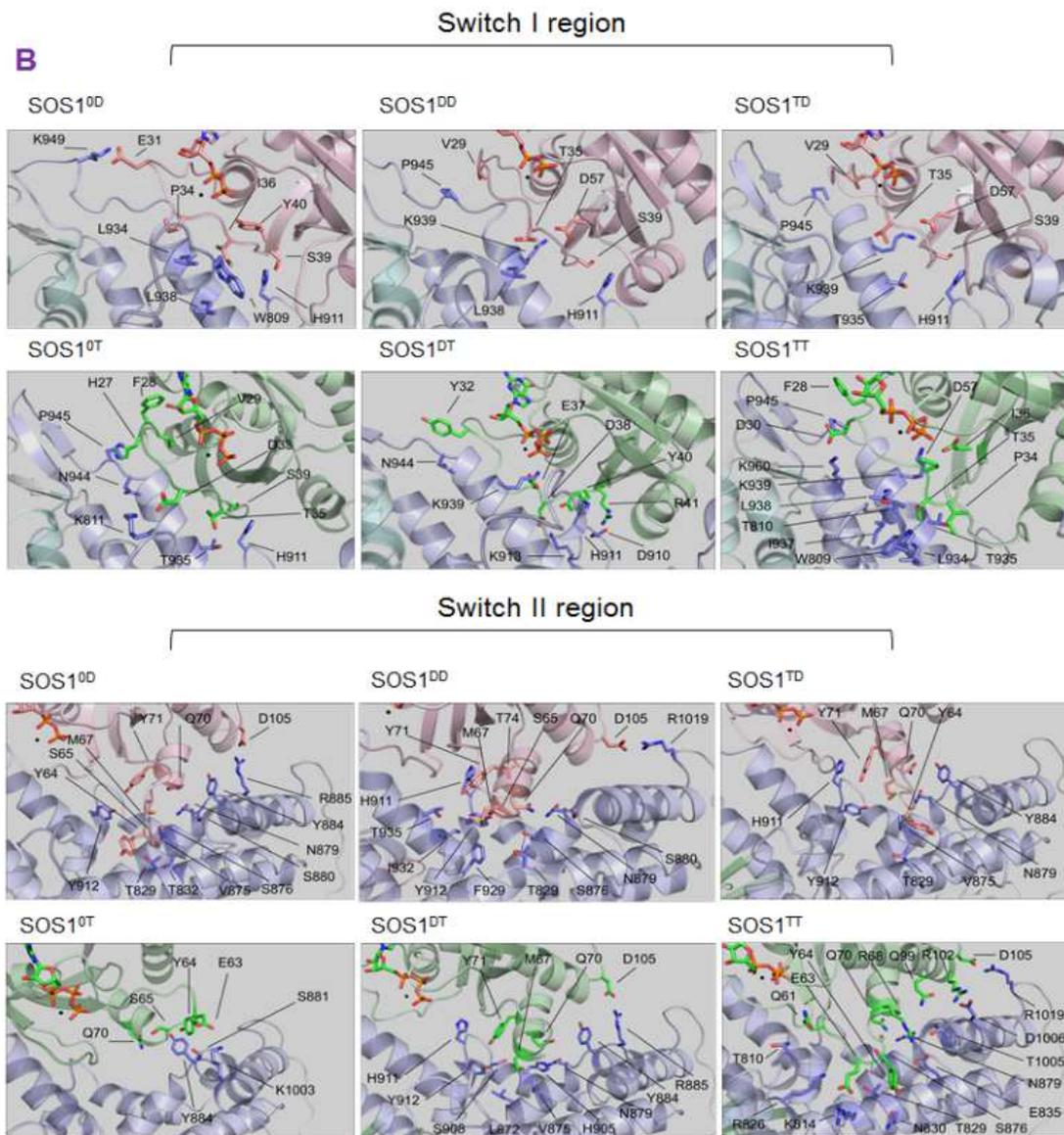
$\alpha 2/\alpha G$  and  $\beta 2/\beta B$  interfaces, and region 2 is formed by the  $\alpha 5/\alpha A$  interface (Fig. 5-1). Hydrophobic contacts drive the tandem domain-domain interaction in region 1 and salt bridge/hydrogen-bond (H-bond)/polar interactions in region 2 (Table D1). All SOS1 systems yield generally similar hydrophobic interacting residue pairs at the interdomain interface (Fig. 5-2A), whereas the residue pairs of the salt bridge/H-bond/polar interactions vary (Fig. 5-2B), indicating that allosteric KRas4B binding mostly affects region 2 rather than region 1 of the interface.

SOS1 function relies on REM dynamic motion that affects the stability of the CDC25 helix-hairpin motif [93, 101]. To observe the REM movement, we superimposed the average structure of CDC25 of apo-SOS1 (Fig. D3) and projected the protein backbone trace onto the two-dimensional XY plane (Fig. 5-3A) to compare the entire SOS1 conformations. In the absence of REM-bound allosteric KRas4B, SOS1 barely activates the catalytic KRas4B at CDC25. SOS1 conformational ensembles for this rare activation cycle follows  $SOS1^{00} \rightarrow SOS1^{0D} \rightarrow SOS1^{0T}$ . For  $SOS1^{0D}$  and  $SOS1^{0T}$ , the center of mass (COM) of REM is projected vertically below the reference position of Leu-670 (top row of Fig. 5-3A and Fig. D3), which is opposite to SOS1 systems containing allosteric KRas4B. While the REM structure in  $SOS1^{0D}$  deviates from that in  $SOS1^{00}$ ,  $SOS1^{0T}$  presents tandem domain conformations as in  $SOS1^{00}$ . The SOS1 system with catalytic KRas4B-GDP represents a configuration in the beginning of the nucleotide exchange, and with catalytic KRas4B-GTP it implicates the completion of nucleotide exchange. The exit of GTP-bound Ras from the catalytic site indicates that SOS1 conformation with catalytic Ras-GTP can be similar to that with Ras-free at the catalytic site. For SOS1 systems with allosteric KRas4B-GDP, the COM of REM is

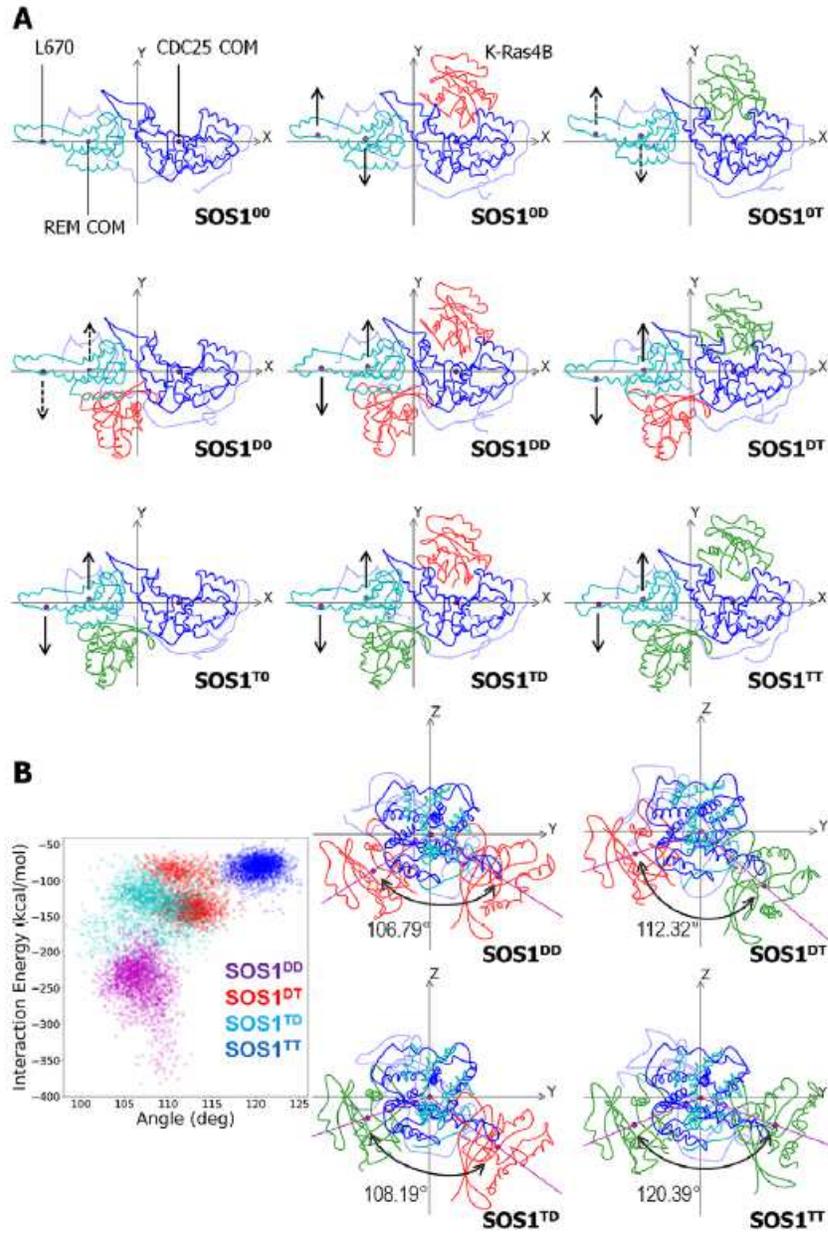
projected slightly above the reference position due to REM dynamic motion (middle row of Fig. 5-3A and Fig. D3). However, in the presence of allosteric KRas4B-GTP, with large conformational changes in SOS1, REM's COM is projected high above the reference position (bottom row of Fig. 5-3A and Fig. D3). The allosteric KRas4B-GDP induces insufficient conformational change in SOS1 to modulate the catalytic CDC25 conformation. In this case, SOS1 delays KRas4B activation in the limited activation cycle through  $SOS1^{D0} \rightarrow SOS1^{DD} \rightarrow SOS1^{DT}$ . In contrast, with allosteric KRas4B-GTP, the catalytic activity of SOS1 is enhanced, leading to an accelerated activation cycle through  $SOS1^{T0} \rightarrow SOS1^{TD} \rightarrow SOS1^{TT}$ . In  $SOS1^{T0}$ , the allosteric KRas4B-GTP induces large conformational change in the REM-CDC25 tandem domains, enabling the CDC25 catalytic site to recruit KRas4B-GDP. The nucleotide exchange begins in  $SOS1^{TD}$ , and conversion of the catalytic Ras to the GTP-bound state is completed in  $SOS1^{TT}$ . SOS1 conformation in  $SOS1^{TT}$  resembles the conformation in  $SOS1^{00}$ , indicating that the backward conformation change in SOS1 facilitates GTP-bound Ras exit from the catalytic site. With allosteric KRas4B-GDP, SOS1 increases the hydrophilic interaction at the tandem domains interface preventing the dynamic motion of REM, while SOS1 decreases the hydrophilic interaction at the interface with the allosteric KRas4B-GTP allowing the movement of REM (Fig. 5-2B). To measure quantitatively the conformational change in the tandem domains, we projected the protein backbone trace onto the YZ plane (middle and right panels of Fig. 5-3B). In the projection, the relative orientation of both Ras proteins due to the movement of REM is clearly visible. For  $SOS1^{DD}$  and  $SOS1^{DT}$  with allosteric KRas4B-GDP, the interaction between REM and CDC25 is relatively stronger (Fig. D4), and the

orientation angle between the two Ras proteins is relatively smaller than the corresponding systems of SOS1<sup>TD</sup> and SOS1<sup>TT</sup> with allosteric KRas4B-GTP. The strong domain-domain interaction between REM and CDC25 may impede SOS1's catalytic activity.





**Figure 5-2.** The interacting residue pairs with high occurrence rate (>50%) for the SOS1 interaction with KRas4B at (A) the allosteric and (B) catalytic sites. The sticks with labels highlight the interacting residues. The intermolecular residue-residue interactions include the hydrophobic, salt bridge, H-bond, and polar interactions. Table S2 summarizes a list of the residue pairs in the intermolecular interaction.



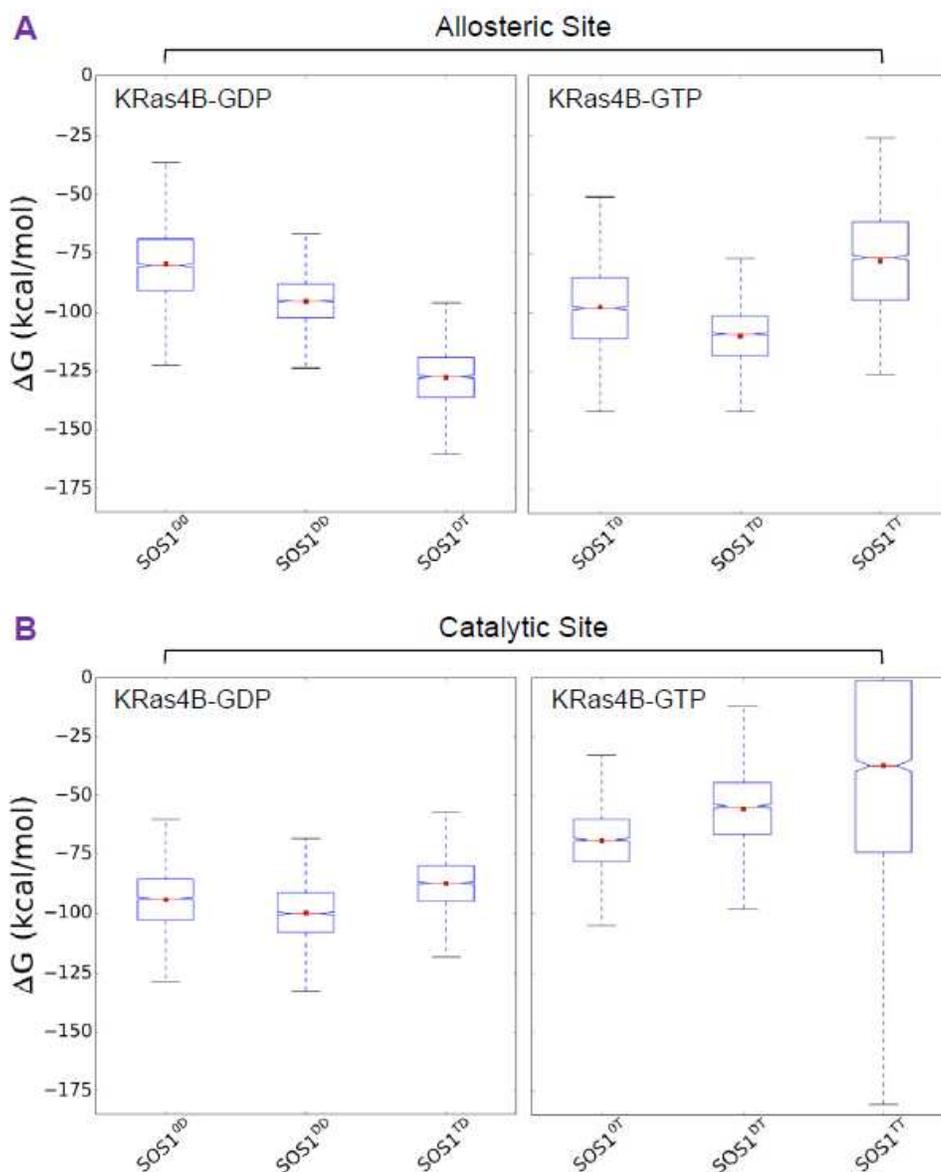
**Figure 5-3.** (A) The projection of protein backbone trace onto the two-dimensional XY plane for average structures of different SOS1 systems. In each panel, the purple dots from left-hand side correspond to the projected positions of the center of mass for the residue Leu-670, REM, and CDC25, respectively. All SOS1 systems were oriented with respect to CDC25 of  $SOS1^{00}$  as a reference with  $Y = 0$  for all dots. The arrows

indicate directions of the relative movement of the center of mass along the Y axis. The solid and dotted arrows denote large and small movements. (B) The rotation of the projected protein backbone trace into the YZ plane for the SOS1 systems on the limited ( $\text{SOS1}^{\text{DD}} \rightarrow \text{SOS1}^{\text{DT}}$ ) and accelerated ( $\text{SOS1}^{\text{TD}} \rightarrow \text{SOS1}^{\text{TT}}$ ) activation cycles. In the trace, the REM and CDC25 domains are colored cyan and blue, respectively, and KRas4B is colored red.

### 5.2.2 Competing interactions of KRas4B at the SOS1 binding sites

KRas4B mainly involves its Switch I and II regions in the interaction with SOS1. The SOS1–KRas4B interacting residue pairs depend on the KRas4B states at the SOS1 binding sites (Table D2). More Switch I and II regions residues of the allosteric KRas4B interact with SOS1 compared to catalytic KRas4B (Fig. D5). To quantify their contributions to the interaction, we calculated the binding free energies of the complex using molecular mechanics combined with the generalized Born and surface area continuum solvation (MMGBSA). The generally lower binding free energy values indicate that the allosteric KRas4B binds SOS1 more strongly than the catalytic KRas4B (Fig. 5-4). The relative changes in the binding free energy among different SOS1 systems suggest that allosteric KRas4B enables recognition of other Ras binding at the catalytic site. At the catalytic site, KRas4B-GDP with low values of the binding free energy interacts with SOS1 more strongly than KRas4B-GTP, reflecting the status of the SOS1 conformations before and after the nucleotide exchange. For example, in  $\text{SOS1}^{\text{TD}}$  the strong interaction of the catalytic KRas4B-GDP with SOS1 promotes nucleotide exchange, while in  $\text{SOS1}^{\text{TT}}$  the weak and unstable interactions of catalytic KRas4B-GTP with SOS1 cause the catalytic Ras to be released. The relative changes

in the binding free energies for catalytic KRas4B among the different possible activation cycles also suggest an allosteric communication between the two Ras binding sites. When SOS1 is recruited to the PM, the allosteric site provides higher binding affinity for KRas4B than the catalytic site, suggesting that KRas4B first binds to the allosteric site. At the allosteric site, KRas4B-GTP with high binding affinity easily competes with KRas4B-GDP as observed in the binding free energies for SOS1<sup>T0</sup> and SOS1<sup>D0</sup>. When SOS1 first recruits allosteric KRas4B-GTP, it enters the accelerated activation cycle with a positive feedback loop [91, 92]. However, in the case of SOS1 with allosteric KRas4B-GDP, the induced catalytic SOS1 conformation prevents exact alignments of Switch I and II of KRas4B in the interaction. As a result, the increasing binding strength at both binding sites due to involvements of other KRas4B regions retard the SOS1 system activation cycle.



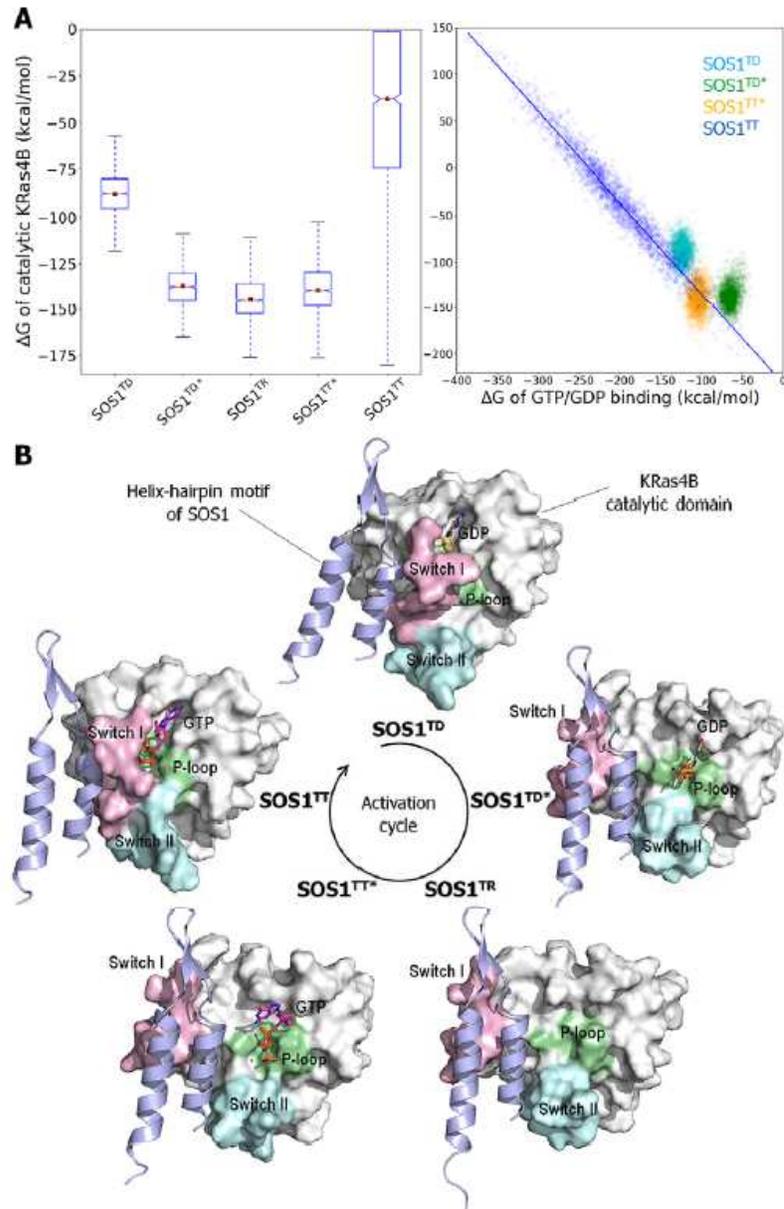
**Figure 5-4.** The binding free energy of KRas4B interacting with SOS1 at the (A) allosteric and (B) catalytic sites. The binding free energy was calculated using the equation,  $\Delta G = \Delta G_{Gas} + \Delta G_{Sol} - T\Delta S$ . The MMGBSA method was used to calculate the gas and solvation contributions, and the quasiharmonic mode analysis was used to calculate the entropic contribution. The mean and median values are denoted by red square and red horizontal line, respectively. The blue boxes represent the quartiles, and

the vertical dashed lines are determined by the 95% confidence intervals. This binding free energy does not reflect the real values of Gibbs free energy.

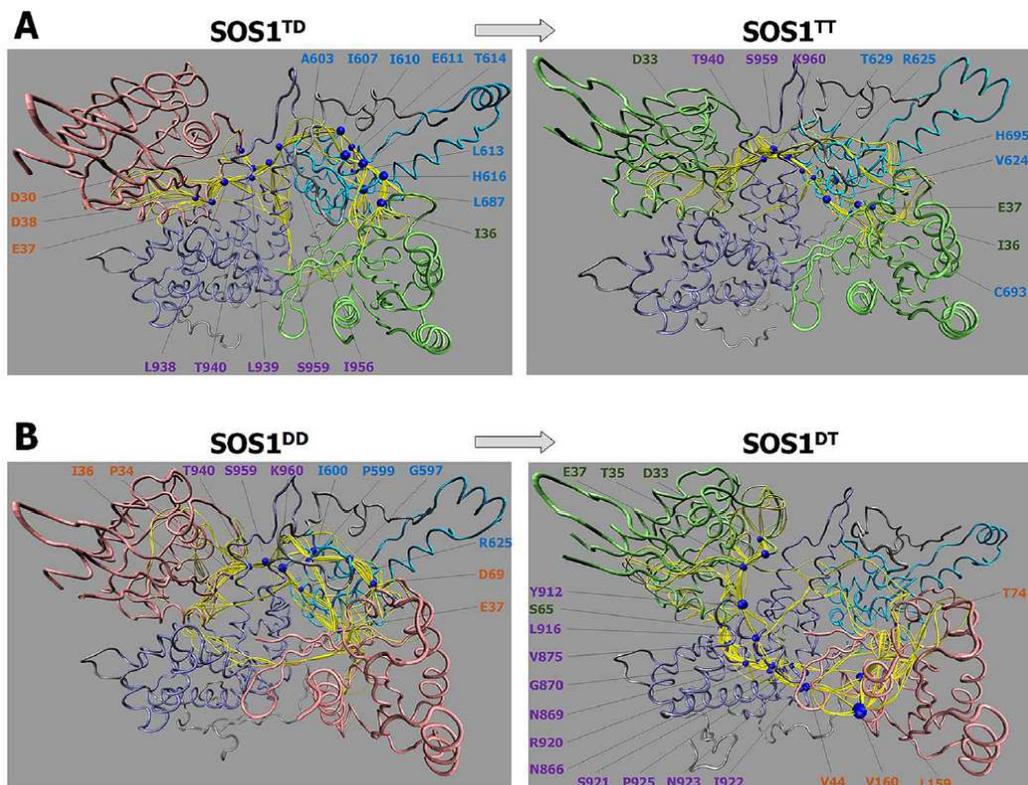
### 5.2.3 Conformational changes of KRas4B at the SOS1 catalytic site during activation

During the nucleotide exchange, catalytic KRas4B dramatically changes its conformation as the exchange event progresses. To reflect the conformational change of the catalytic KRas4B during the exchange, we refined the SOS1 systems in the accelerated activation cycle including the exchange event,  $\text{SOS1}^{\text{T0}} \rightarrow \text{SOS1}^{\text{TD}} \rightarrow \text{SOS1}^{\text{TD}^*} \rightarrow \text{SOS1}^{\text{TR}} \rightarrow \text{SOS1}^{\text{TT}^*} \rightarrow \text{SOS1}^{\text{TT}}$ . At the beginning of the nucleotide exchange, SOS1 begins to pull the Switch I loop of the catalytic KRas4B-GDP ( $\text{SOS1}^{\text{TD}}$ ), exposing the nucleotide binding pocket of KRas4B-GDP with the Switch I in an open conformation ( $\text{SOS1}^{\text{TD}^*}$ ). GDP now escapes, and the catalytic Ras becomes nucleotide-free ( $\text{SOS1}^{\text{TR}}$ ). GTP fills in the void in the binding pocket, and the catalytic Ras turns into a GTP-bound Ras with the Switch I in an open conformation ( $\text{SOS1}^{\text{TT}^*}$ ). Finally, the catalytic Ras closes the Switch I loop and prepares to leave SOS1 catalytic site ( $\text{SOS1}^{\text{TT}}$ ). To grade the conformational ensembles of catalytic KRas4B during the nucleotide exchange, we calculated the binding free energy of the catalytic K-Ras4B with SOS1 (Fig. 5-5A, left panel). We observed that catalytic KRas4B with the Switch I open conformation ( $\text{SOS1}^{\text{TD}^*}$ ,  $\text{SOS1}^{\text{TR}}$ , and  $\text{SOS1}^{\text{TT}^*}$ ) binds SOS1 more strongly than with the Switch I closed conformation ( $\text{SOS1}^{\text{TD}}$  and  $\text{SOS1}^{\text{TT}}$ ). At the catalytic site, the nucleotide-free Ras has a high binding affinity with the dissociation constant,  $K_D$ , thousand times smaller than the nucleotide-loaded Ras [89, 92, 102]. In contrast, catalytic KRas4B with the Switch I in the closed conformation binds the nucleotides, GDP and GTP, more strongly than with the Switch I open conformation. Nucleotide

binding to catalytic KRas4B is strongly correlated to the catalytic KRas4B binding to SOS1 with a correlation coefficient of  $r = -0.98$  (Fig. 5-5A, right panel), suggesting that the increasing binding strength of K-Ras4B to SOS1 compensates for the decrease in nucleotide interaction with the Ras catalytic site. Strong binding of GTP to Ras at the catalytic site suggests that GTP easily competes with GDP when the Ras catalytic site is exposed. To activate Ras, SOS1 uses its  $\alpha$ F helix to intrude into the Switch I loop of KRas4B-GDP generating the open conformation (Fig. 5-5B). For KRas4B in the Switch I closed state (SOS1<sup>TD</sup> and SOS1<sup>TT</sup>), relatively large fluctuations in the Switch II region can be observed (Fig. D6). However, for KRas4B in the Switch I open state, the fluctuations in the Switch II region disappear and instead large fluctuations in the G2 region (residues 26–37) containing Switch I can be observed (SOS1<sup>TD\*</sup>, SOS1<sup>TR</sup>, and SOS1<sup>TT\*</sup>). Essentially, the Switch II of K-Ras4B acts as a pivot when SOS1 opens Switch I [89, 91, 102, 103]; however, for KRas4B in the nucleotide-free state (SOS1<sup>TR</sup>), the fluctuations in the G4 region (residues 117–126) are due to the lack of stability which is established by the guanine group of the nucleotide.



**Figure 5-5.** (A) The binding free energy of catalytic KRas4B at the SOS1 catalytic site during the nucleotide exchange for the systems on the accelerated activation cycle (left panel). The correlation of the binding free energy between the catalytic KRas4B–SOS1 and KRas4B–nucleotide interactions (right panel). (B) Average structures of catalytic KRas4B with the CDC25 helix-hairpin motif depicting intervention of the motif in the Switch I region during the exchange event for the SOS1 systems on the accelerated activation cycle.

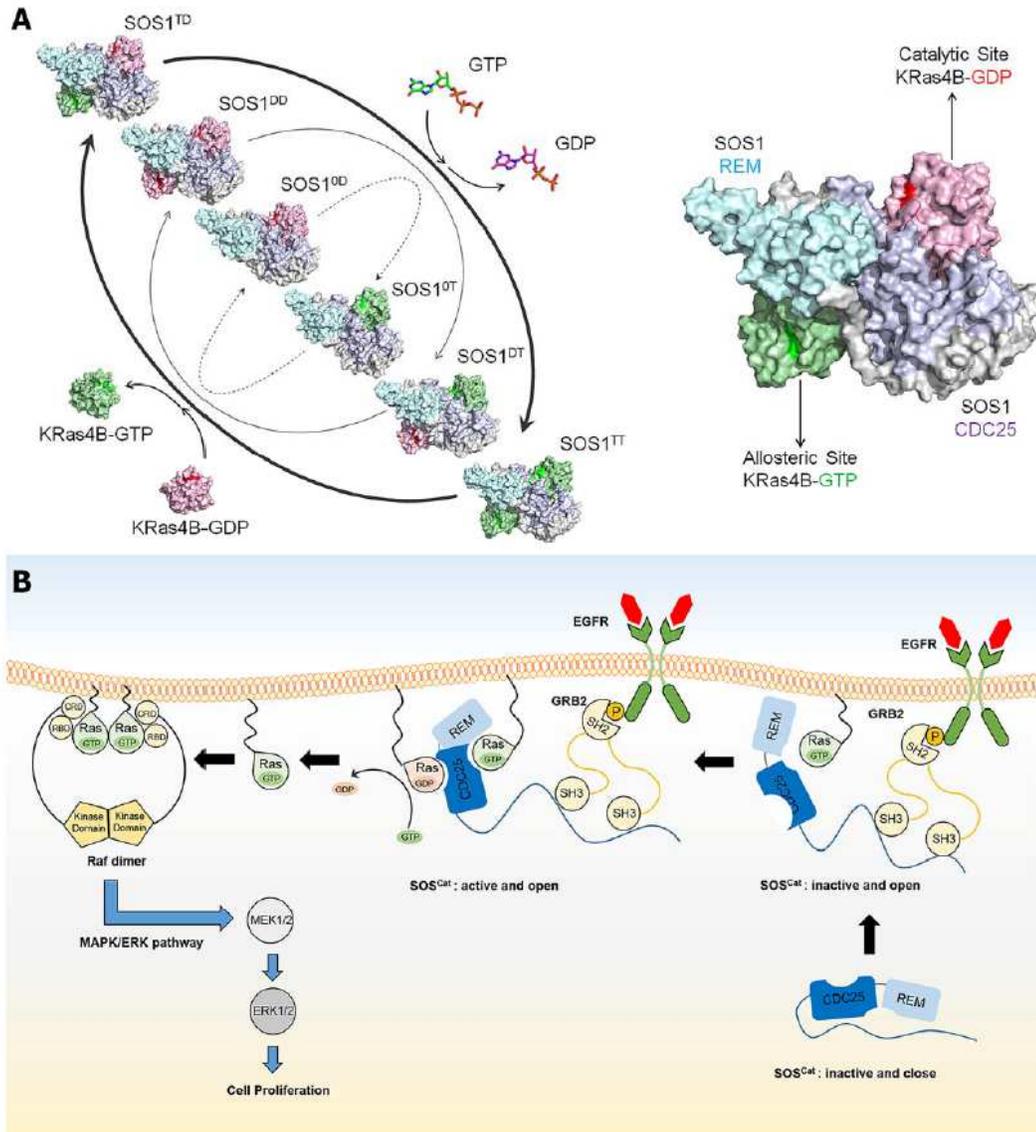


**Figure 5-6.** Allosteric pathways (yellow lines) between the allosteric and catalytic KRas4B propagating through the REM-CDC25 tandem domains for the SOS1 systems on the (A) accelerated (SOS1<sup>TD</sup> → SOS1<sup>TT</sup>) and (B) limited (SOS1<sup>DD</sup> → SOS1<sup>DT</sup>) activation cycles. In tube representation of SOS1, the REM and CDC25 domains are colored cyan and blue, respectively. For K-Ras4B, the GTP- and GDP-bound are colored green and pink, respectively. The blue beads on the pathways represent high occurrence residues (> 50%) in the pathway calculation.

#### 5.2.4 Allosteric signaling from KRas4B at the allosteric site to that at catalytic site in SOS1 activation

We hypothesized that KRas4B binding to the allosteric site may affect the conformation of the catalytic site allosterically, promoting the activation of SOS1. To

test this hypothesis, a dynamical network analysis was conducted using the NetworkView plugin in VMD. The dynamical network analysis can identify signal propagation pathways through the protein by calculating the weighted implementation of suboptimal paths (WISPs) [104]. We calculated a number of optimal and suboptimal pathways between two selected residues on KRas4B, one at the allosteric site and the other at the catalytic site (Table D3), to obtain the best pathways through the REM-CDC25 tandem domains of SOS1. The allosteric signals arose from Switch I and II of allosteric KRas4B, propagate through the SOS1 domains, and terminate at Switch I and II of the catalytic KRas4B. The Switch I and II regions are involved in the binding interface when KRas4B interacts with SOS1. For the accelerated activation cycle (Fig. 5-6A), the allosteric signals mostly propagate through the REM domain (via  $\alpha 1$ ,  $\alpha 2$ , and  $\beta 1$ ). In CDC25, the most frequently occurring residues (Thr-940, Ser-959, and Lys-960) are in the helix-hairpin motif formed by  $\alpha F$ - $\beta A$ - $\beta B$ - $\alpha G$  at the catalytic site. Allosteric dislocation of the helix-hairpin motif results in the  $\alpha F$  helix pushing against the Switch I region of the catalytic KRas4B, inducing the open conformation. For the limited activation cycle (Fig. 4-6B), although SOS1<sup>DD</sup> shows allosteric signals transmitting through the REM domain to the helix-hairpin motif, SOS1<sup>DT</sup> shows allosteric signals passing through the  $\alpha D$  and  $\alpha E$  regions of CDC25, which are far from the helix-hairpin motif at the catalytic site.



**Figure 5-7.** (A) The KRas4B activation cycles by SOS1. The dotted lines denote the rare activation cycle,  $SOS1^{DD} \rightarrow SOS1^{DT}$ , in the absence of allosteric Ras. The thin solid lines represent the limited activation,  $SOS1^{D0} \rightarrow SOS1^{DD} \rightarrow SOS1^{DT}$ . The thick solid lines denote the accelerated activation cycle,  $SOS1^{T0} \rightarrow SOS1^{TD} \rightarrow SOS1^{TT}$ , with high level of Ras activation. (B) A schematic diagram illustrates SOS1 activity at the plasma membrane. SOS1 activates KRas4B, and active KRas4B leads to cell proliferation regulating through the MAPK pathway.

### 5.3 Discussion

Here, we aimed to decipher the structural complexity of KRas4B activation by SOS1. To efficiently activate KRas4B at the CDC25 catalytic site, the allosteric site in the REM domain needs to recruit an active KRas4B [89, 102, 105]. Our simulations illustrate that KRas4B-GTP binding to the REM allosteric site induces movement of the REM domain, translocating the CDC25 helix-hairpin motif (Fig. 5-3), which causes displacement of the Switch I region of KRas4B-GDP at the catalytic site, yielding a Switch I open state [89, 93, 94]. Allosteric pathway analyses show that at the allosteric site, KRas4B-GTP generates signals propagating through SOS1's residues in the helix-hairpin motif (Fig. 5-6). These allosteric signals trigger large conformational changes in the tandem domains which accelerate SOS1 activation with positive feedback through binding cycles of the allosteric active Ras (Fig. 5-7A). By contrast, in the limited activation cycle, the allosteric KRas4B-GDP does not induce proper allosteric signals, impeding SOS1 activation of KRas4B. The binding free energies of the allosteric GDP- and GTP-bound KRas4B are relatively similar. However, the binding free energies of the catalytic KRas4B vary based on the types of nucleotide and the allosteric effects. Due to the allosteric signals, the catalytic KRas4B-GTP becomes less stable when the allosteric KRas4B is free, GDP-bound, and GTP-bound. For all three different activation cycles, the catalytic KRas4B-GDP is more stable than the catalytic KRas4B-GTP. The relative change in the binding free energy of the catalytic KRas4B between  $SOS1^{TD} \rightarrow SOS1^{TT}$  is the largest, followed by  $SOS1^{DD} \rightarrow SOS1^{DT}$  and  $SOS1^{OD} \rightarrow SOS1^{OT}$ .

Our simulations elucidate the mechanism of KRas4B activation by SOS1 in structural detail. They indicate that KRas4B interacts with the REM allosteric site more strongly than with the CDC25 catalytic site, suggesting that SOS1 first recruits allosteric Ras then the catalytic Ras. The binding free energy confirms that apo-SOS1 first loads an active KRas4B-GTP at the allosteric site which primes SOS1 for an inactive KRas4B-GDP at the catalytic site (Fig. 5-4). The interaction of KRas4B-GTP with SOS1 at the allosteric site induces a local conformation change at the catalytic site, facilitating the accommodation of the inactive Ras. The conformational change further involves opening of the nucleotide-binding site of GDP-bound Ras. During this feedback process, the interaction between SOS1 and the catalytic Ras gets tighter, while the interaction between Ras and the nucleotide gets weaker (Fig. 5-5). The nucleotide-free Ras shows the strongest binding strength with SOS1 catalytic site, indicating a widely opened conformation of the nucleotide-binding site. GTP binds Ras more strongly than GDP, and thus is easily loaded to the nucleotide-free Ras. GTP-loaded Ras closes its nucleotide-binding site, becoming the active form. SOS1 releases the active Ras-GTP due to weak interaction at the catalytic site, resetting for the next activation cycle.

The CDC25 domain is found in the family of GEFs for Ras-like small GTPases. Proteins containing this domain include Knkc1, RalGDS, RapGEF1-6, and SOS1-2, which have both REM and CDC25 domains. Our work, which shows how multiple Ras GTPases communicate with each other via forming a complex with GEF's tandem domain, can be applied to most Ras-like small GTPase interacting with GEF. For example, the basic subunits of GEF for the Rho family proteins (RhoGEF) are DH and

PH domains: the DH domain catalyzes Rho, Rac, and Cdc42 by exchanging GDP to GTP, and the PH domain mainly interacts with plasma membrane. However, one of the RhoGEF members, Dbs protein, forms a dimer through its DH-PH interface [106]. Two catalytic Cdc42 may also transmit allosteric signals via a Dbs dimer. Consequently, the PH domain participates in the DH-Cdc42 association and may induce conformational changes, promoting the nucleotide exchanges. Even though there is no allosteric Cdc42 binding to Dbs, this is a similar case, but more complex than our model, where the PH domain may play a role similar to the REM domain in RasGEF.

Oncogenic KRas4B mutants in the GTP-bound state (KRas4B<sup>mut</sup>-GTP) present a conformation with tightly closed Switch I and II regions, while the GDP-bound mutants have large fluctuations in both Switch regions, showing more widely open nucleotide-binding site than wild-type KRas4B-GDP [107, 108]. With distinct Switch I and II conformations, mutant KRas4B interacting with SOS1 can shift the equilibrium of Ras activation. Because oncogenic KRas4B mutants block GTP hydrolysis by GTPase-activating protein (GAP), there is a large population of active KRas4B<sup>mut</sup>-GTP. The highly populated, active KRas4B mutant can occupy the allosteric site of SOS1, sending the activation signal to the catalytic site. SOS1 catalytic site does not favor the tightly closed Switch I and II conformation of active KRas4B<sup>mut</sup>-GTP. Instead, inactive KRas4B<sup>mut</sup>-GDP conformers with pre-opened Switch I and II conformation can easily select and associate with the SOS1 catalytic site. Mutant KRas4B highly accelerates the action of SOS1 in Ras activation cycle, while GTP hydrolysis by GAP to produce the inactive GDP-bound form is abolished or extremely slow. The imbalanced Ras activation/deactivation cycle due to the extremely fast activation by SOS1 and

extremely slow or halted deactivation by GAP, culminate in a highly populated active mutant KRas4B-GTP state in cancer.

The major Ras activation cycle is fast and is the predominant, native function of SOS. To regulate the MAPK pathway (Fig. 5-7B), epidermal growth factor (EGF) first binds to EGF receptor (EGFR) promoting phosphorylation. The phosphorylated EGFR recognizes and interacts with the SH2 domain of growth factor receptor bound protein-2 (Grb2). Subsequently, two SH3 domains of Grb2 bind the proline-rich SOS1 C-terminal tail [3, 27, 30, 33], recruiting SOS1 to the cell membrane. The PH domain in the N-terminal region of SOS1 is responsible for the membrane anchorage, and the activity of the SOS1 catalytic unit is enhanced when interacting with membrane-anchored Ras [109]. The REM allosteric site of SOS1 first binds membrane-anchored GTP-bound KRas4B with lower binding free energy, than the CDC25 site binds the GDP-bound Ras. It was suggested that Ras-GTP binds the allosteric site tenfold tighter than the Ras-GDP [102]. We also observe tighter binding. The binding of KRas4B-GTP to the allosteric site induces the dynamics of REM, with allosterically-promoted conformational changes displacing the CDC25 helix-hairpin motif, resulting in the Switch I open conformation of KRas4B-GDP at the SOS1 catalytic site. A grasp of the allosteric motion of REM-CDC25 tandem domains and identification of allosteric pathways through the SOS1 domains, are crucial for therapeutics to control Ras activation and signaling in cancer. Our comprehensive simulations clarify SOS1 allosterically-linked conformational change, delineate the key steps in KRas4B activation cycle, and provide detailed allosteric signal propagations upon associating with different forms of KRas4B.

## 5.4 Material and Methods

### 5.4.1 Generating initial configurations of KRas4B-SOS1 complex

To construct initial configurations, we obtained the crystal structures of SOS1 (PDB code: 4NYJ) [88], GDP-bound KRas4B<sup>C118S</sup> (PDB code: 4EPT) [97], and GTP-bound KRas4B<sup>Q61H</sup> (PDB code: 3GFT) from the Protein Data Bank. In 4NYJ, both Ras binding sites (allosteric and catalytic) in SOS1 are occupied by HRas. At the allosteric site, the Ras protein is a GNP-bound HRas, and it is a nucleotide-free HRas at the SOS1 catalytic site. After replacing the mutants with the wild-type residues, KRas4B was superimposed onto HRas in complex with SOS1, generating the coordinates for the KRas4B-SOS1 complex. At the catalytic site of SOS1, the Switch I region of both GDP- and GTP-bound KRas4B clashed with the  $\alpha$ F helix, due to the Switch I open conformation of the nucleotide-free Ras. To avoid this steric clash, we re-modeled the Switch I region using the Modeller server [110], ensuring that the modeling of Switch I does not affect the positions of GDP, GTP, and Mg<sup>2+</sup>. A total of twelve SOS1 systems were constructed: an apo-SOS1 monomer, SOS1<sup>00</sup>; four dimeric systems, SOS1<sup>D0</sup>, SOS1<sup>T0</sup>, SOS1<sup>0D</sup>, and SOS1<sup>0T</sup>; four ternary systems, SOS1<sup>DD</sup>, SOS1<sup>DT</sup>, SOS1<sup>TD</sup>, and SOS1<sup>TT</sup>; and four ternary systems in the exchange event, SOS1<sup>TD\*</sup>, SOS1<sup>TR</sup>, and SOS1<sup>TT\*</sup>.

#### 5.4.2 Atomistic molecular dynamics simulations

The initial configurations were subject to MD simulations in an aqueous environment. The modified TIP3P water model [111] was used to create the isomeric unit cell box containing the Ras–SOS complex. The initial systems were neutralized by adding counter ions and soaked additional Na<sup>+</sup> and Cl<sup>-</sup> to satisfy a total ion concentration near 100 mM. The updated CHARMM all-atom additive force field [48] was used to construct the set of starting points and to relax the systems to a production-ready stage, closely following the same protocol as in our previous works [39, 40, 107, 108, 112-118].

A series of minimization steps using steepest decent (SD) and the adopted-basis Newton-Raphson (ABNR) methods and dynamic relaxation cycles were performed for the solvent around the harmonically restrained complex. At the final pre-equilibrium stage, the SOS1 systems were gradually relaxed by removing the harmonic restraints through dynamic cycles with the full particle mesh Ewald (PME) electrostatics calculation. In the production runs, the constant temperature at 310 K was maintained by the Langevin temperature control, and the pressure at 1 atm was sustained by the Nosé-Hoover Langevin piston pressure control. 400 ns production run was performed using the NAMD parallel computing code [47] on a Biowulf cluster at the National Institute of Health (Bethesda, MD). Analysis was performed with the CHARMM programming package [119].

### 5.4.3 Binding free energy calculation

The binding free energy was estimated by the combination of molecular mechanics combined with the generalized Born (GB) and surface area continuum solvation (MMGBSA). In the calculation, we closely followed the protocol reported in our previous studies [39, 116-118]. The average binding free energy was calculated as a sum of the gas phase contribution, the solvation energy contribution, and the entropic contribution,

$$\langle \Delta G_b \rangle = \langle \Delta G_{gas} \rangle + \langle \Delta G_{sol} \rangle - T\Delta S \quad (5-1)$$

, where  $\langle \rangle$  denotes an average along the MD trajectory. The gas phase contribution to the binding free energy is a sum of the internal energy, the vdW interaction, and the electrostatic energy,

$$\Delta G_{gas} = \Delta H_{intra} + \Delta H_{vdW} + \Delta H_{elec}. \quad (5-2)$$

The solvation contribution is a sum of the electrostatic and non-polar contribution,

$$\Delta G_{sol} = \Delta G_{sol}^{elec} + \Delta G_{sol}^{non-polar}. \quad (5-3)$$

The solvation free energy was obtained from the GB calculation using the GBSW module [120] of the CHARMM program [119]. The entropy term can be divided into the translational, rotational, and vibrational contribution,

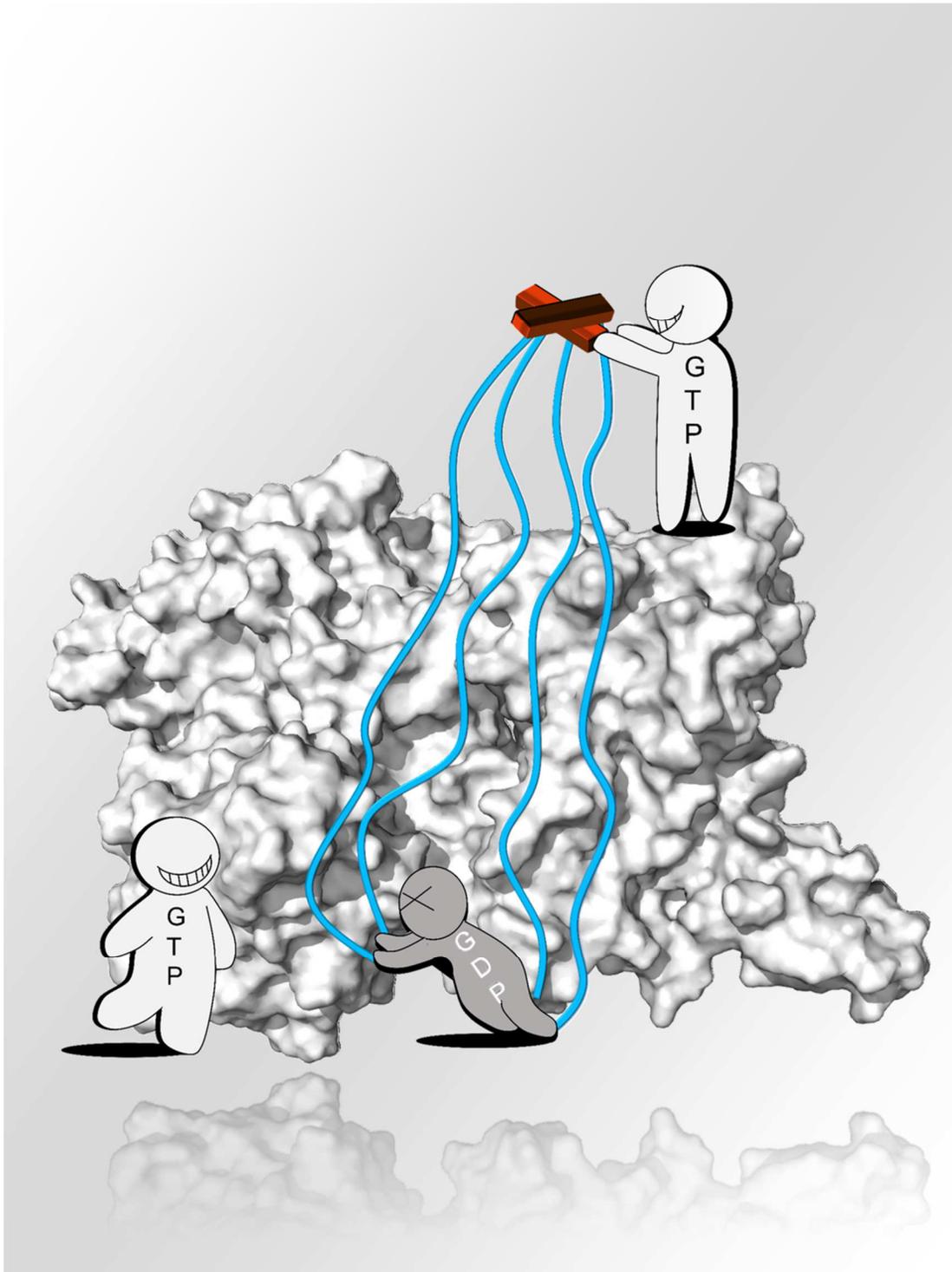
$$T\Delta S = T\Delta S_{trans} + T\Delta S_{rot} + T\Delta S_{vib}. \quad (5-4)$$

The translational and rotational contributions were obtained from the calculation of principal moment of inertia. The vibrational entropy was obtained from the quasi-harmonic mode calculation in the VIBRAN module of the CHARMM program [119].

The number of vectors (mode) to calculate the vibrational analysis was set to  $\text{NATOM} \times 3$ , where  $\text{NATOM}$  denotes the number of atoms. Finally, the change in binding free energy due to the complex formation was calculated using the equation,

$$\Delta G_b = G_b^{\text{complex}} - (G_b^{\text{SOS1}} + G_b^{\text{Ras}}). \quad (5-5)$$

Cover art that submitted to *Biophysical Journal* but not selected



## Chapter 6: RASSF5 is an MST activator and tumor suppressor in vivo but opposite in vitro

Chapter 6 was adapted from: TJ. Liao, CJ. Tsai, H. Jang, D. Fushman, R. Nussinoc: *RASSF5: An MST activator and tumor suppressor in vivo but opposite in vitro*. *Current Opinion in Structural Biology*, 2016, 41:217-224.

### 6.1 Introduction to chapter 6

Ras association domain family member 5 (RASSF5, also known as NORE1A) is an effector of the Ras protein . It also binds to macrophage-stimulating protein 1/2 (MST1/2) kinases in the Hippo pathway. Hippo's signaling promotes phosphorylation of Yes-associated protein 1 (YAP1) [122-124]. Phosphorylation is an essential signal that tags YAP1 for degradation. Overexpression of YAP1 is often observed in cancer, including those cancers where the mitogen-activated protein kinases (MAPK) pathway (Ras/Raf/MEK/ERK) is also mutated [125-129]. Thus, RASSF5 is a key network node linking active, GTP-bound Ras, including the highly oncogenic KRas4B, to the Hippo pathway and YAP1 abundance in cancer. *KRAS4B* is the most abundant mutated oncogene in cancer, particularly of the pancreas (over 95%), lung and colorectal [4, 80, 130]. The clear importance of RASSF5 as a potential drug target has recently led to increasing interest in the community. Signaling through the Hippo pathway, a conserved kinase cascade that controls organ size by regulating cell proliferation, apoptosis, and stem cell self-renewal, is modulated not only by the RASSF5, but also by cell contact inhibition [131]. Dysregulation of the Hippo pathway can result in higher levels of YAP1 and cancer development.

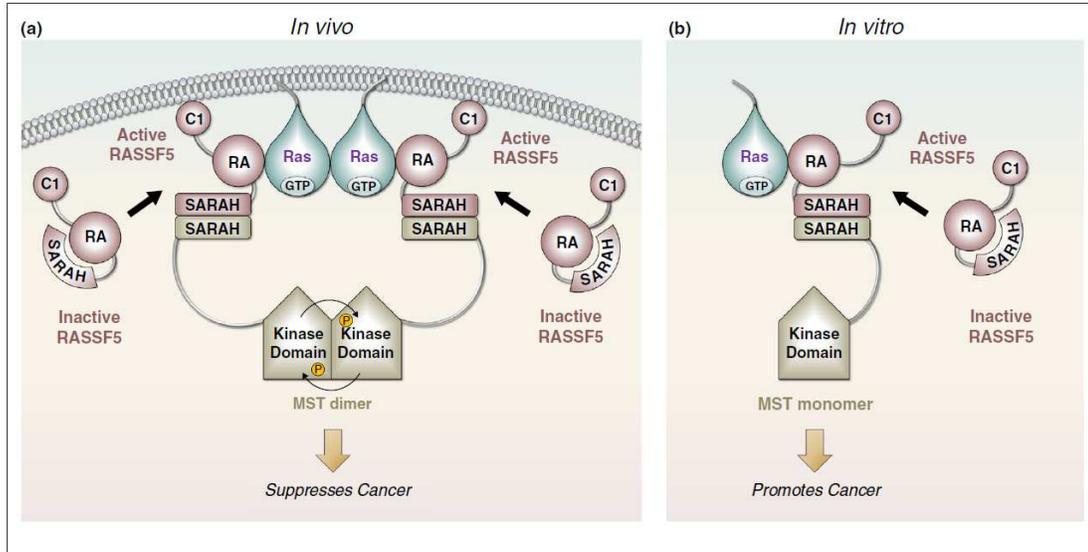
Here we focus on the linkage between KRas and Hippo's signaling through RASSF5. From the mechanistic standpoint, RASSF5 has raised a number of fundamental and perplexing questions. Among these is first, how Ras activates RASSF5; second, can RASSF5 act as both a suppressor and activator of cancer which is what some experimental reports suggest and if so how; third, MST1/2 trans-autophosphorylation requires that the kinase domain homodimerizes. Is the homodimerization promoted by MST1/2 homo- or hetero- SARAH (Sav-RASSF-Hippo) domain dimerization involving RASSF5 and MST1/2? Understanding how RASSF5 links Ras to MST1/2, and via Hippo signaling to YAP1, will provide a good grasp of a key linkage in cancer cell biology, and thus in drug discovery. Here we will comment on these questions from the structural standpoint, as well as through the lens of our views of pathway-driven tumor proliferation.

Below we provide a mechanistic perspective of RASSF5 in *KRAS4B*-driven cancer. We describe how KRas4B activates RASSF5, which in turn can promote – or as some experimental data suggest – inhibit MST1/2 activation. We argue that *in vivo* activated RASSF5 can function to activate MST1/2 and suppress cancer (Fig. 6-1A) whereas *in vitro*, activated RASSF5 can function to inhibit MST1/2 which in a cell scenario would activate cancer (Fig. 6-1B). We also overview KRas pathway-driven tumor initiation and the RASSF5 linkage to the Hippo pathway in this light. The Hippo and MAPK signaling have similar roles in tumor cell proliferation, and the Hippo pathway, as well as YAP1, are frequently mutated or overexpressed in Ras-driven cancers and drug resistance. Hippo and MAPK are independent *core* pathways fulfilling similar roles in the cell cycle [128].

## 6.2 Hypothesis

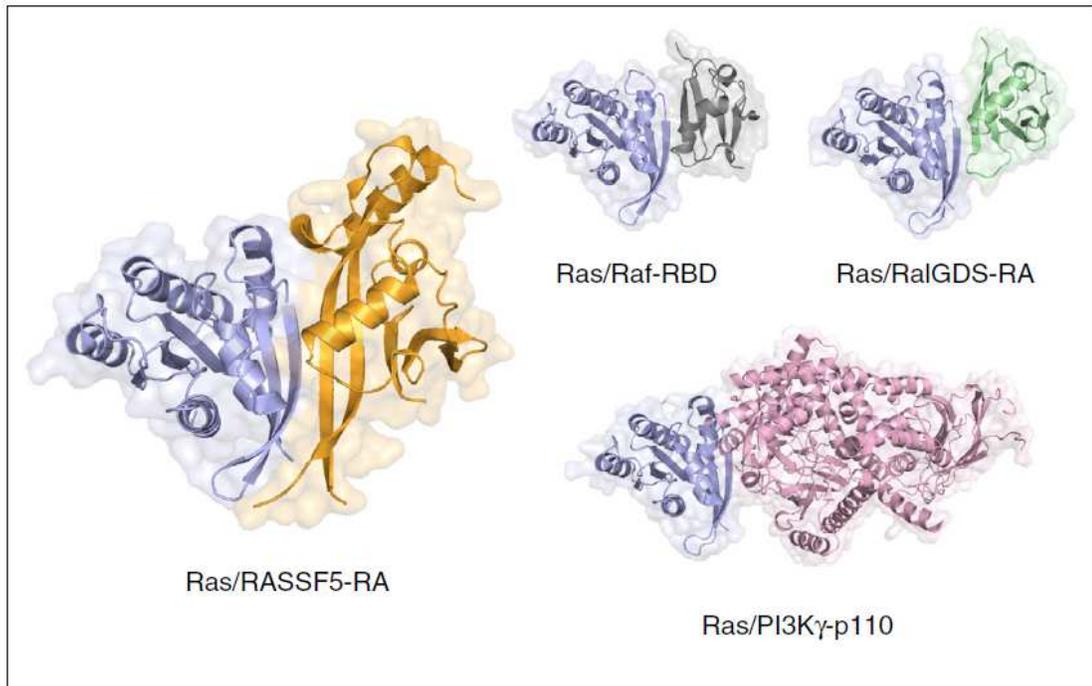
### 6.2.1 RASSF5 is an adaptor protein

*How can RASSF5 act as both suppressor and activator of cancer as some literature reports suggest?* Figure 1 provides a schematic diagram of our thesis. RASSF5 is a Ras effector. It interacts with Ras via its Ras association (RA) domain. The interaction is at the same effector site where other Ras effectors, such as the serine-threonine protein kinase Raf, the lipid kinase phosphatidylinositol-4,5-bisphosphate 3-kinase (PI3K) and Ral guanine nucleotide dissociation stimulator (RalGDS), the exchange factor for other GTPases, also interact [132-135]. The interfaces of Ras/Raf, Ras/PI3K and Ras/RalGDS overlap that of Ras/RASSF5, and the Ras binding domains (RBDs) of Raf, PI3K and the RA domains of RalGDS and RASSF5 are all similar to each other (Fig. 6-2) [4]. GTP-bound Ras activates all its effectors. Among Ras effectors, Raf's activation involves the homodimerization of its catalytic kinase domain and trans-autophosphorylation, where the catalytic domains cross-phosphorylate each other [136, 137]. A long, ~165-residue linker, characteristic to all Raf proteins, connects Raf's RBD with Raf's catalytic domain. Allosteric effects elicited by the binding of the RBD to Ras, and 14-3-3 recognition of a phosphorylated motif in the highly flexible hinge region (Ser259 in human c-Raf), result in the homodimerization of the catalytic kinase domain, Raf's activation, and MAPK – also a phosphorylation cascade – signaling. MAPK signaling acts in the early phase of the G1 (Gap 1) into the S (Synthesis) cell cycle stage [132]. Ras dimers (and nanoclusters) increase the effective local concentration of Raf and may restrain its orientation, preconditioning it for productive binding [54, 138, 139].



**Figure 6-1.** As an adaptor, RASSF5 binds to Ras and MST through its RA domain and SARAH domain respectively. (A) *In vivo*, active Ras is anchored into the membrane, forming dimers and nanoclusters. The C1 domain of RASSF5 attaches to the membrane as well. The interaction with Ras leads to a conformational change in RASSF5, allosterically activating it, shifting the landscape toward an open SARAH domain, which makes it available for interaction with the SARAH domain of MST. The RASSF5/MST SARAH heterodimer has a higher affinity than RASSF5/RASSF5 or MST/MST SARAH homodimers. Membrane-anchored Ras-bound RASSF5 has increased effective local concentration and is preoriented for productive MST homodimerization, with trans-autophosphorylation with RASSF5 acting as an adaptor linking active GTP-bound Ras and MST. The active MST then activates the Hippo pathway which results in phosphorylation of YAP1, leading to its degradation, thus suppressing cancer. By contrast (B), *in vitro*, without the benefits of the membrane, active Ras lacks anchoring and its effective local concentration is low. Due to the stronger interaction of the RASSF5/MST SARAH heterodimer than the MST/MST

SARAH homodimer, the MST/RASSF5 complex is still preferred, which prevents MST kinase domain homodimerization and MST transautophosphorylation. The inactive MST inactivates the Hippo pathway. YAP1 is overexpressed, which promotes cell proliferation.



**Figure 6-2.** Ras binds to its effectors through the same effector binding site of catalytic domain. The interactions are all highly similar. Ras, colored in blue, binds to the Ras association (RA) domain of RASSF5, colored in yellow (PDB ID: 3DDC), to the Ras binding domain (RBD) of Raf, colored in gray (PDB ID: 4G0N), to the RA of RalGDS, colored in green, and to the RBD of PI3K's P110 catalytic subunit.

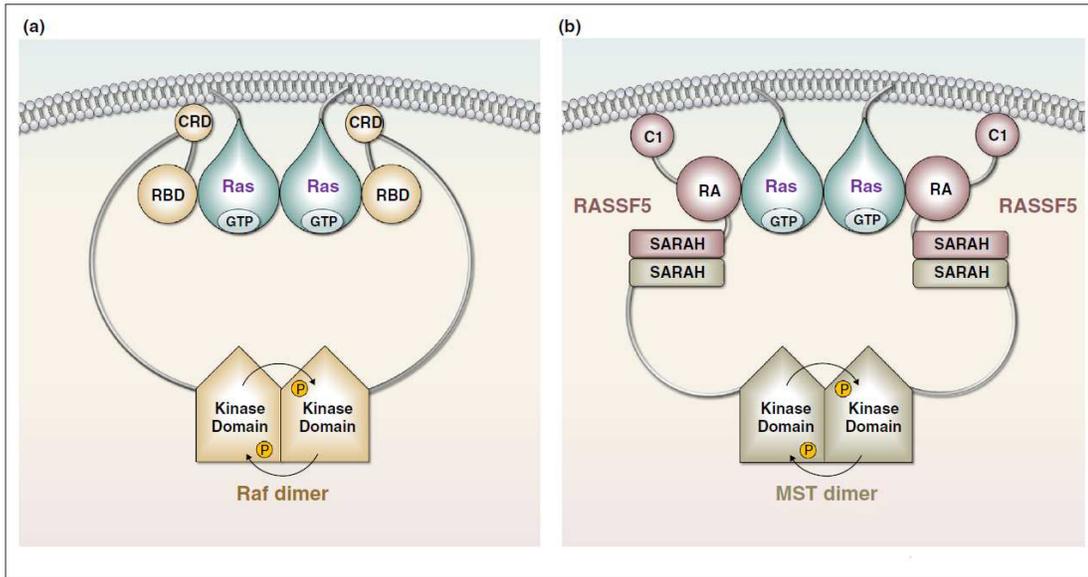
MST1/2 does not have a Ras binding domain; RASSF5 does, and like MST1/2 it has a SARAH domain [140]. The RA domain of RASSF5 is linked to the SARAH coiled-coil domain via a flexible 3-residue segment [141]. Ras activates RASSF5 [142]. Activation allosterically shifts the equilibrium of RASSF5 toward a state where the

SARAH domain shifts away from the Ras-bound RA domain, making it available to bind to the SARAH domain of MST1/2, resulting in a SARAH antiparallel heterodimer (Fig. 6-1). The consequent increased population of the open conformation of the MST1/2 makes the kinase domain available for homodimerization and trans-autophosphorylation. Similar to Raf [54, 133, 135, 138, 143, 144], Ras dimerization and nanocluster formation [82, 145] play a key role in MST1/2 dimerization. Figure 3 depicts the two mechanisms – Raf's and RASSF5/MST1/2's homodimerization – side-by-side, illustrating the mechanistic similarity between them and RASSF5 adaptor role.

As in Raf's dimerization and activation, membrane attachment is crucial to this mechanism of MST1/2 activation. Thus, we believe that this mechanism is the one underlying *in vivo* experimental observations. By contrast, *in vitro*, in the absence of the membrane and Ras dimerization or nanoclustering, and with a higher hetero- than homo- SARAH domain affinity [141], RASSF5 SARAH domain outcompetes MST1/2 SARAH homodimerization to also form a SARAH domain heterodimer; however, in the absence of other nearby favorably preoriented MST1/2 catalytic domains, the chance of productive kinase domain dimerization is low. The outcome is MST1/2 inhibition by RASSF5 *in vitro*. In this scenario, RASSF5 acts as activator of cancer. Suppression of cancer requires MST1/2 activation and Hippo pathway signaling.

*So how can RASSF5 act as both suppressor and activator of cancer as some literature reports suggest?* As an adaptor its action depends on the conditions. *In vivo*, with the membrane present, membrane-anchored Ras dimers/nanoclusters promote SARAH domain heterodimerization, and MST1/2 kinase domain homodimerization and trans-autophosphorylation. By contrast, *in vitro*, no membrane; KRas binding still

releases the RASSF5 SARAH stimulating MST1/2's SARAH heterodimerization; however, without membrane and thus Ras dimers/nanoclusters, no MST1/2 kinase domain dimerization/trans-autophosphorylation.



**Figure 6-3.** Raf and RASSF5 are Ras effectors, binding at the same site. Raf is a kinase; RASSF5 is an adaptor protein which *in vivo* can activate the MST kinase. Here we suggest that RASSF5 plus MST resemble Raf. MST requires RASSF5's help since it does not contain a Ras binding domain. The figure shows the similarity in the mechanisms of Raf and RASSF5 in the cell. (A) Raf binds to Ras and gets activated through its Ras binding domain (RBD). Its cysteine rich domain (CRD), which attaches to the membrane helps in Raf's dimerization. The kinase domain dimerization is required for its autophosphorylation; (B) RASSF5/MST shows similar scenario. Without an RBD, MST cannot interact with Ras. Therefore, RASSF5 plays an adaptor role to bridge Ras and MST. The Ras association (RA) domain and the C1 domain behave like the RBD and the CRD of Raf. RASSF5 first binds to Ras through its RA

domain, and its C1 domain anchors into the membrane. Through the RASSF5/MST SARAH domain heterodimerization, the MST kinase domain is able to link to Ras, inducing the MST kinase domain dimerization and phosphorylation just like Raf's kinase domain. This scheme argues for simplicity: Ras effectors behave similarly, and there is no need to invoke new schemes.

#### 6.2.2 How activated RASSF5 can act to activate MST1/2 and suppress cancer in vivo

MST1/2 can homodimerize through the C-terminal SARAH domain, with activation taking place through autophosphorylation of the activation loop (Thr183 for MST1 and Thr180 for MST2) [146-148] in the N-terminal kinase domain homodimer. Sav1 and RASSF proteins also contain SARAH domains. The MST1/2 SARAH domain can heterodimerize with the RASSF [149] and with the Sav1 SARAH domain. SARAH domain dimerization is necessary for homodimerization and autophosphorylation of the kinase domain. The mechanisms and scenarios of homo- and hetero- SARAH domain dimerization and activation *versus* inhibition have been elusive [146].

RASSFs, such as RASSF1A and RASSF5, have been shown to act as tumor suppressors [150-152]. They were observed to activate MST1/2 [153, 154], for example by coexpression of RASSFs with Ras, which enhances the MST1 kinase activity in cells. However, in what has been seen as a paradoxical behavior, in mammals, RASSF1, RASSF5, and RASSF6 blocked MST1/2 autophosphorylation *in vitro* [146, 155-157]. The kinetics of MST2 activation has been monitored *in vitro*. Addition of purified RASSF5 to full-length MST2 abolished its autophosphorylation and activation. Binding of RASSFs to the already activated MST1/2 did not inhibit their kinase activities [146]. These observations coupled with additional ones, e.g. that

RASSF1A promotes apoptosis through the Hippo pathway [157] and that RASSF2 and MST1/2 form a SARAH heterodimer which activates MST1/2 [158, 159]. Further support for the tumor suppressor role of RASSF5 came from exogenous expression which potently inhibits tumor cell growth [121, 160] and from inhibition of NORE1A expression which enhanced cell proliferation [161]. Cells lacking NORE1A and myocyte-specific enhancer factor (MEF), a transcriptional activating member of the E26 transformation-specific (ETS) family of transcription factors were readily transformed by Ras, which normally require p53 or Rb inactivation for Ras transformation [162]. NORE1A is often inactivated in cancer and its loss has been identified in more aggressive primary tumors [163, 164]. Inactivation of NORE1A has been associated with clear cell renal cell carcinomas [164]. Taken together, these observations led to the assumption that RASSF can both activate and inhibit the Hippo pathway, and various hypotheses have been offered to explain these perplexing dual function observations [165]. Even though it has been noticed that its cancer suppression action takes place *in vivo*, how to explain these observations has been baffling.

Recently the crystal structures of the human MST2 kinase domain and MST2 in complex with the SARAH domain of RASSF5 were obtained permitting in-depth mechanistic studies [146]. SARAH-mediated homodimerization of the MST2 kinase domain is essential for its autophosphorylation and activation. The SARAH domains can homo- and hetero- dimerize. The heterodimer has higher affinity [141], suggesting that the RASSF5 SARAH can outcompete the MST1/2 SARAH homodimer. Figure 1a outlines how Ras-activated RASSF5 can activate MST1/2 and suppress cancer, and Figure 1b illustrates how Ras-activated RASSF5 can inhibit MST1/2

autophosphorylation and thus the Hippo signaling to promote cancer. In both cases the SARAH domains heterodimerize. We propose that the differential activation/suppression action reflects the conditions under which the experiments were carried out. As depicted in the Figure 1 cartoons, in the inactive state of RASSF5, the SARAH domain is likely to be loosely associated with the RA domain. This relatively unstable association is indicated by the unsuccessful attempts to co-crystallize the MST1/2 RA with the SARAH domain and by our modeling. Upon binding to Ras, in both *in vivo* and *in vitro* scenarios, the SARAH domain allosterically shifts to a predominantly open conformation afforded by the 3-residue flexible RA-SARAH junction. *In vitro*, in the absence of the membrane and Ras dimers, MST1/2 and thus Hippo signaling will be inhibited; *in vivo*, due to the increasing proximity of Ras dimers and nanoclusters [4, 54, 133, 135, 138, 143-145, 166], MST1/2 will be activated, now poised to inhibit cancer. Thus in the cell RASSF5 is a cancer suppressor. This mechanistic outline and rational are of fundamental importance and may provide a blueprint for drug discovery efforts. Ras functions when it is membrane-anchored, forming dimers and nanoclusters; this Ras state promotes RASSF5 activation as well as Raf's activation and MAPK signaling. This argues that to enhance RASSF5 cancer suppressor power, drug discovery efforts should focus on restored or increased expression levels. Agents that enhance nanoclustering will promote RASSF5 activation; but will also boost MAPK proliferative action and cancer [167].

We have undertaken modeling and initiated NMR experiments to follow Ras-elicited RASSF5 activation with subsequent MST activation through hetero-SARAH domain

dimerization, as well as other components essential for full understanding of RASSF5 actions.

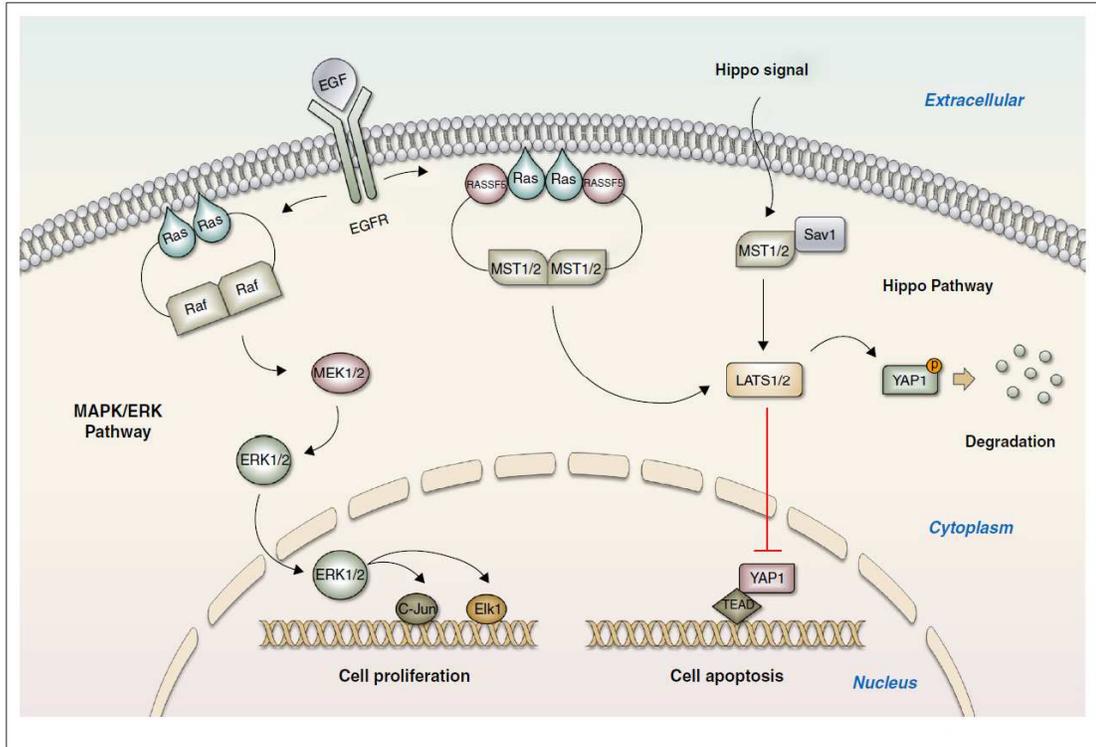
### 6.2.3 The similar roles of MAPK and Hippo signaling in the cell cycle

Like the MAPK, the Hippo pathway [122-124] is a kinase cascade (Fig. 6-4). The complex of MST1/2 kinases and SARAH domains-containing protein Salvador homolog 1 (Sav1, also known as 45 kDa WW domain protein or WW45) phosphorylates and activates the large tumor suppressor 1/2 (LATS1/2) kinases. LATS1/2 activity is also regulated by the cytoskeleton. LATS1/2 phosphorylate YAP1 marking it for degradation. Unphosphorylated YAP1 binds its transcriptional co-activator with PDZ-binding motif (TAZ), translocating the complex to the nucleus to interact with transcription factors such as TEA domain family member 1-4 (TEAD1-4), which turns on expression of genes involved in proliferation and in suppression of apoptosis. MST1/2 and YAP/TAZ phosphorylation are modulated by phosphatases.

MAPK and PI3K pathways act at the G1 phase cell cycle restriction [128, 132, 168-171]. MAPK inhibition is often accompanied by overexpressed YAP1 [125, 126]. Recent observations provide compelling evidence pointing to overexpression of YAP1 in proliferating cells treated with MAPK inhibitors, and of proteins upregulating *MYC* (such as  $\beta$ -catenin [172-174], Notch [175, 176], Hedgehog [177-179], and eIF4E [180-183]) in cancer cells treated with PI3K inhibitors. These and additional data [128, 129] led us to suggest that oncogenic MAPK and PI3K – and YAP1 and c-Myc – play similar roles in cell cycle control in cell proliferation. Overexpression of YAP1 and c-Myc can rescue tumor cells in Ras drug resistance. MAPK/ERK and PI3K/Akt function consecutively in the G1 phase into the S cell cycle restriction point similar to YAP1

and c-Myc, which explains why YAP1 overexpression can rescue MAPK/ERK inhibition and  $\beta$ -catenin, and c-Myc overexpression can rescue PI3K/Akt. The correspondence of the two core pathways clarifies why when combined they can result in more aggressive tumors. Such corresponding pathways can be prophylactically targeted to hinder the emergence of drug resistance [128].

RASSF5 is an important Ras effector because it links the two core pathways (Fig. 6-4). Its action *in vivo* opposes that of Raf which binds at the same Ras site (Fig. 6-2). Raf's activation promotes MAPK signaling thus cell proliferation; by contrast RASSF5 activation and consequently MST1/2 results in Hippo's phosphorylation of YAP1 thus opposing proliferation. To be activated, both Raf and RASSF5 require Ras dimerization (or nanoclustering). Raf's affinity to Ras is in the nanomolar range, and its association with the membrane is further enhanced by the CRD (cystein rich domain); RASSF5 has the C1 domain, fulfilling an analogous CRD membrane-attaching role. Given this tug-of-war competition between Raf and RASSF5, for RASSF5 to suppress cell proliferation its expression level – as well as that of MST1/2 – should be high.



**Figure 6-4.** The two independent core signaling pathways in cancer, MAPK and Hippo, are connected by RASSF5. The two Ras effectors, Raf and RASSF5, are competitive because the RBD of Raf and the RA of RASSF bind to Ras at the same interface. They undergo the activation of the kinase domain of Raf and MST1/2 in a similar way. However, active Raf induces the cell proliferation; by contrast, active MST1/2 leads to cell apoptosis. The point advocated by this review is that the opposite functionalities are fulfilled by similar mechanisms.

### 6.3 Discussion

Ras activates RASSF5 and Raf kinase. Raf activates the MAPK pathway; RASSF5 provides an allosteric docking platform to activate the MST1/2 kinase that activates the Hippo pathway. Raf has a Ras binding domain, a long and highly flexible linker and a kinase domain. Two Raf molecules can attach to two adjacent molecules of Ras through their RBD domains, and their kinase domains can form a homodimer and transphosphorylate each other to transmit the epidermal growth factor receptor (EGFR)-Ras signal down the MAPK pathway. This action cannot be mimicked by MST1/2 which lacks a Ras binding domain; but RASSF5 does and can step in to fulfill this role [184]. Its SARAH domain, connected by a flexible hinge to the RA domain lends the flexibility for the two MST1/2 kinase domains to homodimerize and autophosphorylate, much like Raf. Mechanistically, Raf is equivalent to RASSF5 plus MST1/2. For both, membrane anchoring is essential for activation. The functional outcome differs: Raf promotes proliferation whereas RASSF5 suppresses it. Evolution may have selected this RASSF5+MST1/2 scenario rather than attaching an RA domain to MST1/2 to allow the MST1/2 SARAH domain - Sav1 interaction.

Like Raf, RASSF5 is activated by an incoming EGFR signal. Since both Raf and RASSF5 bind at the same Ras surface, a sufficiently high expression of RASSF5 can provide another cellular control against proliferative growth.

Ras cancer biology is challenging to understand. We believe that a structural grasp together with insight into the cellular pathways [185] may go a long way toward cracking its code and solving its mysteries.

## Chapter 7: The dynamic mechanism of RASSF5 and MST kinase activations by KRas4B

Chapter 7 was adapted from: TJ. Liao, H. Jang, CJ. Tsai, D. Fushman, R. Nussinov: *The dynamics mechanism of RASSF5 and MST kinase activation by Ras*. Physical Chemistry Chemical Physics, 2017, 19, 6470-6480.

### 7.1 Introduction to chapter 7

The classical Ras association domain family (RASSF) proteins such as RASSF1A and RASSF5 (also known as NORE1A) are tumor suppressors, promoting cell apoptosis [151, 186]. RASSF5 activates mammalian sterile 20-like kinase 1/2 (MST1/2) in the Hippo pathway [140, 157-159, 165, 187, 188]. Hippo's signaling stimulates phosphorylation and thereby activation of a core kinase cascade including MST1/2 and large tumor suppressor 1/2 (LATS1/2), leading to phosphorylation of Yes associated protein 1 (YAP1) [122, 189]. YAP1's phosphorylation encodes its degradation, thus abolishing its transcriptional activity [190]. Overexpression amplifies oncogenic signaling through YAP1's association with the TEA domain (TEAD) family of transcription factors [126, 191-193]. RASSF5 links Ras and the Hippo pathway [39]. Ras activates Raf kinase, thus the mitogen-activated protein kinases (MAPK) pathway (Ras/Raf/MEK/ERK). Hippo and MAPK are independent core pathways with similar actions; drug resistance mutations in the MAPK pathway and inactivating mutations in the Hippo pathway lead to similar consequences in tumor cell proliferation [192, 193]. *In vitro*, in the absence of cell membrane, RASSF5 promotes cancer; *in vivo*, it acts as

a tumor suppressor. RASSF5 can be considered as an adaptor protein, connecting Ras to the MST1/2 kinase through a conformational change [39].

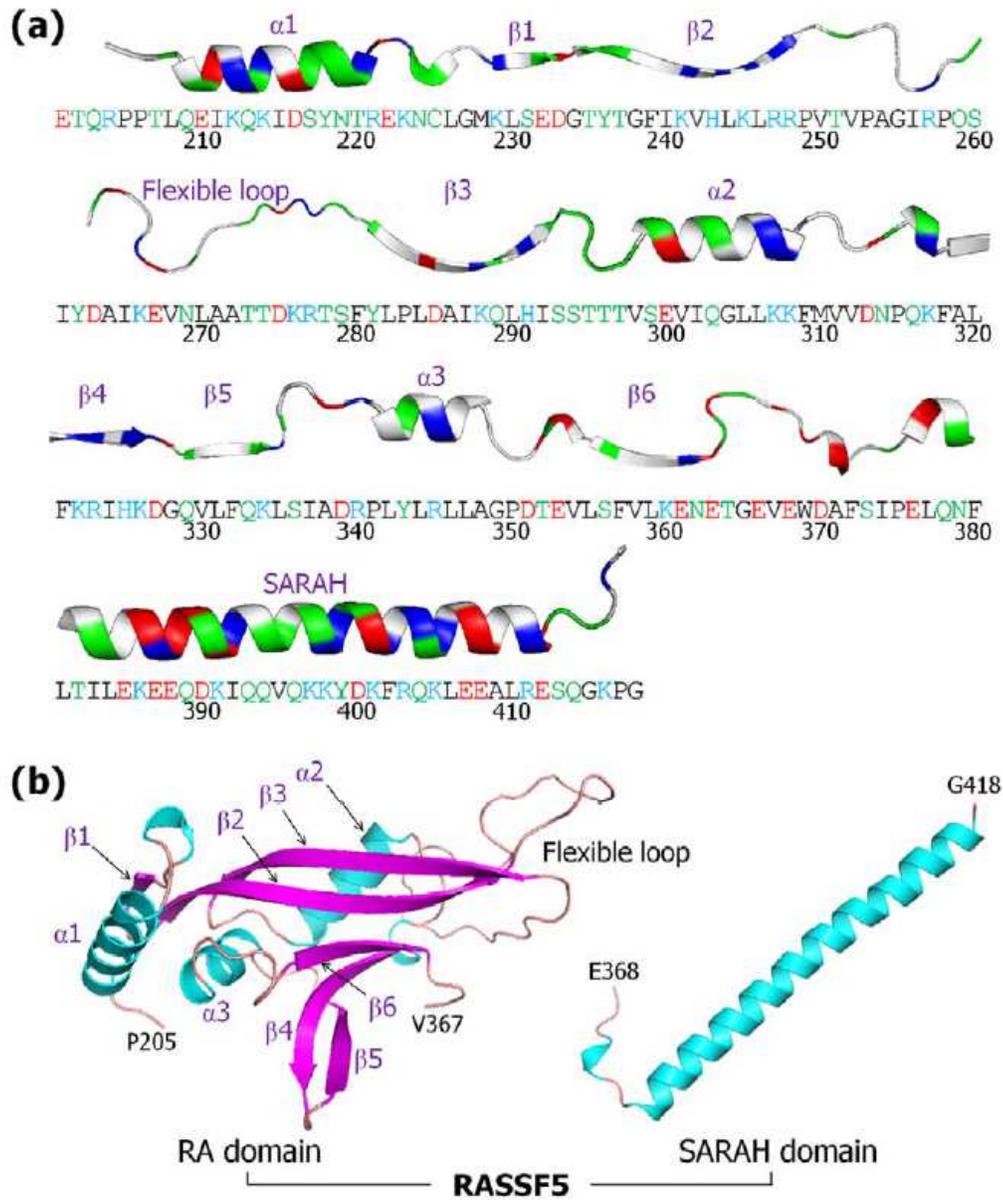
RASSF5 interacts with MST1/2 through heterodimerization of the C-terminal SARAH (Sav-RASSF-Hippo) domains [146, 148]. The RASSF5-MST1/2 SARAH heterodimer promotes homodimerization of the N-terminal kinase domain of MST1/2, followed by trans-autophosphorylation. The C-terminal region of RASSF5 contains the Ras association (RA) and SARAH domains (Fig. 7-1). The structure of the N-terminal region, including the putative membrane binding C1 domain, is currently unavailable. MST1/2 also contains the SARAH domain, structurally similar to that of RASSF5. RASSF5 connects Ras to MST1/2; its RA domain binds the Ras catalytic domain and the SARAH domain forms an antiparallel coiled coil with MST1/2 SARAH. The coiled coil conformation illustrates that SARAH is a long stretched  $\alpha$ -helix. The coiled coil can be also observed in the crystal structures of the RASSF5-RASSF5 and MST-MST SARAH homodimers. The affinity of the RASSF5-MST SARAH heterodimer is higher than those of either homodimers, RASSF5-RASSF5 or MST-MST [141].

RASSF5 is of particular importance since it links the MAPK and the Hippo pathways, two core pathways in tumor cell proliferation [39]. Drug resistant mutations in Ras or MAPK proteins are often accompanied by mutations in the Hippo pathway or YAP1. Ras is still undruggable, currently with no drug in the clinic. There is also a lack of structural information about RASSF5 functional states and thus the mechanism relating to how exactly Ras activates RASSF5 and how RASSF5 activation stimulates activation of the MST1/2 kinase whose action stimulates the Hippo pathway and thus YAP1 degradation. Currently, crystal structures of the functional regions of RASSF5

are only available for the unlinked SARAH and RA domains due to high fluctuations of the domains connected by a flexible linker. The crystal structures of SARAH RASSF5-MST2 heterodimer and the RASSF5-RASSF5 and MST2-MST2 homodimers exhibit the coiled coil motif. As to the RA domain, the available crystal structure is of murine RASSF5 in complex with a GTP analog-bound H-Ras. Thus, we conclude that in the active state, RASSF5 promotes the SARAH and RA domains interaction with their binding partners, Ras and MST – rather than with each other. Data are unavailable for the inactive state; however, we reasoned that RASSF5 retracts both domains, without Ras and MST2, yielding a self-associated conformation. That inactive ‘closed’ RASSF5 conformation can be regarded as the autoinhibited state. The lack of crystal structure data for such closed autoinhibited conformation suggests that the RA-SARAH inter-domain interaction is transient, not sufficiently stable to permit its crystallization; however, we can expect it to be present in the dynamic conformational ensembles, in which case it could be captured by sampling the broadly dispersed structures populating the free energy landscape of the inactive states. Ras can activate RASSF5 and by crossing over the free energy barrier from the closed toward the open, active state, shift the equilibrium to liberate the SARAH domain.

To corroborate RASSF5 structure in the inactive state, we modeled the self-associated RASSF5 conformation with SARAH covalently connected to the RA domain. Since no crystal structure of the combined form is available, we exploited docking programs to predict possible modes of the interaction between SARAH and the RA domain. In the initial prediction, both long straight  $\alpha$ -helical SARAH extracted from the crystal coiled coil, and kinked SARAH sampled from replica-exchange

molecular dynamics (REMD) simulations were used. The prediction programs generated multiple decoys of self-associated RASSF5, and a screening process based on the energy score was performed in order to sample the best initial configurations for standard atomistic molecular dynamics (MD) simulations in solution. Our results verify that in the inactive state RASSF5 persists in the self-associated conformation, in which SARAH contacts the RA domain through various interfaces. However, when those same RASSF5 structures interact with GTP-bound K-Ras4B, SARAH tends to disengage from the RA domain, demolishing the weakly self-associated conformation. RASSF5 activation by GTP-bound Ras points to a shift of the RASSF5 ensemble from the inactive to the active state toward the strongly associated SARAH heterodimer. Variants of the RASSF5 conformation suggest that the protein acts as an adaptor between Ras and MST1/2. Scaffolding proteins typically control regulation dynamically [184]. Our simulated RASSF5 structures in atomic detail reveal how the protein accomplishes the conformational adjustments which are required for MST activation and Hippo pathway signaling, and offer the rational and target for Ras-driven cancer.



**Figure 7-1.** RASSF5 sequence and structure. Human RASSF5 contains 418 residues involving the RA domain (residues 274-364) and SARAH (residues 366-413). (A) The sequence of N-terminal truncated RASSF5 protein with each domain structure embedded in a cartoon representation. In the sequence, hydrophobic/glycine, polar, positively charged, and negatively charged residues are colored black, green, blue, and red, respectively. Similarly, in the ribbon representation for the secondary structures,

the same colors were used, except for the hydrophobic/glycine residues which are colored white. (B) Crystal structures of murine RA domain (PDB code: 3DDC) and human SARAH domain (PDB code: 4LGD) as a long straight  $\alpha$ -helix defined in a coiled coil with MST2 SARAH domain. In the RA domain, the missing flexible loop (residues 254-278) was constructed by using the Modeller server.

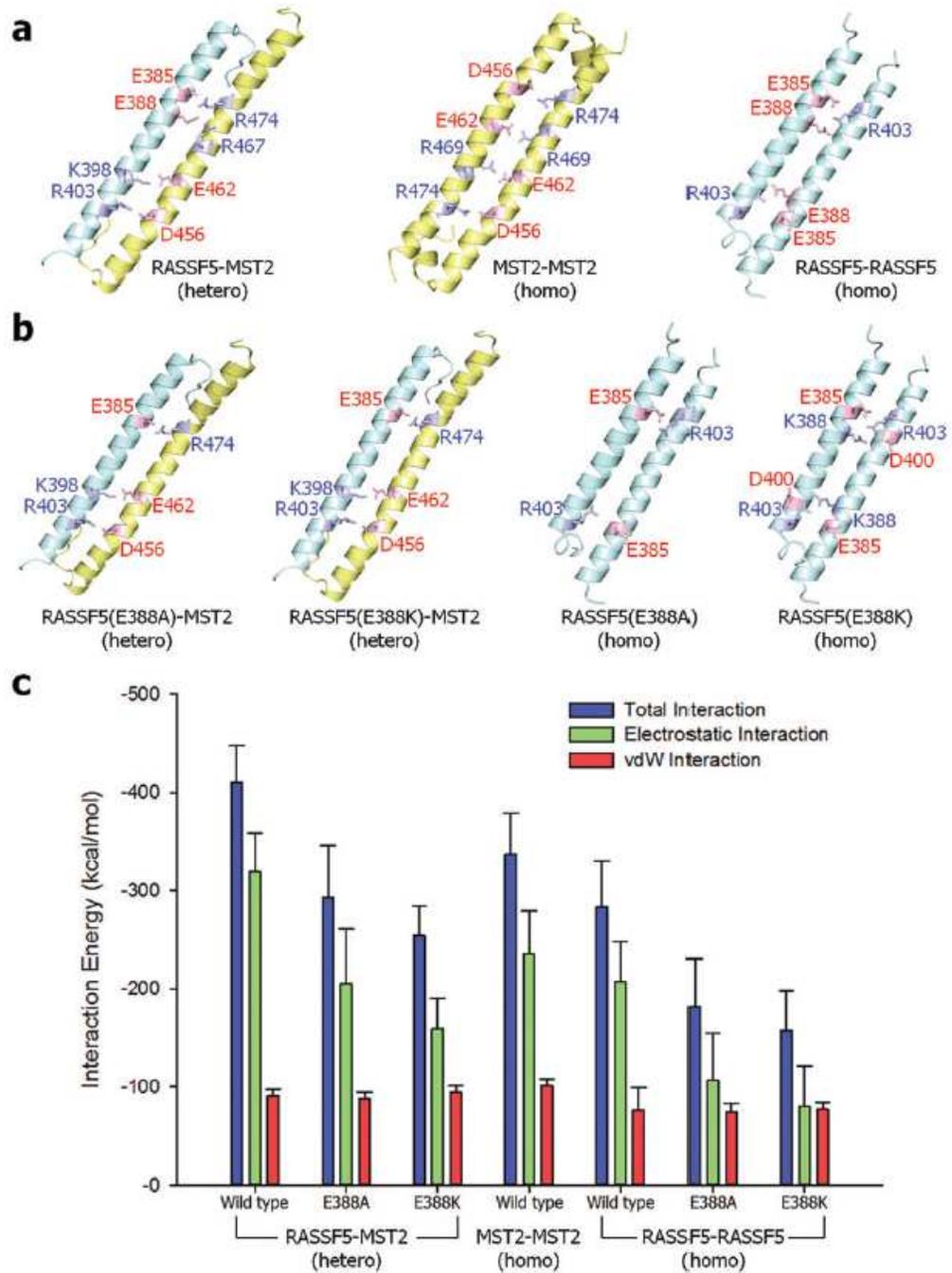
## 7.2 Results

### 7.2.1 Strong interaction of the RASSF5-MST2 SARAH heterodimer compared to the SARAH homodimers

The SARAH domain plays a key role in RASSF5 and MST2 association. The formation of RASSF5-MST2 SARAH heterodimer facilitates the MST2 kinase domain dimerization and trans-autophosphorylation, leading to cell apoptosis. The SARAH-SARAH interaction involves an antiparallel coiled coil formation. Both the RASSF5 and MST2 SARAHs can form heterodimeric and homodimeric coiled coils, but the heterodimeric coiled coil has a stronger dimeric interface [141]. The RASSF5 SARAH domain is very similar to the MST2 SARAH; both are long straight  $\alpha$ -helices of ~50 residues in a coiled coil organization (Fig. 7-2A, B). To decipher the relative interface interaction strength for different SARAH dimers, we simulated three SARAH dimers observed in crystals: RASSF5-MST2 SARAH heterodimer (PDB code: 4LGD), MST2-MST2 SARAH homodimer (PDB code: 4OH9), and RASSF5-RASSF5 SARAH homodimer (PDB code: 2YMY). As expected, we observed that RASSF5 SARAH strongly interacts with MST2 SARAH, while its interaction with the same RASSF5 SARAH is weaker (Fig. 7-2C). Similarly, the MST2 SARAH interaction with

the same SARAH is weaker than the heterodimeric SARAH interaction, but it is relatively stronger than the homodimeric RASSF5 SARAH interaction. The strong heterodimeric SARAH interaction is due to strong electrostatic contribution, since RASSF5 SARAH is acidic with the isoelectric point,  $pI < 7$ , while MST2 SARAH is basic with  $pI > 7$ . These bipolar characteristics favor heterodimer formation with strong electrostatic attraction. To corroborate the SARAH dimer interface, we examined key residues involved in dimer formation. For the RASSF5-MST2 SARAH heterodimer, salt bridge interactions between residues E385-R474, E388-R467, E388-R474, K398-E462, and R403-D456 (former and latter residues denote RASSF5 and MST2, respectively) strongly retain the dimer interface. Similarly, the salt bridge interactions between the residues, D456-R474, E462-R469, R469-E462, and R474-D456 for the MST2-MST2 SARAH homodimer and E385-R403, E388-R403, R403-E388, and R403-E385 for the RASSF5-RASSF5 SARAH homodimer hold the dimer interface. In the salt bridge interactions, we found that residues E385, E388, and R403 in the RASSF5 SARAH domain, and R474 and D456 in the MST2 SARAH domain are frequently involved in the dimer interface. In contrast, for both MST2-MST2 and RASSF5-RASSF5 SARAH homodimers, relatively strong hydrophobic and hydrophilic interactions can be observed (Table E1), although the contributions from these interactions to the total interaction energy are relatively weak. The E388A and E388K mutations in RASSF5 weaken both the heterodimeric and homodimeric SARAH interactions (Fig. 7-2B, C). The E388A mutation reduces the salt bridge interactions, and the E388K mutation introduces unfavorable electrostatic repulsion with its binding partners (Table E1). Of particular note, based on the data from the

catalogue of somatic mutations in cancer (COSMIC), the mutation E388K in RASSF5 can cause cancer.

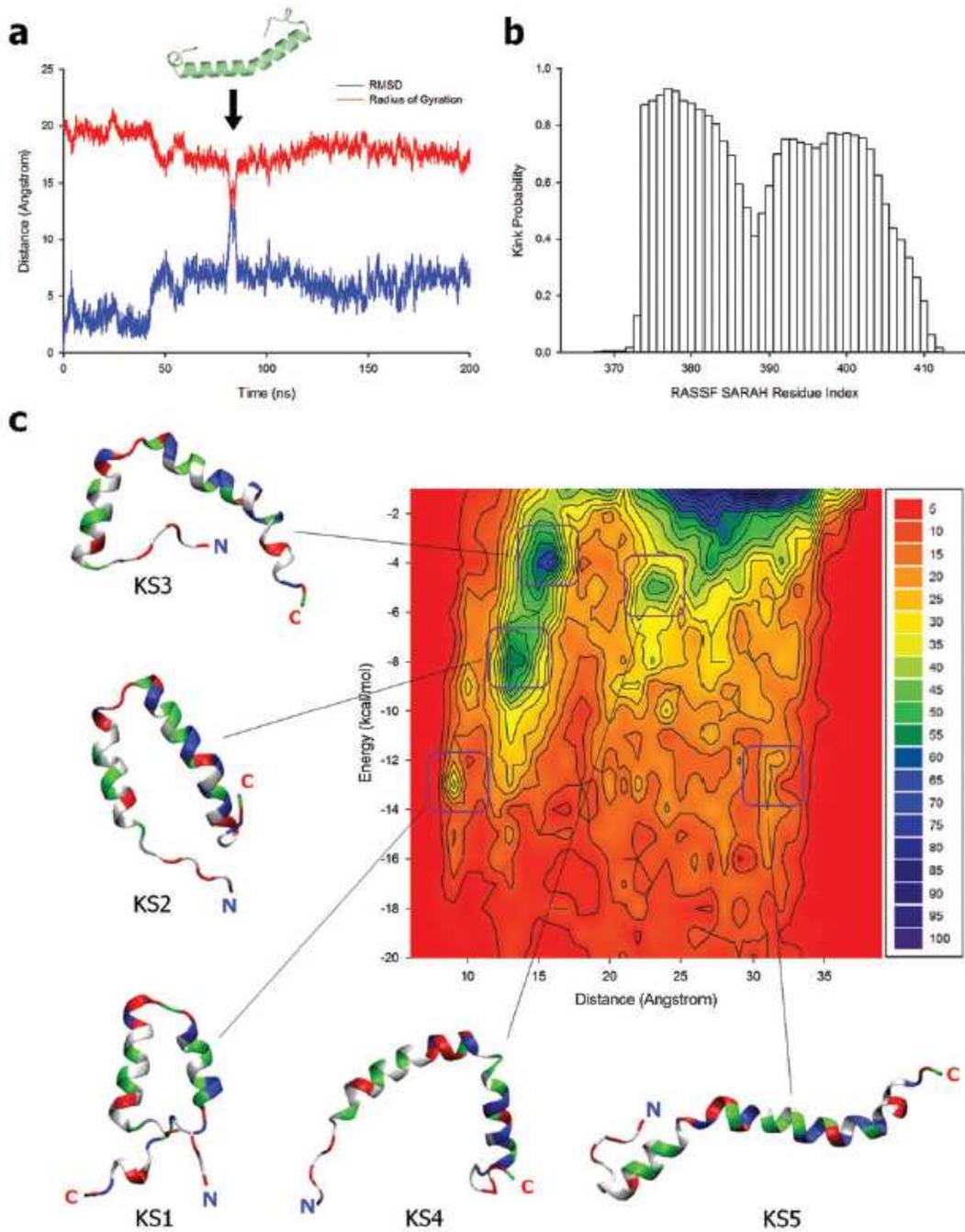


**Figure 7-2.** The SARAH-SARAH interaction in the coiled coil dimer. (A) Averaged structures of RASSF5-MST2 SARAH heterodimer (left), MST2-MST2 (center) and RASSF5-RASSF5 (right) homodimers. (B) Averaged structures of RASSF5-MST2 SARAH heterodimers with the E388A (far left) and E388K (second left) mutations in RASSF5, and RASSF5-RASSF5 SARAH homodimers with the E388A (second right) and E388K (far right) mutations in RASSF5. In the ribbon representation, RASSF5 and MST2 are colored cyan and yellow, respectively. The salt bridge pairs are shown in the average structures of each SARAH-SARAH dimer. (C) Interaction energy gauging the SARAH-SARAH association in the coiled coil dimer. Averaged total interaction energy (blue bars), and the contributions from the electrostatic (green bars) and vdW (red bars) interactions for three different SARAH dimers are shown.

#### 7.2.2 A kinked helical motif predominantly observed in the RASSF5 SARAH monomer

In the antiparallel coiled coil motif, SARAH exists as a long straight  $\alpha$ -helix. When dissociated from the coiled coil, monomeric SARAH does not persist in a straight  $\alpha$ -helical motif. We observed incidents of a kinked helical motif of monomeric RASSF5 SARAH emerging during the simulation (Fig. 7-3A). To verify the kinked SARAH structure as a representative of the monomeric conformational ensemble and its relative population, we employed HingeProt [194], an algorithm for protein hinge prediction using elastic network models, to predict the position of the kink. The results showed that a kink takes place at Q389. To obtain ensembles of monomeric SARAH conformations, we performed replica-exchange molecular dynamics (REMD) simulations of the RASSF5 SARAH monomer. During the simulations, we observed

that about 90% of the SARAH domain structures contain zero and one kink (Table E2). The location of the kink can be identified by analyzing the secondary structure, since residues at the kink have a low  $\alpha$ -helical probability. Using STRIDE [195], we found that the kink occurs at the region involving residues 387-389 (Fig. 7-3B), which is consistent with the predicted location at Q389 by HingeProt [194]. We sampled SARAH conformations over trajectories generated from the REMD simulations and sorted SARAH into several groups with similar conformations based on population maps for different SARAH topologies (Fig. 7-3C). To classify the conformations, we monitored the occurrence frequency of the interaction energy of two  $\alpha$ -helical segments, separated at the hinge point, as a function of the distance between two residues located at both ends of SARAH. Thus highly kinked SARAH has a shorter residue pair-distance and stronger interaction. The two-dimensional contour surface of the population map provides highly populated states enclosed by many contour lines. The highly populated state located at the upper right on the map represents the monotonic conformation of straight SARAH, while widely dispersed kinked SARAH states indicate various SARAH conformations. From the contour surface, we extracted five different SARAH conformations with a single kink, from a highly kinked SARAH 1 (KS1) to less kinked SARAH 5 (KS5) (Fig. 7-3C). The sampled kinked SARAH conformations, as well as the straight SARAHs, were used in the construction of the self-associated RASSF5 model.



**Figure 7-3.** The analysis of RASSF5 SARAH monomer. (A) Time series of the  $C\alpha$  atoms root-mean-squared deviation (RMSD) from the starting point and the radius of gyration of SARAH monomer in the standard MD simulation. The SARAH structure in the graph reflects the kinked motif occurred at the event indicated by arrow. (B)

Probability of the secondary structure analyzed by the STRIDE [195] along the SARAH residues for the ensembles of SARAH monomer generated from the REMD simulation. (C) Two-dimensional contour map representing the occurrence frequency of the residue pair-distance dependence of the interaction energy between two  $\alpha$ -helical segments in the kinked SARAH conformations sampled from the REMD simulation. Five selected kinked SARAH conformations from the population map. The SARAH configurations from KS1 with highly kinked to KS5 with less kinked, where KS denotes kinked SARAH, are marked on the map. In the ribbon representation of SARAH, hydrophobic/glycine, polar, positively charged, and negatively charged residues are colored white, green, blue, and red, respectively.

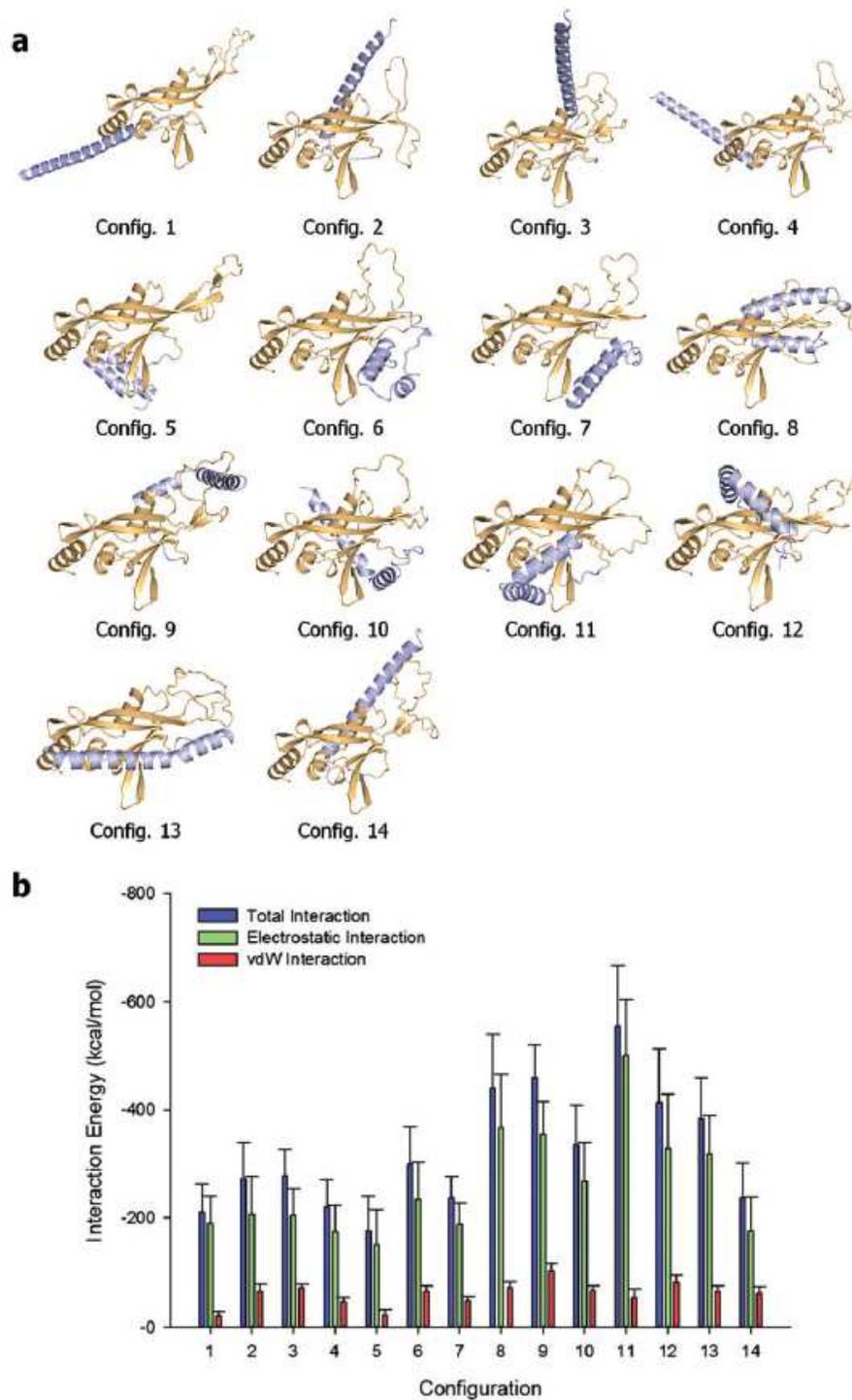
### 7.2.3 The self-associated RASSF5 conformation in the inactive state

We constructed the N-terminal truncated RASSF5 structure (residues 205-418) with SARAH covalently connected to the RA domain. Since the structure of the human RASSF5 RA domain is currently unavailable, we adopted the murine RASSF5 RA domain (PDB code: 3DDC) converting its sequence to human. The murine and human RA have 96% sequence identity. The disordered region with missing coordinates for the residues 254-278 was constructed as a flexible loop by using the Modeller server [196]. In the covalently connected SARAH with the RA domain, we considered two possible SARAH conformations; straight and kinked  $\alpha$ -helix. Since the linker between the SARAH and RA domains can serve as a hinge point, there are a limited number of modes of SARAH binding to the RA domain. To obtain the self-associated inactive RASSF5 conformation, we exploited a powerful template-based protein-protein complex structure prediction algorithm (PRISM) [197, 198]. By using a straight  $\alpha$ -

helical SARAH, defined from the crystal structure, we obtained four possible modes of the SARAH interaction with the RA domain (Fig. E1A). With the kinked SARAH models (KS1-5 from the REMD simulations in Fig. 7-3C), we were able to obtain many possible decoys representing the RA-SARAH association by using the Patchdock server.[199, 200] For each kinked SARAH, the docking program generated 12 decoys of the RASSF5 structure (Fig. E1B, C, D, E, F). Based on the energy scoring function, we selected the first two decoys for each kinked SARAH model, gathering 10 additional configurations. Thus, a total of 14 configurations representing the possible ensembles of self-associated RASSF5 were subjected to an all-atom MD simulation in aqueous environment. During the simulations, no immediate dissociation of SARAH from the RA domain was observed. Occasionally, we observed a structural convergence between the configurations. For example, configurations 6 and 7 present similar conformational ensembles, and the structure in configuration 2 resembles configuration 14 (Fig. 7-4A). To quantify the RA-SARAH interaction, we calculated the interaction energy of SARAH with the RA domain (Fig. 7-4B). Here, we found that the kinked SARAH interaction with the RA domain is relatively stronger than the straight SARAH interaction. The straight SARAH weakly interacts with the RA domain with fewer residues involved in the interface, while the kinked SARAH with conformational change folds onto the interface increasing the interaction surface. The weaker interaction of straight SARAH indicates that the RA domain does not have a comparable long helix. The long straight  $\alpha$ -helical SARAH favors a coiled coil.

To grade the self-associated RASSF5 conformations, we investigated the interacting residue types at the RA-SARAH interface. Unlike the SARAH residues involved in

coiled coil formation, we found different residues participating in the RA-SARAH association (Table E3). In particular, for all 14 configurations we observed that the SARAH residues, E366, D370, E376, and E387, and the RA residues, R323, K334, and E353, are commonly involved in strong salt bridge interactions. A number of residue pairs are involved in hydrophobic interactions, additionally supporting the self-associated conformation, but no common residue pairs were observed, nor were hydrophilic interacting residue pairs. However, in the RA domain, hydrophobic residues, P283 and I337, and the hydrophilic residues, Y281, H325, Q329, and Q333, significantly contribute to the interaction with SARAH. Hydrogen bonding (H-bond) residue pairs were rarely observed during the simulations. To delineate the RA domain residues interacting with SARAH, we calculated the backbone amide NMR chemical shift perturbations (CSPs) for the RA domain using the ShiftX2 server[201] (Fig. 7-5). These CSPs represent the difference in predicted chemical shifts for the trajectory of RA domain alone and as part of the various RASSF5 configurations shown in Figure 4b. For all configurations, strong CSPs are common in the region containing residues 254-278, which represent the flexible loop of the RA domain. In addition, we observed large CSPs for RA residues which are in contact with SARAH, consistent with the interaction pair analysis. We noticed that each configuration has distinct residue pair interactions as reflected in the inhomogeneous distributions of the strong CSPs among the configurations, suggesting that the SARAH interaction with the RA domain might be transient, and multiple modes of interaction are possible, in agreement with the absence of a crystal structure of the self-associated state.

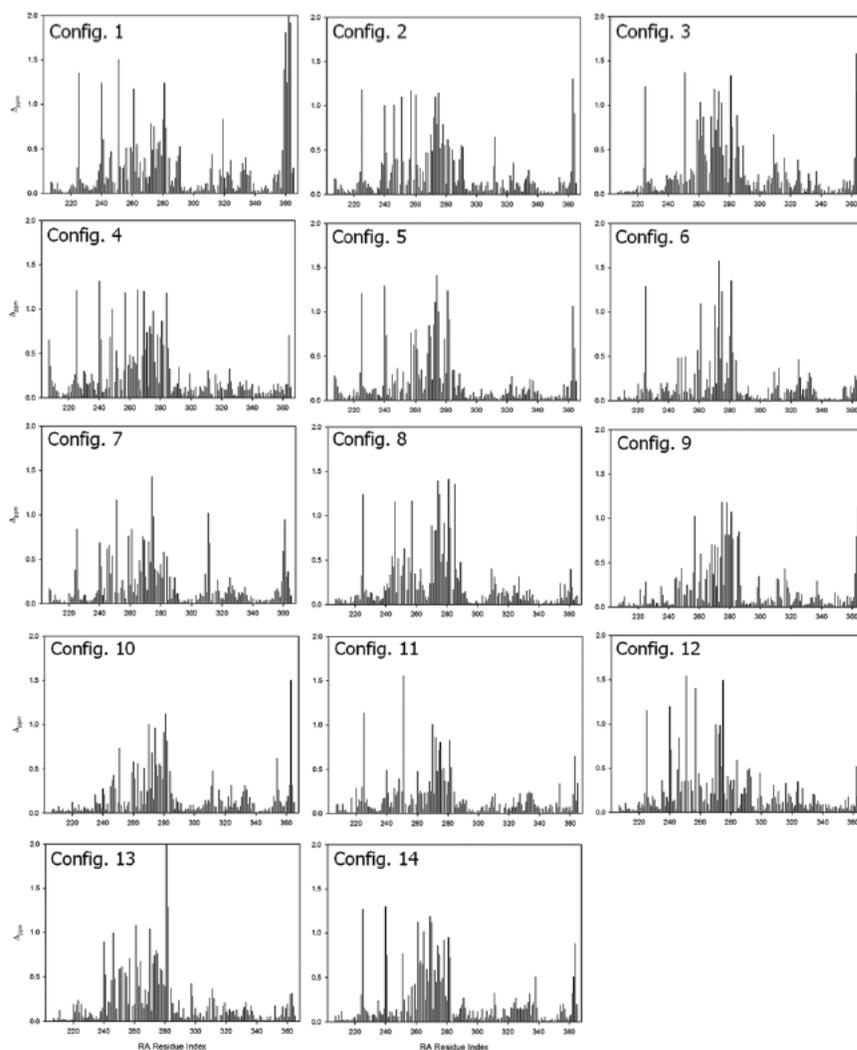


**Figure 7-4.** Self-associated RASSF5 conformation. (A) Averaged structures of RASSF5 after 200 ns molecular dynamics simulations. RASSF5 configurations 1 – 4 contain straight SARAH, and configurations 5 – 14 were modeled with kinked

SARAH. In the cartoon, the RA and SARAH domains are colored orange and light blue, respectively. (B) Interaction energy of SARAH with the RA domain for the RASSF5 configurations. Averaged total interaction energy (blue bars), and the contributions from the electrostatic (green bars) and vdW (red bars) interactions for the RASSF5 configurations are shown.

Despite the fact that the interaction of SARAH with the RA domain is transient, we attempted to capture the most reliable models of the self-associated inactive RASSF5. Based on the combined residue pair interaction and CSP analyses, we selected configurations 1, 4, 6, 9, 10, 11, and 14 for the representative RASSF5 conformation (Fig. 7-4A). These configurations ensured that the SARAH domain does not block the Ras binding site at the  $\beta$ 3 region. Without blocking the Ras binding site, we were able to construct initial models of RASSF5/KRas4B complex at a later stage. For the straight SARAH model, only configurations 1 and 4 were selected, since in configurations 2 and 3 the SARAH domain interferes with Ras binding. For the kinked SARAH model, configurations 8, 12, and 13 were not considered for the same reason. Configurations 5 and 7 were also omitted, since they have the weakest SARAH interaction among the kinked models (Fig. 7-4B). We noticed that SARAH dominantly interacts with the flexible loop (residues 254-278) in the RA domain, which may contribute to the RA-SARAH association. Among the representative RASSF5 models, configurations 6, 9, and 10 show strong interaction between the loop and SARAH (Fig. 7-6). In the configuration 6, the loop marginally holds the SARAH domain through the salt bridge interaction of R277 with D370 and the cation- $\pi$  interaction of Y262 with K416. However, RA residue Y281, near the loop, sequesters SARAH by forming H-bonds

with E408 and L410 (Table E3). In the RA-SARAH interaction, configuration 6 contains most of the common salt bridge residues, involving R323, K334, and E353 in the RA domain and E366, E368, E370, K386, and E387 in SARAH. Configuration 9 shows the most abundant interactions between the loop and SARAH, including the strong salt bridge interactions of E267-K391, E267-K398, K276-D400, and K276-E407, and the H-bond formation between T274 and R403 (Table E3). The hydrophobic and hydrophilic interactions also stabilize the loop interaction with SARAH. Similarly, configuration 10 contains abundant hydrophobic and hydrophilic interactions, as well as strong salt bridge interactions, K266-E407 and E267-R403, but no H-bond formation. In configurations 11 and 14, the loop interacting with SARAH only involves weak hydrophobic interactions. The residue pair interactions illustrate that the flexible loop commonly interacts with the C-terminal portion of SARAH, except configuration 11. We speculate that the loop may act as a hook that transiently holds SARAH onto the RA domain. Figure 6 helps deduce the self-associated inactive RASSF5 conformation. As we noted above, the autoinhibitory RA-SARAH interaction is transient, since SARAH has to be released when RASSF5 is activated by Ras, forming a coiled coil with another SARAH. Thus, the 'hooked' SARAH configurations can be candidates RASSF5 conformations in the inactive state.



**Figure 7-5.** Calculated amide chemical shift perturbations (CSPs) for RA residues in the RASSF5 configurations shown in Fig. 4. The ShiftX2 program[201] calculated  $^1\text{H}$ ,  $^{13}\text{C}$ , and  $^{15}\text{N}$  chemical shifts. The combined amide CSPs for the RASSF5 configurations with respect to the RA domain alone trajectory were calculated during the simulations using the equation  $\Delta_{\text{ppm}} = \sqrt{(\Delta\delta_{\text{HN}})^2 + (\Delta\delta_{\text{N}}\alpha_{\text{N}})^2}$ , where  $\Delta\delta_{\text{HN}}$  and  $\Delta\delta_{\text{N}}$  denote the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shift differences, respectively, between the RASSF5 and RA domain trajectories. The predicted chemical shifts were averaged over 1500 trajectory snapshots. A scaling factor  $\alpha_{\text{N}} = 0.17$  was applied to the  $^{15}\text{N}$  chemical shift difference.

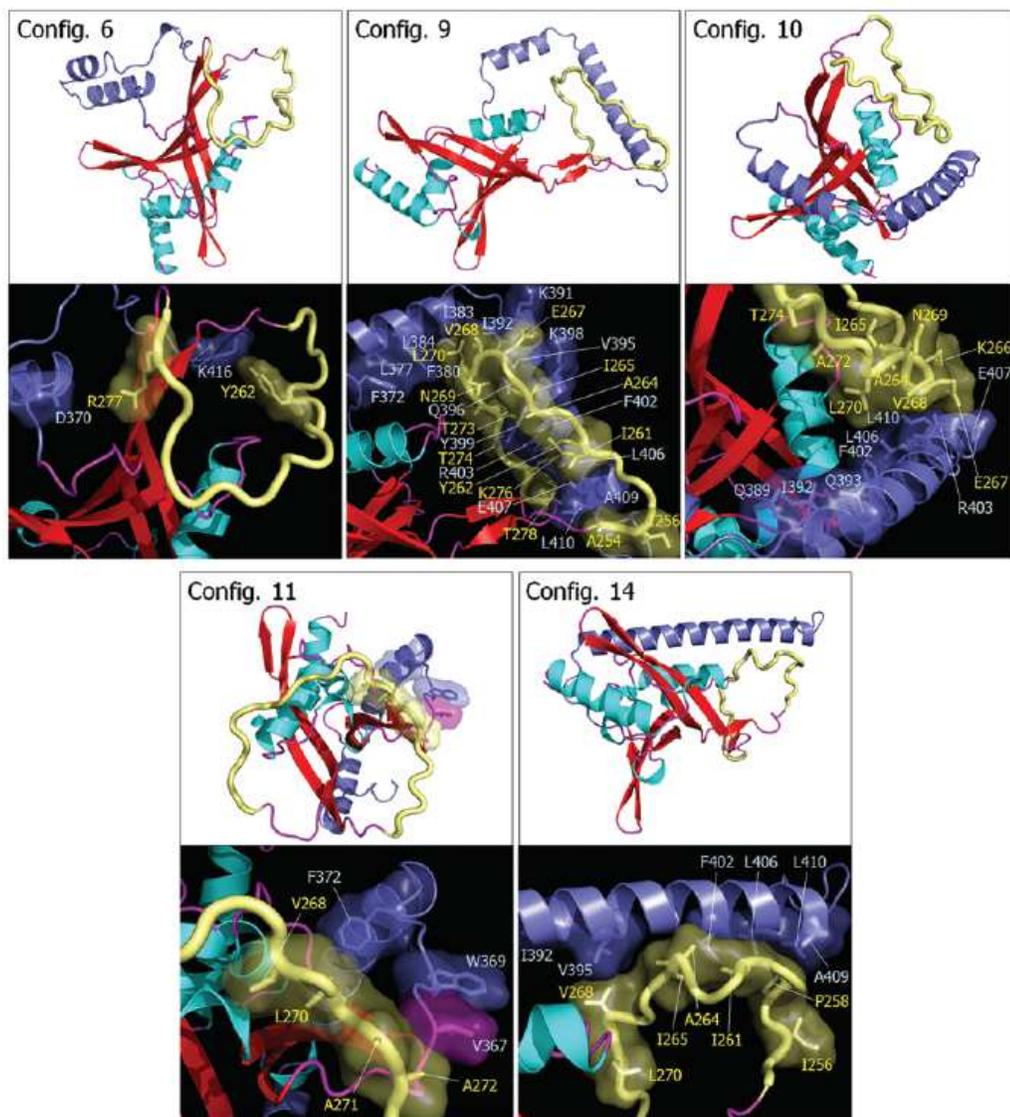
#### 7.2.4 RASSF5 in complex with GTP-bound KRas4B

Active Ras recruits RASSF5, resulting in a dimeric complex at the plasma membrane (PM). This conduces to the MST1/2 kinase being able to translocate to the PM [39]. RASSF5 connects between Ras and MST1/2. The RA domain of RASSF5 binds the catalytic domain of Ras, and at the same time the SARA domain forms an antiparallel coiled coil with the MST1/2 SARA. Ras dimerization would further facilitate the dimerization of MST1/2 kinase domains and the trans-autophosphorylation. Each Ras molecule requires an active RASSF5 to connect to MST1/2. In inactive RASSF5, SARA clings to the RA domain, biding its time until the RA domain is enlisted to reside on the Ras catalytic domain. SARA is released once RASSF5 binds and gets activated by Ras. To understand the mechanism of the conformational aptness of RASSF5, we simulated it in complex with the GTP-bound KRas4B. We adopted the crystal structure of murine RASSF5 RA domain in complex with HRas (PDB code: 3DDC) as a template for the MD simulations. In the initial construction, those self-associated inactive RASSF5 conformations (configurations 1, 4, 6, 9, 10, 11, and 14 in Fig. 4a) were used to generate dimeric complexes with K-Ras4B-GTP. This resulted in seven different dimeric configurations, DC1(1), DC2(4), DC3(6), DC4(9), DC5(10), DC6(11), and DC7(14), where DC represents dimeric configuration and the numbers in the parenthesis denote the self-associated RASSF5 configuration. Remarkably, we observed that the SARA domain tends to dissociate from the RA domain upon binding to KRas4B (Fig. 7-7A). The release generates straight SARA conformations, which suggests a clan of coiled coils. The interaction strength of SARA with the RA

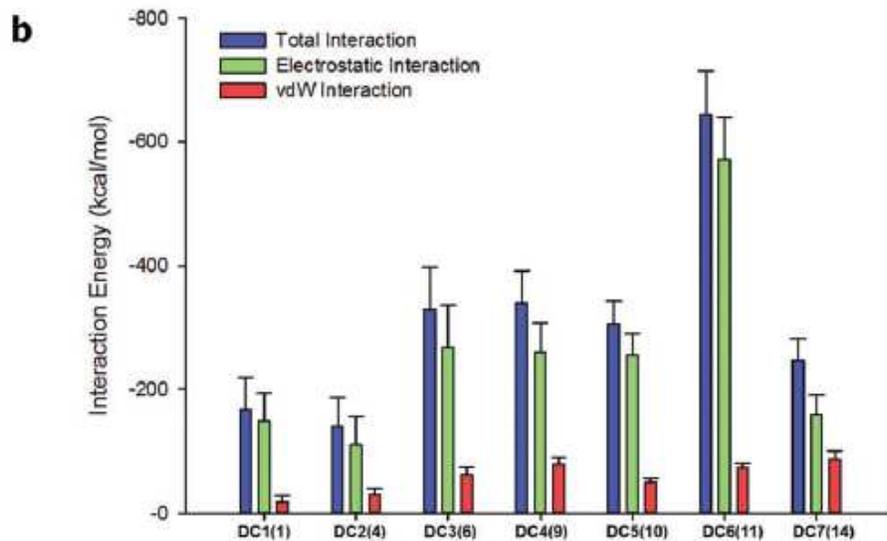
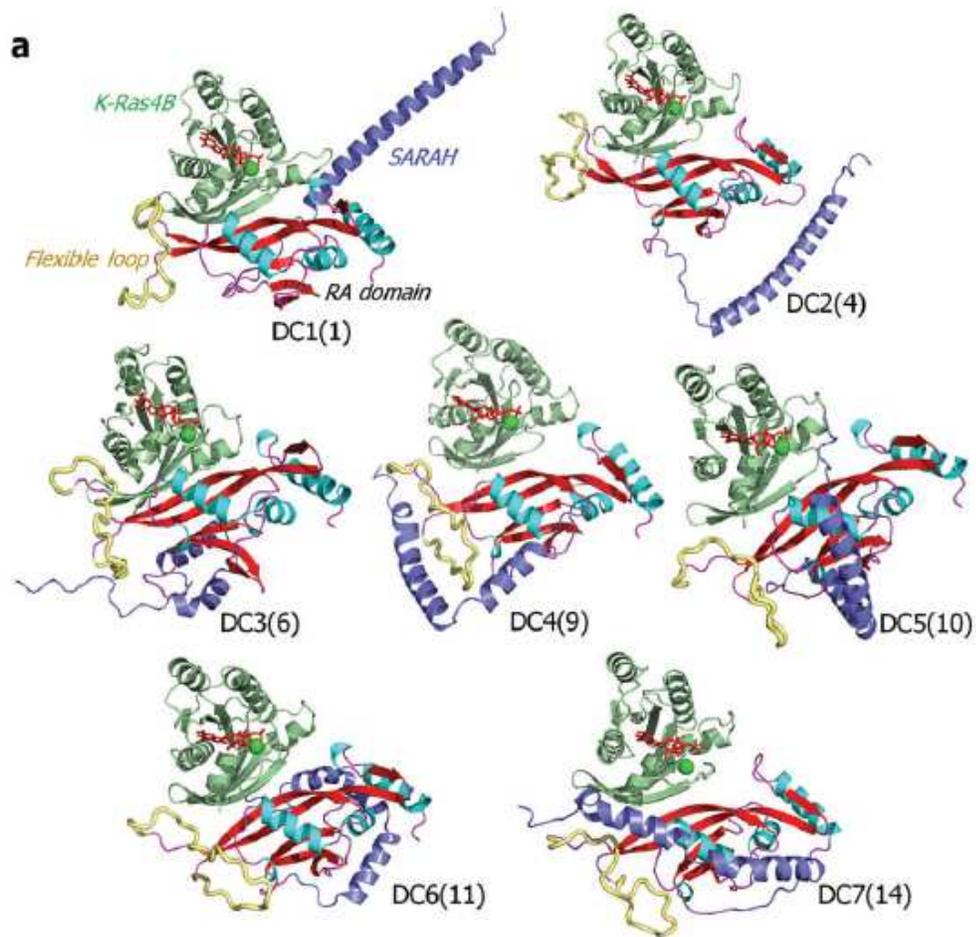
domain is significantly reduced as compared to that of the Ras-free RASSF5 configurations (Fig. 7-7B). In DC6(11), SARAH still holds the RA domain strongly even when bound with K-Ras4B, suggesting that the configuration may not have reached relaxation in the simulation. Nevertheless, SARAH loses a number of interacting residue pairs including DC6(11), as RA residues participate in the interaction with K-Ras4B (Table E4). The mechanism of SARAH release suggests that K-Ras4B promotes RASSF5 activation, allosterically shifting the protein from the fluctuating loosely associated inactive to the strongly bound active state.

To elucidate how K-Ras4B allosterically affects the RASSF5 conformation, we conducted a dynamical network analysis using the NetworkView plugin in VMD. The dynamical network analysis can identify the shortest signal propagation pathway through nodes, represented as residues in a protein, and edges connecting between nodes. In this analysis, weighted implementation of suboptimal paths (WISPs) [104] for the allosteric signal transmitted from KRas4B to RASSF5 were calculated based on the sum of the shortest edge distances defined by the pairwise dynamic cross-correlation. We calculated over 100 optimal and suboptimal pathways between two selected residues on K-Ras4B and RASSF5 SARAH, and provided two best pathways for the dimeric configurations with the kinked SARAH model (Fig. 7-8). It can be seen that most of the signaling pathways are connected through the effector binding region of Ras (residues 32-40 of the Ras catalytic domain, Table E5). Interestingly, some pathways pass through GTP, implying that only the GTP-bound state induces conformational changes in RASSF5, facilitating its activation. We noted that the GTP-bound, not the GDP-bound, Ras preferentially binds its effectors, such as Raf and PI3K.

The dynamical cross-correlation map (DCCM) reveals apparent residue motions across the K-Ras4B and RASSF5 domains in the highly-kinked SARAH dimeric complex model, with positive ( $C(i,j) \rightarrow 1$ ) and negative ( $C(i,j) \rightarrow -1$ ) correlations (Fig. E2). In contrast, the dimeric complex with the straight SARAH model yields uncorrelated residue motions with  $C(i,j) \rightarrow 0$ , suggesting that no allosteric signal is required for the unconstrained straight SARAH. The dimeric configurations can be ranked from more to less kinked SARAH in the order of DC3(6) > DC6(11) > DC4(9) > DC5(10) > DC7(14) > DC1(1) > DC2(4), which is consistent with the highly correlated residue motions.

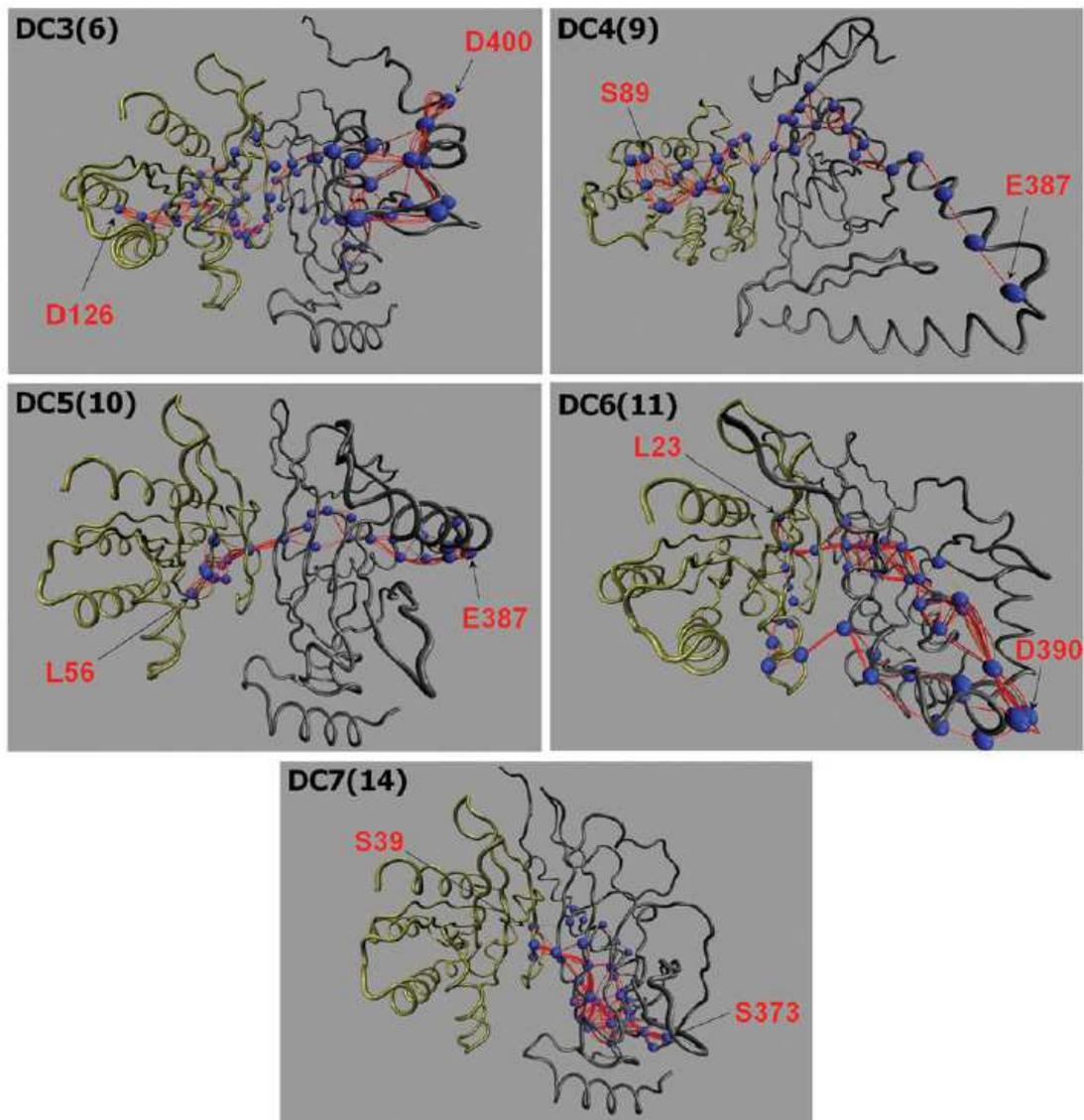


**Figure 7-6.** Snapshots representing average conformations of RASSF5. Selected RASSF5 configurations 6, 9, 10, 11, and 14 are shown on the top of each panel. In the cartoon representing the secondary structure, the  $\alpha$ -helix and  $\beta$ -sheet structures are colored cyan and red, respectively. The SARAH domain is highlighted by blue, and the flexible loop is shown as yellow tube. Highlighted interfaces between SARAH and flexible loop are shown in the bottom of each panel. Residues involving the interactions are marked.



**Figure 7-7.** RASSF5 in complex with the GTP-bound K-Ras4B. (A) Averaged structures of RASSF5 interacting with K-Ras4B-GTP after 200 ns molecular dynamics

simulations. Seven dimeric complexes, DC1(1), DC2(4), DC3(6), DC4(9), DC5(10), DC6(11), and DC7(14), where DC denotes dimeric configuration and the numbers in the parenthesis correspond to the self-associated RASSF5 configuration, were selected for the simulations. DC1(1) and DC2(4) contain straight SARA, and other DCs have the kinked SARA model. In the cartoon representing the secondary structure, the  $\alpha$ -helix and  $\beta$ -sheet structures are colored cyan and red, respectively. The SARA domain is highlighted by blue, and the flexible loop is shown as yellow tube. (B) Interaction energy of SARA with the RA domain for the dimeric configurations of RASSF5/K-Ras4B-GTP complex. Averaged total interaction energy (blue bars), and the contributions from the electrostatic (green bars) and vdW (red bars) interactions for the dimeric configurations are shown.



**Figure. 7-8.** The allosteric pathways induced by the GTP-bound KRas4B. The allosteric pathways were calculated by the weighted implementation of suboptimal path (WISP) [104] methods for the dimeric configurations, DC3(6), DC4(9), DC5(10), DC6(11), and DC7(14) with kinked SARAH. Blue beads on the paths highlight the residues with high occurrence rate, >50% (Table E5). The source residue on KRas4B-GTP (left) and sink residue on RASSF5 (right) are marked.

### 7.3 Discussion

RASSF5 is a mediator that links Ras and MST1/2 [187]. Its SARAH domain binds its partners, including the RA and other SARAH domains with varied affinities, thereby playing a critical role in RASSF5's functions [141, 188]. RASSF5's SARAH can interact with the RA domain resulting in autoinhibition and also associate with a cognate SARAH [39]. The SARAH-SARAH interaction forms an antiparallel coiled coil, since SARAH adopts a long  $\alpha$ -helical motif [202]. The RASSF5-MST2 coiled coil SARAH heterodimer has the highest binding affinity, compared to both MST-MST and RASSF5-RASSF5 SARAH homodimers [141]. Thus, RASSF5 SARAH can easily compete with the MST-MST SARAH homodimer. Our MD simulations also verified the strong interaction of RASSF5 SARAH with MST2 SARAH, compared to the homo-dimeric SARAH-SARAH interactions. We suspect that the high binding affinity of the SARAH heterodimer is due to the strong electrostatic attraction originating from the bipolar characteristics between RASSF5 and MST2 SARAHs. The isoelectric point calculations of the SARAH domains illustrated that RASSF5 SARAH is acidic, while MST2 SARAH is basic. In the SARAH-SARAH interactions, we observed that RASSF5 SARAH residues, E385, E388, and R403, are commonly involved in both the heterodimer and homodimer. Of particular note is E388, the residue at the kink in the  $\alpha$ -helix when dissociated from the coiled coil. COSMIC data showed that the E388K mutation in RASSF5 can cause cancer. We speculate that E388K can disrupt the RASSF5-MST2 SARAH heterodimer, since the opposite basic residue can introduce unfavorable electrostatic repulsion in the SARAH-SARAH interaction. Without

activation of MST1/2, Hippo signaling can be suppressed, and as a result, the oncogenic YAP1 signal is amplified.

It is clear that RASSF5 assists in MST1/2 kinase activation [39, 140, 149]. In the inactive state, MST1/2 SARAH blocks its kinase domain, preventing dimerization of the kinase domain with other MST1/2. For MST1/2, kinase domain dimerization is required to conduct trans-autophosphorylation, leading to phosphorylation of YAP1. RASSF5 enables kinase domain dimerization, since RASSF5 SARAH can sequester MST1/2 SARAH, thereby shifting the equilibrium to an exposed kinase domain. The dynamic mechanism of the MST1/2 SARAH can be also applied to RASSF5 SARAH.[39] In the inactive state, similar to MST1/2, the RASSF5 SARAH engages the RA domain, resulting in autoinhibition of RASSF5. The engagement of RASSF5 SARAH is dislodged when Ras interacts with the RA domain. To elucidate the mechanism, we modeled the inactive self-associated RASSF5 conformation in atomic detail. We obtained several possible RASSF5 structures with multiple modes of interaction of SARAH with the RA domain, suggesting that the inactive RASSF5 conformation is transient. In the RASSF5 structures, the SARAH domain was dominantly defined as a kinked  $\alpha$ -helix. It is reasonable that the long  $\alpha$ -helical motif of SARAH can be stabilized in a coiled coil with another SARAH, and that the RA domain does not contain a compatible  $\alpha$ -helix to form the coiled coil. Our modeling suggests that the RASSF5 RA domain marginally holds the SARAH domain through the salt bridge interactions. Especially, the interactions of the flexible loop in the RA domain with the C-terminal portion of the SARAH domain constrain the fluctuations of

SARAH's C-terminus. The loop acts a hook that sequesters SARAH to the RA domain.

The autoinhibited RASSF5 becomes active when Ras binds to the RA domain. Ras induces a conformational change in RASSF5, liberating SARAH from the RA domain. To capture the event, we simulated RASSF5 interacting with the GTP-bound K-Ras4B and monitored allosteric pathways propagating from KRas4B-GTP to the RASSF5 SARAH domain. With KRas4B-GTP, we observed that RASSF5 generally released the SARAH, removing the autoinhibition. The interaction strength between SARAH and the RA domain is dramatically reduced, compared to that of RASSF5 without KRas4B-GTP. We reasoned that KRas4B-GTP transmits the allosteric signal releasing SARAH. This shifts the energy landscape of RASSF5 from the inactive to the active state. For KRas4B-GTP/RASSF5 complex with kinked SARAH, the DCCM calculations showed that the residue motions between KRas4B and RASSF5 are highly correlated. The allosteric signals are promoted over long range, stretching from the KRas4B allosteric lobe to the SARAH domain. In contrast, for the complex with straight SARAH, the DCCM shows less correlated residue motions, with weak allosteric signals transmitted over the short distance from the KRas4B effector lobe. This suggests that in the case of straight SARAH domain RASSF5 is more likely to be found in or near the active state.

To conclude, here we aim to figure out the mechanism on the conformational level of how Ras binding results in MST1/2 activation, making RASSF5 a tumor suppressor. Ras activates RASSF5, and the activated RASSF5 activates MST1/2. The mechanism of RASSF5 activation deploys the conformational dynamics of the SARAH domain,

switching it between states. SARAH's dynamics characterizes its conformational plasticity: in the active state of RASSF5, SARAH is an elongated  $\alpha$ -helix stretching to form a coiled coil with MST1/2 SARAH; in the inactive state, SARAH becomes flexible converting to a kinked  $\alpha$ -helix and receding toward the RA domain. The closed, self-associated RASSF5 conformation with the retracted SARAH implies an autoinhibition of the protein; Ras can however easily compete with SARAH, allosterically altering the autoinhibited conformation. Our structural modeling of RASSF5 in atomic detail provides insight into the mechanism of how the structural dynamics of RASSF5 with SARAH can be related to its functional role - connecting oncogenic Ras proteins to cancer, here by suppressing the Hippo pathway. Noteworthy, like other Ras binding proteins, such as Raf, PI3K, and RalGDS, RASSF5 contains a largely undefined N-terminal portion, including the putative membrane binding C1 domain [4, 203]. A more complete understanding should involve these segments, as well as a mechanistic structural grasp of how MST activates the Hippo pathway and exactly how the Hippo pathway and YAP1 can rescue Ras and MAPK inhibition. Mutations in YAP1 are frequently observed in Ras-driven cancers [122, 125-127, 189], and more [192, 193]. RASSF5 tumor suppressor is currently among the key drug targets in Ras-related cancers due to its role in linking Ras/MAPK and YAP1.

Taken together, we suggest that when RASSF5 binds KRas4B-GTP, the equilibrium shifts. SARAH's fluctuations increase predominantly toward the open state and heterodimer formation with MST's SARAH. Since the RA/SARAH affinity is relatively low, whereas that of the SARAH heterodimer is in the nM range, the emerging mechanism of RASSF5 action as a tumor suppressor is through competitive

association with the Ras-GTP effector interaction site, and recruitment to the membrane, where it supports MST kinase domain dimerization and trans-autoactivation. Such a conformational dynamic view suggests how RASSF5 can suppress cell proliferation. Its two domains link Ras and the MAPK and the Hippo pathways, the two major pathways in tumor cell proliferation – the first promoting the second suppressing – thereby enabling RASSF5’s tumor suppressing action. At the same time, its N-terminal region, including its putative C1 domain, anchor it in the membrane, further stimulating MST1/2 activation, ensuring RASSF5 action emerges as an important cancer drug discovery strategy.

#### 7.4 Materials and Methods

For the SARAH dimer simulations, the crystal structures of SARAH dimers, RASSF5-MST2 SARAH heterodimer (PDB code: 4LGD), MST2-MST2 SARAH homodimer (PDB code: 4OH9), and RASSF5-RASSF5 SARAH homodimer (PDB code: 2YMY) were obtained from the Protein Data Bank (PDB). The CHARMM program [48] was used to construct the set of initial conformations and to relax the systems for the atomistic molecular dynamics (MD) simulations. Using the same MD protocols as in our previous work [54, 107, 108, 112-115], the dimer systems were solvated by the modified TIP3P water model and subsequently minimized using the steepest descent (SD) and adopted-basis Newton-Raphson (ABNR) methods, followed by consecutive cycles of dynamics in the pre-equilibrium stage. During the pre-equilibrium simulations, the initial crystal structures of SARAH dimers were gradually relaxed after solvation.

To sample ensembles of monomeric SARAH, the coordinates of RASSF5 SARAH were extracted from the RASSF5-MST2 SARAH heterodimer (PDB code: 4LGD). The isolated RASSF5 SARAH monomer was subjected to replica-exchange molecular dynamics (REMD) simulations for 50 ns. Twelve replicas of the SARAH monomer were simulated at temperatures in the range from 300 K to 360 K, incremented gradually for each replica. During the simulation, temperatures were exchanged between trajectories according to a Metropolis criterion. After the simulations, ensemble configurations of the SARAH monomer with the kinked  $\alpha$ -helical motif were determined by analyzing replica trajectories, discarding the high-temperature replicas. Based on the population distribution of the interaction energy as a function of residue pair-distance, five initial kinked SARAH configurations (KS1 to KS5) were obtained.

To construct the N-terminal truncated RASSF5 (residues 205 – 418) structure, the SARAH domain was covalently connected to the RA domain. The crystal RA domain structure of murine RASSF5 (PDB code: 3DDC) was obtained from the PDB server and subsequently used to model the human RA domain. Residue modifications were made to obtain the human sequence, and the flexible loop with missing coordinates (residues 254 – 278) was constructed by using the Modeller server [196]. Both straight  $\alpha$ -helical SARAH extracted from the crystal structure and kinked SARAH defined by the REMD simulations were used to predict the self-associated RASSF5 conformation by using the PRISM [197, 198] and Patchdock [199, 200] servers. All-atom MD simulations were prepared for the best fourteen configurations of the initial RASSF5 structure predicted from the docking programs. During the simulations, ensembles of the self-associated RASSF5 conformations in an aqueous environment were obtained.

For the simulations of RASSF5 in complex with Ras, seven different RASSF5 conformations among the fourteen configurations were selected to model the dimeric RASSF5/Ras complex. The crystal structure of 3DDC contains the murine RASSF5 RA domain in complex with the GNP-bound HRas<sup>D30E/E31K</sup>. By replacements of HRas with the GTP-bound K-Ras4B (PDB code: 3GFT) and the murine RASSF5 RA domain with the selected RASSF5 configurations, seven different dimeric configurations for the RASSF5/K-Ras4B-GTP complex, DC1(1), DC2(4), DC3(6), DC4(9), DC5(10), DC6(11), and DC7(14), where DC represents dimeric configuration and the numbers in the parenthesis denote the monomeric RASSF5 configuration, were obtained. The structure of 3GFT contains a point mutation, Q61H, which was replaced with the wild-type residue, and the GNP in 3GFT was replaced by GTP.

The initial fourteen configurations of monomeric RASSF5 and seven dimeric configurations of RASSF5 in complex with KRas4B-GTP were solvated by the TIP3P water model and subsequently minimized with the protein backbone rigid for 10,000 steps, followed by a dynamic cycle of 50,000 time steps. All systems were subsequently neutralized, with the number of counter ions needed to neutralize the systems. In addition, a number of ions, Na<sup>+</sup> and Cl<sup>-</sup>, were added to the systems to satisfy a total cation concentration near 100 mM, followed by the same minimization and dynamic protocols of previous stage. In the pre-equilibrium stages, harmonic restraints were applied to the heavy atoms ( $k = 5 \text{ kcal/mol/\AA}^2/\text{atom}$ ) and gradually relaxed to  $k = 0$  with a full particle mesh Ewald (PME) calculation for long-range electrostatic interactions and a constant temperature (Nosé–Hoover) thermostat/barostat at 310 K. Following pre-equilibrium, a 200 ns production run was performed with the NAMD

2.10 code [47] and the CHARMM program [48] with version 36 [53] on the Biowulf cluster at the National Institutes of Health (Bethesda, MD). Averages were taken after 30 ns, discarding initial transients.

## Chapter 8: Conclusion

In this study, we comprehensively investigated the interaction of KRas4B with its upstream effectors (SOS1, Grb2, and EGFR) and downstream effector RASSF5 of Hippo pathway. We used NMR experiments to measure the affinity of Grb2-SOS1 via the isolated Grb2 nSH3/cSH3 domains and the selected SOS1 PR segments. We successfully uncovered a novel cSH3 binding site PKLPPKTYKREH on SOS1 PR domain. This alternative cSH3 binding site leads to the multiple binding modes of Grb2-SOS1 interaction. Subsequently, we used MD simulation to search all other probable PR segments on SOS1. The simulation revealed that nSH3 would perform open/closed conformation when binding to a weak/strong binding partner while cSH3 always retains a consistent and stable conformation. This implies nSH3 and cSH3 might have distinct biological roles: nSH3 adapts the association with SOS1, and cSH3 helps Grb2 dimerization.

The full length Grb2 performs much stronger binding affinities for the selected one-site SOS1 peptides than the isolated nSH3/cSH3. Interestingly, these affinities have the same magnitude for both nSH3 and cSH3. We also tested the interaction of Grb2 with the two-site SOS1 peptides. These bindings follow the pairwise affinities of nSH3/cSH3 with one-site SOS1 segments. In order to evaluate the EGFR effect, the phosphorylated EGFR segment was added to the full length Grb2. The EGFR-Grb2 complex and Grb2 perform the same interactions with one-site and two-site SOS1 peptides. This suggests EGFR has limited influence on Grb2-SOS1 interactions.

SOS1 activate KRas4B at the CDC25 catalytic site by associating with KRas4B at the REM allosteric site. Either GTP or GDP bound KRas4B at the REM domain is

capable to fully activate SOS1. The allosteric signals are roused from switch I/II of allosteric KRas4B and ended at the switch I/II of catalytic KRas4B. At the REM domain, KRas4B-GDP renders limited activation cycle while KRas4B-GTP induces accelerated activation cycle.

Regarding Ras/RASSF5 interactions, we reviewed the previous publications and proposed that membrane-anchored Ras is essential for RASSF5 activation. Dimeric Ras activates RASSF5 by releasing its SARAH domain. RASSF5 SARAH heterodimerize with MST1/2 SARAH, transmitting Hippo signaling. Without membrane, MST1/2 kinase domains is unable to dimerize and undergo self-phosphorylation. By simulating KRas4B with the modeled RASSF5 auto-inhibited complex, we suggested KRas4B binding to RASSF5 RA domain releases RASSF5 SARAH domain.

To conclude, chapter 2 to 6 decipher and summarize the interaction mechanism of EGFR-Grb2, Grb2-SOS1, SOS1-KRas4B, and KRas4B-RASSF5. We elucidated the wild type KRas4B with part of its upstream and downstream effectors. While my work temporarily ends here due to the limited time, the Ras research is endless. As one of the most oncogenic proteins, inhibiting Ras signaling is the major attempt that many scientists have been working on. In the future, I would suggest to look into the detail of Grb2-SOS1 complex. The complex interaction of Grb2-SOS1 is dynamic. The wild-type Grb2 interacting with wild-type SOS1 PR may only reflect early stage of Ras signaling. In the subsequent Ras signaling, SOS1 PR undergoes multiple residue phosphorylation and enforce the nSH3/cSH3 shift to other binding sites. Understanding the Grb2-SOS1 interactions in different stage of Ras signaling will be beneficial for the development of peptide inhibitors.

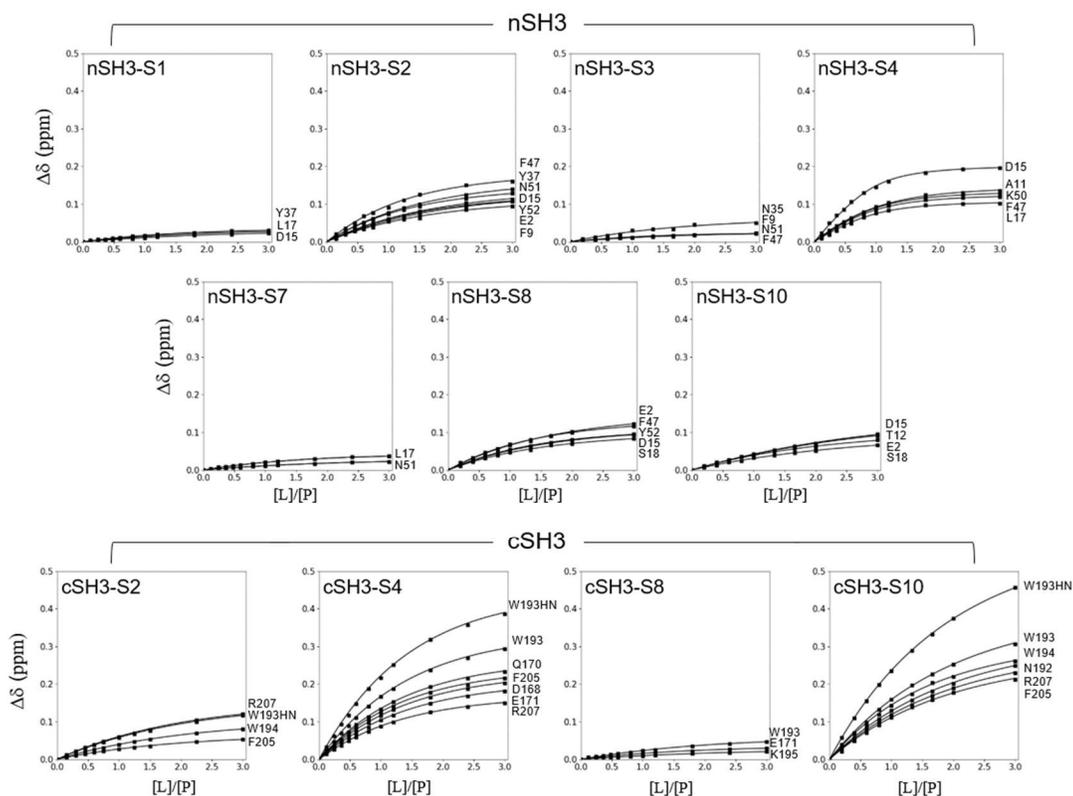
# Appendices

## Appendix A

**A**

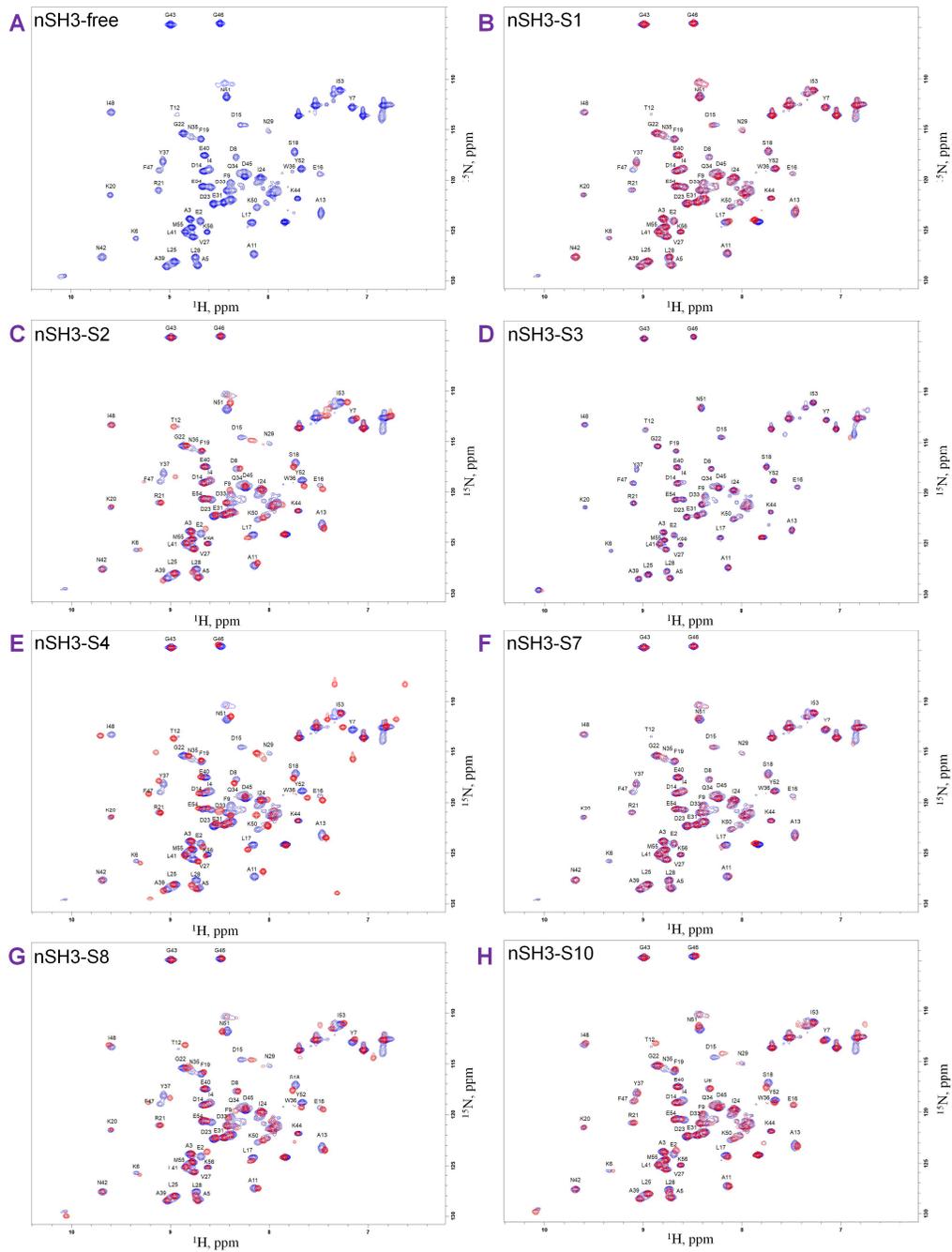


**B**

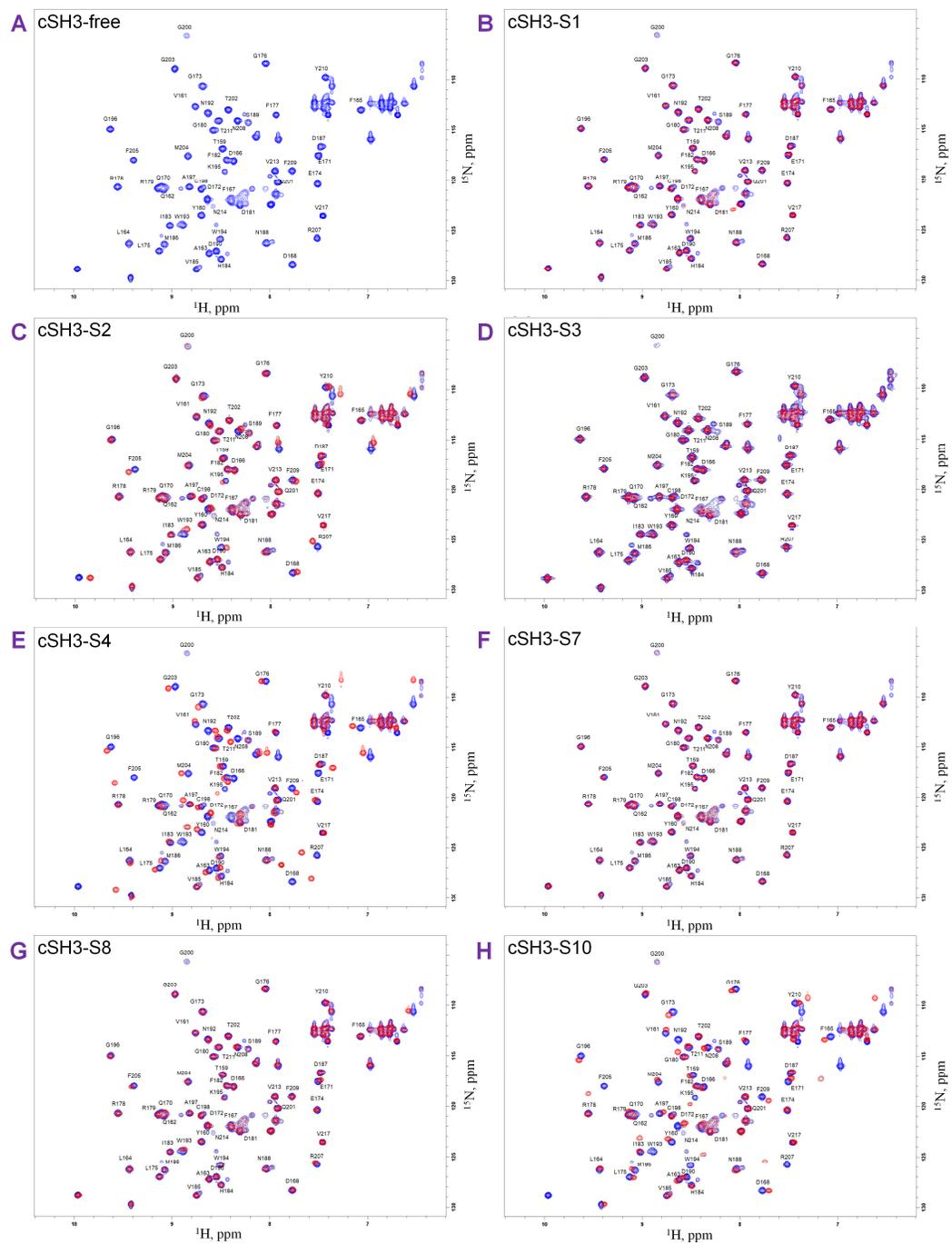


**Figure A1.** (A) Sequence alignment of Grb2 nSH3 and cSH3 domains. The interface residues Y7, F9, D33, W36, F47, and Y52 of nSH3 are aligned with F165, F167, D190, W193, F205, and Y209 of cSH3. The unaligned critical residues D15 and E171 result in the nSH3/cSH3's distinct binding motifs. (B) Titration curves for the selected

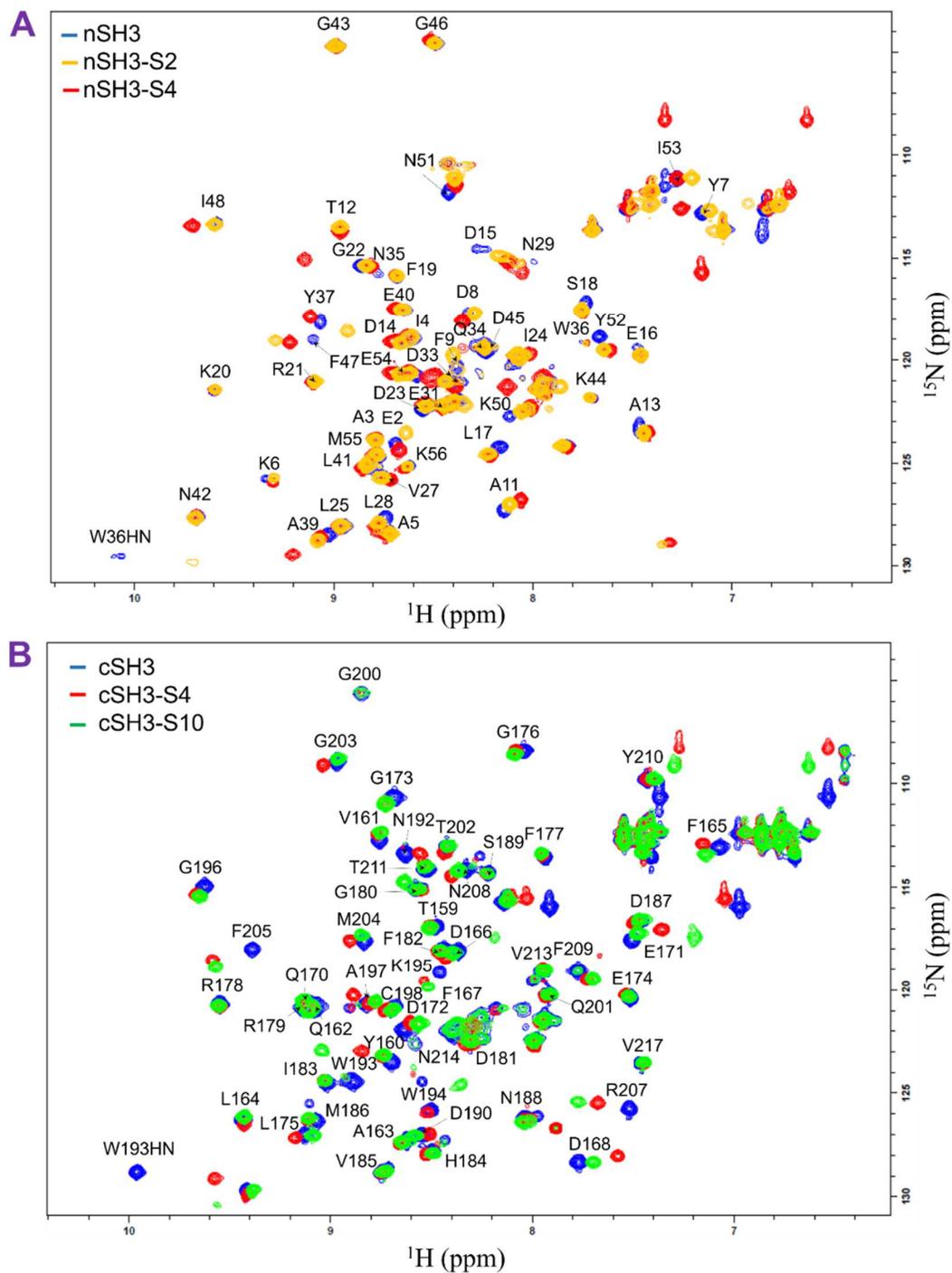
residues of nSH3/cSH3 ([P]) upon addition of various SOS1 peptides ([L]): the observed CSPs are plotted as a function of [L]/[P] ratio and fitted to a single-site binding model. In  $K_D$  calculations of nSH3, 3 residues (D15, L17, and Y37) are used for S1, 4 residues (F9, N35, F47, and N51) for S3, 2 residues (L17 and N51) for S7, and 7 residues (E2, D15, L17, S18, F47, N51, and Y52) for S8. For cSH3-S8, 3 residues (E171, W193, and K195) are used to calculate  $K_D$ . For nSH3-S2/S4/S10 and cSH3-S2/S4/S10, the selected residues are indicated in the figure 2 and 3.



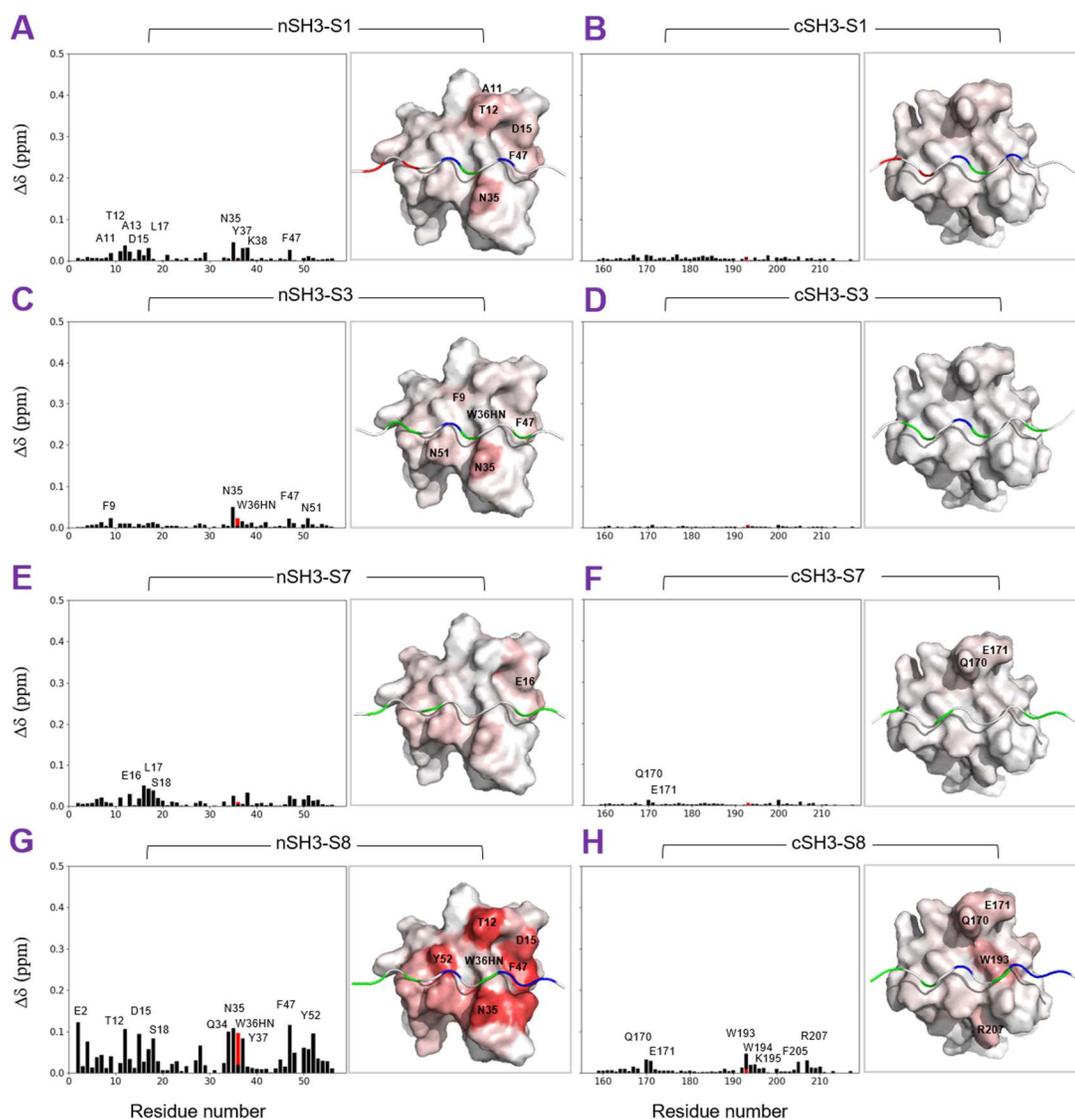
**Figure A2.**  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC spectra of nSH3 in the unbound state (blue) and at the endpoint of titration (red) with the SOS1 peptides S1, S2, S3, S4, S7, S8, and S10. (A) Spectrum of unbound nSH3. (B-H) Overlay of the spectra of free nSH3 (blue) and of (B) nSH3–S1, (C) nSH3–S2, (D) nSH3–S3, (E) nSH3–S4, (F) nSH3–S7, (G) nSH3–S8, and (H) nSH3–S10. The peptide:nSH3 ratio was 3:1 at endpoint.



**Figure A3.**  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC spectra of cSH3 in the unbound state (blue) and at the endpoint of titration (red) with the SOS1 peptides: S1, S2, S3, S4, S7, S8, and S10. (A) Spectrum of unbound cSH3. (B-H) Overlay of the spectra of free nSH3 (blue) and of (B) cSH3–S1, (C) cSH3–S2, (D) cSH3–S3, (E) cSH3–S4, (F) cSH3–S7, (G) cSH3–S8, and (H) cSH3–S10. The peptide: cSH3 molar ratio at endpoint was 3:1.

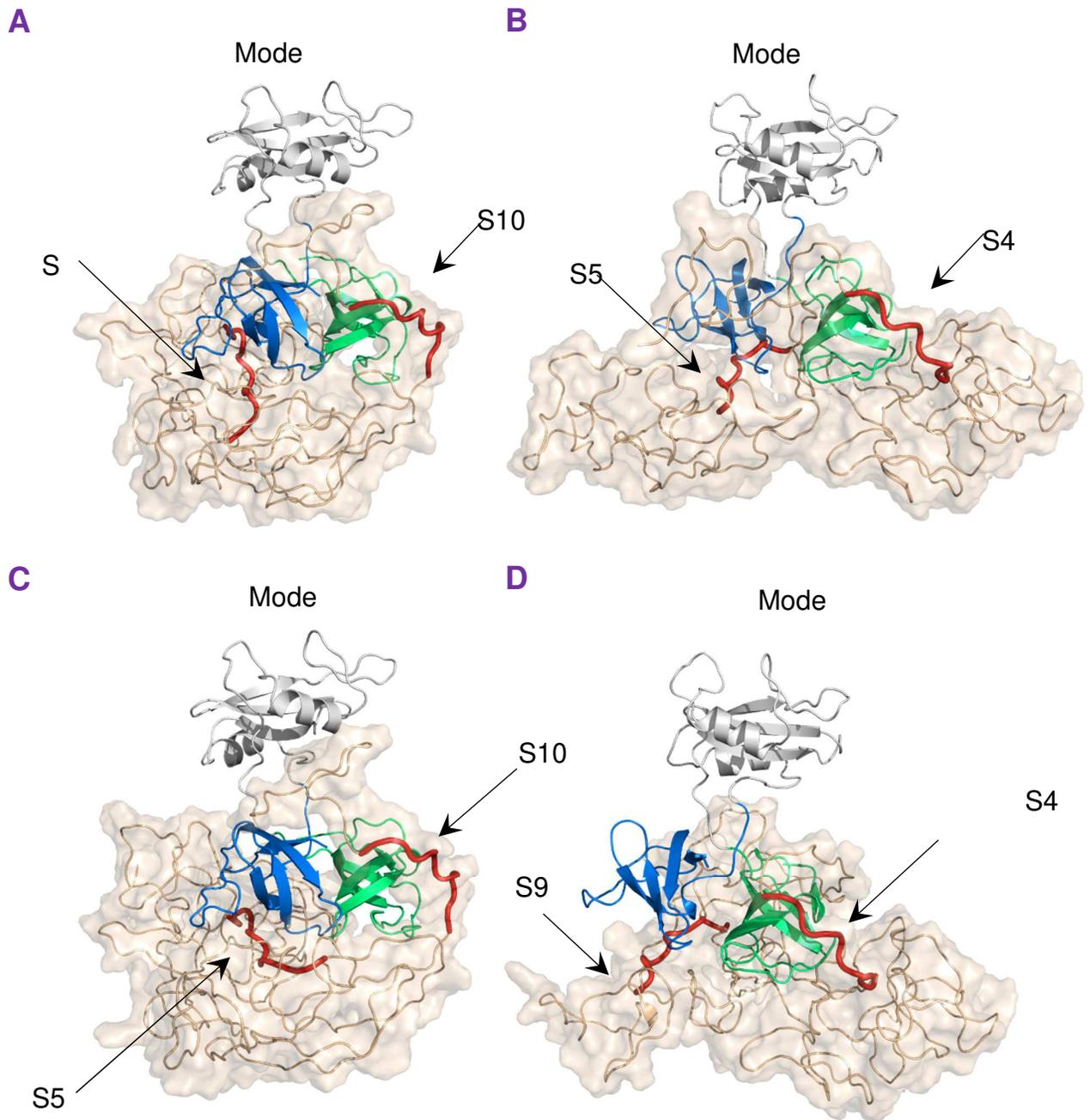


**Figure A4.** The overlaid  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC spectra of (A) nSH3 in the unbound state (blue) and at the endpoint of titration with the SOS1 peptides S2 (yellow) and S4 (red) and (B) cSH3 in the unbound state (blue) with S4 (red) and S10 (green). The molar ratio of peptide to nSH3/cSH3 at the titration endpoint was 3:1.



**Figure A5.** Amide chemical shift perturbations (CSPs,  $\Delta\delta$ ) in nSH3/cSH3 at the endpoint of titration with various SOS1 peptides: (A) nSH3-S1, (B) cSH3-S1, (C) nSH3-S3, (D) cSH3-S3, (E) nSH3-S7, (F) cSH3-S7, (G) nSH3-S8, and (H) cSH3-S8. Structure images to the right of the plots show CSPs mapped onto surfaces of nSH3/cSH3 (PDB: 1GRI) in complexes with the corresponding modeled peptides, respectively. The red bar indicates the signal shifts of W36/W193 side chain (W36HN/W193HN). The surface coloring is based on the observed CSP values and ranges from white (0 ppm) to red (maximal  $\Delta\delta = 0.15$ ); select residues are indicated.

The nSH3- or cSH3-bound peptides are colored based on non-polar (white), polar (green), negatively-charged (red), and positively-charged (blue) residues. The residues with shifts greater than one standard deviation are indicated.



**Figure A6.** The initial structures of four Grb2-SOS1 complex models: (A) nSH3-S4/cSH3-S10, (B) nSH3-S5/cSH3-S4, (C) nSH3-S5/cSH3-S10, and (D) nSH3-S9/cSH3-S4. Grb2 nSH3, SH2, cSH3 domains, and SOS1 PR domain are colored blue, white, green, and yellow, respectively. The SOS1 binding sites are highlighted by the thick red tubes and pointed out by the black arrows.

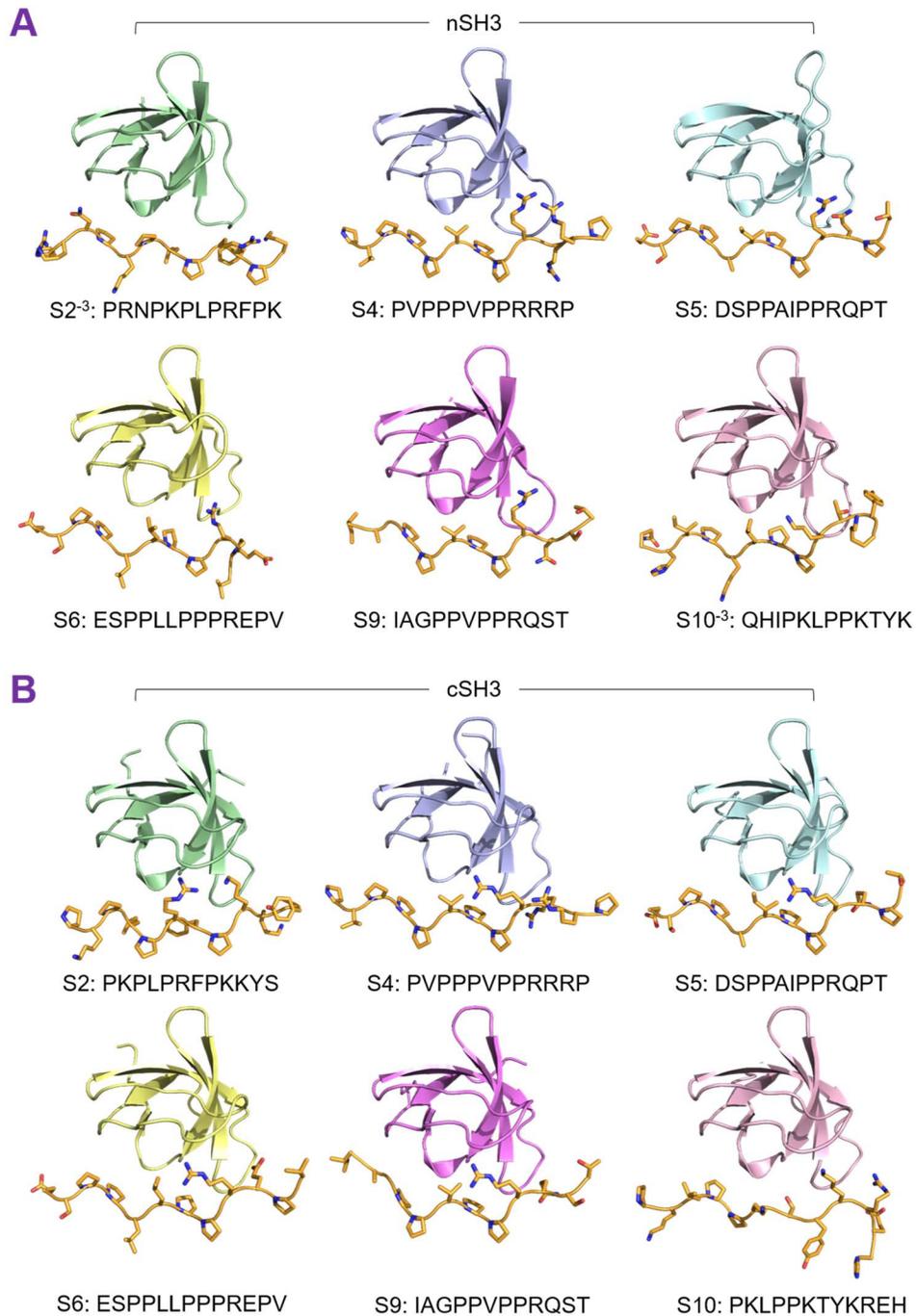
**Table A1**

Pairwise affinities of nSH3–S2/S4/S5/S6/S9 and cSH3–S2/S4/S10

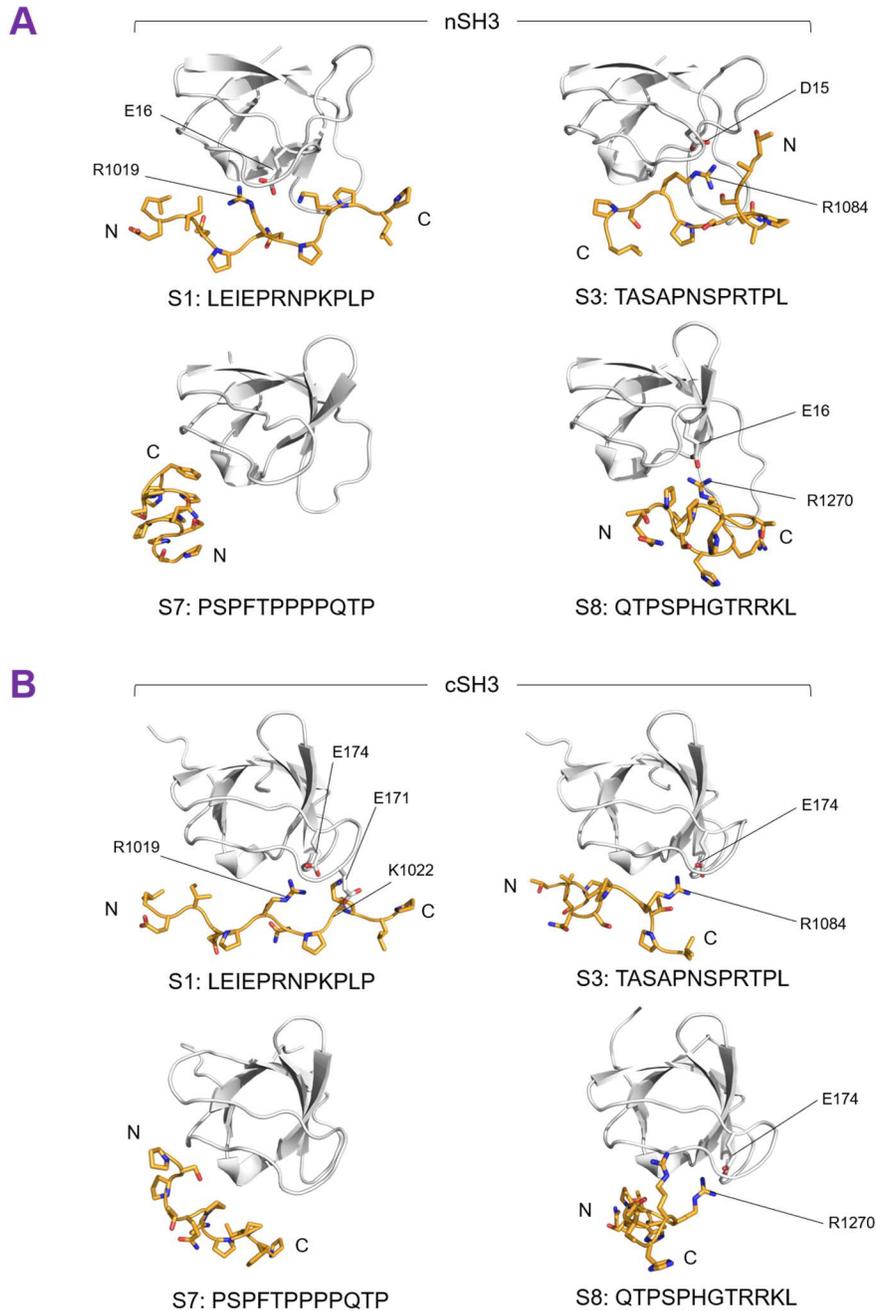
cSH3/nSH3	S2 250 ( $\mu\text{M}$ )	S4 35 ( $\mu\text{M}$ )	S5 56 ( $\mu\text{M}$ )	S6 117 ( $\mu\text{M}$ )	S9 82 ( $\mu\text{M}$ )
S2 400 ( $\mu\text{M}$ )	100,000	14,000	22,400	46,800	32,800
S4 140 ( $\mu\text{M}$ )	35,000	X	7840	16,380	11,480
S10 150 ( $\mu\text{M}$ )	37,500	5250	8400	17,550	12,300

Summing up the  $K_D$  values for nSH3–S2/S4/S5/S6/S9 and cSH3–S2/S4/S10 suggests the probable combination of nSH3/cSH3 with SOS1 peptides. The smaller values may represent the higher probabilities. The nSH3–S4/cSH3S10 has the smallest values, followed by nSH3–S5/cSH3–S4, nSH3–S5/cSH3–S10, nSH3–S9/cSH3–S4, nSH3–S9/cSH3–S10, nSH3–S6/cSH3–S4, nSH3–S6/cSH3–S10, nSH3–S2/cSH3–S4, nSH3–S2/cSH3–S10, nSH3–S4/cSH3–S2, nSH3–S5/cSH3–S2, nSH3–S9/cSH3–S2, and nSH3–S6/cSH3–S2.

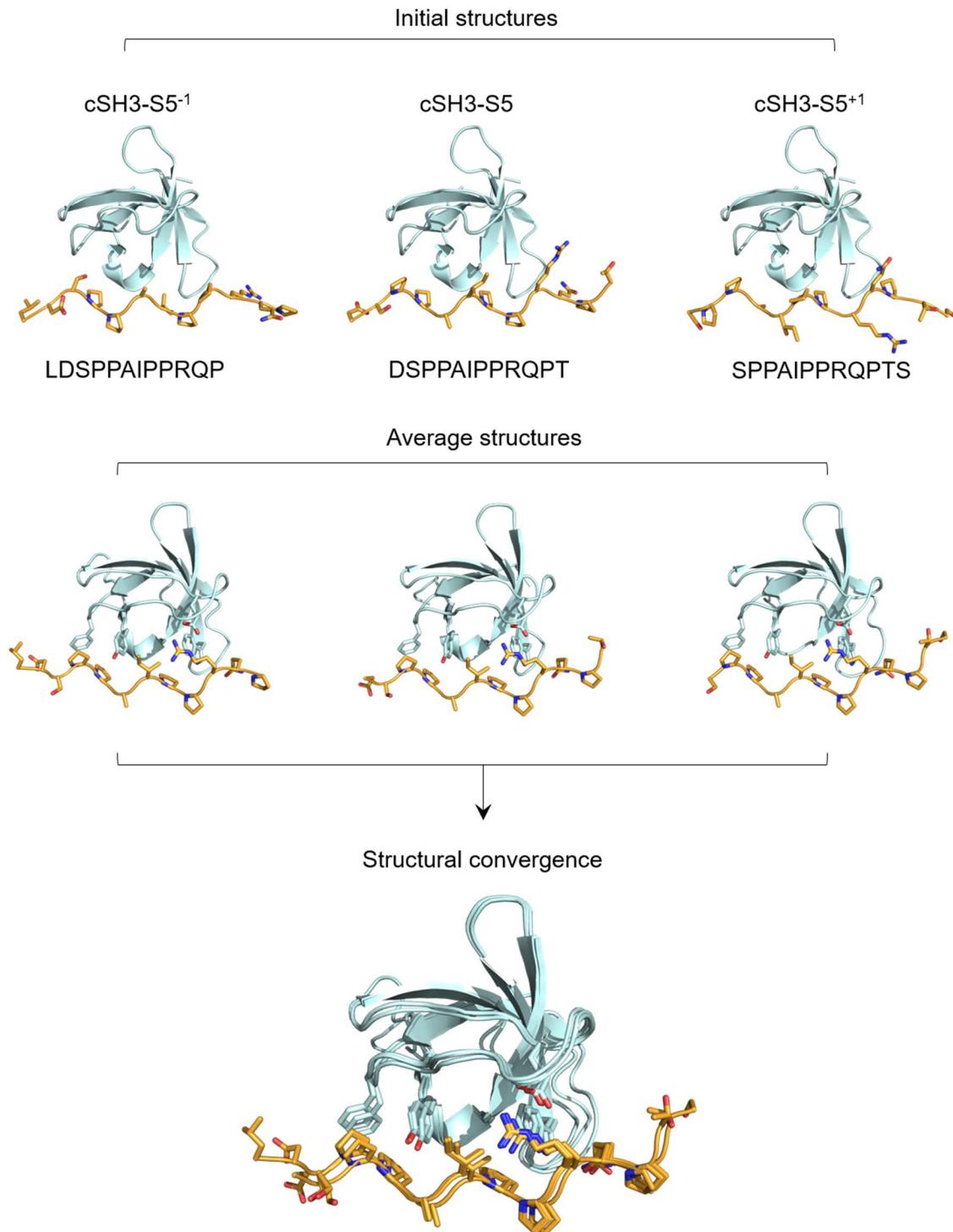
*Appendix B*



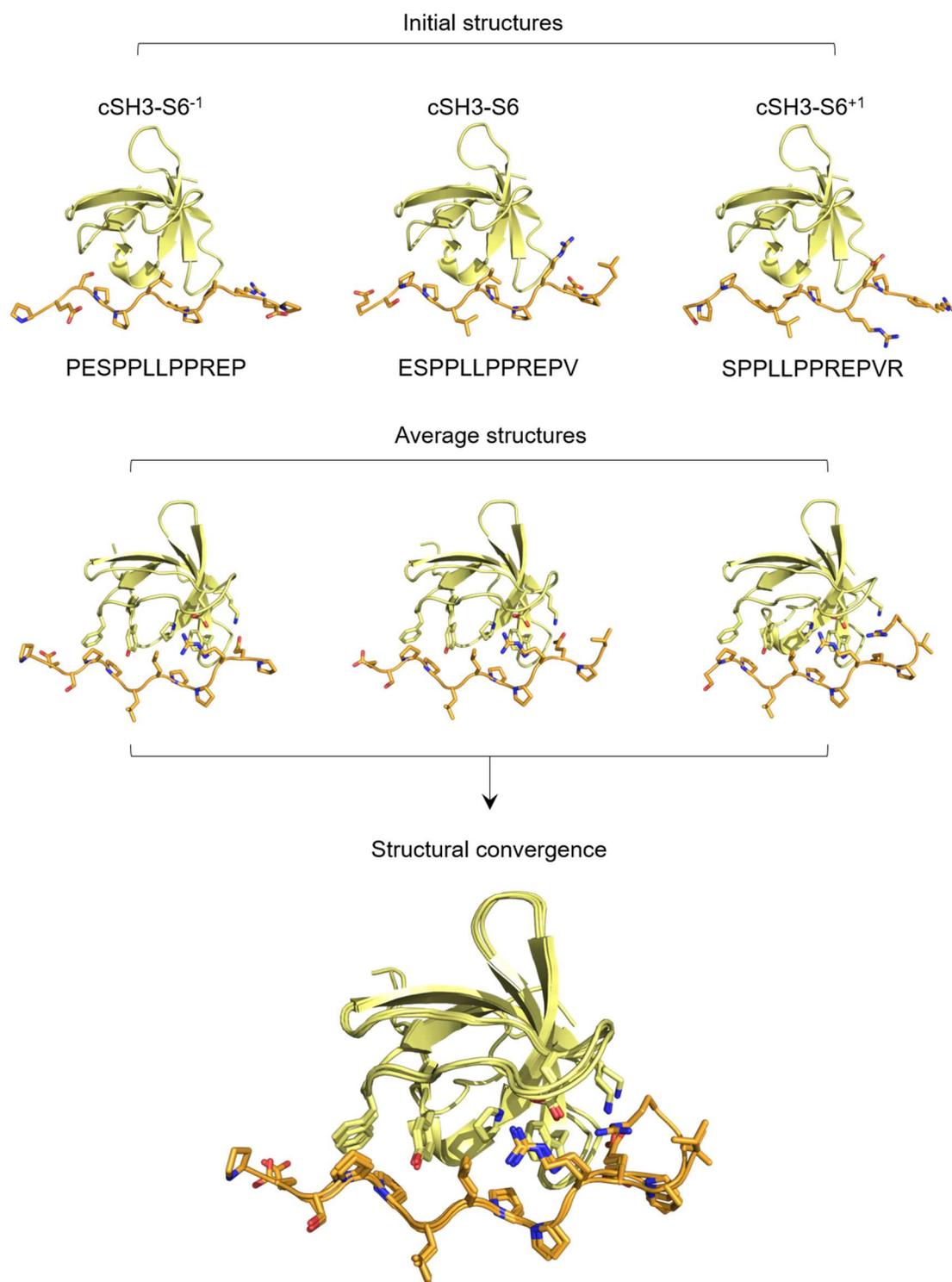
**Figure B1.** The average structures of (A) nSH3–S2<sup>-3</sup>/S4/S5/S6/S9/S10<sup>-3</sup> and (B) cSH3–S2/S4/S5/S6/S9/S10 complexes. Each SOS1 PR peptide (orange) contains 12 residues; the structures of residue side chains are shown.



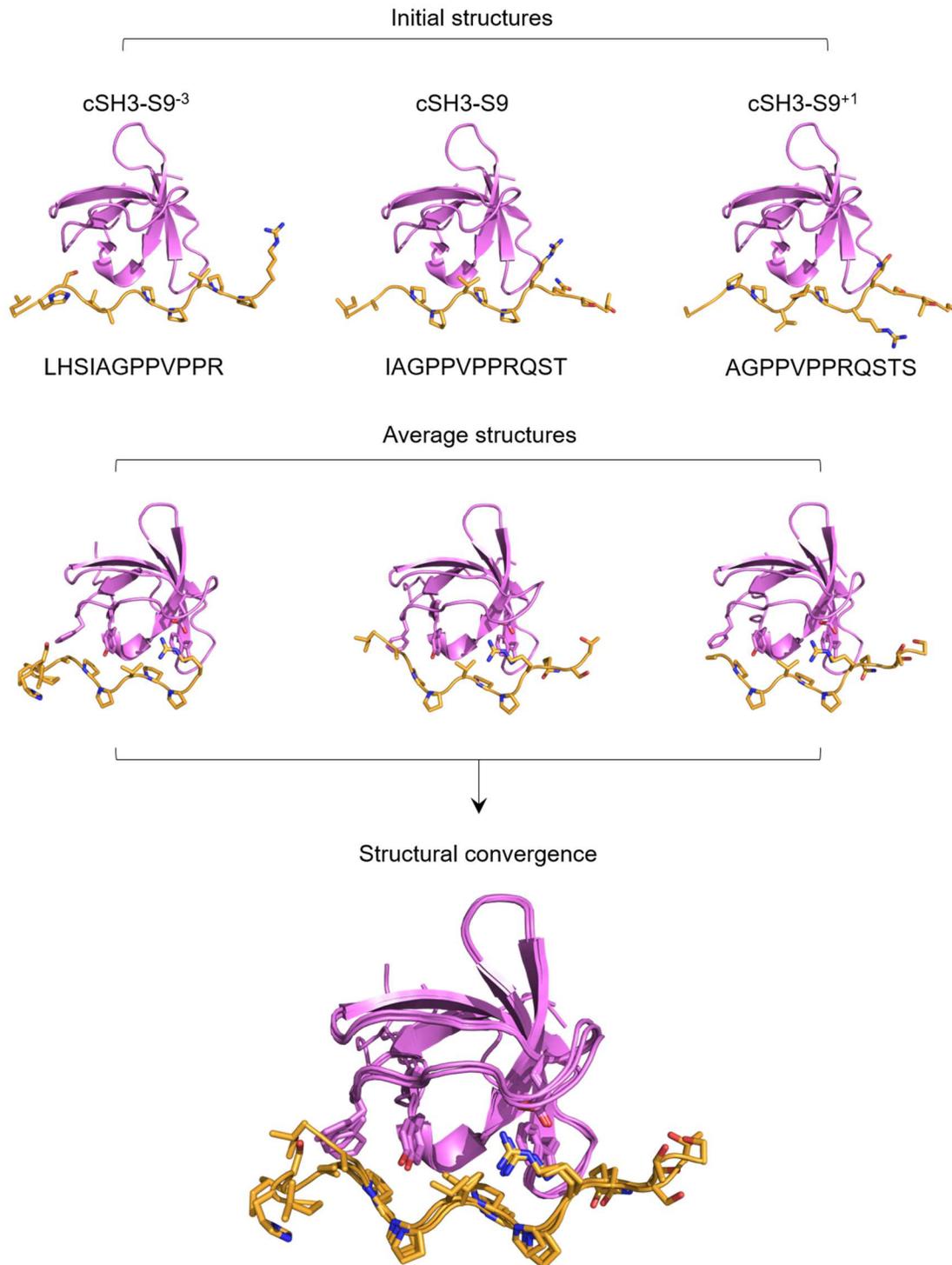
**Figure B2.** The average structures of (A) nSH3–S1/S3/S7/S8 and (B) cSH3–S1/S3/S7/S8 complexes. S1, S3, and S8 form a salt bridge with nSH3/cSH3. Due to the lack of interactions with nSH3/cSH3 hydrophobic pocket, S1, S3, and S8 hardly retain the secondary structures and partially associate with nSH3/cSH3. S7 is totally dissociated from nSH3/cSH3.



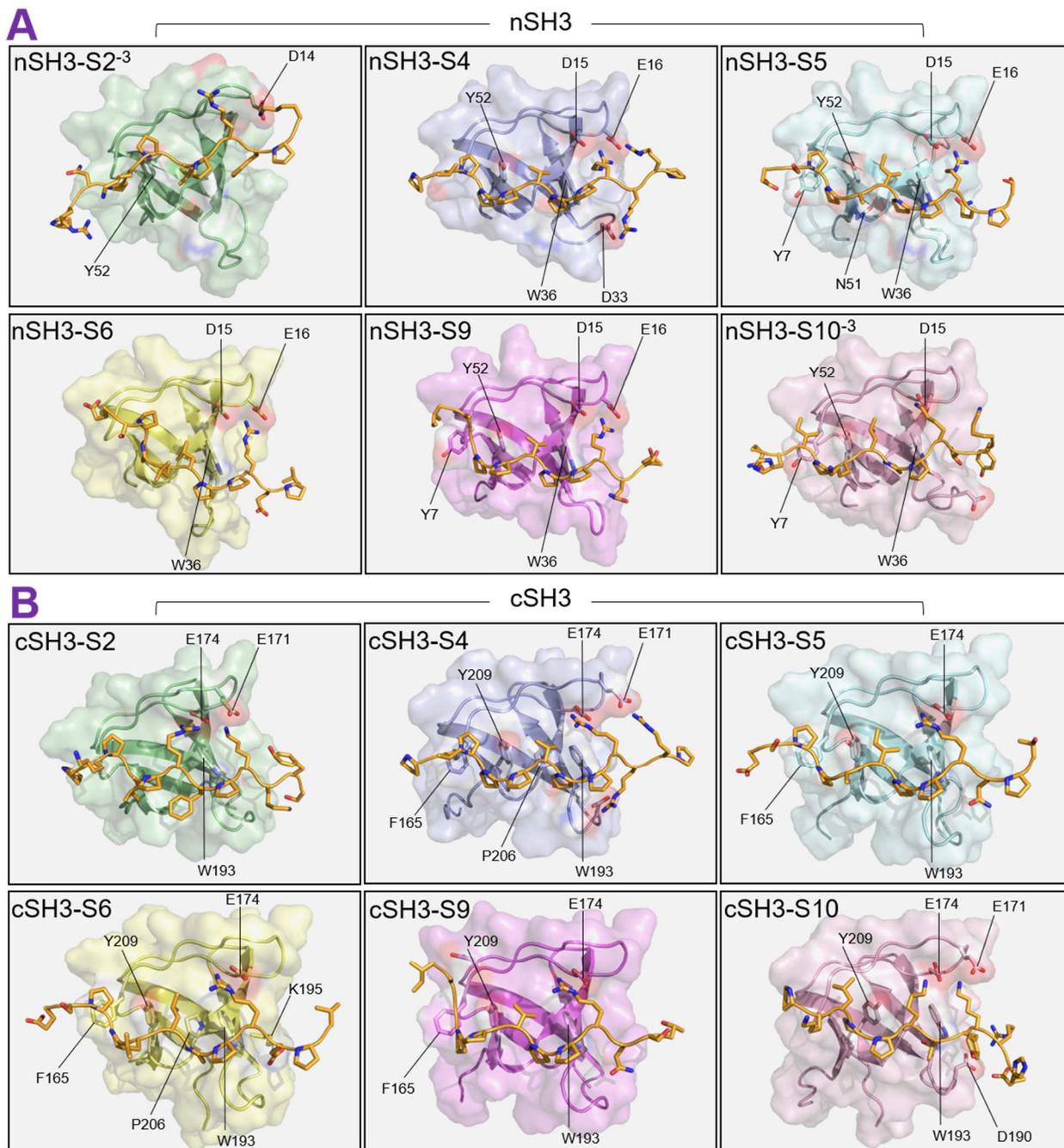
**Figure B3.** The average structures of cSH3-S5<sup>-1</sup>, cSH3-S5 and cSH3-S5<sup>+1</sup> are highly similar while S5<sup>-1</sup>, S5, and S5<sup>+1</sup> have different initial structural models. The converged structure suggests the probable model of the nSH3-S5 complex.

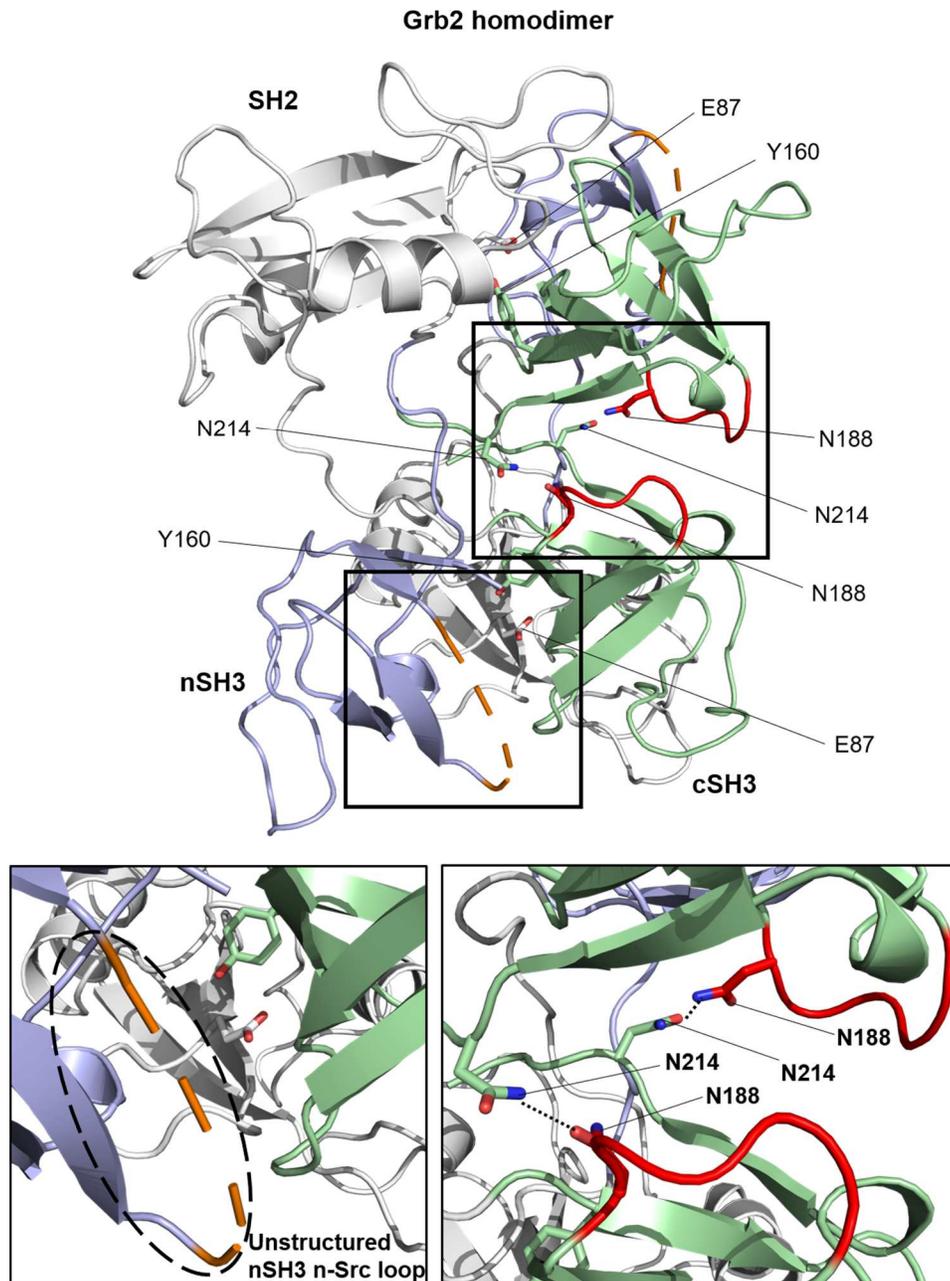


**Figure B4.** The average structures of cSH3-S6<sup>-1</sup>, cSH3-S6 and cSH3-S6<sup>+1</sup> are highly similar while S6<sup>-1</sup>, S6, and S6<sup>+1</sup> have different initial structural models. The converged structure suggests the probable model of the nSH3-S6 complex.

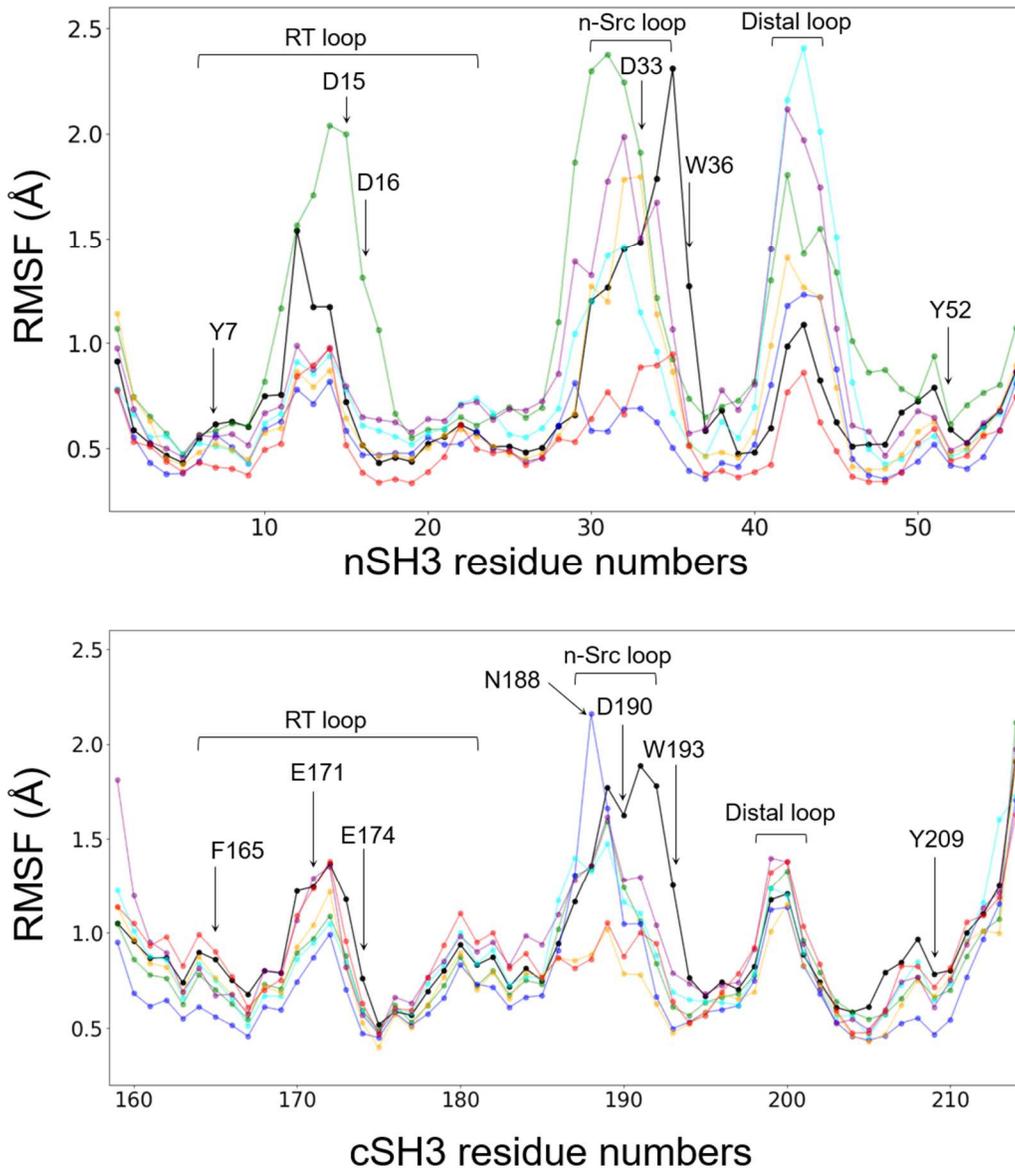


**Figure B5.** The average structures of cSH3-S9<sup>-3</sup>, cSH3-S9 and cSH3-S9<sup>+1</sup> are highly similar while S9<sup>-1</sup>, S9, and S9<sup>+1</sup> have different initial structural models. The converged structure suggests the probable model of the nSH3-S9 complex.



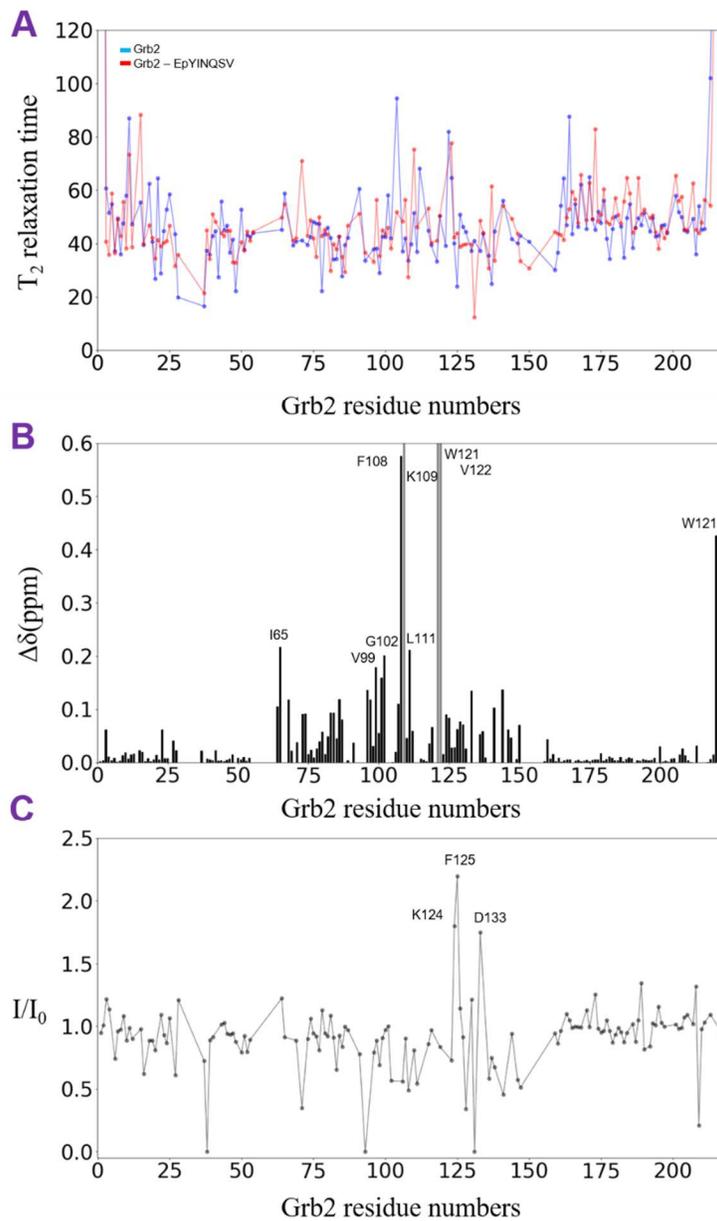


**Figure B7.** The crystal structure of Grb2 homodimer (PDB: 1GRI) contains nSH3 (blue), SH2 (white), and cSH3 (green). (A) The orange dashed line represents the unstructured nSH3 n-Src loop (LNEECD, residues 28-33). (B) Grb2 is dimerized through the hydrogen bond (H-bond) formations of E87-Y160 and N188-N214. The cSH3 n-Src loop is colored red.

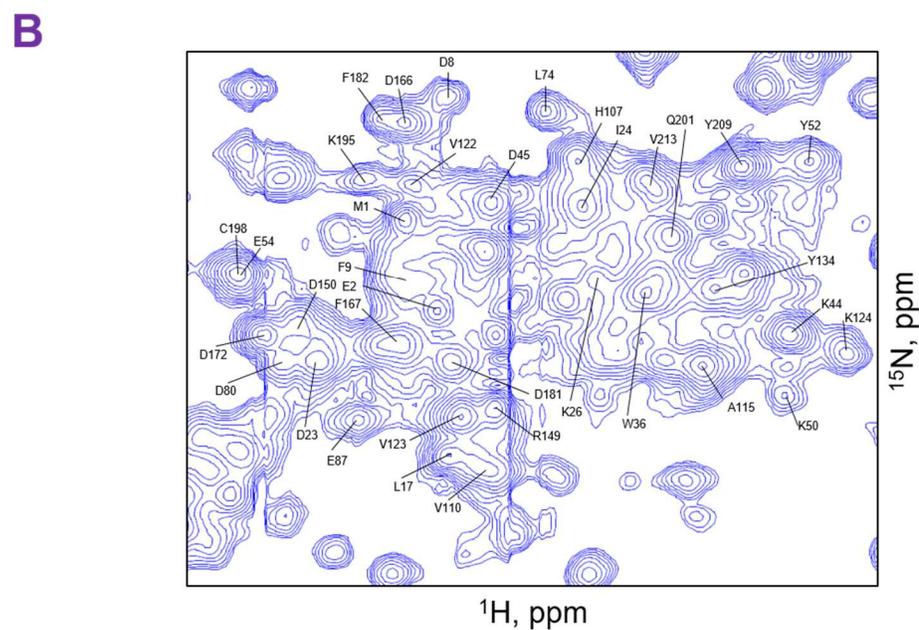
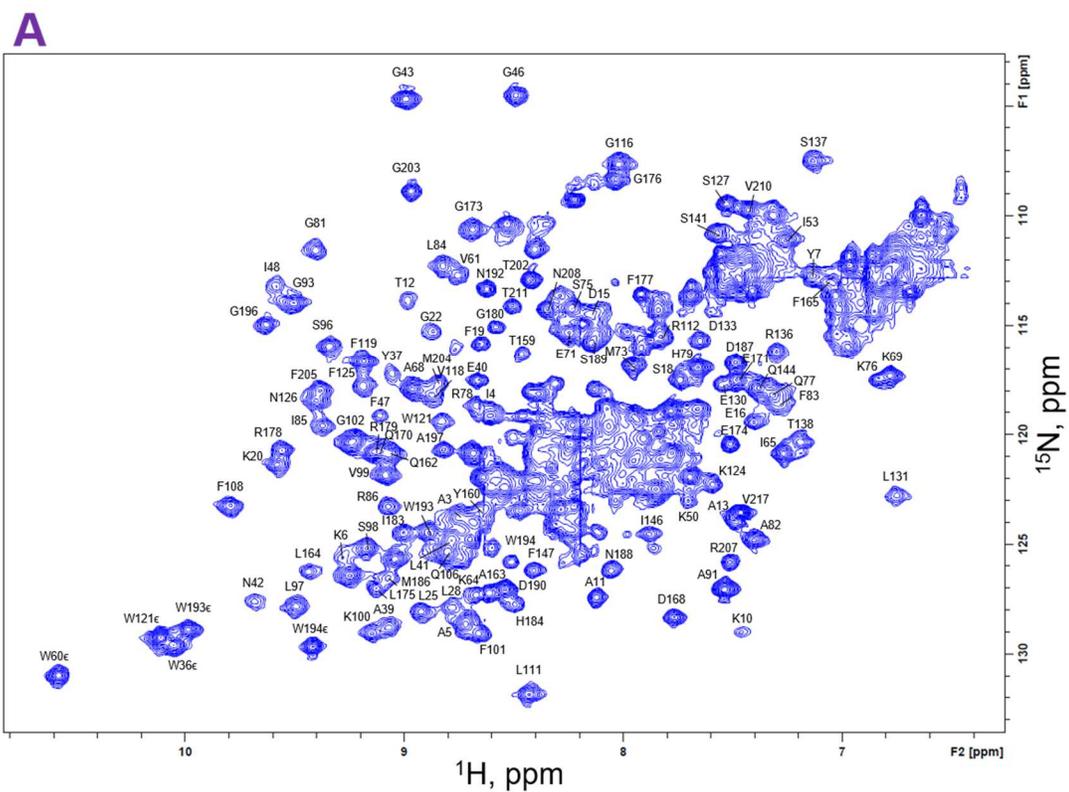


**Figure B8.** The root-mean-square fluctuations (RMSF) of (A) nSH3-S2<sup>-3</sup>/S4/S5/S6/S9/S10<sup>-3</sup> and (B) cSH3-S2/S4/S5/S6/S9/S10 complexes. The nSH3/cSH3 bound to S2<sup>-3</sup>/S2, S4, S5, S6, S9, and S10<sup>-3</sup>/S10 are colored green, blue, cyan, yellow, purple, and red. The RT, n-Src, and distal loops as well as the critical residues involved in the interaction for SOS1 PR domain are indicated. The cSH3 N188 which is responsible for Grb2 homo-dimerization shows higher fluctuation when bound to S4.

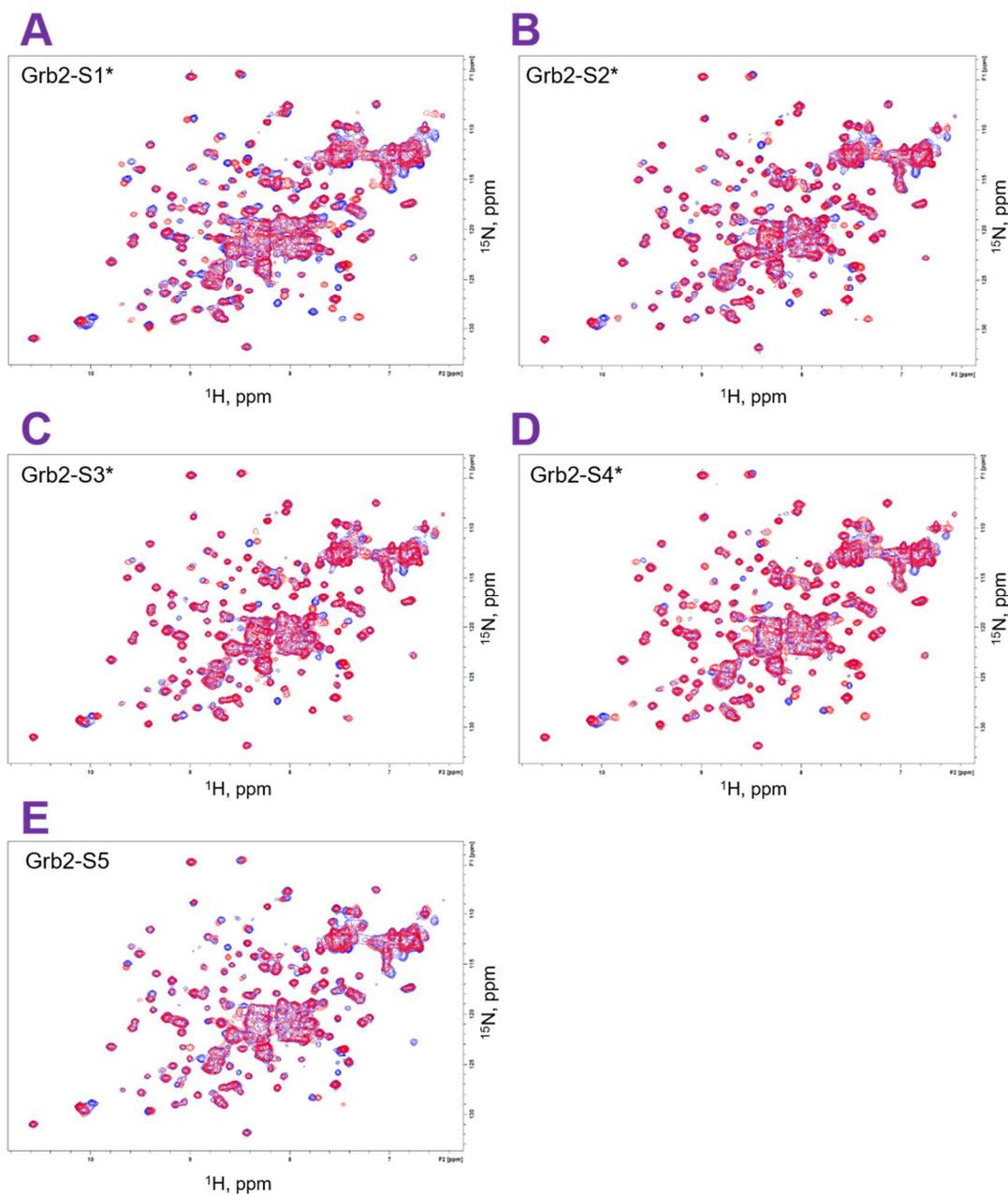
Appendix C



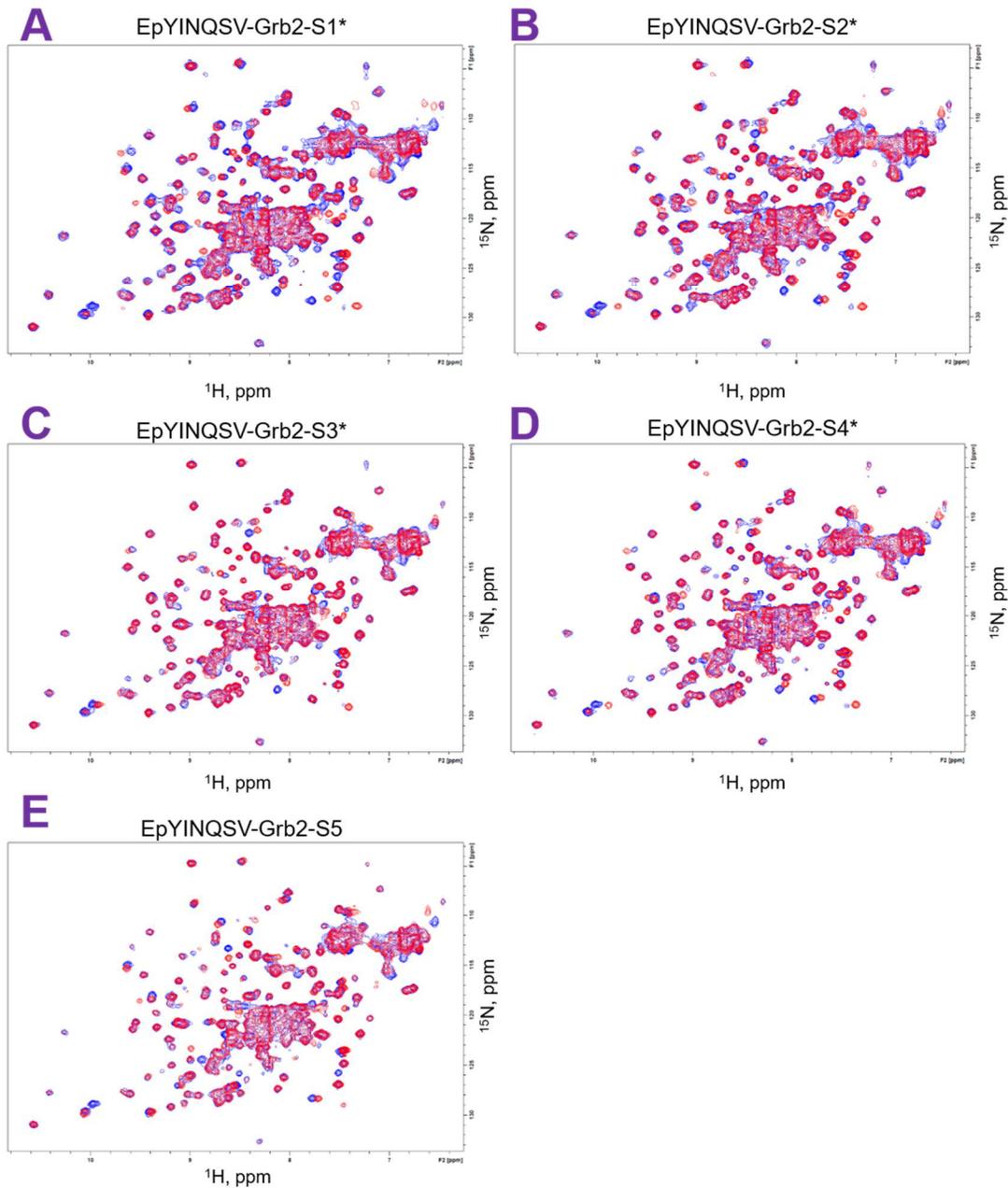
**Figure C1.** (A)  $T_2$  relaxation of Grb2 and EpYINQSV–Grb2. The averaged  $T_2$  relaxation times (unit in ms) of Grb2 (blue) and EpYINQSV–Grb2 (red) are 52.4 and 53.2 ms, respectively. (B) CSP of Grb2 interacting with EpYINQSV. Residues with large CSPs are indicated. (C) Signal intensity ratio of residues intensity in EpYINQSV–Grb2 over Grb2. Residues with  $I/I_0 > 1.2$  are pointed out.



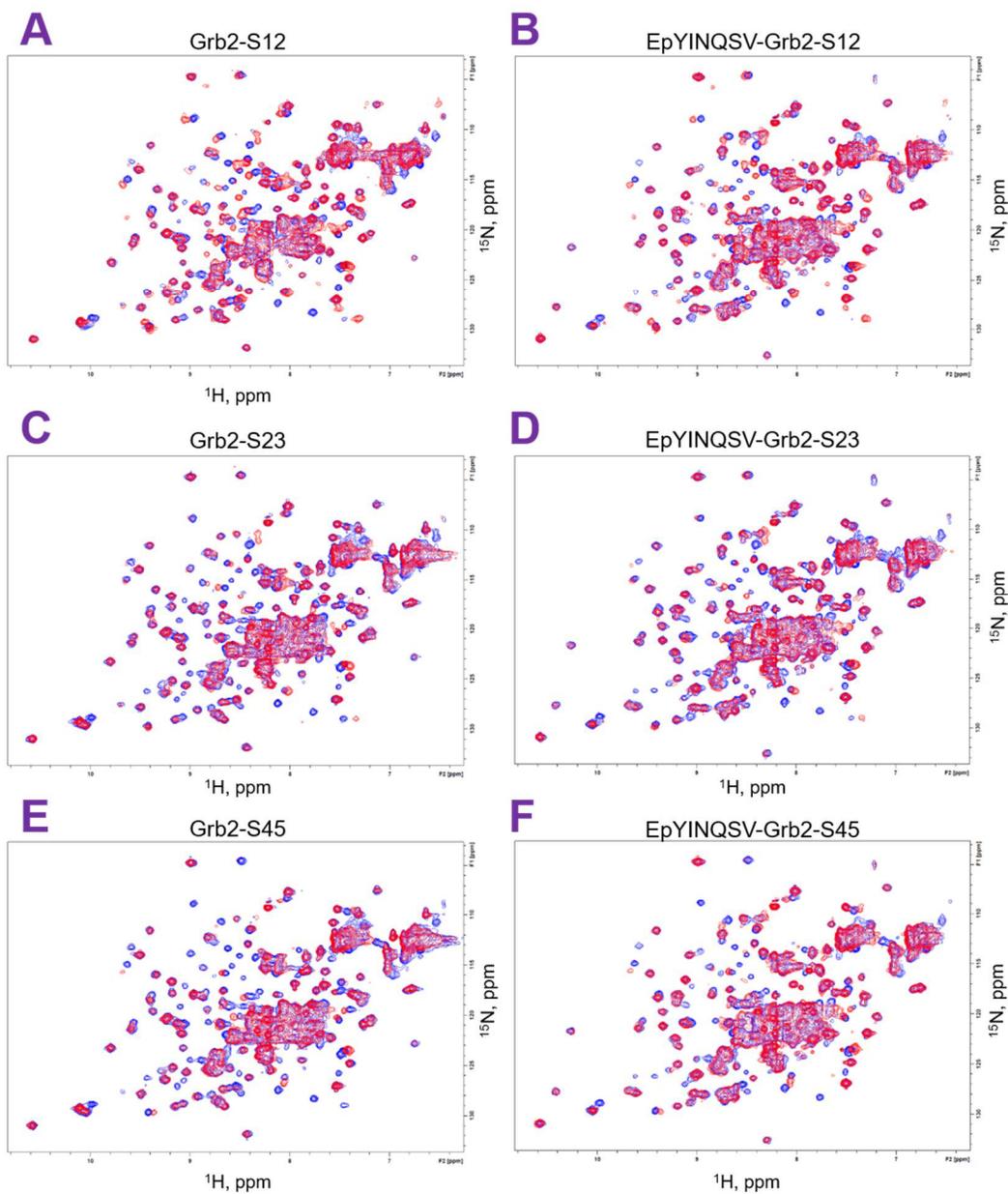
**Figure C2.**  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC spectra of (A) full length Grb2 and (B) the middle signal-crowded region. The signal assignments are based on the published data. The corresponding residues are indicated above the signals.



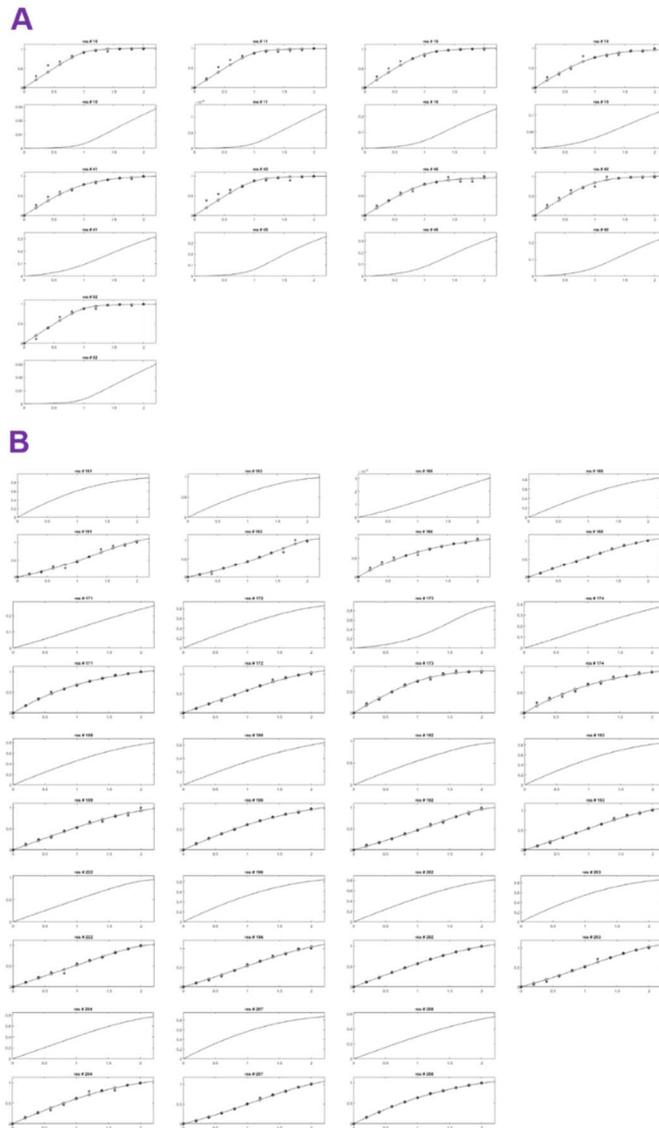
**Figure C3.**  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC spectra of Grb2 in the unbound state (blue) and at the endpoint of titration (red) with the SOS1 peptides S1\*, S2\*, S3\*, S4\*, and S5. Overlay of the spectra of free Grb2 (blue) and of (A) Grb2-S1\*, (B) Grb2-S2\*, (C) Grb2-S3\*, (D) Grb2-S4\*, and (E) Grb2-S5. The molar ratio of SOS1/Grb2 is 2:1 at the titration endpoint.



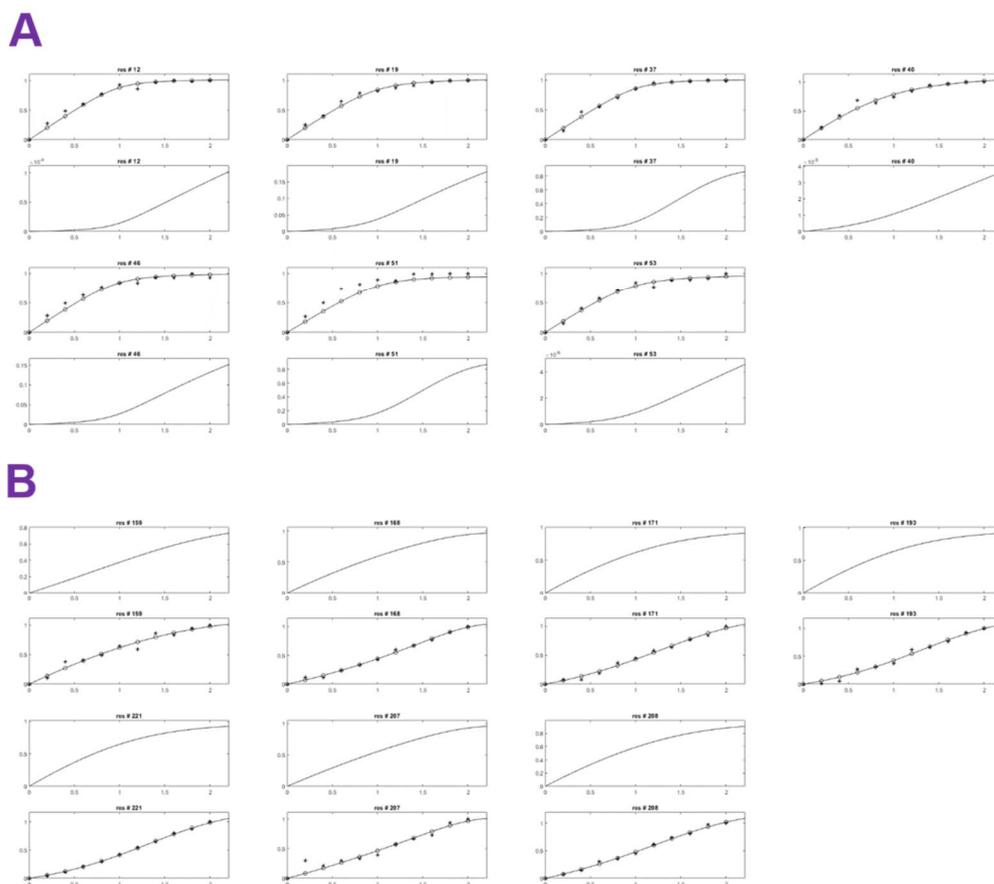
**Figure C4.**  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC spectra of EpYINQSV–Grb2 (blue) and at the endpoint of titration (red) with the SOS1 peptides S1\*, S2\*, S3\*, S4\*, and S5. Overlay of the spectra of EpYINQSV–Grb2 (blue) and of (A) Grb2–S1\*, (B) Grb2–S2\*, (C) Grb2–S3\*, (D) Grb2–S4\*, and (E) Grb2–S5. EpYINQSV binds to Grb2 with 1:1 stoichiometry, and the molar ratio of SOS1/Grb2 is 2:1 at the titration endpoint.



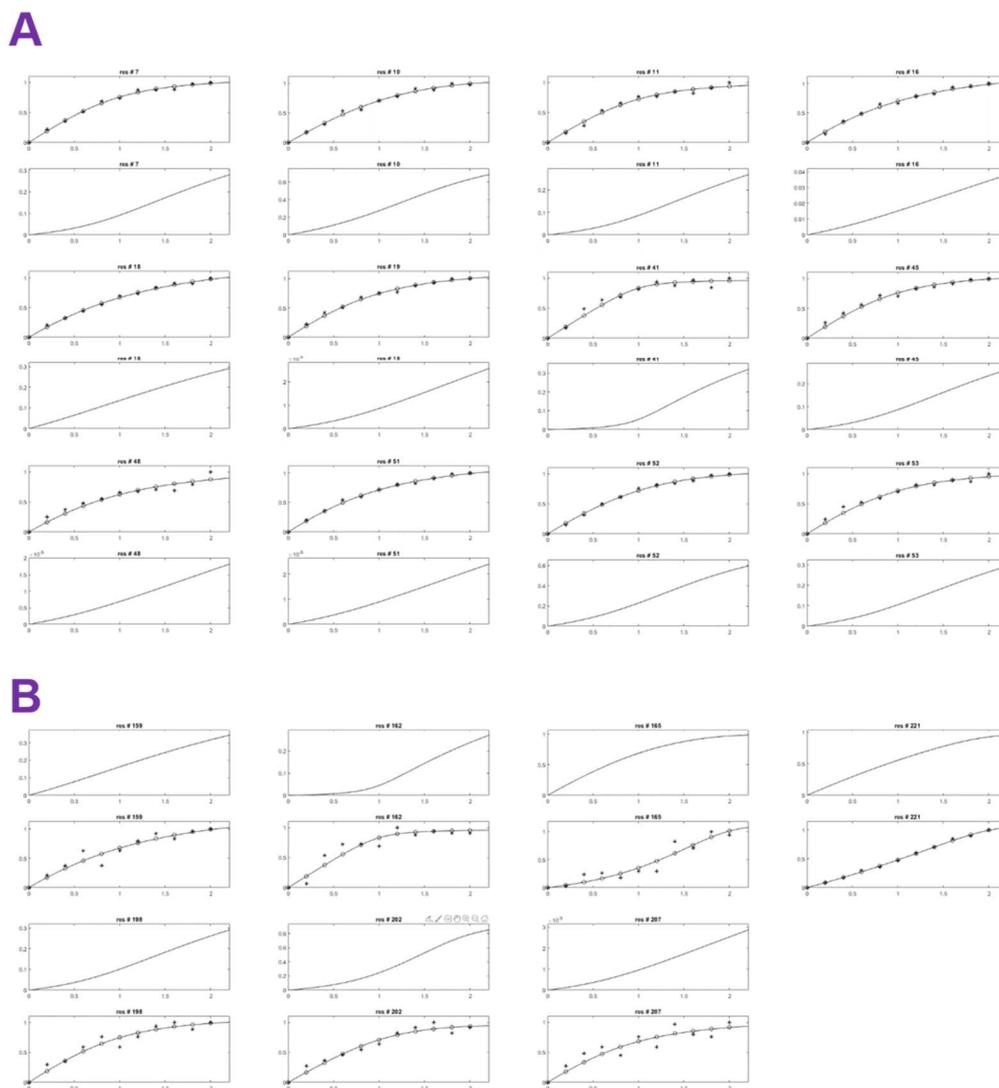
**Figure C5.**  $^1\text{H}$ - $^{15}\text{N}$  SOFAST-HMQC spectra of Grb2 and EpYINQSV-Grb2 interacting with SOS1 two-site peptides. Overlay of the spectra of Grb2 (blue) with (A) Grb2-S12, (C) Grb2-S23, and (E) Grb2-S45; the spectra of EpYINQSV-Grb2 (blue) superimposes (B) EpYINQSV-Grb2-S12, (D) EpYINQSV-Grb2-S23, and (F) EpYINQSV-Grb2-S45. SOS1 peptides bind to Grb2 and complex of EpYINQSV-Grb2 with 2:1 stoichiometry.



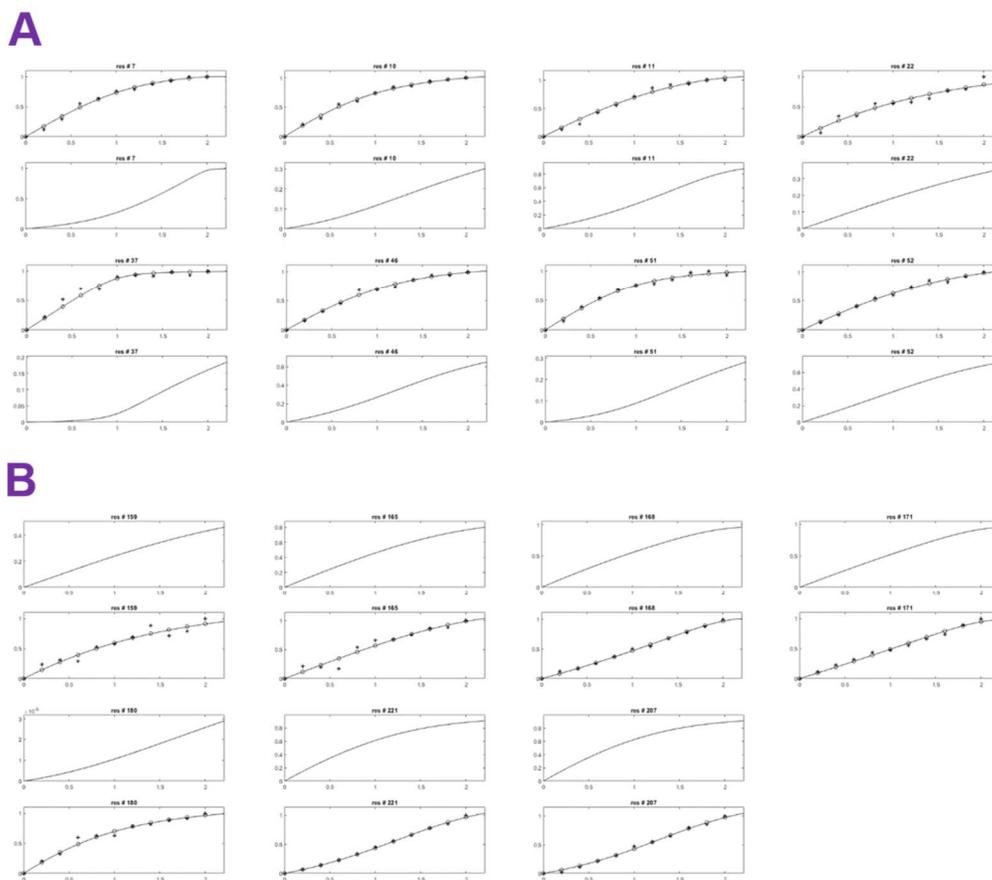
**Figure C6.** Titration curves for the selected residues of Grb2 (A) nSH3 and (B) cSH3 upon addition of S1\* peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).



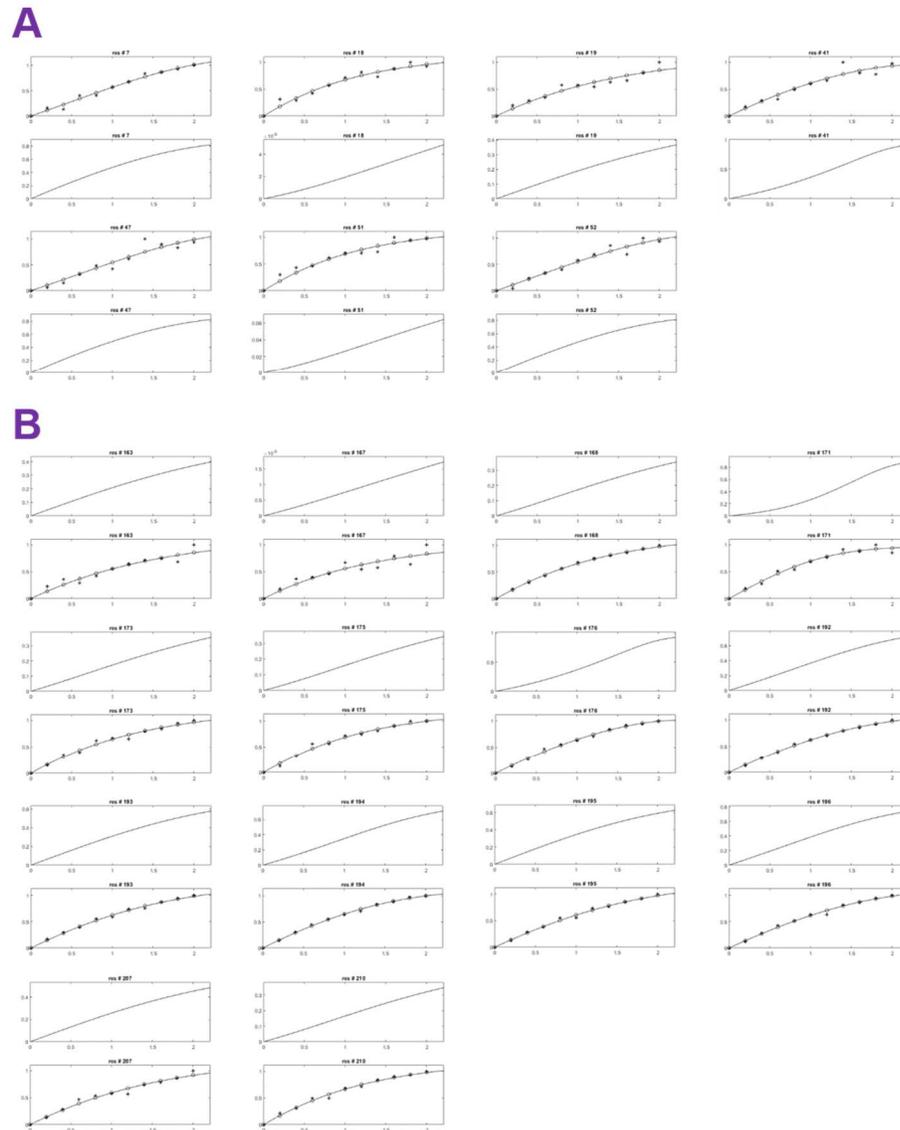
**Figure C7.** Titration curves for the selected residues of Grb2 (A) nSH3 and (B) cSH3 upon addition of S2\* peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).



**Figure C8.** Titration curves for the selected residues of Grb2 (A) nSH3 and (B) cSH3 upon addition of S3\* peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).

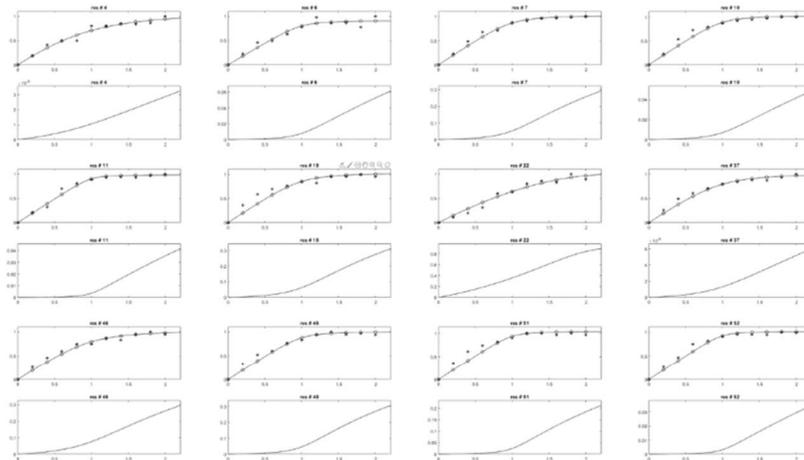


**Figure C9.** Titration curves for the selected residues of Grb2 (A) nSH3 and (B) cSH3 upon addition of S4\* peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).

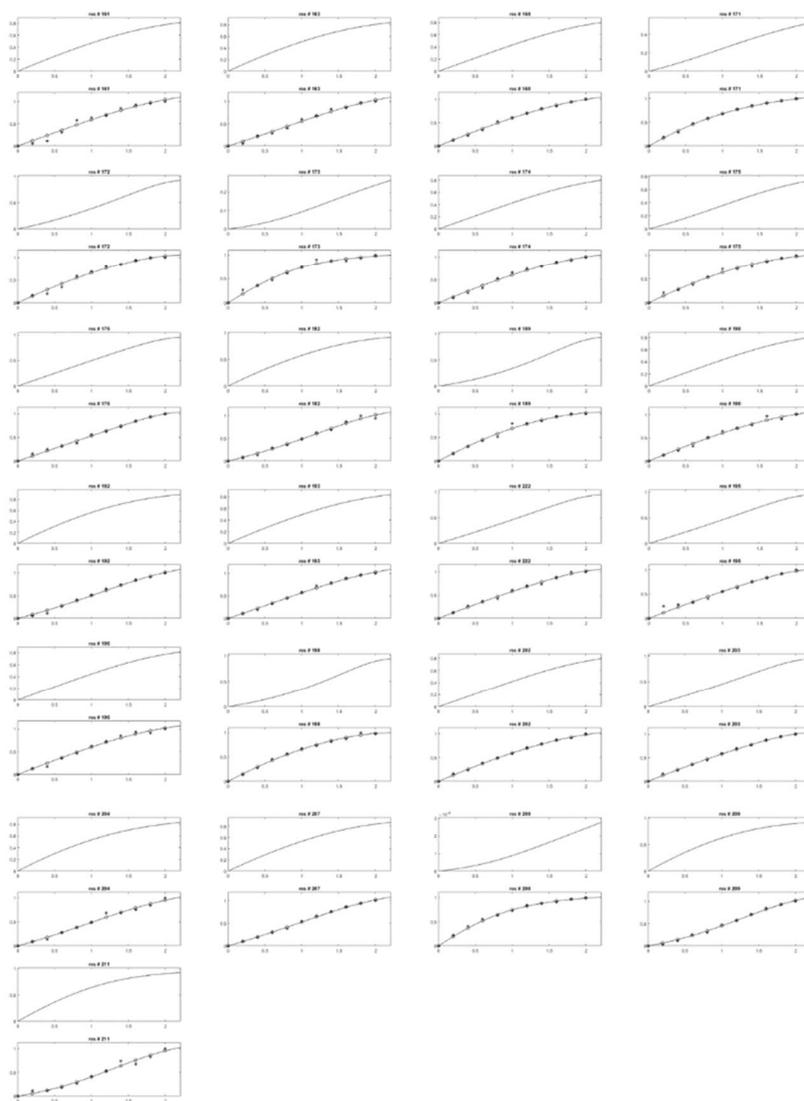


**Figure C10.** Titration curves for the selected residues of Grb2 (A) nSH3 and (B) cSH3 upon addition of S5 peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) The fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).

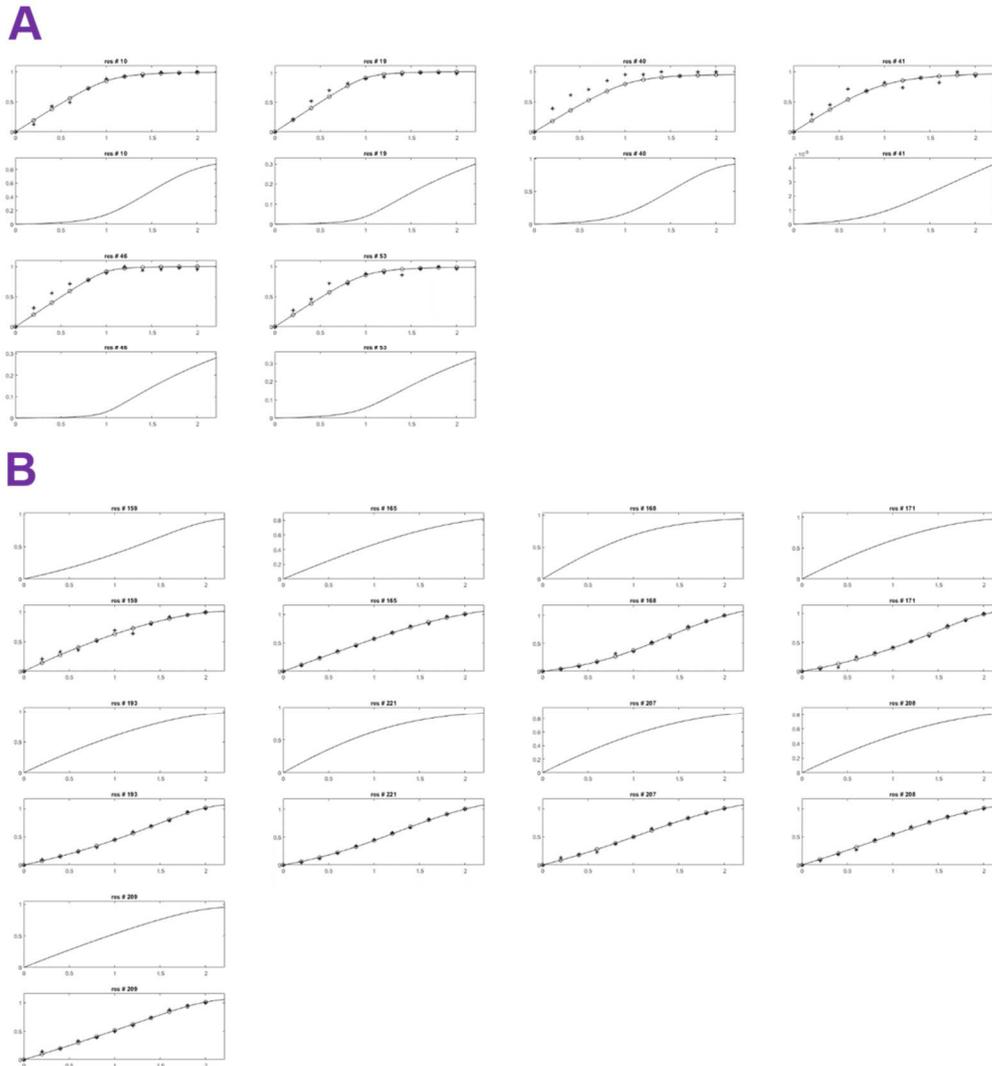
**A**



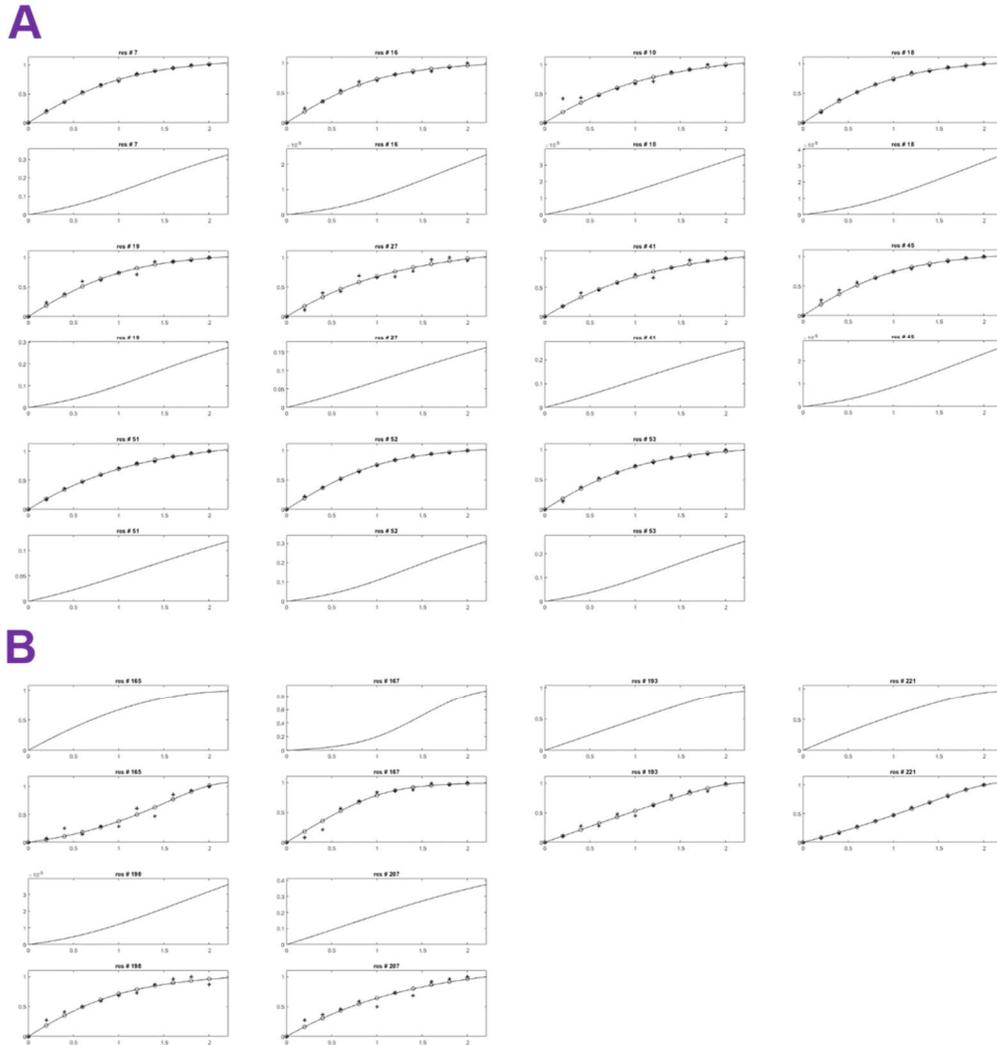
**B**



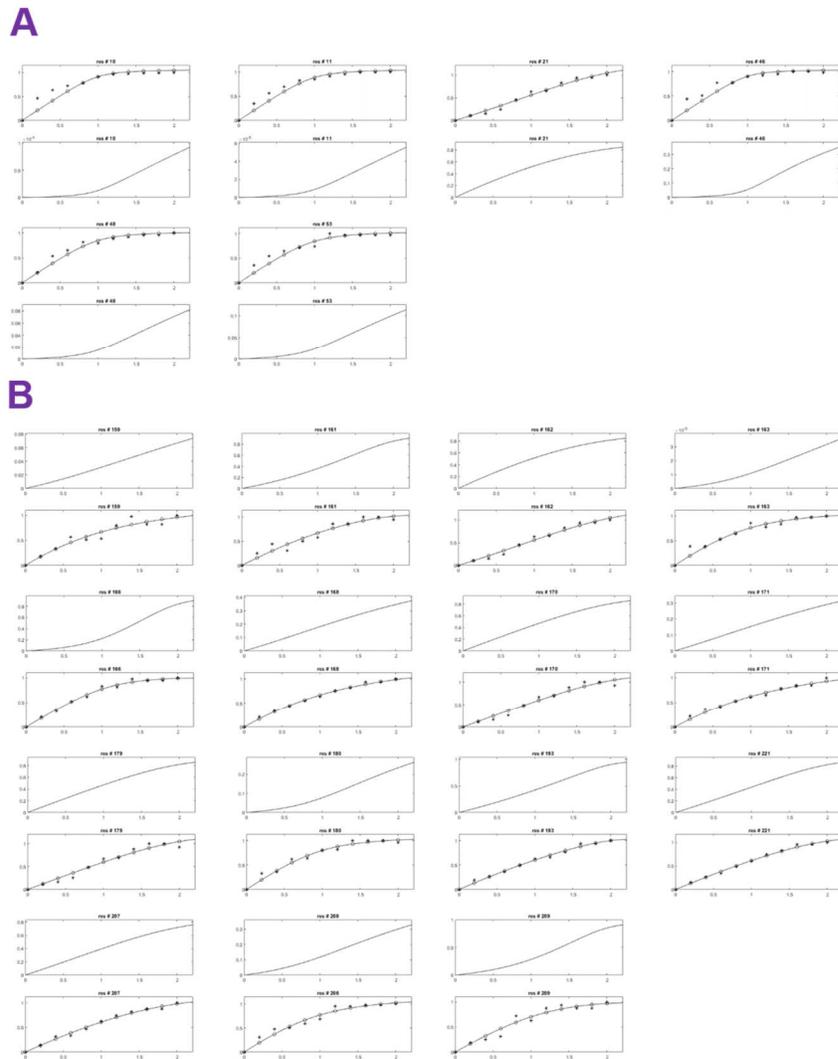
**Figure C11.** Titration curves for the selected residues of EpYINQSV–Grb2 (A) nSH3 and (B) cSH3 upon addition of S1\* peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).



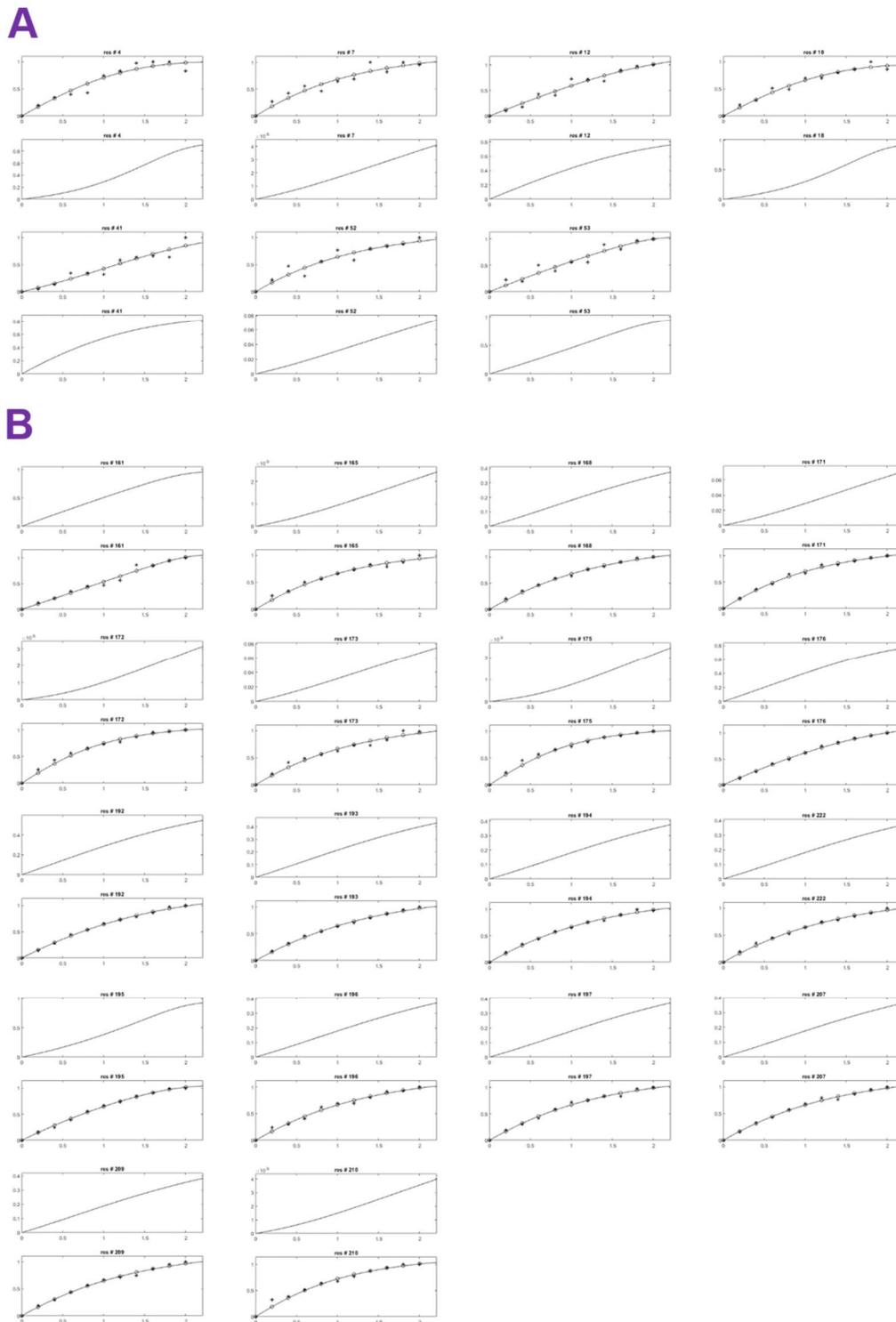
**Figure C12.** Titration curves for the selected residues of EpYINQSV–Grb2 (A) nSH3 and (B) cSH3 upon addition of S2\* peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).



**Figure C13.** Titration curves for the selected residues of EpYINQSV–Grb2 (A) nSH3 and (B) cSH3 upon addition of S3\* peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).



**Figure C14.** Titration curves for the selected residues of EpYINQSV–Grb2 (A) nSH3 and (B) cSH3 upon addition of S4\* peptide and fitted to a two-site binding model. The observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).



**Figure C15.** Titration curves for the selected residues of EpYINQSV-Grb2 (A) nSH3 and (B) cSH3 upon addition of S5 peptide and fitted to a two-site binding model. The

observed CSPs are plotted as a function of peptide/protein ratio. The vertical axis is fractional bound of protein. The corresponding residue number is indicated above each subplot. (A) The fits of CSPs observed in nSH3 domain (upper) reflect the corresponding binding events for cSH3 domain (lower), which are not directly observed. (B) Similarly, the fits for cSH3 domain (lower) also reflect the indirect bindings for nSH3 domain (upper).

Table C1. Signal attenuation while the titration of SOS1 peptides with Grb2 and EpYINQSV-Grb2

SOS1	Grb2	EpYINQSV-Grb2
S1*	K38, K50	W36ε, K38, F47, K50
S2*	K38, F47, K50	K38, F47, K50
S3*	Y37, K38, F47, K50	D15, Y37, K38, F47, K50
S4*	K38, K50	D15, K38, F47, K50
S5	W36ε, K38	D15, W36ε, K38
S12	A11, W36ε, Y37, K38, F47, K50, R207	W36ε, Y37, K38, F47, K50, R207
S23	A11, W36ε, Y37, K38, F47, R207	A11, W36ε, Y37, K38, F47, K50, D168, W193, R207
S45	K10, W36ε, K38, F47, K50, L164, D168, W194, R207, N208	K10, T12, W36ε, Y37, K38, F47, I48, K50, L175, H184, N192, W193, W194

Table C2. <sup>15</sup>N T2 relaxation measurements for Grb2 and EpYINQSV–Grb2

	Grb2	EpYINQSV–Grb2		Grb2	EpYINQSV–Grb2		Grb2	EpYINQSV–Grb2
M1	439.75	297.78	R78	22.19	42.73	D150	40.66	30.81
E2	358.11	525.31	H79	43.62	45.18	T159	30.06	44.30
A3	60.64	40.74	D80	45.98	43.46	Y160	36.63	43.75
I4	51.44	35.84	G81	42.10	29.77	V161	54.24	43.26
A5	54.77	58.76	A82	34.02	39.82	Q162	64.53	41.42
K6	37.18	36.51	F83	34.30	37.91	A163	46.92	49.78
Y7	49.50	48.92	L84	42.68	42.61	L164	87.65	52.94
D8	35.90	42.86	I85	27.75	35.08	F165	43.53	59.46
F9	47.58	55.57	R86	39.52	29.31	D166	54.83	56.55
K10	57.98	38.23	E87	42.19	46.77	F167	46.23	47.74
A11	86.94	73.45	A91	60.57	51.26	D168	62.17	65.79
T12	47.21	38.68	G93	33.57	36.66	Q170	45.46	48.87
D15	55.47	88.25	S96	37.93	33.18	E171	65.02	62.81
D16	39.44	39.92	L97	38.25	56.49	D172	49.16	49.08
S18	62.38	46.85	S98	28.96	35.37	G173	45.21	82.86
F19	40.65	42.20	V99	42.59	45.05	E174	51.94	48.30
K20	26.71	34.34	K100	42.39	43.50	L175	47.70	50.87
R21	64.53	41.42	F101	58.11	45.92	G176	56.17	60.39
G22	28.84	38.92	G102	42.04	38.17	F177	41.78	48.00
D23	44.74	40.22	D104	94.50	51.76	R178	34.28	47.26
I24	52.80	40.86	Q106	37.00	48.26	R179	45.46	49.53
L25	58.53	46.74	H107	41.94	56.42	F180	50.24	57.11
V27	43.59	31.52	F108	33.64	27.45	F182	46.40	47.49
L28	19.81	35.79	K109	39.75	-	I183	34.75	55.79
Y37	16.47	21.38	V110	51.45	75.31	H184	49.55	64.71
K38	37.43	45.05	L111	36.89	46.08	V185	54.77	58.76
A39	35.82	34.20	R112	68.12	-	M186	38.39	44.06
E40	42.79	51.06	A115	44.91	53.17	D187	46.00	45.87
L41	44.76	48.18	G116	39.42	40.28	N188	49.63	64.71
G43	55.83	43.96	Y118	33.29	41.12	S189	46.95	50.90
K44	45.02	42.82	F119	50.37	50.36	D190	51.53	52.72
D45	46.67	44.90	W121	39.10	-	N192	44.47	49.58
G46	36.58	44.90	V122	81.98	-	W193	49.59	50.58
F47	41.54	32.99	V123	64.63	77.72	W194	42.56	44.17
I48	22.15	32.85	K124	40.13	42.35	K195	42.90	38.06
K50	52.79	40.51	F125	24.00	43.96	G196	46.79	45.97
N51	37.78	37.43	N126	50.87	38.82	A197	47.04	41.95
Y52	43.18	44.57	S127	46.20	39.42	C198	43.85	44.45
I53	42.43	41.05	L128	44.41	39.74	Q201	57.93	65.46
E54	43.85	44.45	E130	37.21	39.80	T202	51.81	55.89
K64	45.14	49.78	L131	40.97	12.40	G203	49.91	57.55
I65	58.79	54.77	D133	37.28	48.63	M204	44.96	45.40
A68	39.32	41.12	Y134	43.81	44.11	F205	44.37	44.97
K69	40.80	41.95	R136	35.52	30.74	R207	49.24	62.64
E71	41.23	71.03	S137	24.89	61.48	N208	35.89	45.13
M73	39.50	42.95	T138	44.52	33.60	Y209	54.00	43.96
L74	42.53	48.85	S141	56.08	54.15	V210	45.17	47.83
S75	48.13	41.92	Q144	41.72	49.35	T211	45.46	56.50
K76	47.57	34.98	I146	40.01	43.71	V213	102.08	54.13
Q77	47.36	49.95	F147	42.89	33.36	V217	336.90	298.29

Table C3. Direct binding affinities ( $K_D$  in unit  $\mu\text{M}$ ) of individual residue for interaction of Grb2 with one-site SOS1 peptides

nSH3	S1*	S2*	S3*	S4*	S5	cSH3	S1*	S2*	S3*	S4*	S5
Y7		0.8	6.8	0.01	8.9	T159		7.8	26.8	30.6	
K10	1.2	0.9	3.3	11.1		V161	12.2				
A11	1.4		6.7	0.9		A163	1.3				43.3
T12		1.6				F165			1.3	9.2	
E16	2.1		30.7			D166	51.3				
S18			30.2		36.8	F167					63.8
F19	6.9	2.4	16.8		45.3	D168	9.5	1.2		1.1	29.7
G22				43.4		E171	35.8	10.2		1.0	0.3
Y37		0.1		1.0		D172	5.9				
E40		10.6				G173	0.2				31.4
L41	5.2		1.0		0.6	E174	30.2				
D45	1.1		7.5			L175					24.7
G46	2.7	1.8		4.4		G176					0.5
F47					9.8	G180				24.3	
I48	2.9		27.6			S189	10.9				
N51		0.1	25.9	6.1	34.1	D190	24.8				
Y52	1.3		4.0	7.4	9.1	N192	1.1				7.5
I53		3.6	8.8			W193	11.9	11.4			22.5
						W193 $\epsilon$		11.8	1.3	10.6	
						W194					7.0
						W194 $\epsilon$	1.0				
						K195					21.0
						G196	13.9				7.9
						C198			8.1		
						T202	9.4		0.4		
						G203	12.1				
						M204	8.1				
						R207	14.1	1.2	16.6	13.3	32.5
						N208	26.6	7.9			
						V210					27.8

Table C4. Indirect binding affinities ( $K_D$  in unit  $\mu\text{M}$ ) of individual residue for interaction of Grb2 with one-site SOS1 peptides

nSH3	S1*	S2*	S3*	S4*	S5	cSH3	S1*	S2*	S3*	S4*	S5
Y7		107	137	0.1	7.0	T159		12.7	115	66	
K10	2522	142	14.7	129		V161	3.6				
A11	$6 \times 10^5$		147	2.7		A163	0.5				95.6
T12		$6 \times 10^5$				F165			0.3	8.2	
E16	189		1891			D166	7.1				
S18			154		$2 \times 10^6$	F167					$5 \times 10^6$
F19	540	249	$3 \times 10^6$		113	D168	221	0.5		0.7	112
G22				116		E171	5.1	2.8		0.8	2.2
Y37		2.8		240		D172	2.0				
E40		$2 \times 10^6$				G173	111				113
L41	130		99		1.6	E174	9.6				
D45	96		154			L175					115
G46	112	313		18.7		G176					1.4
F47					6.8	G180				$3 \times 10^6$	
I48	204		$4 \times 10^6$			S189	32				
N51		2.6	$3 \times 10^6$	137	$1 \times 10^3$	D190	0.7				
Y52	2418		24.7	13.1	7.4	N192	7.0				13.1
I53		$1 \times 10^6$	130			W193	5.6	2.5			36.0
						W193 $\epsilon$		2.5	0.8	3.0	
						W194					14.3
						W194 $\epsilon$	7.2				
						K195					27.5
						G196	8.5				12.0
						C198			131		
						T202	5.6		3.5		
						G203	11.2				
						M204	5.5				
						R207	45.3	0.7	$3 \times 10^6$	3.2	60.5
						N208	1.0	2.9			
						V210					116

Table C5. Direct binding affinities ( $K_D$  in unit  $\mu\text{M}$ ) of individual residue for interaction of EpYINQSV–Grb2 with one-site SOS1 peptides

nSH3	S1*	S2*	S3*	S4*	S5	cSH3	S1*	S2*	S3*	S4*	S5
I4	21.1				0.4	T159		0.6		42.6	
K6	1.6					V161	12.7			0.7	1.0
Y7	1.8		11.7		43.1	Q162				10.6	
K10	2.1	0.1	40.1	1.8		A163	14.0			13.6	
A11	0.7			2.4		F165		9.3	1.3		36.3
T12					14.5	D166				0.2	
E16			13.5			F167			0.2		
S18	2.2		16.7		0.3	D168	8.7	12.3		31.2	30.2
F19		0.7	12.1			E171	17.0	1.3		41.0	35.2
R21				10.6		D172	0.7				17.7
G22	1.0					G173	13.0				45.2
V27			38.7			E174	8.6				
Y37	6.5					L175	7.2				15.7
E40		0.1				G176	1.0				8.1
L41		5.1	34.2		30.1	G180				5.9	
D45			17.6			F182	6.7				
G46	4.4	0.4		0.9		S189	0.4				
I48	1.1			2.8		D190	8.6				
N51	0.8		37.8			N192	11.5				25.7
						W193	11.7	1.3	1.0	0.6	31.3
						W193 $\epsilon$		13.2	1.1	2.8	
						W194					31.1
						W194 $\epsilon$	0.9				33.4
						K195	0.9				0.6
						G196	8.8				30.9
						A197					30.1
						C198	0.4		20.1		
						T202	8.7				
						G203	0.8				
						M204	23.9				
						R207	10.6	9.6	35.0	8.0	32.4
						N208	19.9	15.5		9.2	
						Y209	13.1	1.3		0.3	33.5
						V210					30.9
						T211	14.8				

Table C6. Indirect binding affinities ( $K_D$  in unit  $\mu\text{M}$ ) of individual residue for interaction of EpYINQSV–Grb2 with one-site SOS1 peptides

nSH3	S1*	S2*	S3*	S4*	S5	cSH3	S1*	S2*	S3*	S4*	S5
I4	$3 \times 10^6$				2.1	T159		1.4		1080	
K6	1288					V161	10.4			2.1	0.9
Y7	165		128		$2 \times 10^6$	Q162				6.3	
K10	1680	2.6	$2 \times 10^6$	$8 \times 10^5$		A163	8.5			$2 \times 10^6$	
A11	1956			$1 \times 10^6$		F165		7.8	0.3		$4 \times 10^6$
T12					13.3	D166				2.1	
E16			$3 \times 10^6$			F167			2.7		
S18	148		$2 \times 10^6$		1.6	D168	10.4	1.9		113	112
F19		126	169			E171	57.6	0.4		160	1100
R21				6.3		D172	1.8				$3 \times 10^6$
G22	3.0					G173	220				1080
V27			418			E174	10.5				
Y37	$2 \times 10^6$					L175	15.3				$3 \times 10^6$
E40		1.5				G176	1.0				11.3
L41		$2 \times 10^6$	224		10.1	G180				171	
D45			$3 \times 10^6$			F182	2.9				
G46	167	141		95.6		S189	1.5				
I48	151			768		D190	10.2				
N51	271		619			N192	5.1				47.8
						W193	8.4	0.6	1.0	1.1	86.7
						W193 $\epsilon$		3.3	0.6	3.9	
						W194					112
						W194 $\epsilon$	1.1				116
						K195	1.2				1.5
						G196	9.6				114
						A197					114
						C198	1.3		$2 \times 10^6$		
						T202	11.2				
						G203	1.2				
						M204	10.2				
						R207	5.8	4.2	115	11.9	119
						N208	$4 \times 10^6$	8.6		122	
						Y209	3.5	1.0		1.8	110
						V210					$2 \times 10^6$
						T211	3.4				

Table C7. Signal intensity ratios of Grb2-S1\*

Residue	I/I <sub>0</sub>								
1	0.81	47	0.97	102	0.92	164	0.77	203	0.80
2	0.96	48	1.02	106	0.84	165	0.54	204	1.02
3	0.95	50	0.00	107	0.97	167	0.78	205	0.40
4	0.83	51	0.91	108	0.75	168	0.49	207	0.56
6	0.95	52	0.68	110	0.65	170	0.58	208	0.89
7	0.71	53	0.79	111	0.72	171	0.75	209	0.62
8	1.19	65	0.78	115	0.82	173	0.62	210	0.74
9	0.81	68	0.84	116	0.75	174	0.89	211	0.84
10	2.05	69	0.62	118	0.86	175	0.83	213	0.84
11	0.71	73	0.79	119	0.69	176	0.72	217	0.86
12	1.15	74	0.75	123	0.85	177	0.82	60ε	0.82
15	3.09	75	0.65	124	0.73	178	0.78	221	0.11
16	0.89	76	0.65	125	0.94	179	0.6	222	0.85
18	1.08	78	1.28	126	0.76	180	0.74		
19	1.04	79	0.84	127	0.70	181	0.75		
20	0.92	80	0.8	130	0.57	183	0.81		
22	0.95	81	0.73	131	0.73	184	0.97		
23	0.83	82	0.73	133	0.74	186	0.88		
24	0.81	84	0.82	134	0.81	187	0.73		
25	0.99	85	0.73	136	0.72	188	0.91		
27	0.77	86	0.73	137	0.81	189	1.15		
28	1.22	87	0.98	138	0.84	190	0.75		
37	1.51	91	0.72	141	0.76	192	0.85		
38	0.00	93	0.78	144	0.86	193	0.76		
39	0.96	96	0.76	146	0.74	194	0.90		
40	0.98	97	0.92	147	0.69	195	0.94		
43	0.81	98	0.77	150	1.17	196	0.86		
44	0.78	99	0.82	159	1.18	197	1.02		
45	0.76	100	0.87	160	0.77	201	0.88		
46	0.86	101	0.76	161	1.24	202	0.89		

Table C8. Signal intensity ratios of Grb2-S2\*

Residue	I/I <sub>0</sub>								
1	0.80	51	0.69	107	0.89	165	0.63	207	0.95
2	0.80	52	0.66	108	0.85	167	0.9	208	1.10
3	1.37	53	0.70	110	1.03	168	0.82	209	0.76
4	0.52	64	0.98	111	0.70	170	0.78	210	0.97
6	1.03	65	0.83	115	0.77	171	0.92	211	0.98
7	0.76	68	0.85	116	0.90	173	0.91	213	0.78
9	0.93	69	0.81	118	0.94	174	0.93	217	0.70
10	3.19	71	0.77	119	0.79	175	0.99	219	0.83
12	1.39	73	0.80	123	0.84	176	0.90	221	0.35
15	2.80	74	0.75	124	0.80	178	0.88	222	0.90
16	1.47	75	0.76	125	0.86	179	0.80		
18	1.07	76	0.88	126	0.91	180	0.90		
19	1.16	78	1.21	127	0.86	181	0.95		
20	1.12	79	0.82	130	0.89	183	0.93		
22	1.06	80	0.97	131	0.92	184	1.08		
23	0.86	81	0.91	133	0.79	186	0.97		
24	0.81	82	0.88	134	0.83	187	1.04		
25	0.89	84	0.88	136	0.83	189	1.14		
27	1.04	85	1.05	137	0.89	190	0.90		
28	0.94	86	0.95	138	0.91	192	0.99		
37	1.43	87	0.89	141	0.83	193	0.9		
38	0.00	91	0.78	144	0.94	194	1.10		
40	0.88	93	0.88	146	0.87	195	0.84		
43	0.89	96	0.71	147	0.90	196	0.84		
44	0.91	97	0.82	150	0.91	197	0.80		
45	0.83	98	0.84	159	1.34	201	1.00		
46	0.92	99	0.86	160	1.59	202	0.93		
47	0.00	101	0.90	161	1.01	203	0.88		
48	0.81	102	0.90	163	0.91	204	0.96		
50	0.00	106	0.89	164	1.01	205	0.88		

Table C9. Signal intensity ratios of Grb2-S3\*

Residue	I/I <sub>0</sub>								
1	0.87	48	0.88	102	0.94	163	1.18	203	1.02
2	0.78	50	0.00	106	0.83	164	0.99	204	1.10
3	1.29	51	0.57	107	0.98	165	0.90	205	1.10
4	0.64	52	0.96	108	0.92	167	0.93	207	1.14
6	0.84	53	0.66	110	0.83	168	1.02	208	1.24
7	0.69	64	0.99	111	0.87	171	1.05	209	0.94
9	1.11	65	0.88	115	0.84	172	1.16	210	1.13
10	2.35	68	0.93	116	0.87	173	1.10	211	0.99
11	0.58	69	0.92	118	0.90	174	1.13	213	0.87
12	1.55	73	0.95	119	0.85	175	1.00	217	0.82
15	2.11	74	0.80	123	0.97	176	0.95	218	0.34
16	1.01	75	0.87	124	0.94	177	1.01	219	0.82
18	1.00	76	1.08	125	0.90	178	1.03	220	0.90
19	1.02	78	1.17	126	1.03	180	1.09	221	0.90
20	0.93	79	0.88	127	0.82	181	1.12	222	0.92
22	0.88	80	0.90	130	0.98	183	1.08		
23	0.92	81	0.92	131	0.86	184	1.17		
24	0.86	82	0.89	133	0.82	186	0.98		
25	0.87	84	0.92	134	0.86	187	1.13		
27	0.72	85	0.93	136	0.88	188	1.32		
28	0.88	86	0.93	137	1.00	189	1.04		
37	0.00	87	0.81	138	1.03	190	1.06		
38	0.00	91	0.91	141	0.84	192	1.05		
39	0.42	93	0.93	144	0.98	193	1.09		
40	0.82	96	1.01	146	0.92	194	0.91		
43	0.94	97	0.78	147	1.11	195	1.04		
44	0.88	98	0.92	150	0.98	196	1.02		
45	0.84	99	0.97	159	1.39	197	0.95		
46	0.88	100	0.88	160	1.58	201	1.12		
47	0.00	101	0.92	161	1.19	202	0.97		

Table C10. Signal intensity ratios of Grb2-S4\*

Residue	I/I <sub>0</sub>								
1	0.93	48	1.06	102	0.96	163	0.83	203	1.01
2	1.07	50	0.00	106	0.92	164	0.88	204	1.13
3	1.24	51	0.98	107	1.07	165	0.83	205	0.78
4	0.94	52	0.89	108	0.90	167	0.86	207	0.88
6	0.77	53	0.82	110	1.06	168	0.76	208	1.11
7	0.70	64	1.11	111	0.95	171	0.87	209	0.93
9	0.88	65	0.91	115	0.99	172	1.15	210	1.00
10	4.86	68	0.99	116	1.04	173	1.10	211	0.79
11	0.96	69	0.98	118	1.06	174	1.04	213	1.01
12	2.19	73	0.96	119	0.94	175	1.08	217	0.74
15	3.66	74	0.96	123	0.85	176	0.93	219	1.08
16	2.25	75	0.83	124	1.03	177	0.98	221	0.34
18	1.34	76	0.90	125	1.01	178	1.06	222	0.95
19	1.21	78	1.38	126	0.97	180	0.35		
20	1.19	79	0.95	127	1.02	181	0.86		
22	0.62	80	0.94	130	0.95	183	0.93		
23	0.95	81	0.93	131	1.17	184	0.98		
24	0.99	82	1.03	133	0.93	186	1.22		
25	1.03	84	1.03	134	0.99	187	1.04		
27	0.87	85	1.07	136	1.02	188	1.24		
28	1.21	86	1.06	137	0.98	189	1.22		
37	1.84	87	1.06	138	1.04	190	0.95		
38	0.00	91	0.85	141	0.85	192	0.98		
39	0.94	93	0.98	144	1.05	193	0.88		
40	1.04	96	1.06	146	0.96	194	1.10		
43	1.02	97	0.89	147	0.99	195	0.92		
44	1.01	98	1.04	150	1.09	196	1.02		
45	1.01	99	1.00	159	0.92	197	0.96		
46	1.05	100	1.06	160	0.50	201	1.08		
47	1.30	101	0.99	161	1.12	202	0.92		

Table C11. Signal intensity ratios of Grb2-S5

Residue	I/I <sub>0</sub>								
1	0.93	48	0.85	102	0.88	164	0.91	204	0.74
2	1.14	50	0.80	106	0.80	165	0.54	205	0.36
3	0.93	51	0.68	107	0.81	166	0.81	207	0.14
4	0.81	52	0.93	108	0.79	167	0.75	208	0.88
6	0.97	53	0.79	110	0.76	168	0.68	209	0.64
7	0.71	64	1.01	111	0.72	171	0.72	210	0.82
8	1.31	65	0.75	115	0.88	173	0.64	211	0.85
9	0.93	68	0.85	116	0.78	174	0.86	213	0.81
10	1.63	69	0.70	118	0.76	175	0.72	217	0.78
11	0.95	73	0.76	119	0.68	176	0.70	218	0.70
12	1.06	74	0.81	123	0.83	177	0.80	219	0.83
16	0.77	75	0.80	124	0.85	178	0.90	222	0.72
18	0.90	76	0.68	125	0.70	180	0.87		
19	1.01	78	0.95	126	0.86	181	0.76		
20	1.00	79	0.80	127	0.72	182	0.85		
22	1.06	80	0.84	130	0.64	183	0.94		
23	0.86	81	0.78	131	0.53	184	0.90		
24	0.90	82	0.72	133	0.90	186	0.88		
25	0.91	84	0.78	134	0.80	187	0.57		
27	0.71	85	0.84	136	0.71	188	0.77		
28	0.99	86	0.78	137	0.81	190	0.73		
37	1.01	87	0.94	138	0.82	192	0.66		
38	0.00	91	0.69	141	0.69	193	0.58		
39	0.71	93	0.76	144	0.85	194	0.48		
40	0.86	96	0.79	146	0.78	195	0.88		
43	0.91	97	0.81	147	0.76	196	0.78		
44	0.82	98	0.81	159	0.95	197	0.79		
45	0.77	99	0.83	160	0.93	201	0.93		
46	0.72	100	0.78	161	0.83	202	0.77		
47	0.59	101	0.82	163	0.88	203	0.76		

Table C12. Signal intensity ratios of Grb2-S12

Residue	I/I <sub>0</sub>								
1	0.76	69	0.78	123	0.79	182	0.72		
2	0.82	73	0.75	124	0.75	183	0.67		
3	1.24	74	0.66	127	0.76	184	0.68		
4	0.63	75	0.69	130	0.77	186	0.56		
6	0.87	76	0.79	131	0.79	187	0.67		
9	0.85	77	0.77	133	0.7	188	1.36		
10	2.12	78	1.23	134	0.67	190	0.64		
11	0.00	79	0.71	136	0.67	192	0.60		
12	0.69	80	0.71	137	0.78	194	0.73		
16	0.89	81	0.61	138	0.66	196	0.46		
18	0.73	82	0.63	141	0.81	197	0.67		
19	0.78	83	0.7	144	0.82	201	0.8		
20	0.78	84	0.7	146	0.84	202	0.74		
22	0.72	85	0.63	147	0.76	203	0.53		
23	0.72	86	0.69	150	1.38	207	0.00		
24	0.75	87	0.76	159	1.61	208	0.78		
25	0.56	91	0.74	160	0.50	210	0.7		
27	0.24	96	0.57	161	0.73	211	0.73		
38	0.00	97	0.78	164	0.63	213	0.57		
40	0.85	98	0.69	167	0.64	217	0.89		
43	0.74	99	0.67	168	0.00	219	0.90		
44	0.8	101	0.75	171	0.47	222	0.78		
45	0.67	102	0.78	173	0.39				
46	0.64	106	0.66	174	0.78				
47	0.00	107	0.75	175	0.52				
50	0.00	108	0.57	176	0.55				
51	0.48	111	0.62	177	0.68				
53	0.54	115	0.78	178	0.68				
65	0.85	116	0.75	180	0.32				
68	0.76	119	0.67	181	0.64				

Table C13. Signal intensity ratios of Grb2-S23

Residue	I/I <sub>0</sub>								
1	0.88	51	0.31	110	0.59	172	0.69	218	0.39
2	0.93	53	0.28	111	0.56	174	0.67	219	0.66
3	0.90	65	0.64	115	0.70	175	0.42	222	0.73
4	0.43	68	0.59	116	0.57	176	0.49		
6	0.65	69	0.62	118	0.54	177	0.39		
7	0.67	73	0.60	119	0.54	178	0.58		
8	0.22	74	0.54	123	0.69	180	0.40		
9	0.68	75	0.55	124	0.65	181	0.68		
10	0.56	76	0.54	125	0.55	183	0.44		
11	0.00	77	0.65	127	0.58	184	0.41		
12	0.53	78	0.78	130	0.42	186	0.60		
15	1.22	79	0.63	131	0.55	187	0.51		
16	0.58	80	0.66	133	0.52	188	1.35		
18	0.41	81	0.54	134	0.61	189	0.86		
19	0.49	82	0.54	136	0.50	190	0.49		
20	0.55	83	0.49	137	0.51	192	0.57		
22	0.52	84	0.59	138	0.53	194	0.39		
23	0.59	85	0.62	141	0.53	195	0.45		
24	0.50	86	0.61	144	0.69	196	0.54		
25	0.42	87	0.58	146	0.55	197	0.47		
37	0.00	91	0.61	147	0.55	201	0.71		
38	0.00	93	0.57	150	0.71	202	0.48		
40	0.54	96	0.52	159	1.62	203	0.32		
43	0.50	97	0.50	160	0.58	204	0.42		
44	0.54	98	0.51	161	0.70	207	0.00		
45	0.48	99	0.54	164	0.45	208	0.23		
46	0.42	101	0.53	165	1.22	210	0.39		
47	0.00	102	0.67	167	0.67	211	0.45		
48	0.43	107	0.61	168	0.15	213	0.87		
50	0.50	108	0.57	171	0.28	217	0.68		

Table C14. Signal intensity ratios of Grb2-S45

Residue	I/I <sub>0</sub>								
1	0.89	51	0.53	108	0.53	176	0.41		
2	0.80	52	0.27	110	0.60	177	0.32		
6	0.69	53	0.37	111	0.60	178	0.84		
7	0.75	65	0.56	115	0.74	181	0.45		
8	0.26	68	0.61	116	0.61	183	0.36		
9	0.57	69	0.59	119	0.58	186	0.34		
10	0.00	73	0.62	123	0.50	187	0.26		
11	0.45	74	0.55	124	0.70	188	0.74		
12	0.19	75	0.65	125	0.53	189	0.78		
15	0.66	76	0.66	127	0.67	192	0.00		
16	0.67	77	0.67	130	0.63	194	0.00		
18	0.42	79	0.60	131	0.51	195	0.23		
20	0.61	80	0.59	133	0.63	196	0.38		
22	0.42	81	0.60	134	0.60	197	0.46		
23	0.68	82	0.61	136	0.60	201	0.57		
24	0.56	83	0.55	137	0.64	202	0.38		
25	0.55	85	0.51	138	0.56	203	0.23		
27	0.51	86	0.50	141	0.63	207	0.00		
28	0.59	87	0.67	144	0.69	208	0.00		
37	0.39	91	0.67	146	0.57	209	0.54		
38	0.00	93	0.58	147	0.65	210	0.71		
39	0.50	96	0.56	159	0.41	211	0.25		
40	0.47	97	0.56	164	0.00	213	0.68		
43	0.61	98	0.63	165	1.51	217	0.35		
44	0.62	99	0.58	167	0.59	218	0.33		
45	0.51	100	0.54	168	0.00	219	0.49		
46	0.17	101	0.63	171	0.20	222	0.36		
47	0.00	102	0.64	173	0.31				
48	0.23	106	0.60	174	0.60				
50	0.00	107	0.64	175	0.00				

Table C15. Signal intensity ratios of Grb2-EGFR

Residue	I/I <sub>0</sub>								
1	0.95	47	0.95	99	0.91	161	0.96	196	1.03
2	1.01	48	0.88	100	0.97	163	1.10	197	1.00
3	1.22	50	0.79	101	1.00	164	1.05	201	1.01
4	1.14	51	0.92	102	0.57	165	0.99	202	0.98
6	0.74	52	0.80	106	0.56	166	1.00	203	0.99
7	0.96	53	0.89	107	0.90	167	0.99	204	1.07
8	0.98	64	1.22	108	0.49	168	0.99	205	1.09
9	1.08	65	0.91	110	0.81	170	1.13	207	1.02
10	0.89	69	0.89	111	0.54	171	0.99	208	1.32
11	0.99	71	0.35	115	0.86	173	1.25	209	0.21
12	0.90	73	0.90	116	0.97	174	0.98	210	0.98
15	0.98	74	1.06	119	0.84	175	0.95	211	1.03
16	0.62	75	0.95	123	0.73	176	0.96	213	1.09
18	0.89	76	0.92	124	1.80	177	1.05	217	0.95
19	0.89	77	0.81	125	2.20	178	0.96	218	0.80
20	0.81	78	1.13	126	1.14	179	0.87	219	0.91
22	1.09	79	0.94	127	0.91	180	0.93	220	0.15
23	0.93	80	0.92	128	0.34	181	0.99	221	0.92
24	0.87	81	1.09	130	1.21	182	0.95	222	0.92
25	1.07	82	0.91	131	0.00	183	0.87		
27	0.61	83	0.66	133	1.75	184	0.95		
28	1.21	84	0.93	136	0.58	186	1.02		
37	0.73	85	0.84	137	0.75	187	0.88		
38	0.00	86	1.00	138	0.67	188	1.05		
39	0.89	87	0.97	141	0.46	189	1.34		
40	0.91	91	0.78	144	0.94	190	0.82		
43	1.02	93	0.00	146	0.57	192	0.84		
44	1.03	96	0.79	147	0.51	193	1.02		
45	0.94	97	0.89	159	0.94	194	1.01		
46	0.93	98	0.69	160	0.86	195	1.16		

Table C16. Signal intensity ratios of EpYINQSV–Grb2–S1\*

Residue	I/I <sub>0</sub>								
1	0.86	47	0.00	97	0.88	147	0.53	197	1.04
2	0.99	48	1.18	98	0.71	159	1.50	198	1.01
3	1.25	50	0.00	99	0.81	160	0.95	201	1.18
4	0.81	51	1.1.0	100	1.08	161	1.30	202	0.96
6	0.94	52	0.92	101	1.02	162	1.08	203	0.87
7	0.67	53	0.87	102	0.52	163	1.06	204	0.91
8	1.47	54	0.50	106	0.45	164	1.06	205	0.53
9	0.98	64	1.07	107	0.91	165	0.57	207	0.73
10	5.97	65	0.87	108	0.38	167	0.85	208	1.20
11	1.05	68	0.35	110	0.54	171	0.89	209	0.22
12	2.01	69	0.86	111	0.42	173	0.93	210	0.98
15	3.94	71	0.47	115	0.86	174	1.09	211	1.08
16	1.75	73	0.87	116	0.82	175	0.82	213	1.14
18	1.31	74	0.90	118	0.80	176	0.93	217	0.91
19	1.36	75	0.76	119	0.79	177	1.04	218	0.38
20	0.93	76	0.79	123	0.78	178	0.90	219	1.00
22	1.00	77	0.79	124	1.37	180	0.88	220	0.16
23	0.79	78	1.35	125	2.16	181	0.89	221	0.00
24	0.73	79	0.79	126	1.13	183	1.07	222	0.93
25	1.17	80	0.88	127	0.98	184	1.19		
27	0.27	81	0.85	128	0.42	186	3.39		
28	1.44	82	0.88	130	1.04	187	0.97		
37	2.05	83	0.45	131	0.30	188	1.28		
38	0.00	84	1.10	133	1.81	189	1.43		
39	0.98	85	0.82	136	0.59	190	0.74		
40	1.02	86	0.91	137	0.72	192	1.03		
43	1.06	87	0.83	138	0.70	193	0.87		
44	0.97	91	0.68	141	0.39	194	1.03		
45	0.99	93	0.59	144	0.90	195	1.16		
46	0.95	96	0.81	146	0.70	196	0.89		

Table C17. Signal intensity ratios of EpYINQSV–Grb2–S2\*

Residue	I/I <sub>0</sub>								
1	0.80	50	0.00	100	0.99	163	1.34	203	1.06
2	0.87	51	0.89	101	1.05	164	1.25	204	1.20
3	1.89	52	0.95	102	0.51	167	0.89	205	0.84
4	0.80	53	0.97	106	0.52	168	0.83	207	1.12
6	0.92	54	0.66	107	0.93	170	0.94	208	1.35
9	1.01	64	1.15	108	0.47	171	0.98	209	0.26
10	4.25	65	0.93	110	0.85	173	1.25	210	1.08
12	2.66	69	0.84	111	0.43	174	1.15	211	1.22
15	5.16	71	0.21	115	0.80	175	0.95	213	1.04
16	1.68	73	0.86	116	0.94	176	1.01	217	0.76
18	1.28	74	0.98	119	0.85	177	1.06	218	0.26
19	1.47	75	0.68	123	0.83	178	1.18	219	1.11
20	1.02	76	0.93	124	1.47	179	0.86	220	0.15
21	0.75	77	0.78	125	2.10	180	1.10	221	0.32
22	1.42	78	1.48	126	1.05	181	1.02	222	0.97
23	0.88	79	0.87	127	0.92	183	1.11		
24	0.78	80	0.91	128	0.34	184	1.29		
25	1.12	81	1.01	130	1.12	186	1.21		
27	0.93	82	0.86	133	1.72	187	1.19		
28	1.05	83	0.57	136	0.51	189	1.50		
37	1.59	84	1.08	137	0.88	190	0.84		
38	0.00	85	1.09	138	0.69	192	1.08		
39	0.26	86	0.91	141	0.44	193	1.18		
40	1.05	87	1.08	144	0.89	194	1.29		
43	1.08	91	0.72	146	0.64	195	1.21		
44	1.12	93	0.68	147	0.60	196	1.15		
45	0.92	96	0.81	159	1.72	197	1.10		
46	0.97	97	0.92	160	1.99	198	0.44		
47	0.00	98	0.82	161	1.29	201	1.18		
48	0.91	99	0.81	162	0.81	202	1.06		

Table C18. Signal intensity ratios of EpYINQSV–Grb2–S3\*

Residue	I/I <sub>0</sub>								
1	0.89	50	0.00	101	0.97	163	1.26	197	1.06
2	0.80	51	0.62	102	0.54	164	1.15	198	0.96
3	1.62	52	0.85	106	0.48	165	0.83	201	1.11
4	0.81	53	0.91	107	1.12	167	0.89	202	1.01
5	0.88	54	0.30	108	0.42	168	0.99	203	1.01
6	0.99	65	0.87	110	0.43	170	0.75	204	1.29
7	0.74	69	0.89	111	0.47	171	1.15	205	1.18
9	1.42	71	0.30	115	0.75	173	1.20	207	1.13
10	3.05	73	0.89	116	0.82	174	1.16	208	1.63
11	0.59	74	1.13	119	0.84	175	0.93	209	0.24
12	1.87	75	0.76	123	0.85	176	0.99	210	0.99
16	1.14	76	0.85	124	1.37	177	1.07	211	1.13
18	0.93	77	0.84	125	2.08	178	1.10	213	0.97
19	1.08	78	1.01	126	1.22	179	0.81	217	0.79
20	0.95	79	0.85	127	0.95	180	1.17	218	0.28
22	0.99	80	0.96	128	0.45	181	1.05	219	1.05
23	0.91	81	1.01	130	1.18	182	1.13	220	0.15
24	0.84	82	0.82	131	0.28	183	1.18	221	0.79
25	1.14	83	0.60	133	1.84	184	1.16	222	1.01
27	0.86	84	1.10	136	0.52	185	0.95		
37	0.00	85	1.03	137	0.92	186	1.07		
38	0.00	86	0.87	138	0.64	187	1.09		
39	0.57	87	1.06	141	0.42	188	1.30		
40	0.94	91	0.70	144	0.96	189	1.51		
43	1.07	93	0.50	146	0.54	190	0.85		
44	0.96	96	0.85	147	0.56	192	1.03		
45	0.90	97	0.98	150	0.62	193	1.07		
46	1.02	98	0.79	159	1.50	194	1.04		
47	0.00	99	0.78	160	1.36	195	1.17		
48	0.77	100	0.91	161	1.42	196	0.98		

Table C19. Signal intensity ratios of EpYINQSV–Grb2–S4\*

Residue	I/I <sub>0</sub>								
1	0.86	50	0.00	101	1.04	164	0.73	203	0.93
2	0.98	51	1.08	102	0.53	165	0.64	204	1.13
3	1.38	52	0.81	106	0.38	167	0.75	205	0.70
4	0.83	53	0.85	108	0.37	168	0.74	207	0.93
5	0.74	54	0.31	110	0.73	171	0.82	208	1.26
6	0.64	64	1.18	111	0.43	173	1.06	209	0.20
7	0.63	65	0.94	115	0.92	174	0.99	210	1.01
9	0.87	69	0.86	116	0.91	175	0.93	211	0.84
10	2.63	73	0.89	119	0.91	176	0.91	213	1.06
11	0.95	74	0.92	123	0.68	177	0.88	217	0.73
12	2.29	75	0.68	124	1.56	178	0.95	218	0.33
16	1.55	76	0.86	125	2.03	180	0.39	219	1.07
18	1.35	77	0.88	126	1.00	181	0.86	220	0.14
19	1.32	78	1.15	127	1.00	183	1.20	221	0.29
20	1.07	79	0.84	128	0.45	184	0.88	222	0.92
22	0.62	80	0.82	130	1.08	185	0.93		
23	0.85	81	0.89	131	0.28	186	1.15		
25	1.19	82	0.79	133	1.80	187	0.82		
27	0.97	83	0.64	136	0.47	188	1.14		
28	1.35	84	1.08	137	0.83	189	1.30		
37	0.00	85	0.84	138	0.59	190	0.79		
38	0.00	86	0.67	141	0.44	192	0.92		
39	0.93	87	0.84	144	0.93	193	0.89		
40	1.04	91	0.75	146	0.63	194	0.82		
43	1.04	93	0.50	147	0.63	195	1.16		
44	0.98	96	0.74	150	1.49	196	1.06		
45	0.93	97	0.81	159	0.98	197	0.98		
46	0.94	98	0.83	160	0.75	198	0.79		
47	0.00	99	0.77	161	1.13	201	1.15		
48	1.02	100	0.93	163	0.99	202	0.92		

Table C20. Signal intensity ratios of EpYINQSV–Grb2–S5

Residue	I/I <sub>0</sub>								
1	0.79	51	0.82	102	0.52	167	0.71	205	0.40
2	1.07	52	0.78	106	0.49	168	0.74	207	0.27
3	1.25	53	0.92	108	0.31	171	0.94	208	1.12
4	0.72	54	0.26	110	0.36	173	0.69	209	0.23
6	0.69	65	0.77	111	0.41	174	0.90	210	0.91
7	0.83	69	0.79	115	0.78	175	0.93	211	0.77
8	1.17	71	0.28	116	0.76	176	0.83	213	0.81
9	1.07	73	0.74	119	0.88	177	0.88	217	0.88
10	1.55	74	0.70	123	0.62	178	0.88	219	0.86
11	0.92	75	0.89	124	1.80	180	0.88	220	0.11
12	0.86	76	0.80	125	1.81	181	0.88	222	0.76
16	1.17	77	0.77	126	0.79	182	0.73		
18	0.96	78	1.05	127	0.77	183	0.83		
19	0.97	79	0.82	128	0.39	184	0.98		
20	0.91	80	0.77	130	0.80	186	0.71		
22	1.05	81	0.86	133	1.39	187	0.80		
23	0.88	82	0.75	136	0.51	188	0.90		
25	0.93	83	0.45	137	0.64	189	1.18		
27	0.62	84	0.99	138	0.67	190	0.74		
37	0.89	85	0.89	141	0.35	192	0.63		
38	0.00	86	0.81	144	0.82	193	0.49		
39	0.90	87	0.87	146	0.71	194	0.41		
40	1.01	91	0.67	147	0.40	195	0.94		
43	0.90	93	0.42	159	0.88	196	0.74		
44	0.95	96	0.88	160	1.13	197	0.79		
45	0.94	97	0.75	161	1.17	198	0.88		
46	0.85	98	0.68	163	0.94	201	0.90		
47	0.98	99	0.75	164	0.99	202	0.85		
48	0.89	100	0.91	165	0.64	203	0.83		
50	0.84	101	0.91	166	0.78	204	0.93		

Table C21. Signal intensity ratios of EpYINQSV–Grb2–S12

Residue	I/I <sub>0</sub>								
1	0.86	51	0.32	104	0.69	165	1.35	210	0.65
2	0.83	52	0.60	106	0.48	167	0.57	211	0.56
3	0.95	53	0.47	107	0.72	168	0.00	213	0.80
4	0.69	65	0.72	108	0.48	171	0.93	217	0.70
5	0.67	69	0.62	110	0.77	173	0.49	218	0.50
6	0.48	71	0.64	111	0.55	174	0.79	219	0.71
7	0.55	73	0.67	115	0.88	175	0.40	220	0.57
9	0.76	74	0.62	116	0.65	176	0.58	221	0.00
10	2.97	75	0.57	119	0.66	177	0.60	222	0.73
12	1.07	76	0.59	123	0.74	178	0.48		
16	1.93	77	0.68	124	0.63	180	0.31		
18	0.87	78	0.84	125	0.61	181	0.55		
19	0.63	79	0.60	127	0.59	183	0.55		
20	0.75	80	0.67	128	0.71	184	0.52		
22	0.52	81	0.61	130	0.77	185	0.60		
23	0.71	82	0.62	131	0.69	187	0.54		
24	0.57	83	0.49	133	0.65	190	0.49		
25	0.56	84	0.69	134	0.72	192	0.00		
27	0.59	85	0.71	136	0.71	193	0.16		
37	0.00	86	0.62	137	0.63	194	0.32		
38	0.00	87	0.59	138	0.71	195	0.80		
39	0.84	91	0.68	141	0.64	196	0.47		
40	0.64	93	0.65	144	0.65	197	0.51		
43	0.64	96	0.58	146	0.62	201	0.80		
44	0.67	97	0.53	147	0.70	202	0.65		
45	0.65	98	0.59	159	1.17	203	0.52		
46	0.51	99	0.56	160	1.96	204	0.42		
47	0.00	100	0.55	161	1.08	207	0.00		
48	0.37	101	0.62	163	0.36	208	0.51		
50	0.00	102	0.66	164	0.49	209	0.84		

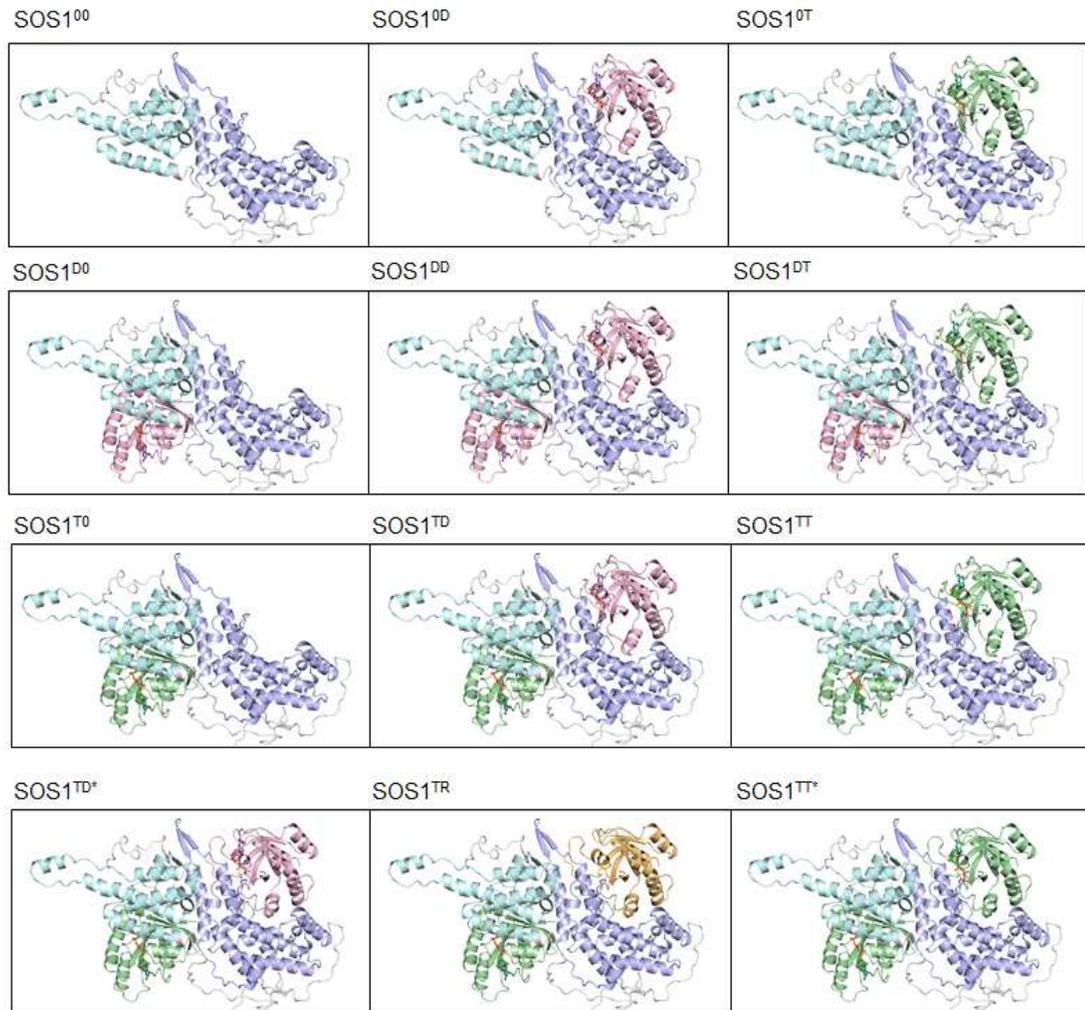
Table C22. Signal intensity ratios of EpYINQSV–Grb2–S23

Residue	I/I <sub>0</sub>								
1	0.87	51	0.38	104	0.70	165	0.84	208	0.51
2	0.86	53	0.43	106	0.51	167	0.62	210	0.56
4	0.70	54	0.25	107	0.74	168	0.00	211	0.54
6	0.64	65	0.68	108	0.50	173	0.51	213	0.82
7	0.73	68	0.61	110	0.72	174	0.74	217	0.68
9	0.84	69	0.70	111	0.51	175	0.40	218	0.60
10	0.91	73	0.68	115	0.79	176	0.54	219	0.74
11	0.00	74	0.47	116	0.69	177	0.44	220	0.75
12	1.08	75	0.51	118	0.61	178	0.58	221	0.00
15	2.15	76	0.64	119	0.60	180	0.35	222	0.76
16	1.63	77	0.67	123	0.79	181	0.65		
18	0.57	78	0.80	124	0.62	183	0.60		
19	0.76	79	0.63	125	0.61	184	0.67		
20	0.58	80	0.77	126	0.73	186	0.65		
22	0.50	81	0.69	127	0.67	187	0.57		
23	0.62	82	0.60	128	0.71	188	1.13		
24	0.57	83	0.59	130	0.74	189	0.69		
25	0.65	84	0.58	133	0.71	190	0.49		
27	0.74	85	0.79	134	1.01	192	0.61		
37	0.00	86	0.59	136	0.70	194	0.18		
38	0.00	87	0.65	137	0.58	195	0.55		
39	1.01	91	0.69	138	0.84	196	0.46		
40	0.53	93	0.64	141	0.52	197	0.41		
43	0.60	96	0.63	146	0.54	198	0.45		
44	0.66	97	0.53	147	0.60	201	0.74		
45	0.57	98	0.69	150	0.49	202	0.63		
46	0.47	99	0.46	159	1.33	203	0.37		
47	0.00	100	0.54	161	0.80	204	0.66		
48	0.50	101	0.57	163	0.51	205	0.73		
50	0.00	102	0.61	164	0.39	207	0.00		

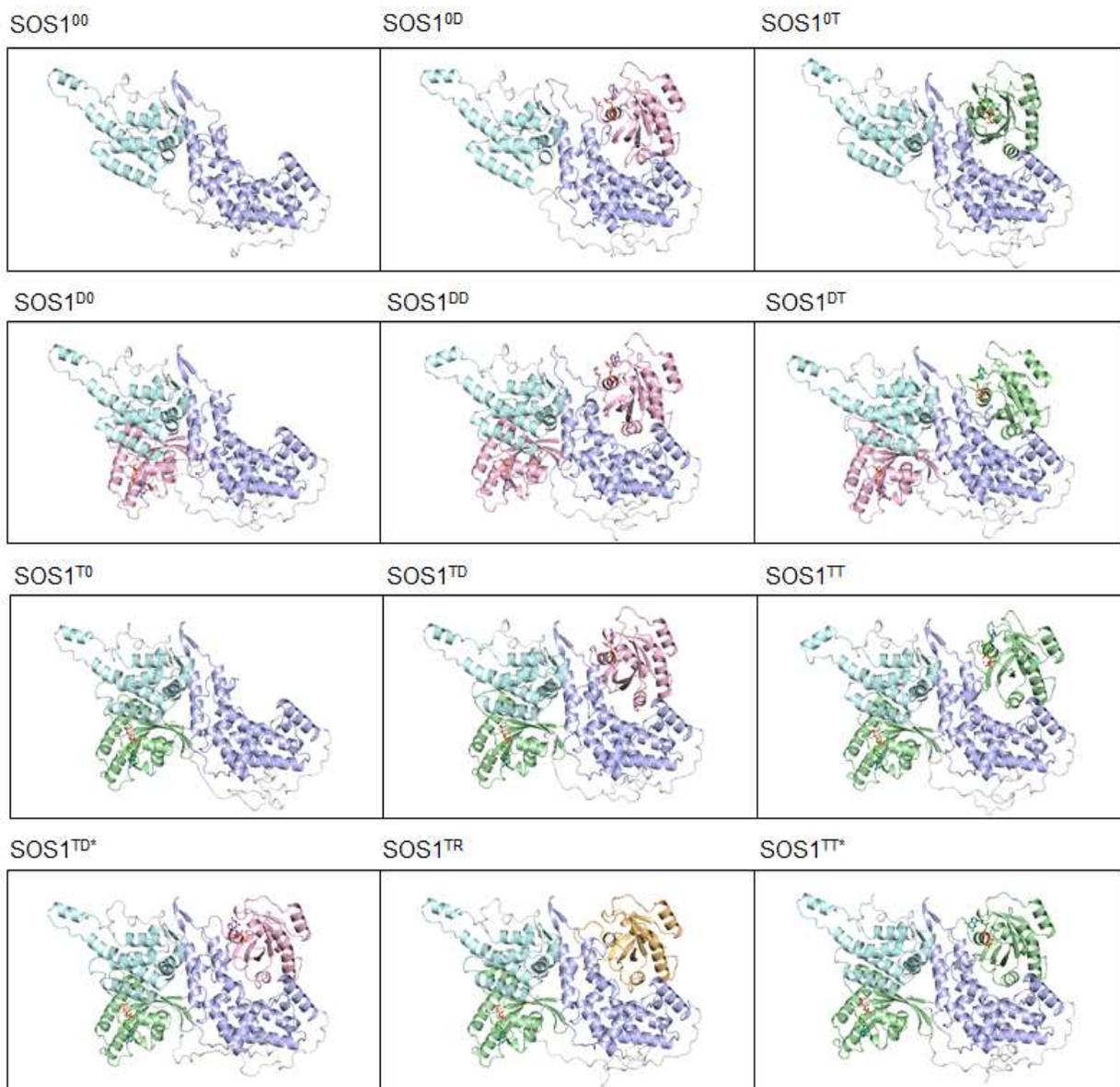
Table C23. Signal intensity ratios of EpYINQSV–Grb2–S45

Residue	I/I <sub>0</sub>								
1	0.83	53	0.38	119	0.62	189	0.80		
2	0.84	65	0.52	123	0.37	192	0.00		
6	0.32	69	0.60	124	1.00	194	0.00		
7	0.48	71	0.29	125	1.34	195	0.96		
8	0.21	73	0.56	127	0.62	196	0.30		
9	0.48	74	0.58	128	0.19	201	0.59		
10	0.00	75	0.58	130	0.78	202	0.32		
11	0.35	76	0.63	133	1.05	203	0.15		
15	0.66	77	0.63	136	0.35	207	0.00		
16	0.51	79	0.55	137	0.55	208	0.88		
18	0.44	81	0.60	138	0.57	209	0.08		
19	0.59	82	0.55	141	0.28	210	0.39		
20	0.45	83	0.31	146	0.29	211	0.32		
22	0.44	85	0.48	147	0.34	213	0.83		
24	0.54	86	0.60	159	0.54	217	0.27		
25	0.55	87	0.70	164	0.26	218	0.48		
27	0.51	91	0.58	165	1.30	219	0.46		
37	0.00	96	0.54	167	0.53	220	0.12		
38	0.00	97	0.51	168	0.14	221	0.00		
39	0.38	98	0.45	173	0.42	222	0.28		
40	0.54	99	0.47	174	0.52				
43	0.61	100	0.51	176	0.42				
44	0.64	102	0.43	177	0.27				
45	0.58	106	0.34	178	0.42				
46	0.18	107	0.69	180	0.23				
47	0.00	108	0.24	181	0.52				
48	0.00	110	0.64	183	0.38				
50	0.00	111	0.32	186	0.41				
51	0.51	115	0.76	187	0.30				
52	0.33	116	0.56	188	0.70				

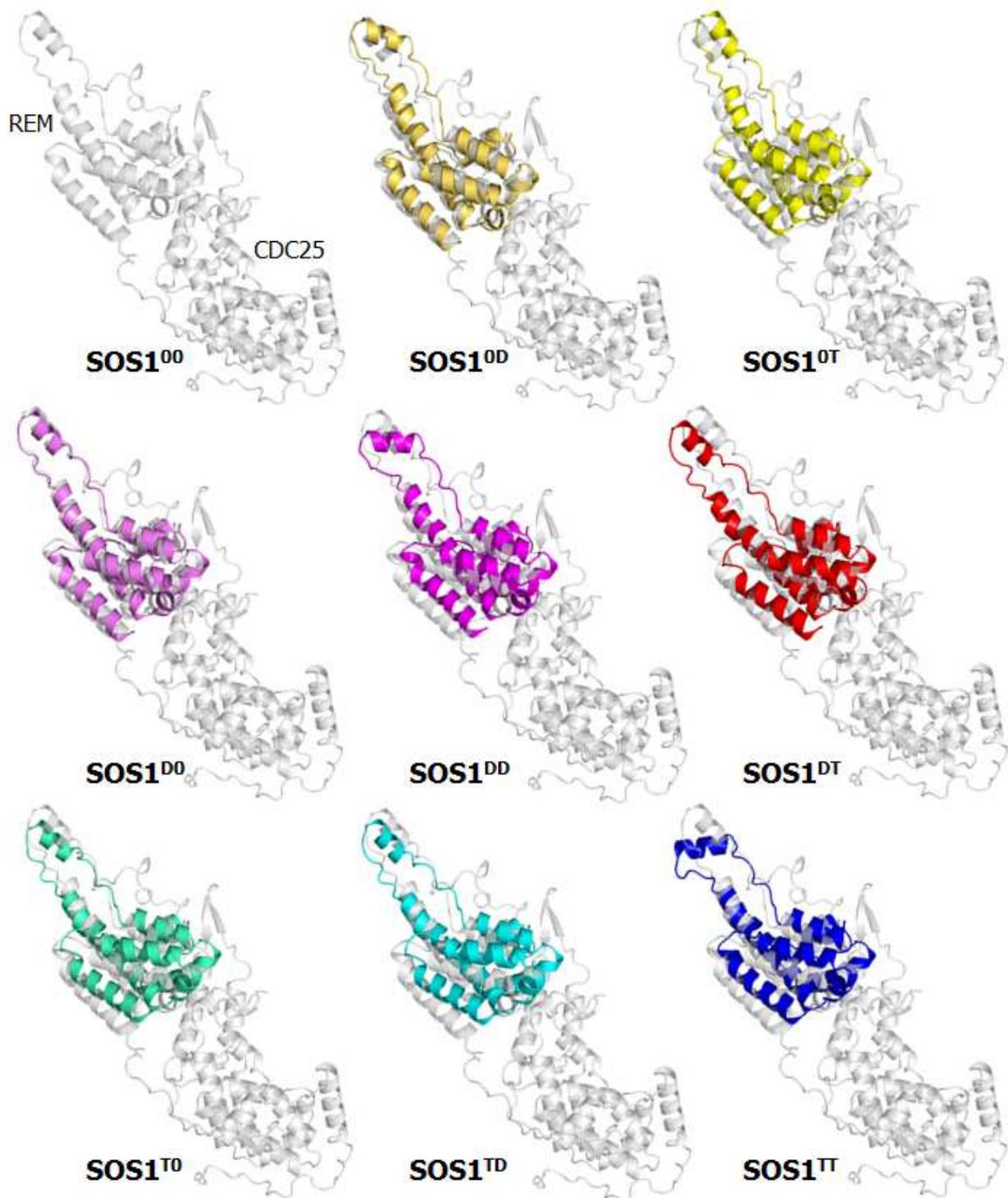
*Appendix D*



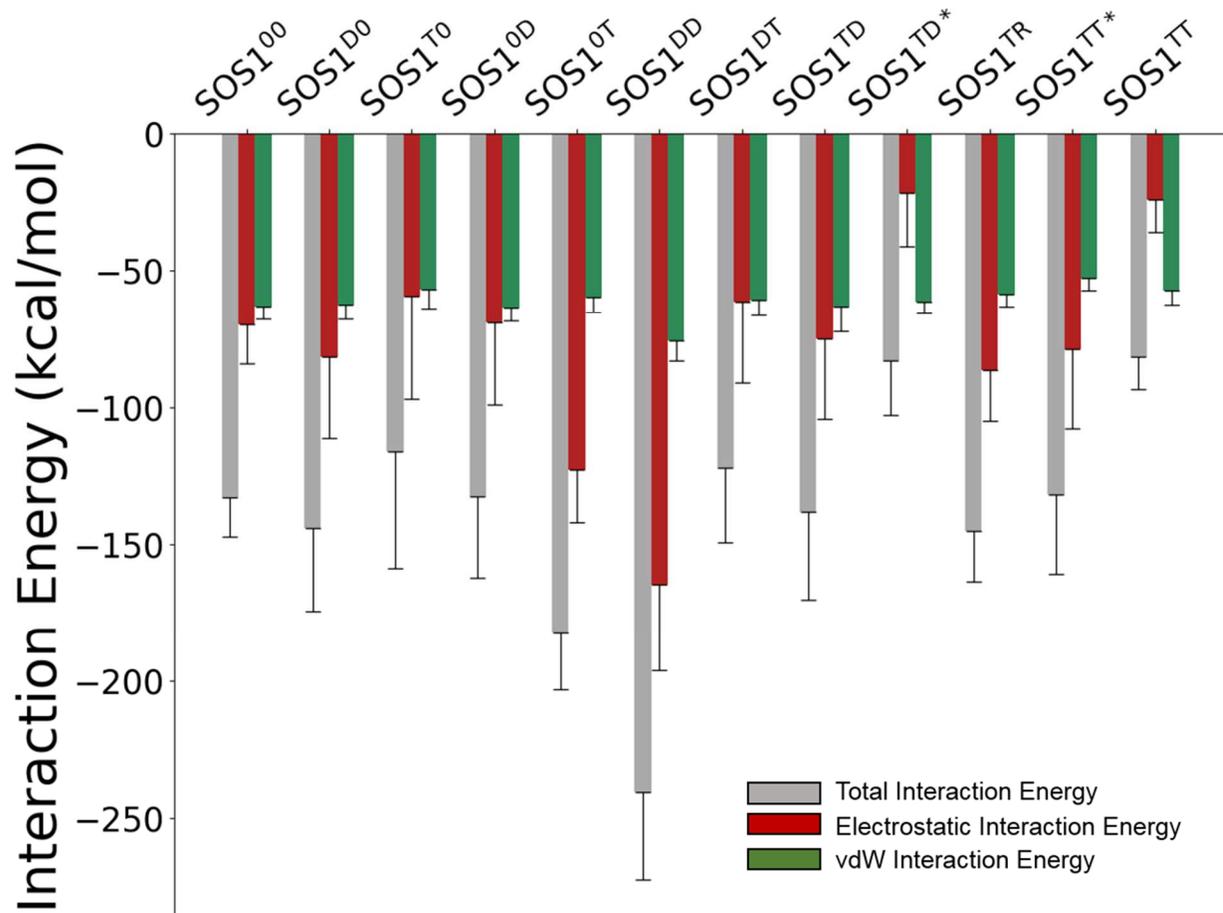
**Figure D1.** Initial structures of KRas4B–SOS1 complex. In the structures, the REM and CDC25 domains are shown in cyan and blue. KRas4B-GTP, KRas4B-GDP, and nucleotide-free KRas4B are represented in green, red, and yellow, respectively. The rest regions of SOS1 is shown in gray.



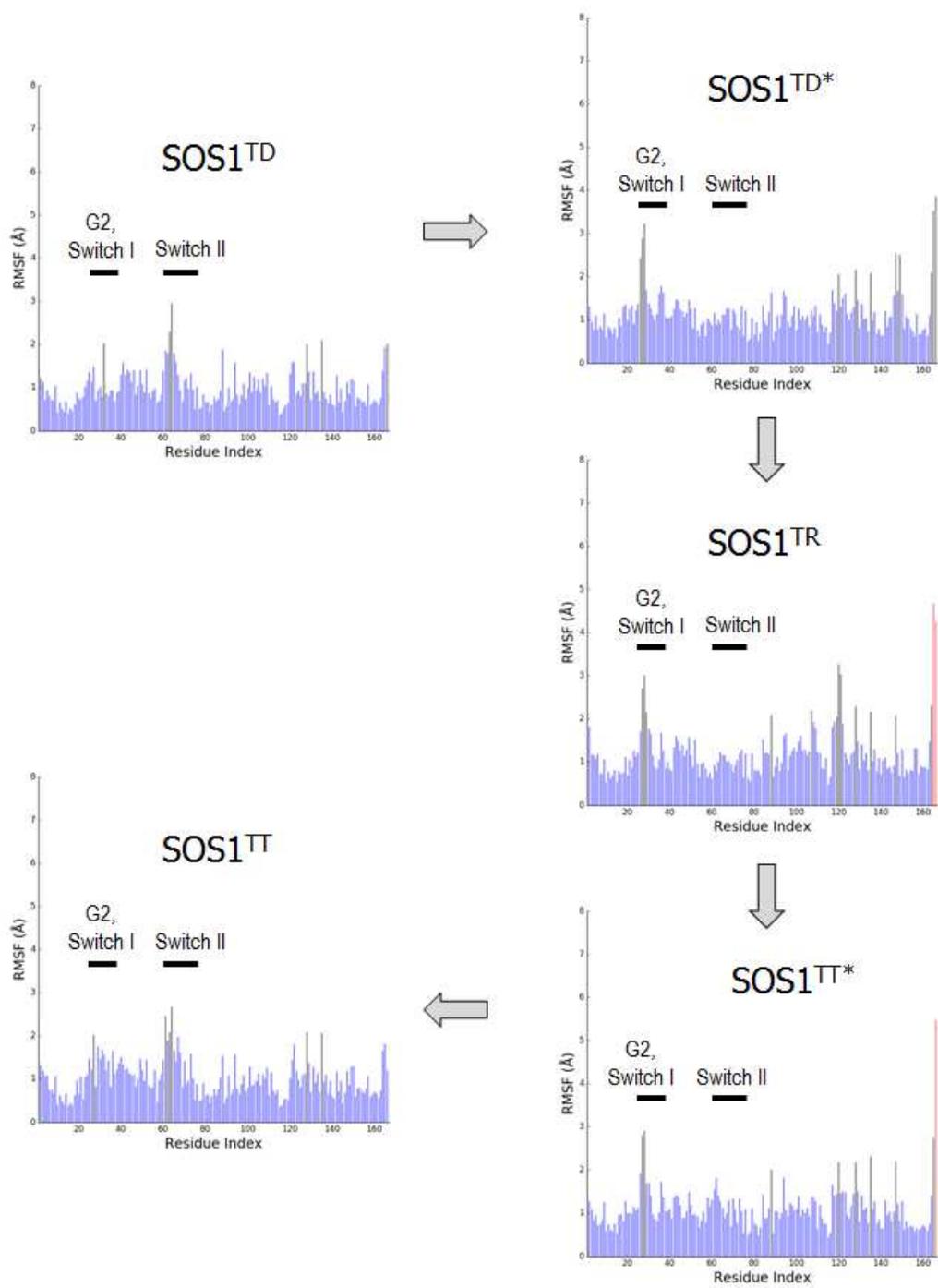
**Figure D2.** Average structures of KRas4B–SOS1 complex. In the structures, the REM and CDC25 domains are shown in cyan and blue. KRas4B-GTP, KRas4B-GDP, and nucleotide-free KRas4B are represented in green, red, and yellow, respectively. The rest regions of SOS1 is shown in gray.



**Figure D3.** Superimpositions of average structure of SOS1 systems with respect to CDC25 of the apo-SOS1<sup>00</sup> system. The averaged REM domains (colored cartoon) are only shown and superimposed onto the reference system (gray cartoon).



**Figure D4.** Interaction energy between the REM and CDC25 domains. The vdW interaction is mainly conserved, while the electrostatic interaction varies among different systems.



**Figure D5.** The root mean squared fluctuation (RMSF) for each residue of KRas4B at the catalytic site.

**Table D1.** The intramolecular interacting residue pairs between the REM and CDC25 domains. The percentage in parenthesis denotes the occurrence rate based on cumulative occurrence of the interaction for the residue pairs during the simulations. The pairs with high occurrence rate (>50%) are listed.

— Region I      — Region II

Configuration	Hydrophobic	Salt bridge	H-bond	Polar
SOS1 <sup>00</sup>	I601 - F958 (66%), A603 - V948 (82%) A603 - V955 (58%), A603 - I956 (72%) A603 - F958 (94%), V606 - I956 (56%) V609 - I956 (58%), V609 - F958 (67%) F623 - F958 (71%), F627 - F958 (62%) F634 - I956 (53%), F634 - F958 (97%)	R706 - E803 (99%)	G604 - I956 (90%)	T629 - S802 (55%) T629 - T968 (55%) Y702 - Q800 (72%) Y702 - S802 (81%)
SOS1 <sup>0D</sup>	I601 - F958 (76%), A603 - V948 (61%) A603 - V955 (99%), A603 - I956 (70%) A603 - F958 (91%), V606 - I956 (56%) V609 - I956 (73%), V609 - F958 (53%) F623 - F958 (93%), F627 - F958 (88%) F634 - I956 (62%), F634 - F958 (54%)	R706 - E803 (55%)	G604 - I956 (90%)	T629 - S802 (53%) T629 - T968 (80%) H700 - S802 (51%) Y702 - Q800 (55%) Y702 - S802 (52%)
SOS1 <sup>0T</sup>	I601 - F958 (58%), A603 - V948 (78%) A603 - V955 (51%), A603 - I956 (56%) A603 - F958 (94%), V606 - I956 (53%) V609 - I956 (51%), V609 - F958 (63%) F623 - F958 (70%), F627 - F958 (76%) F634 - I956 (53%), F634 - F958 (74%)	R706 - E803 (51%)	G604 - I956 (89%)	T629 - S802 (54%) T629 - T968 (70%) H700 - S802 (54%) Y702 - Q800 (55%)
SOS1 <sup>D0</sup>	I601 - F958 (76%), A603 - V948 (72%) A603 - V955 (54%), A603 - I956 (52%) A603 - F958 (95%), V606 - I956 (56%) V609 - I956 (52%), V609 - F958 (51%) F623 - F958 (53%), F627 - F958 (68%) F634 - I956 (54%), F634 - F958 (54%)	R706 - E803 (53%)	G604 - I956 (82%)	T629 - S802 (61%) T629 - T968 (58%) H700 - S802 (74%) Y702 - Q800 (57%) Y702 - S802 (52%)
SOS1 <sup>DD</sup>	I601 - F958 (73%), A603 - V948 (87%) A603 - V955 (51%), A603 - I956 (51%) A603 - F958 (97%), V606 - I956 (56%) V609 - I956 (56%), V609 - F958 (59%) F623 - F958 (62%), F627 - F958 (69%) F634 - I956 (68%), F634 - F958 (72%)	R706 - E803 (75%) R706 - D813 (75%)	G604 - I956 (87%)	T629 - S802 (76%) T629 - T968 (62%) T630 - S802 (62%) Y631 - S802 (65%) H699 - Q800 (54%) H700 - Q800 (69%) H700 - S802 (61%) Y702 - Q800 (61%) Y702 - N817 (59%) Y702 - S818 (63%)
SOS1 <sup>DT</sup>	I601 - F958 (67%), A603 - V948 (72%) A603 - V955 (50%), A603 - I956 (70%) A603 - F958 (95%), V606 - I956 (52%) V609 - I956 (54%), V609 - F958 (51%) F623 - F958 (62%), F627 - F958 (77%) F634 - I956 (50%), F634 - F958 (97%)	R706 - E803 (61%)	G604 - I956 (90%)	T629 - T968 (74%) H700 - Q800 (54%) Y702 - Q800 (57%) Y702 - S802 (61%)
SOS1 <sup>T0</sup>	I601 - F958 (51%), A603 - V948 (73%) A603 - V955 (53%), A603 - I956 (56%) A603 - F958 (91%), V606 - I956 (52%) V609 - I956 (58%), V609 - F958 (60%) F623 - F958 (50%), F627 - F958 (75%) F634 - I956 (70%), F634 - F958 (57%)	R706 - E803 (55%)	G604 - I956 (88%)	T629 - S802 (54%) T629 - T968 (50%) H700 - S802 (58%) Y702 - Q800 (50%) Y702 - S802 (52%)
SOS1 <sup>TD</sup>	I601 - F958 (87%), A603 - V948 (78%) A603 - V955 (54%), A603 - I956 (60%) A603 - F958 (93%), V606 - I956 (60%) V609 - I956 (54%), V609 - F958 (52%) F623 - F958 (65%), F627 - F958 (68%) F634 - I956 (59%), F634 - F958 (56%)	R962 - D620 (74%)	G604 - I956 (84%)	H700 - S802 (61%) Y702 - N817 (51%)

<b>SOS1<sup>TT</sup></b>	I601 - F958 (61%), A603 - V948 (78%) A603 - V955 (55%), A603 - I956 (63%) A603 - F958 (93%), V606 - I956 (77%) V609 - I956 (58%), V609 - F958 (65%) F623 - F958 (56%), F627 - F958 (66%) F634 - I956 (60%), F634 - F958 (97%)	N/A	G604 - I956 (94%)	T629 - T968 (51%) Y702 - S802 (52%)
--------------------------	--	-----	-------------------	--

**Table D2.** The intermolecular interacting residue pairs between SOS1 and KRas4B.

The percentage in parenthesis denotes the occurrence rate based on cumulative occurrence of the interaction for the residue pairs during the simulations. The pairs with high occurrence rate (>50%) are listed.

— CDC25/switch I      — CDC25/switch II      — Others  
 — REM/switch I      — REM/switch II

Configuration	Site	Hydrophobic	Salt bridge	H-bond	Polar
SOS1 <sup>0D</sup>	Catalytic	W809 - I36 (58%) V875 - M67 (66%) V934 - I36 (53%) V938 - P34 (92%) V938 - I36 (68%)	R885 - D105 (56%) K949 - E31 (64%)	T829 - S65 (90%)	T829 - Y64 (94%), T829 - S65 (82%) T832 - S65 (94%), S876 - S65 (81%) N879 - Q70 (56%), S880 - Q70 (71%) Y884 - Q70 (55%), H911 - S39 (53%) H911 - Y40 (68%), Y912 - Y71 (56%)
SOS1 <sup>0T</sup>	Catalytic	P945 - F28 (70%) P945 - V29 (59%)	K811 - D33 (53%) K1003 - E63 (55%)	N/A	S881 - Y64 (64%), S881 - S65 (63%) Y884 - Y64 (51%), Y884 - Q70 (54%) H911 - S39 (60%), T935 - T35 (79%) N944 - H27 (53%)
SOS1 <sup>D0</sup>	Allosteric	A917 - V45 (66%)	R688 - E37 (56%) R694 - D33 (87%) K735 - D33 (71%) R739 - E31 (68%)	N/A	Y618 - Q70 (51%), N691 - T35 (59%) H699 - Y40 (52%), S732 - T35 (96%) Q973 - Q25 (52%), N976 - Q25 (52%) N976 - N26 (61%), N976 - H27 (54%) Q977 - N26 (61%), Q977 - H27 (50%)
SOS1 <sup>DD</sup>	Allosteric	M617 - A66 (63%) M617 - M67 (67%) A619 - M67 (59%) P621 - V56 (87%) P684 - M67 (76%)	K595 - E76 (69%) R625 - D38 (96%) R694 - D33 (90%)	Y974 - Q43 (51%)	Y618 - Q70 (76%), Y618 - T74 (90%) N622 - S39 (57%), N691 - T35 (69%) S732 - T35 (98%), Q973 - Q43 (51%) Y974 - Q43 (69%), N976 - Q25 (59%) Q977 - Q25 (54%), Q977 - N26 (59%)
	Catalytic	F929 - M67 (64%) I932 - M67 (50%) P945 - V29 (62%)	K939 - D57 (81%) R1019 - D105 (66%)	H911 - S39 (58%) V938 - T35 (96%)	T829 - S65 (59%), S876 - Q70 (53%) N879 - Q70 (57%), S880 - Q70 (56%) H911 - S39 (76%), H911 - Y71 (53%) H911 - T74 (56%), Y912 - Q70 (83%) Y912 - Y71 (62%), T935 - Y71 (84%)
SOS1 <sup>DT</sup>	Allosteric	M617 - A66 (57%) M617 - M67 (82%) A619 - M67 (51%) P684 - M67 (57%) I922 - V45 (66%) P924 - V44 (77%) P924 - V45 (87%)	R625 - D38 (90%) K679 - E63 (56%) R694 - D33 (67%) K735 - D33 (50%)	A596 - Q70 (52%) G749 - H27 (66%) T753 - N26 (94%) Y919 - Q43 (54%) N923 - V45 (57%)	Y618 - Q70 (57%), S732 - T35 (87%) N751 - H27 (53%), T753 - Q22 (70%) T753 - N26 (56%), T753 - H27 (54%) S921 - Q43 (53%), S921 - T50 (50%) N923 - Q43 (60%), Q973 - Q43 (50%) Y974 - Q43 (65%), N976 - Q25 (54%) Q977 - Q25 (56%), Q977 - Q43 (52%)
	Catalytic	V872 - M67 (55%) V875 - M67 (67%)	R885 - D105 (79%) K913 - D38 (75%) K939 - E37 (53%) D910 - R41 (56%)	N/A	N879 - Q70 (56%), Y884 - Q70 (65%) H905 - Q70 (57%), S908 - Y71 (60%) H911 - Y40 (51%), H911 - Y71 (78%) Y912 - Y71 (66%), N944 - Y32 (50%)
SOS1 <sup>T0</sup>	Allosteric	P684 - A66 (66%), P684 - M67 (69%) V687 - I36 (51%), V687 - M67 (50%) V690 - I36 (90%), W729 - P34 (65%) W729 - I36 (82%), W729 - A59 (62%) P924 - V44 (62%), P924 - V45 (54%)	R625 - D38 (60%) R688 - E37 (100%) R694 - D38 (88%) K728 - E62 (50%) K735 - E31 (65%) K735 - D33 (59%) R739 - E31 (64%)	A619 - Q70 (56%)	Y618 - Q70 (57%), N622 - S39 (52%) H695 - S39 (63%), S921 - T50 (65%) N923 - Q43 (53%), N923 - T50 (50%) Q973 - Q43 (51%), Y974 - Q43 (66%) N976 - Q25 (52%), Q977 - N26 (62%)
SOS1 <sup>TD</sup>	Allosteric	M617 - A66 (72%), M617 - M67 (52%) P684 - M67 (55%), V687 - I36 (56%) V687 - M67 (50%), V690 - I36 (91%) W729 - P34 (51%), W729 - I36 (72%) P924 - V44 (52%)	K595 - E76 (81%) R625 - D38 (52%) R688 - E37 (100%) R694 - D33 (55%) R739 - E31 (63%)	N/A	Y618 - Q70 (72%), Y618 - Y71 (65%) Y618 - T74 (70%), S921 - T50 (58%) N923 - Q43 (64%), Q973 - Q43 (63%) Y974 - Q43 (65%), N976 - Q25 (58%) Q977 - N26 (51%)
	Catalytic	V875 - M67 (64%) P945 - V29 (70%)	K939 - D57 (94%)	N/A	T829 - Y64 (53%), N879 - Q70 (66%) Y884 - Q70 (53%), H911 - S39 (56%) H911 - Y71 (58%), Y912 - Y71 (50%) T935 - T35 (52%)
SOS1 <sup>TT</sup>	Allosteric	M617 - A66 (53%), M617 - M67 (51%) A619 - M67 (53%), P684 - M67 (56%) W729 - P34 (52%), W729 - I36 (51%) P924 - V44 (53%)	R625 - D38 (82%) R694 - D33 (73%) R739 - E31 (53%)	V919 - Q43 (95%) I922 - Q43 (85%) Y974 - Q43 (81%)	Y618 - Q70 (52%), S921 - Q43 (67%) S921 - T50 (58%), N923 - Q43 (99%) Y974 - Q43 (59%), N976 - Q25 (54%) Q977 - N26 (54%)

	Catalytic	V934 - I36 (50%) I937 - I36 (53%) V938 - P34 (67%) V938 - I36 (53%) P945 - F28 (60%)	K814 - E63 (53%) R826 - E63 (99%) K939 - D57 (100%) K960 - D30 (99%) R1019 - D105 (57%) E836 - R68 (56%) D1007 - R102 (97%)	W809 - T35 (75%) T935 - I36 (53%)	T810 - T35 (95%), T810 - Q61 (80%) T829 - Y64 (68%), N830 - Y64 (60%) S876 - Q70 (54%), N879 - Q70 (65%) T935 - T35 (61%), T1006 - Q99 (63%)
--	-----------	--	---	--------------------------------------	---

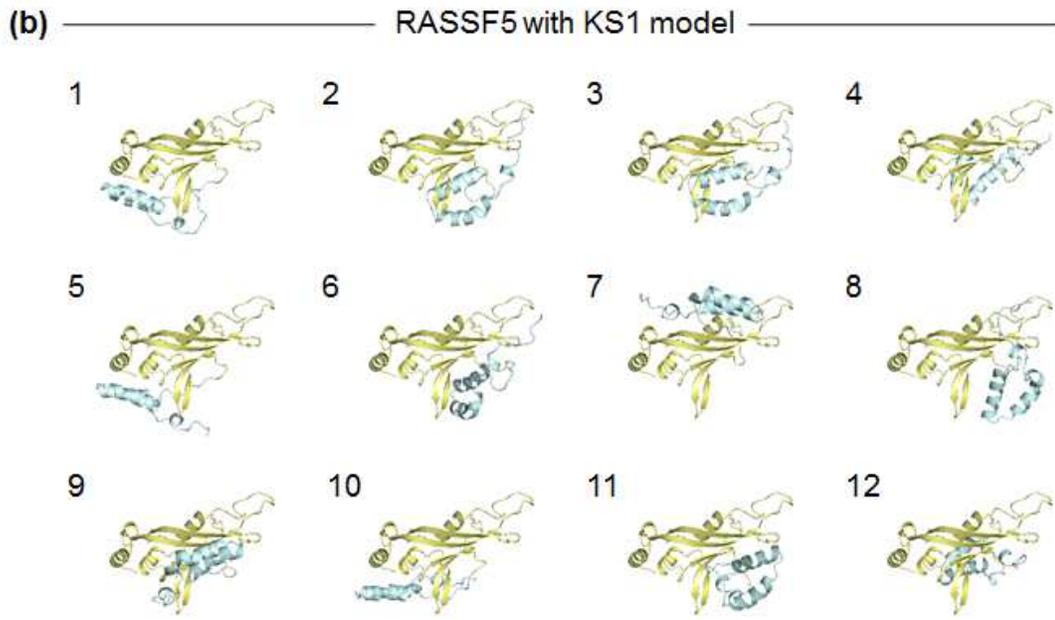
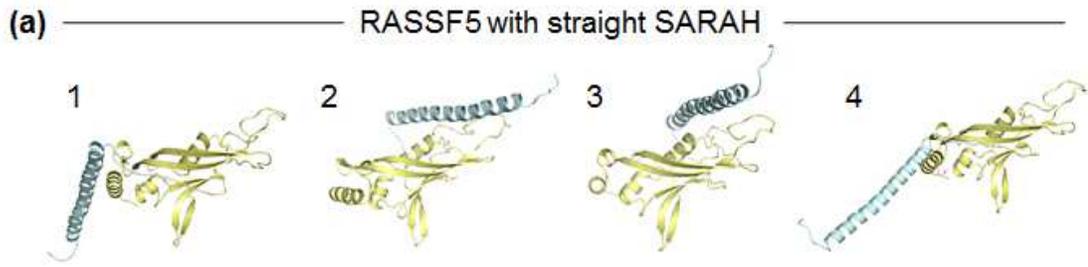
**Table D3.** The allosteric pathways between the KRas4B proteins at the allosteric and catalytic sites. The percentage in parenthesis denotes the occurrence rate based on cumulative occurrence for the residues involved in the optimal pathways. The residues with high occurrence rate (>50%) are listed.

— switch I    — switch II    — REM    — CDC25    — Others

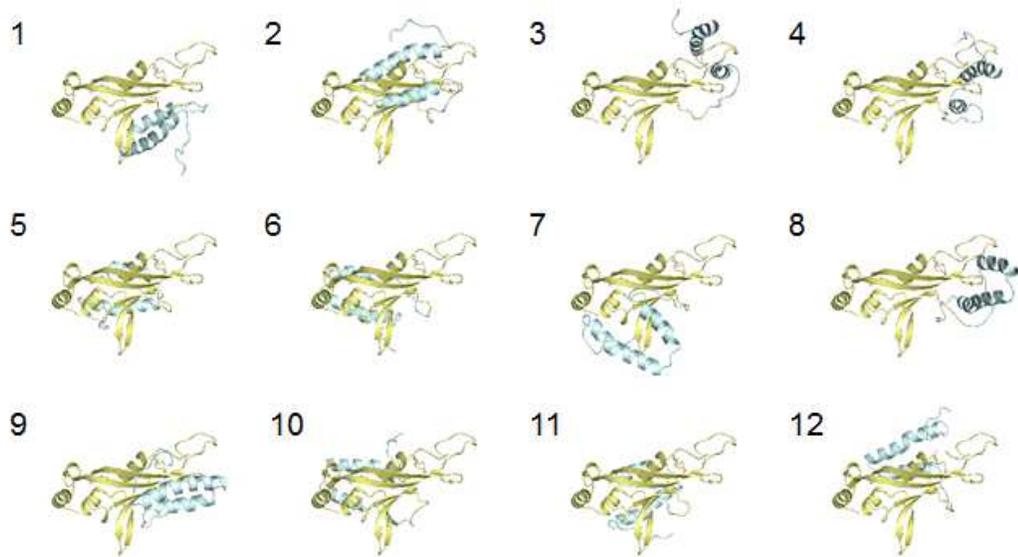
Configuration		Allosteric	→	SOS1	→	Catalytic	
SOS1 <sup>DD</sup>	S1	E37 (69%)		R625 (63%), T940 (78%) S959 (78%), K960 (78%)		P34 (55%)	S1
	S1	E37 (79%)		R625 (63%), T940 (51%) S959 (51%), K960 (57%)		I36 (55%)	S2
	S2	D69 (57%)		G597 (69%), P599 (69%) I600 (55%), T940 (97%) S959 (97%)		P34 (55%)	S1
	S2	N/A		G597 (51%), S959 (50%)		N/A	S2
SOS1 <sup>DT</sup>	S1	N/A		N/A		N/A	S1
	S1	N/A		N866 (65%), N869 (53%) G870 (51%), V875 (81%)		S65 (82%)	S2
	S2	V44 (93%), V45 (75%) T74 (58%), L159 (52%) V160 (52%)		Y912 (100%), L916 (89%) R920 (89%), S921 (85%) I922 (100%)		D33 (66%), T35 (57%) E37 (95%)	S1
	S2	V44 (88%), V45 (80%) T74 (57%), L159 (67%)		N866 (76%), G870 (65%) V875 (72%), I922 (92%) N923 (64%), P925 (64%)		S65 (65%)	S2
SOS1 <sup>TD</sup>	S1	I36 (61%)		A603 (70%), I607 (62%) I610 (64%), L613 (90%) L687 (99%), T940 (66%) I956 (89%), S959 (95%)		N/A	S1
	S1	N/A		A603 (76%), I607 (65%) I610 (65%), L613 (92%) L687 (100%), L938 (100%) L939 (100%), I956 (94%) S959 (98%)		N/A	S2
	S2	N/A		A603 (55%), T614 (84%) H616 (63%), T940 (77%) I956 (80%), S959 (90%)		D30 (51%)	S1
	S2	N/A		A603 (64%), I607 (56%) E611 (57%), T614 (100%) H616 (86%), L938 (100%) K939 (98%), I956 (96%) S959 (96%)		E37 (53%), D38 (51%)	S2
SOS1 <sup>TT</sup>	S1	N/A		V624 (68%), T940 (72%) K960 (99%)		N/A	S1
	S1	N/A		T629 (69%), C693 (54%) H695 (63%), T940 (67%) K960 (83%)		N/A	S2

	S2	E37 (61%)	V624 (61%), R625 (61%) T940 (69%), S959 (60%) K960 (72%)	N/A	S1
	S2	I36 (61%), E37 (78%)	V624 (78%), R625 (78%) T940 (74%), S959 (53%) K960 (83%)	D33 (68%)	S2

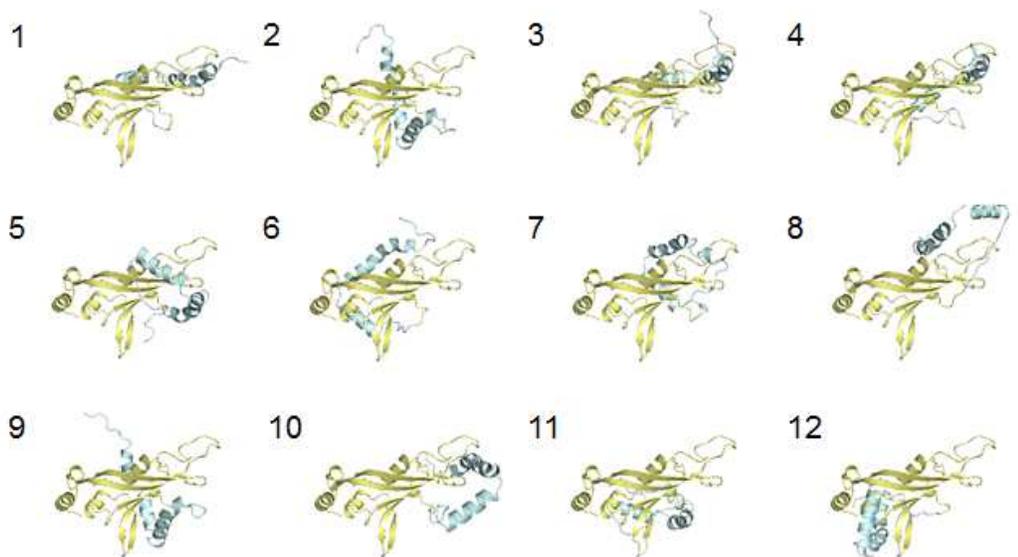
*Appendix E*

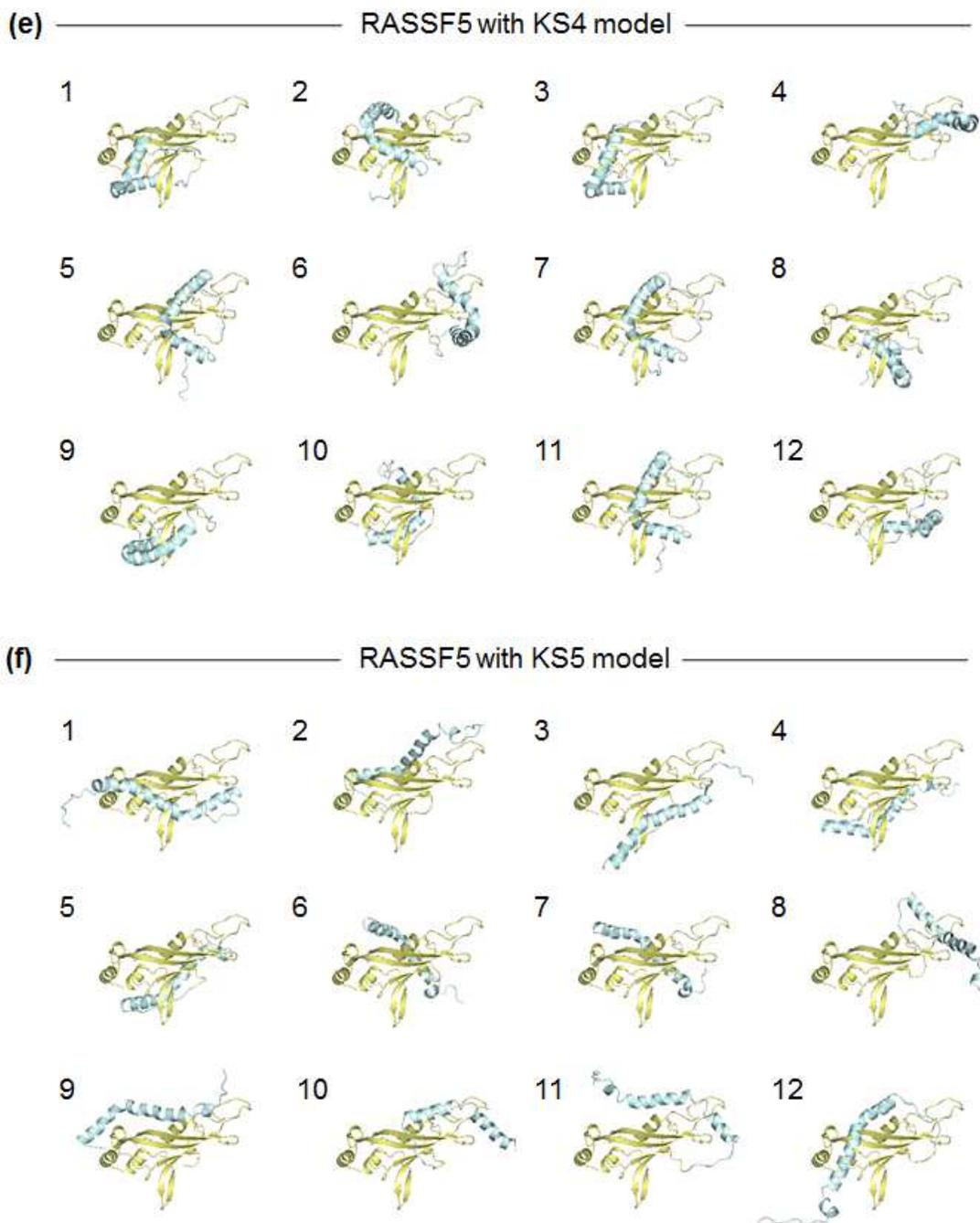


(c) ————— RASSF5 with KS2 model —————

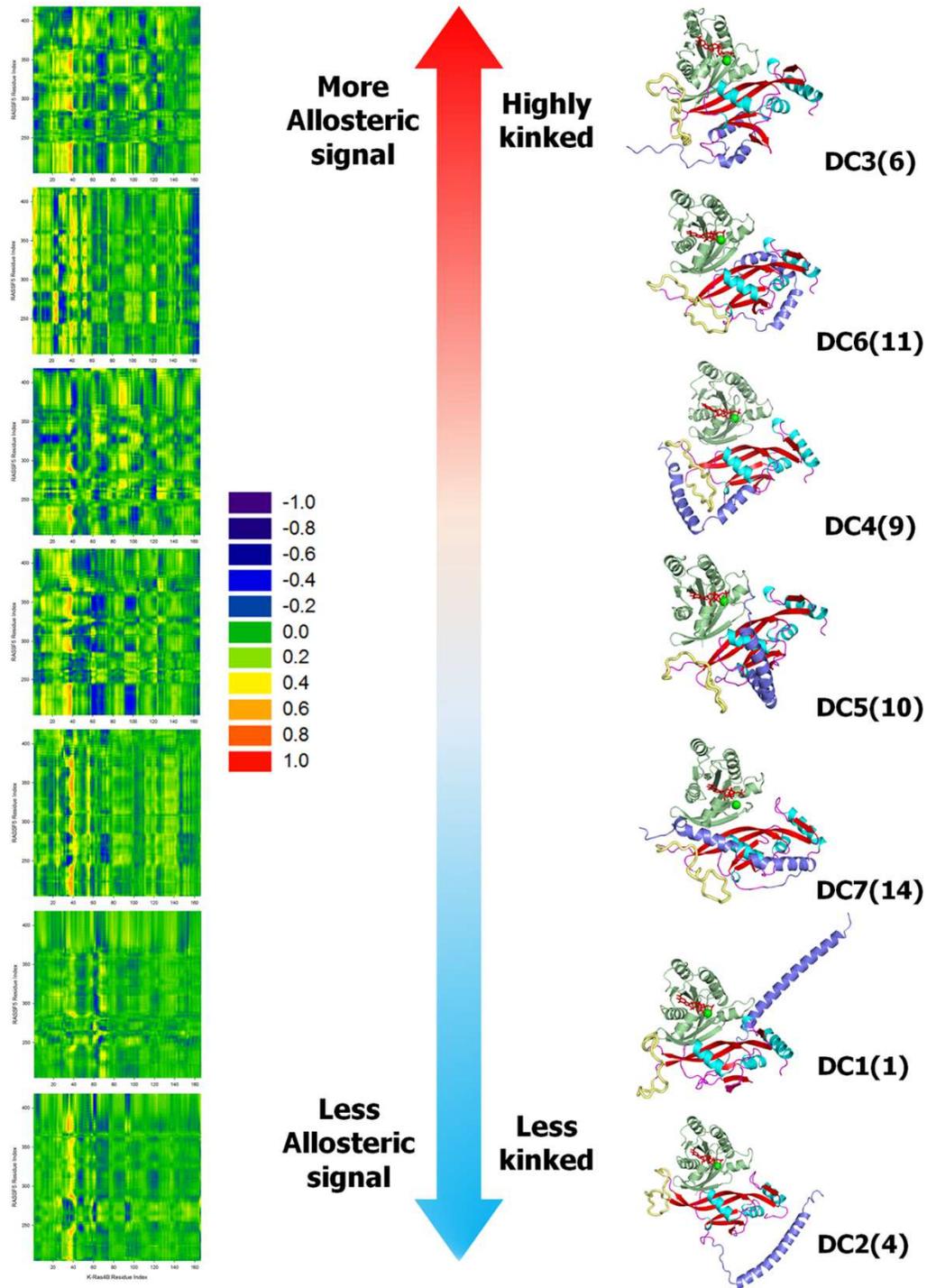


(d) ————— RASSF5 with KS3 model —————





**Figure E1.** Predicted RASSF5 decoys from (a) the PRISM server with straight SARAH and the Patchdock server with the kinked SARAH models of (b) KS1, (c) KS2, (d) KS3, (e) KS4, and (f) KS5.



**Figure E2.** The configurations with highly kinked SARAH show the strong residue correlations with K-Ras4B-GTP. In contrast, straight SARAH (the least kinked) has less correlation with K-Ras4B-GTP. The RASSF5 structures with K-Ras4B-GTP are shown in the order of highly kinked on the left (DC3(6)) to less kinked or straight on

the right (DC1(1), DC2(4)) SARAH. For the kinked SARAH models, the dynamical cross-correlation maps (DCCMs) shows strong residue correlations between K-Ras4B allosteric lobe and SARAH, which suggests that the allosteric signals are transmitted to release SARAH, activating RASSF5. With less kinked or straight SARAH, the signals from either allosteric lobe or effective lobe become weaker and eventually disappear when SARAH is fully stretched.

**Table E1.** Types of atomic interactions in the SARAH dimers. The atomic pair indicates monomer 1 – monomer 2 residue pair in the interface. The pairs will be viewed effective and counted when the residue pairs have more than 50 % occurrence rate.

Configuration	Salt bridge	Nonpolar	Polar	H-bond
<b>RASSF5-MST2 hetero-SARAH</b>	E385 - R474, E388 - R467, E388 - R474, K398 - E462, R403 - D456	I374 - M481, L377 - I477, L377 - A480, L377 - M481, F380 - I477, L381 - I477, L381 - L478, L381 - M481, L384 - I477, I392 - M459, I392 - I463, I392 - L466, V395 - M459, V395 - I463, V395 - L466, F402 - F437, F402 - F439, F402 - L440, F402 - L455, F402 - M458, L406 - F437, L406 - L440, L406 - L443, L406 - L448, L406 - L452, L406 - L455, A409 - L440, A409 - L443, A409 - L448	S373 - Q489	NA
<b>MST2-MST2 homo-SARAH</b>	D456 - R474, E462 - R469 R469 - E462, R474 - D456	F437 - P476, F437 - I477, F437 - A480, L440 - P476, L440 - I477, L440 - A480, L443 - I477, L443 - A480, L445 - M481, L448 - I477, L448 - A480, L448 - M481, L452 - I477, L452 - L478, L452 - M481, L455 - I477, M459 - I463, M459 - L466, I463 - M459, I463 - I463, I463 - L466, L466 - M458, L466 - M459, L466 - I463, L466 - L466, P476 - F437, P476 - L440, I477 - F437, I477 - L440, I477 - L443, I477 - L448, I477 - L452, I477 - L455, L478 - L452, A480 - F437, A480 - L440, A480 - L443, A480 - L448, M481 - L445, M481 - L448, M481 - L452	NA	NA

<b>RASSF5-RASSF5 homo-SARAH</b>	E385 - R403, E388 - R403, R403 - E388	L377 - L406, L377 - A409, F380 - F402, F380 - L406, F380 - A409, L381 - L406, L384 - F402, L384 - L406, I392 - I392, I392 - V395, V395 - I392, V395 - V395, F402 - F380, F402 - L384	Q389 - Q396, Q396 - Q389	NA
<b>RASSF5(E388A)- MST2 hetero-SARAH</b>	E385 - R474, K398 - E462, R403 - D456	I374 - M481, L377 - I477, L377 - A480 L377 - M481, F380 - I477, L381 - I477, L381 - L478, L381 - M481, L384 - I477, A388 - L466, I392 - M459, I392 - I463, I392 - L466, V395 - M459, V395 - I463, V395 - L466, F402 - F437, F402 - F439, F402 - L440, F402 - L455, F402 - M458, L406 - F437, L406 - F439, L406 - L440, L406 - L443, L406 - L448, L406 - L452, L406 - L455, A409 - L440, A409 - L443, A409 - L448, L410 - L445, L410 - L448, L410 - L452	S373 - N490	NA
<b>RASSF5(E388K)- MST2 hetero-SARAH</b>	E385 - R474, K398 - E462, R403 - D456	I374 - M481, L377 - I477, L377 - A480, L377 - M481, F380 - I477, L381 - I477, L381 - L478, L381 - M481, L384 - I477, I392 - M459, I392 - I463, I392 - L466, V395 - M459, V395 - I463, V395 - L466, F402 - F437, F402 - F439, F402 - L440, F402 - L455, F402 - M458, L406 - F437, L406 - L440, L406 - L443, L406 - L448, L406 - L452, L406 - L455, A409 - L440, A409 - L448, L410 - L445, L410 - L448, L410 - L452	S373 - N490	NA
<b>RASSF5(E388A)- RASSF5(E388A) homo-SARAH</b>	R403 - E385	I374 - L410, L377 - L406, L377 - L410, F380 - F402, F380 - L406, F380 - L410, L381 - L406, L381 - L410, L384 - F402, L384 - L406, A388 - V395, I392 - I392, I392 - V395, V395 - A388, V395 - I392, V395 - V395, F402 - F380, F402 - L384, L406 - I374, L406 - L377, L406 - F380, L406 - L381, L406 - L384, A409 - I374, A409 - L377, A409 - F380, L410 - I374, L410 - P375, L410 - L377, L410 - F380, L410 - L381	Q389 - Q396, Q396 - Q389	NA

<b>RASSF5(E388K)- RASSF5(E388K) homo-SARAH</b>	E385 - R403, D400 - K388, K388 - D400	I374 - L406, I374 - A409, I374 - L410, P375 - L410, L377 - L406, L377 - A409, L377 - L410, F380 - F402, F380 - L406, F380 - L410, L381 - L406, L381 - L410, L384 - F402, L384 - L406, I392 - I392, I392 - V395, V395 - I392, V395 - V395, F402 - F380, F402 - L384, L406 - I374, L406 - L377, L406 - F380, L406 - L381, L406 - L384, A409 - L377, A409 - F380, L410 - I374, L410 - P375, L410 - L377, L410 - F380, L410 - L381	Q389 - Q396, Q396 - Q389	NA
--	--	---	-----------------------------	----

**Table E2.** Types of atomic interactions in the SARAH dimers. The results of replica exchange molecular dynamics (REMD) simulations of SARAH monomer show that the sum of straight and one kinked SARAH domain have about 90% population across 300 K to 360 K. To consider the most probable RASSF5 configurations, we used straight and one kinked SARAH for docking and modeling.

Number of kink	Number of structure	Percentage	kink residue
Fully unfolded	646	1.077%	NA
0	23945	39.908%	NA
1	30512	50.853%	E388
2	4677	7.795%	E388 K398
3	219	0.365%	Q378 E388 Y399
4	1	0.002%	NA

**Table E3.** Types of atomic interactions in the RA-SARAH association for the RASSF5 configurations 1-14. The atomic pair indicates RA – SARAH residue pair in the interface. The pairs will be viewed effective and counted when the residue pairs have more than 50 % occurrence rate. Letters in italic denote the cation- $\pi$  interaction.

RASSF5 configuration	Salt bridge	Nonpolar	Polar	H-bond
<b>Config. 1</b>	K214 - E387, R221 - D370, R221 - E376, K223 - D370, K223 - E376, K241 - E366, K241 - E368	I215 - L381, I215 - L384, M228 - L377, L230 - I383, L230 - L384, I287 - V367, V354 - F380	Q209 - Q394, Q213 - Q394, T220 - Q378, T220 - N379, T220 - T382	NA
<b>Config. 2</b>	R257 - D400, R257 - E407, K307 - E387, K307 - D390, K334 - E366, K334 - E368	V252 - F402, V252 - L406, P253 - F402, P253 - L406, A254 - F402, A254 - L406, I256 - L406, I256 - A409, I256 - L410, P258 - F402, P258 - L406, P258 - L410, I261 - I392, I261 - V395, V298 - W369, I302 - W369, M310 - V395, M310 - F402, V311 - V395, V312 - I392, V312 - V395, P315 - F380, P315 - L384, A319 - V367, A319 - W369, I337 - V367, I337 - W369, I337 - A371, I337 - F372, I337 - F380, A338 - W369, A338 - A371, A338 - F372, A338 - F380	Q259 - Q396, Q259 - Y399, S260 - Q396, S260 - Y399, Y262 - Y399	NA
<b>Config. 3</b>	R257 - E388, K266 - E385, K266 - E388, K307 - E387, K308 - E387, K334 - E368	I256 - I392, I256 - V395, P258 - I392, P258 - V395, V268 - F372, V268 - I374, V268 - L377, V268 - F380, V268 - L381, V268 - L384, L270 - V367, L270 - W369, L270 - A371, L270 - F372, L270 - I374, L270 - L377, A272 - V367, A272 - W369, V298 - W369, I302 - W369, L306 - W369, L306 - F380, L306 - L384, M310 - L384, V311 - F380, V311 - L384, V312 - F380, V312 - L381, V312 - L384, P315 - W369, P315 - A371, P315 - F372, P315 - L377, P315 - F380, F318 - W369, A319 - W369, L320 - W369, I337 - W369	NA	NA

<b>Config. 4</b>	K212 - E387, E232 - K391, D233 - K391, K334 - E368, R340 - D370, R340 - E376	L208 - L377, L208 - F380, L208 - L381, L208 - I383, L208 - L384, I211 - I383, L230 - I383, V298 - W369, I302 - V367, I302 - W369, P315 - W369, F318 - V367, A319 - V367, L331 - W369, I337 - V367, I337 - W369, I337 - A371, I337 - F372, A338 - W369, A338 - A371, A338 - F372, A338 - L377	T295 - N379, Y343 - N379	NA
<b>Config. 5</b>	K214 - E387, K322 - D370, K326 - E376, K334 - D370, R340 - E387	F321 - V367, F321 - W369, F321 - P417, F332 - V367, L335 - W369, I337 - W369, I337 - F372, I337 - L377, I337 - F380, I337 - I383, I337 - L384, I337 - V395, A338 - F380, A338 - I383, A338 - L384, L347 - F380, L347 - I383, L347 - F402, A348 - I383, P350 - I383, V354 - I383	Q213 - Q389, S217 - Q389, T297 - Q389, H325 - N379, Q329 - Q414, Y343 - Y399	NA
<b>Config. 6</b>	K245 - E376, R247 - D370, R277 - D370, R323 - E387, K334 - E366, K334 - E368, K334 - D370, K360 - E376, E353 - K386	L246 - W369, L246 - P375, V250 - L410, F280 - A409, F280 - L410, L282 - A409, L282 - L410, P283 - I374, P283 - P375, P283 - L406, P283 - A409, P283 - L410, L284 - P375, F321 - W369, F321 - F372, V330 - F372, V330 - F380, V330 - I383, V330 - L384, V330 - F402, L331 - W369, L331 - F372, L331 - F380, L331 - F402, F332 - V367, F332 - W369, F332 - F372, F332 - F380, F332 - I383, F332 - F402, V358 - W369	H243 - N379, T251 - S413, Y262 - K416, Y281 - S373, Y281 - Q404, H325 - N379, H325 - Q389, Q329 - Q389, Q329 - Y399	Y281 - E408, Y281 - L410
<b>Config. 7</b>	K245 - E407, R247 - E366, R323 - D400, K334 - E385, K360 - E366	I265 - F372, V268 - F372, L270 - A371, L270 - F372, L270 - I374, L270 - P375, A271 - P375, A272 - I374, A272 - P375, F280 - A409, L282 - A409, P283 - L406, P283 - A409, F321 - F380, L331 - I392, I337 - L381	T273 - Q378, T274 - Q378, Q329 - Q393, Q329 - Q394, Q333 - Q389, Q333 - Q393, Q333 - Q396, Q333 - Y399, N362 - Q378, N362 - Y399, T364 - S373, T364 - Q378	F332 - Q396
<b>Config. 8</b>	K223 - E388, R247 - E368, R248 - E407, R257 - E366, R257 - D370, R323 - E385, E353 - K386	V250 - F372, V250 - L406, V250 - A409, V250 - L410, V252 - F372, V252 - L406, V252 - A409, V252 - L410, P253 - F372, P253 - L406, P253 - A409, P253 - L410, A254 - F372, A254 - A409, A254 - L410, I256 - W369, P258 - W369, P258 - A409, F280 - W369, F280 - A371, F280 - F372, F280 - L406, F280 - L410, L282 - A371,	H243 - Q378, Q289 - Q389, Q289 - Q393	Y281 - D370, Y281 - F372, P283 - Y399

		L282 - F372, L282 - L377, L282 - F402, L282 - L406, L282 - L410, P283 - A371, P283 - F372, P283 - I374, P283 - L377, P283 - F380, P283 - L381, L284 - F380, L284 - F402, A286 - F380, A286 - L381, A286 - L384, A286 - I392, I287 - L381, I287 - L384, I287 - I392		
<b>Config. 9</b>	E267 - K391, E267 - K398, K276 - D400, K276 - E407, K307 - E376, K334 - E366	P249 - F402, P249 - L406, P249 - L410, V250 - L406, V250 - L410, V252 - F402, V252 - L406, V252 - A409, V252 - L410, P253 - L406, P253 - A409, P253 - L410, A254 - L410, I256 - A409, I256 - L410, I261 - F402, I261 - L406, I261 - A409, I261 - L410, A264 - F402, I265 - V395, I265 - F402, V268 - F372, V268 - F380, V268 - I383, V268 - L384, V268 - I392, V268 - V395, L270 - F372, L270 - L377, L270 - F380, L270 - L381, L270 - I383, L270 - L384, L270 - I392, L270 - V395, A271 - F380, A271 - L381, A271 - I383, A271 - L384, A271 - I392, A271 - V395, A272 - F380, F280 - L406, F280 - L410, V298 - W369, I302 - W369, I302 - A371, L306 - F372, V312 - F380, P315 - W369, P315 - A371, P315 - F372, F318 - W369, A319 - W369, L320 - W369, I337 - V367, I337 - W369, I337 - A371	Y262 - Y399, N269 - Q396, N269 - Y399, T273 - Q396, T273 - Y399, T274 - Q396, T274 - Y399, T278 - Q404, S279 - Q404, Y281 - Y399, T297 - S373, S299 - S373, Q303 - S373, N314 - Y399	T274 - R403, Q316 - W369, N362 - E366
<b>Config. 10</b>	K245 - D370, K245 - E376, R247 - E366, K266 - E407, E267 - R403, E300 - K405, R323 - E376, K334 - E387, K360 - E376,	L246 - W369, A264 - L406, A264 - L410, I265 - F402, I265 - L406, I265 - L410, V268 - L406, V268 - L410, L270 - I392, L270 - L406, L270 - L410, A272 - I392, F280 - V367, F280 - F372, L282 - F372, P283 - W369, P283 - F372, P315 - F402, P315 - L406, A319 - L384, F321 - F380, F321 - I383, F321 - L384, V330 - P375, L331 - I383, F332 - F380, F332 - I383, I337 - V395, I337 - F402, A338 - F402, V358 - F380	T235 - Q414, N269 - Q393, T274 - Q389, Y281 - S373, T295 - Q414, S299 - Y399, Q303 - Y399, N314 - Q396, N314 - Y399, Q316 - Q396, Q316 - Y399, Q329 - N379, Q333 - N379, N362 - Q389, N362 - Q396, T364 - Q389, T364 - Q393	NA

<p><b>Config. 11</b></p>	<p>K214 - E385, R221 - E388, K241 - E412, R247 - E366, R323 - E412, K326 - E376, K326 - E387, D327 - K386, K334 - E366, K334 - D370, D351 - K391, D351 - K398, D351 - K401, E353 - K391, E353 - K397, E353 - K398, E353 - K401, E353 - K405, K360 - E366, K360 - E368</p>	<p>P206 - L381, V268 - W369, L270 - V367, L270 - W369, L270 - F372, A271 - V367, A271 - F372, A272 - V367, I287 - F402, I287 - L406, I287 - A409, I287 - L410, V298 - W369, I324 - L377, I324 - F380, I324 - I383, I324 - L384, L331 - W369, L331 - F380, L331 - I383, F332 - V367, F332 - W369, L335 - F372, L335 - F380, I337 - W369, L344 - L377, L344 - F380, L344 - L381, L344 - L384, L347 - L381, L347 - L384, A348 - L384, P350 - V395, V354 - I383, V354 - L384, L355 - F380, L355 - L384, F357 - F380</p>	<p>Q333 - S373</p>	<p>T364 - E368</p>
<p><b>Config. 12</b></p>	<p>K245 - E412, K308 - E385, K308 - E388, R323 - E412, K334 - E366, E222 - K398, E353 - K405</p>	<p>L226 - V395, F239 - I383, F239 - L384, V268 - I374, L270 - A371, L270 - F372, A271 - V367, A272 - V367, A272 - W369, A286 - F402, A286 - L406, I287 - F402, I287 - L406, I287 - L410, L290 - L377, L290 - F380, L290 - L381, L290 - L384, I292 - L377, I292 - F380, V301 - L377, I302 - W369, L305 - L381, L306 - W369, P315 - W369, P315 - A371, P315 - I374, F318 - W369, A319 - W369, F332 - L410, I337 - V367, I337 - W369, I337 - F372, A338 - F372, V358 - L410</p>	<p>N224 - Q394, C225 - Q394, Q289 - Y399, Q303 - S373, H325 - S413, Q329 - S413, T352 - S413</p>	<p>Q316 - W369</p>
<p><b>Config. 13</b></p>	<p>K223 - E408, K245 - E387, R247 - D370, R257 - E385, R323 - E387, R323 - E388, E222 - K398, D285 - K386, D351 - K401, E353 - K401</p>	<p>P253 - P375, A254 - P375, I256 - I374, I256 - P375, P258 - I383, I261 - I383, A272 - V367, F280 - P375, L282 - F380, L282 - I383, P283 - F380, P283 - I383, L284 - I383, V330 - V395</p>	<p>N224 - S413, H243 - Q394, Q259 - N379, Q259 - T382, Y281 - N379, H325 - Q394, H325 - Y399, T352 - Q394, S356 - Q394</p>	<p>Y281 - E376, Y281 - N379</p>
<p><b>Config. 14</b></p>	<p>K307 - E387, K307 - E388, K307 - D390, K334 - E366, R340 - E376</p>	<p>I256 - L406, I256 - A409, I256 - L410, P258 - F402, P258 - L406, P258 - A409, P258 - L410, I261 - V395, I261 - F402, I261 - L406, I261 - A409, A264 - V395, A264 - F402, I265 - V395, I265 - F402, V268 - I392, V268 - V395, L270 - L384, L270 - I392, L270 - V395, V312 - V395, P315 - L384, L335 - V367, L335 - W369, I337 - V367, I337 - W369, I337 - F372, I337 - L377, I337 - F380, I337 - L381, I337 - L384, A338 - W369, A338 - A371, A338 - F372, A338 - L377, A338 - F380, A338 - L381</p>	<p>T297 - N379</p>	<p>NA</p>

**Table E4.** Types of atomic interactions in the RA-SARAH association for the K-Ras4B-GTP/RASSF5 dimeric configurations. The atomic pair indicates RA – SARAH residue pair in the interface. The pairs will be viewed effective and counted when the residue pairs have more than 50 % occurrence rate.

Dimeric configuration	Salt bridge	Nonpolar	Polar	H-bond
<b>DC1(1)</b>	K223 - E376, R323 - E366, R323 - E368, R323 - D370, K360 - E368	I287 - V367, I287 - W369, V330 - V367, V330 - W369, F332 - V367, P350 - F372	N219 - Q378, T220 - Q378, N224 - S373	NA
<b>DC2(4)</b>	K322 - E376, K334 - E368, E210 - R411	P205 - F372, P205 - I374, P205 - I392, P206 - F372, P206 - I374, L208 - L381, L270 - V367, L270 - W369, A272 - V367, V298 - W369, I302 - W369, F318 - V367, A319 - V367, A319 - W369, I337 - V367, I337 - W369, I337 - A371, I337 - F372, I337 - I374, A338 - W369, A338 - A371, A338 - F372, A338 - I374, L344 - P375, L347 - P375, L347 - L377, L347 - L381, A348 - L377	T207 - Q389, T207 - Q393, T207 - Q396, Q209 - Q389, Q209 - Q393, Q209 - Q396, Q209 - Q404, S336 - S373, S336 - Q378, Y343 - Q378, Y343 - T382	NA
<b>DC3(6)</b>	K276 - E408, K276 - E412, R323 - E387, K334 - D370, K360 - E376, D327 - K398	P283 - P375, I287 - I383, F321 - F372, V330 - F380, V330 - L384, V330 - V395, L331 - F380, L331 - V395, F332 - F372, F332 - L377, F332 - F380, V358 - F372	H243 - N379, T278 - S413, T278 - Q414, S279 - Q414, H325 - Q389, Q329 - Q394, Q333 - Y399	F332 - Y399
<b>DC4(9)</b>	K276 - E407, K276 - E408, K317 - E368, K334 - E366	P249 - L410, V250 - L410, V252 - A409, V252 - L410, A254 - A409, A254 - L410, I256 - A409, I261 - A409, I261 - L410, A264 - F402, A264 - L406, I265 - F402, I265 - L406, V268 - F372, V268 - F380, V268 - I383, L270 - F380, L270 - I383, L270 - L384, L270 - I392, L270 - V395, A271 - F380, A271 - I392, A271 - V395, F280 - L410, I302 - A371, P315 - A371, P315 - F372, I337 - W369, I337 - A371, I337 - F372, I337 - I374, A338 - I374	T251 - S413, T251 - Q414, S260 - S413, Y262 - Y399, N269 - Q396, N269 - Y399, T273 - Q396, T273 - Y399, T274 - Q396, T274 - Y399, T297 - S373, S299 - S373, Q303 - S373, N314 - Y399	T274 - Y399

<b>DC5(10)</b>	K245 - E366, K245 - D370, K245 - E376, R323 - E376, K334 - E387, K360 - E366, K360 - E376	L282 - V367, P283 - V367, P283 - F372, P315 - F402, F321 - F380, F321 - I383, V330 - F372, V330 - P375, F332 - F380, F332 - I383, I337 - V395	S299 - Y399, Q303 - Y399, N314 - Q396, N314 - Y399, Q316 - Q396, Q316 - Y399, Q333 - N379, Q333 - T382, N362 - Q389,	NA
<b>DC6(11)</b>	K214 - E385, K214 - E388, K241 - E407, K245 - E366, K326 - D370, K326 - E376, K360 - E366	L226 - L406, I287 - L406, I287 - A409, F321 - W369, I324 - L377, I324 - F380, L331 - W369, L331 - A371, F332 - V367, F332 - W369, L344 - L377, L344 - F380, L344 - L381, L344 - L384, L347 - F380, L347 - L381, L347 - L384, A348 - F380, A348 - L381, A348 - L384, G349 - L384, G349 - I392, P350 - I392, V354 - F380, V354 - I383, V354 - L384, L355 - L377, L355 - F380, L355 - L381, L355 - L384, F357 - L406	NA	NA
<b>DC7(14)</b>	K307 - D390, K334 - E366, R340 - D370	P205 - A371, P205 - F372, P205 - I374, P205 - P375, L208 - P375, I211 - P375, P253 - L410, P253 - P417, A254 - P417, P258 - F402, P258 - L406, P258 - A409, P258 - L410, I261 - V395, I261 - F402, I261 - L406, I261 - L410, V268 - I392, L306 - I392, M310 - V395, M310 - F402, V311 - V395, V312 - I392, V312 - V395, P315 - I392, L335 - V367, I337 - V367, I337 - W369, A338 - W369, A338 - A371, A338 - F380, L344 - W369	T235 - S373, T235 - N379, T251 - S413, T251 - Q414, Q259 - Q414, S260 - Y399, S294 - S373, S294 - N379, T295 - S373, T295 - N379	S336 - W369

**Table E5.** The WISP calculates 100 desirable allosteric pathways between two selected residues for K-Ras4B-GTP/RASSF5 dimeric configurations, DC3(6), DC4(9), DC5(10), and DC7(14). The table shows the occurrence rate of each residue on the allosteric pathway. Except the source and sink residue which have the 100% occurrence rate, based on the results, the higher occurrence rate may imply the more importance.

Dimeric configuration	K-Ras4B		RASSF5			
	Residue	Rate (%)	RA residue	Rate (%)	SARAH residue	Rate (%)
DC3(6)	G13	12				
	V14	20				
	K16	3				
	S17	1				
	Q22	4				
	L23	4				
	F28	1				
	V29	4				
	Y32	44				
	D33	66	K245	5	N379	34
	P34	8	L246	5	L381	34
	T35	24	D285	29	T382	4
	D38	25	V301	66	L384	14
	S39	25	G304	66	E385	13
	R41	4	K308	66	E387	58
	K42	4	M310	5	E388	20
	D57	1	V311	5	Q389	30
	A83	25	R323	66	D390	30
	I84	58	Q329	7	I392	59
	N85	23	V330	7	V395	19
	N116	1	P341	66	Q396	100
	K117	79	R345	66	D400	100
	D119	2	L355	66		
	T124	15				
	V125	99				
	D126	100				
	S145	3				
	A146	1				
	K147	1				
	GTP	81				
	Mg <sup>2+</sup>	15				

<b>DC4(9)</b>	G13 V14 G15 S17 Y32 P34 T35 I36 D38 A83 N86 S89 N116 K117 GTP Mg <sup>2+</sup>	28 63 4 44 54 55 45 55 100 62 2 100 2 37 78 44	T237 G238 F239 Q289 L290 H291 I292 S293 T296 T297 E300	58 85 87 100 13 2 68 75 100 32 68	S373 P375 N379 I383 E387	100 100 100 100 100
<b>DC5(10)</b>	G15 K16 S17 Y40 I55 L56 D57 T58 GTP Mg <sup>2+</sup>	8 30 78 100 2 100 77 18 16 48	L306 F309 V311 V312 D313 N314 P315 Q316	6 100 97 33 94 4 98 100	E387 E388 Q389 D390 K391 I392 Q394 V395 Q396	100 98 2 18 93 9 28 91 11
<b>DC6(11)</b>	S17 T20 L23 T35 Y40 A59 R68 D69 Q70 Y71 Mg <sup>2+</sup>	7 100 100 7 93 7 4 1 7 3 7	V242 H243 L244 D285 A286 I287 R323 D351 T352 E353 V354 L355 S356 F357 V358 L359	29 67 15 14 86 4 1 19 16 37 59 76 86 63 54 5	D390 K391 I392 Q393 Q394 V395 Q396 Y399 D400 R403 Q404	100 93 2 3 8 4 3 4 5 2 7
<b>DC7(14)</b>	D38 S39 Y40	1 100 4	Y236 T237 G238 I240 V242 H243 I287 K288 L290 H291 I292 S293 S294 T295 T296 T297 V301 F309 M310 L342	17 25 31 50 86 2 86 11 41 4 67 31 92 8 12 1 9 3 3 18	S373 P375 E376 Q378 N379 T382 I383 K386 E388 K391 I392 V395	100 14 100 1 6 3 1 3 3 3 3 3

## Bibliography

1. Margolis, B. and E.Y. Skolnik, *Activation of Ras by receptor tyrosine kinases*. J Am Soc Nephrol, 1994. **5**(6): p. 1288-99.
2. Yamazaki, T., et al., *Role of Grb2 in EGF-stimulated EGFR internalization*. J Cell Sci, 2002. **115**(Pt 9): p. 1791-802.
3. Mayer, B.J., *The discovery of modular binding domains: building blocks of cell signalling*. Nat Rev Mol Cell Biol, 2015. **16**(11): p. 691-8.
4. Lu, S., et al., *Ras Conformational Ensembles, Allostery, and Signaling*. Chem Rev, 2016. **116**(11): p. 6607-65.
5. Chardin, P., et al., *The Grb2 adaptor*. FEBS Lett, 1995. **369**(1): p. 47-51.
6. Cherfils, J. and P. Chardin, *GEFs: structural basis for their activation of small GTP-binding proteins*. Trends Biochem Sci, 1999. **24**(8): p. 306-11.
7. Chardin, P., et al., *Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2*. Science, 1993. **260**(5112): p. 1338-43.
8. Okkenhaug, K. and R. Rottapel, *Grb2 forms an inducible protein complex with CD28 through a Src homology 3 domain-proline interaction*. J Biol Chem, 1998. **273**(33): p. 21194-202.
9. Young, A., D. Lou, and F. McCormick, *Oncogenic and wild-type Ras play divergent roles in the regulation of mitogen-activated protein kinase signaling*. Cancer Discov, 2013. **3**(1): p. 112-23.
10. Samatar, A.A. and P.I. Poulikakos, *Targeting RAS-ERK signalling in cancer: promises and challenges*. Nat Rev Drug Discov, 2014. **13**(12): p. 928-42.
11. Zhang, M.Z., H. Jang, and R. Nussinov, *The mechanism of PI3K activation at the atomic level*. Chemical Science, 2019. **10**(12): p. 3671-3680.
12. Morris, V. and S. Kopetz, *BRAF inhibitors in clinical oncology*. F1000Prime Rep, 2013. **5**: p. 11.
13. McNamara, C.R. and A. Degterev, *Small-molecule inhibitors of the PI3K signaling network*. Future Med Chem, 2011. **3**(5): p. 549-65.
14. Lu, S., et al., *Inhibitors of Ras-SOS Interactions*. ChemMedChem, 2016. **11**(8): p. 814-21.
15. Cussac, D., et al., *A Sos-derived peptidimer blocks the Ras signaling pathway by binding both Grb2 SH3 domains and displays antiproliferative activity*. FASEB J, 1999. **13**(1): p. 31-8.
16. Ye, Y.B., et al., *The cytotoxicity of a Grb2-SH3 inhibitor in Bcr-Abl positive K562 cells*. Biochem Pharmacol, 2008. **75**(11): p. 2080-91.
17. Yu, Y., et al., *Targeted Covalent Inhibition of Grb2-Sos1 Interaction through Proximity-Induced Conjugation in Breast Cancer Cells*. Mol Pharm, 2017. **14**(5): p. 1548-1557.
18. Oneyama, C., H. Nakano, and S.V. Sharma, *UCS15A, a novel small molecule, SH3 domain-mediated protein-protein interaction blocking drug*. Oncogene, 2002. **21**(13): p. 2037-50.
19. Nguyen, J.T., et al., *Exploiting the basis of proline recognition by SH3 and WW domains: design of N-substituted inhibitors*. Science, 1998. **282**(5396): p. 2088-92.

20. Gril, B., et al., *Grb2-SH3 ligand inhibits the growth of HER2+ cancer cells and has antitumor effects in human cancer xenografts alone and in combination with docetaxel*. Int J Cancer, 2007. **121**(2): p. 407-15.
21. Maignan, S., et al., *Crystal structure of the mammalian Grb2 adaptor*. Science, 1995. **268**(5208): p. 291-3.
22. Kessels, H.W., A.C. Ward, and T.N. Schumacher, *Specificity and affinity motifs for Grb2 SH2-ligand interactions*. Proc Natl Acad Sci U S A, 2002. **99**(13): p. 8524-9.
23. Ettmayer, P., et al., *Structural and conformational requirements for high-affinity binding to the SH2 domain of Grb2(1)*. J Med Chem, 1999. **42**(6): p. 971-80.
24. Nussinov, R., et al., *Autoinhibition in Ras effectors Raf, PI3Kalpha, and RASSF5: a comprehensive review underscoring the challenges in pharmacological intervention*. Biophys Rev, 2018. **10**(5): p. 1263-1282.
25. Liao, T.J., et al., *Allosteric KRas4B Can Modulate SOS1 Fast and Slow Ras Activation Cycles*. Biophys J, 2018. **115**(4): p. 629-641.
26. Saksela, K. and P. Permi, *SH3 domain ligand binding: What's the consensus and where's the specificity?* FEBS Lett, 2012. **586**(17): p. 2609-14.
27. Terasawa, H., et al., *Structure of the N-terminal SH3 domain of GRB2 complexed with a peptide from the guanine nucleotide releasing factor Sos*. Nat Struct Biol, 1994. **1**(12): p. 891-7.
28. Lewitzky, M., et al., *Mona/Gads SH3C binding to hematopoietic progenitor kinase 1 (HPK1) combines an atypical SH3 binding motif, R/KXXX, with a classical PXXP motif embedded in a polyproline type II (PPII) helix*. J Biol Chem, 2004. **279**(27): p. 28724-32.
29. Schaeper, U., et al., *Coupling of Gab1 to c-Met, Grb2, and Shp2 mediates biological responses*. J Cell Biol, 2000. **149**(7): p. 1419-32.
30. Harkiolaki, M., et al., *Distinct binding modes of two epitopes in Gab2 that interact with the SH3C domain of Grb2*. Structure, 2009. **17**(6): p. 809-22.
31. Harkiolaki, M., et al., *Structural basis for SH3 domain-mediated high-affinity binding between Mona/Gads and SLP-76*. EMBO J, 2003. **22**(11): p. 2571-82.
32. Chook, Y.M., et al., *The Grb2-mSos1 complex binds phosphopeptides with higher affinity than Grb2*. J Biol Chem, 1996. **271**(48): p. 30472-8.
33. McDonald, C.B., et al., *SH3 domains of Grb2 adaptor bind to PXXpsiPXR motifs within the Sos1 nucleotide exchange factor in a discriminate manner*. Biochemistry, 2009. **48**(19): p. 4074-85.
34. Nishida, M., et al., *Novel recognition mode between Vav and Grb2 SH3 domains*. EMBO J, 2001. **20**(12): p. 2995-3007.
35. Arya, R., et al., *Grb2 carboxyl-terminal SH3 domain can bivalently associate with two ligands, in an SH3 dependent manner*. Sci Rep, 2017. **7**(1): p. 1284.
36. Vidal, M., et al., *Molecular and cellular analysis of Grb2 SH3 domain mutants: interaction with Sos and dynamin*. J Mol Biol, 1999. **290**(3): p. 717-30.
37. Kohda, D., et al., *Solution structure and ligand-binding site of the carboxy-terminal SH3 domain of GRB2*. Structure, 1994. **2**(11): p. 1029-40.

38. McDonald, C.B., et al., *Structural basis of the differential binding of the SH3 domains of Grb2 adaptor to the guanine nucleotide exchange factor Sos1*. Arch Biochem Biophys, 2008. **479**(1): p. 52-62.
39. Liao, T.J., et al., *RASSF5: An MST activator and tumor suppressor in vivo but opposite in vitro*. Curr Opin Struct Biol, 2016. **41**: p. 217-224.
40. Liao, T.J., et al., *The dynamic mechanism of RASSF5 and MST kinase activation by Ras*. Phys Chem Chem Phys, 2017. **19**(9): p. 6470-6480.
41. McDonald, C.B., et al., *Allostery mediates ligand binding to Grb2 adaptor in a mutually exclusive manner*. J Mol Recognit, 2013. **26**(2): p. 92-103.
42. Sastry, L., et al., *Quantitative analysis of Grb2-Sos1 interaction: the N-terminal SH3 domain of Grb2 mediates affinity*. Oncogene, 1995. **11**(6): p. 1107-12.
43. Yuzawa, S., et al., *Backbone assignments of Grb2 complexed with ligand peptides for SH3 and SH2 domains*. J Biomol NMR, 2003. **27**(2): p. 185-6.
44. Lin, C.C., et al., *Inhibition of basal FGF receptor signaling by dimeric Grb2*. Cell, 2012. **149**(7): p. 1514-24.
45. Varadan, R., et al., *Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling*. J Biol Chem, 2004. **279**(8): p. 7055-63.
46. Fiser, A. and A. Sali, *Modeller: generation and refinement of homology-based protein structure models*. Methods Enzymol, 2003. **374**: p. 461-91.
47. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD*. J Comput Chem, 2005. **26**(16): p. 1781-802.
48. Brooks, B.R., et al., *CHARMM: the biomolecular simulation program*. J Comput Chem, 2009. **30**(10): p. 1545-614.
49. Liao, T.J., et al., *High-Affinity Interactions of the nSH3/cSH3 Domains of Grb2 with the C-Terminal Proline-Rich Domain of SOS1*. J Am Chem Soc, 2020. **142**(7): p. 3401-3411.
50. Bartelt, R.R., et al., *Regions outside of conserved PxxPxR motifs drive the high affinity interaction of GRB2 with SH3 domain ligands*. Biochim Biophys Acta, 2015. **1853**(10 Pt A): p. 2560-9.
51. Hosoe, Y., et al., *Structural and functional properties of Grb2 SH2 dimer in CD28 binding*. Biophys Physicobiol, 2019. **16**: p. 80-88.
52. McDonald, C.B., et al., *Binding of the cSH3 domain of Grb2 adaptor to two distinct RXXK motifs within Gab1 docker employs differential mechanisms*. J Mol Recognit, 2011. **24**(4): p. 585-96.
53. Klauda, J.B., et al., *Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types*. J Phys Chem B, 2010. **114**(23): p. 7830-43.
54. Jang, H., et al., *Membrane-associated Ras dimers are isoform-specific: K-Ras dimers differ from H-Ras dimers*. Biochem J, 2016. **473**(12): p. 1719-32.
55. Jang, H., et al., *The Structural Basis of the Farnesylated and Methylated KRas4B Interaction with Calmodulin*. Structure, 2019. **27**(11): p. 1647-1659 e4.
56. Jang, H., M. Zhang, and R. Nussinov, *The quaternary assembly of KRas4B with Raf-1 at the membrane*. Comput Struct Biotechnol J, 2020. **18**: p. 737-748.

57. Corbalan-Garcia, S., et al., *Identification of the mitogen-activated protein kinase phosphorylation sites on human Sos1 that regulate interaction with Grb2*. Mol Cell Biol, 1996. **16**(10): p. 5674-82.
58. Porfiri, E. and F. McCormick, *Regulation of epidermal growth factor receptor signaling by phosphorylation of the ras exchange factor hSOS1*. J Biol Chem, 1996. **271**(10): p. 5871-7.
59. Kamioka, Y., et al., *Multiple decisive phosphorylation sites for the negative feedback regulation of SOS1 via ERK*. J Biol Chem, 2010. **285**(43): p. 33540-8.
60. Lake, D., S.A. Corrêa, and J. Müller, *Negative feedback regulation of the ERK1/2 MAPK pathway*. Cell Mol Life Sci, 2016. **73**(23): p. 4397-4413.
61. Nussinov, R., C.J. Tsai, and H. Jang, *Does Ras Activate Raf and PI3K Allosterically?* Front Oncol, 2019. **9**: p. 1231.
62. Saha, M., et al., *RSK phosphorylates SOS1 creating 14-3-3-docking sites and negatively regulating MAPK activation*. Biochem J, 2012. **447**(1): p. 159-66.
63. Zhou, H., et al., *Toward a comprehensive characterization of a human cancer cell phosphoproteome*. J Proteome Res, 2013. **12**(1): p. 260-71.
64. Bian, Y., et al., *An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome*. J Proteomics, 2014. **96**: p. 253-62.
65. Nussinov, R., et al., *Phosphorylation and Driver Mutations in PI3K $\alpha$  and PTEN Autoinhibition*. Mol Cancer Res, 2020.
66. Brummer, T., et al., *Phosphorylation-dependent binding of 14-3-3 terminates signalling by the Gab2 docking protein*. EMBO J, 2008. **27**(17): p. 2305-16.
67. Wöhrle, F.U., R.J. Daly, and T. Brummer, *How to Grb2 a Gab*. Structure, 2009. **17**(6): p. 779-81.
68. Ding, C.B., et al., *Structure and function of Gab2 and its role in cancer (Review)*. Mol Med Rep, 2015. **12**(3): p. 4007-4014.
69. McDonald, C.B., et al., *Bivalent binding drives the formation of the Grb2-Gab1 signaling complex in a noncooperative manner*. FEBS J, 2012. **279**(12): p. 2156-73.
70. Cussac, D., M. Frech, and P. Chardin, *Binding of the Grb2 SH2 domain to phosphotyrosine motifs does not change the affinity of its SH3 domains for Sos proline-rich motifs*. EMBO J, 1994. **13**(17): p. 4011-21.
71. Lemmon, M.A., et al., *Independent binding of peptide ligands to the SH2 and SH3 domains of Grb2*. J Biol Chem, 1994. **269**(50): p. 31653-8.
72. Liao, T.J., et al., *SOS1 interacts with Grb2 through regions that induce closed nSH3 conformations*. J Chem Phys, 2020. **153**(4): p. 045106.
73. Sanches, K., et al., *The dynamics of free and phosphopeptide-bound Grb2-SH2 reveals two dynamically independent subdomains and an encounter complex with fuzzy interactions*. Sci Rep, 2020. **10**(1): p. 13040.
74. Thornton, K.H., et al., *Nuclear magnetic resonance solution structure of the growth factor receptor-bound protein 2 Src homology 2 domain*. Biochemistry, 1996. **35**(36): p. 11852-64.
75. Papaioannou, D., et al., *Structural and biophysical investigation of the interaction of a mutant Grb2 SH2 domain (W121G) with its cognate phosphopeptide*. Protein Sci, 2016. **25**(3): p. 627-37.

76. Pinet, L., et al., [Formula: see text]H, [Formula: see text]C and [Formula: see text]N assignments of human Grb2 free of ligands. *Biomol NMR Assign*, 2020. **14**(2): p. 323-327.
77. Goddard, T.D. and D.G. Kneller, *SPARKY 3*.
78. Varadan, R., et al., *Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain*. *Mol Cell*, 2005. **18**(6): p. 687-98.
79. Hall, J.B. and D. Fushman, *Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G*. *J Biomol NMR*, 2003. **27**(3): p. 261-75.
80. Prior, I.A., P.D. Lewis, and C. Mattos, *A comprehensive survey of Ras mutations in cancer*. *Cancer Res*, 2012. **72**(10): p. 2457-67.
81. Pylayeva-Gupta, Y., E. Grabocka, and D. Bar-Sagi, *RAS oncogenes: weaving a tumorigenic web*. *Nat Rev Cancer*, 2011. **11**(11): p. 761-74.
82. Zhou, Y., et al., *Signal integration by lipid-mediated spatial cross talk between Ras nanoclusters*. *Mol Cell Biol*, 2014. **34**(5): p. 862-76.
83. Nussinov, R., C.J. Tsai, and H. Jang, *Is Nanoclustering essential for all oncogenic KRas pathways? Can it explain why wild-type KRas can inhibit its oncogenic variant?* *Semin Cancer Biol*, 2018. DOI: p. 10.1016/j.semcancer.2018.01.002.
84. Lepri, F., et al., *SOS1 mutations in Noonan syndrome: molecular spectrum, structural insights on pathogenic effects, and genotype-phenotype correlations*. *Hum Mutat*, 2011. **32**(7): p. 760-72.
85. Tartaglia, M., et al., *Gain-of-function SOS1 mutations cause a distinctive form of Noonan syndrome*. *Nat Genet*, 2007. **39**(1): p. 75-9.
86. Gureasko, J., et al., *Role of the histone domain in the autoinhibition and activation of the Ras activator Son of Sevenless*. *Proc Natl Acad Sci U S A*, 2010. **107**(8): p. 3430-5.
87. Tong, J., et al., *Epidermal growth factor receptor phosphorylation sites Ser991 and Tyr998 are implicated in the regulation of receptor endocytosis and phosphorylations at Ser1039 and Thr1041*. *Mol Cell Proteomics*, 2009. **8**(9): p. 2131-44.
88. Burns, M.C., et al., *Approach for targeting Ras with small molecules that activate SOS-mediated nucleotide exchange*. *Proc Natl Acad Sci U S A*, 2014. **111**(9): p. 3401-6.
89. Boriack-Sjodin, P.A., et al., *The structural basis of the activation of Ras by Sos*. *Nature*, 1998. **394**(6691): p. 337-43.
90. Rojas, J.M., J.L. Oliva, and E. Santos, *Mammalian son of sevenless Guanine nucleotide exchange factors: old concepts and new perspectives*. *Genes Cancer*, 2011. **2**(3): p. 298-305.
91. Margarit, S.M., et al., *Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS*. *Cell*, 2003. **112**(5): p. 685-95.
92. Vo, U., et al., *Monitoring Ras Interactions with the Nucleotide Exchange Factor Son of Sevenless (Sos) Using Site-specific NMR Reporter Signals and Intrinsic Fluorescence*. *J Biol Chem*, 2016. **291**(4): p. 1703-18.

93. Freedman, T.S., et al., *A Ras-induced conformational switch in the Ras activator Son of sevenless*. Proc Natl Acad Sci U S A, 2006. **103**(45): p. 16692-7.
94. Pierre, S., A.S. Bats, and X. Coumoul, *Understanding SOS (Son of Sevenless)*. Biochem Pharmacol, 2011. **82**(9): p. 1049-56.
95. Ford, B., et al., *Structure of the G60A mutant of Ras: implications for the dominant negative effect*. J Biol Chem, 2005. **280**(27): p. 25697-705.
96. Bos, J.L., H. Rehmann, and A. Wittinghofer, *GEFs and GAPs: critical elements in the control of small G proteins*. Cell, 2007. **129**(5): p. 865-77.
97. Sun, Q., et al., *Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation*. Angew Chem Int Ed Engl, 2012. **51**(25): p. 6140-3.
98. Maurer, T., et al., *Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity*. Proc Natl Acad Sci U S A, 2012. **109**(14): p. 5299-304.
99. Evelyn, C.R., et al., *Rational design of small molecule inhibitors targeting the Ras GEF, SOS1*. Chem Biol, 2014. **21**(12): p. 1618-28.
100. Winter, J.J., et al., *Small molecule binding sites on the Ras:SOS complex can be exploited for inhibition of Ras activation*. J Med Chem, 2015. **58**(5): p. 2265-74.
101. Freedman, T.S., et al., *Differences in flexibility underlie functional differences in the Ras activators son of sevenless and Ras guanine nucleotide releasing factor 1*. Structure, 2009. **17**(1): p. 41-53.
102. Sondermann, H., et al., *Structural analysis of autoinhibition in the Ras activator Son of sevenless*. Cell, 2004. **119**(3): p. 393-405.
103. Hall, B.E., et al., *Structure-based mutagenesis reveals distinct functions for Ras switch 1 and switch 2 in Sos-catalyzed guanine nucleotide exchange*. J Biol Chem, 2001. **276**(29): p. 27629-37.
104. Van Wart, A.T., et al., *Weighted Implementation of Suboptimal Paths (WISP): An Optimized Algorithm and Tool for Dynamical Network Analysis*. J Chem Theory Comput, 2014. **10**(2): p. 511-517.
105. Iversen, L., et al., *Molecular kinetics. Ras activation by SOS: allosteric regulation by altered fluctuation dynamics*. Science, 2014. **345**(6192): p. 50-4.
106. Rossman, K.L., et al., *A crystallographic view of interactions between Dbs and Cdc42: PH domain-assisted guanine nucleotide exchange*. EMBO J, 2002. **21**(6): p. 1315-26.
107. Lu, S., et al., *The Structural Basis of Oncogenic Mutations G12, G13 and Q61 in Small GTPase K-Ras4B*. Sci Rep, 2016. **6**: p. 21949.
108. Lu, S., et al., *GTP Binding and Oncogenic Mutations May Attenuate Hypervariable Region (HVR)-Catalytic Domain Interactions in Small GTPase K-Ras4B, Exposing the Effector Binding Site*. J Biol Chem, 2015. **290**(48): p. 28887-900.
109. Gureasko, J., et al., *Membrane-dependent signal integration by the Ras activator Son of sevenless*. Nat Struct Mol Biol, 2008. **15**(5): p. 452-61.
110. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER*. Curr Protoc Protein Sci, 2016. **86**: p. 2.9.1-2.9.37.

111. Durell, S.R., B.R. Brooks, and A. Bennaïm, *Solvent-Induced Forces between 2 Hydrophilic Groups*. Journal of Physical Chemistry, 1994. **98**(8): p. 2198-2202.
112. Jang, H., et al., *Mechanisms of membrane binding of small GTPase K-Ras4B farnesylated hypervariable region*. J Biol Chem, 2015. **290**(15): p. 9465-77.
113. Chavan, T.S., et al., *High-Affinity Interaction of the K-Ras4B Hypervariable Region with the Ras Active Site*. Biophys J, 2015. **109**(12): p. 2602-13.
114. Jang, H., et al., *The higher level of complexity of K-Ras4B activation at the membrane*. FASEB J, 2016. **30**(4): p. 1643-55.
115. Chakrabarti, M., H. Jang, and R. Nussinov, *Comparison of the Conformations of KRAS Isoforms, K-Ras4A and K-Ras4B, Points to Similarities and Significant Differences*. J Phys Chem B, 2016. **120**(4): p. 667-79.
116. Zhang, M., et al., *Phosphorylated Calmodulin Promotes PI3K Activation by Binding to the SH2 Domains*. Biophys J, 2017. **113**(9): p. 1956-1967.
117. Muratcioglu, S., et al., *PDEdelta Binding to Ras Isoforms Provides a Route to Proper Membrane Localization*. J Phys Chem B, 2017. **121**(24): p. 5917-5927.
118. Jang, H., et al., *Flexible-body motions of calmodulin and the farnesylated hypervariable region yield a high-affinity interaction enabling K-Ras4B membrane extraction*. J Biol Chem, 2017. **292**(30): p. 12544-12559.
119. Brooks, B.R., et al., *Charmm - a program for macromolecular energy, minimization, and dynamics calculations*. J. Comp. Chem., 1983. **4**(2): p. 187-217.
120. Im, W., M.S. Lee, and C.L. Brooks, 3rd, *Generalized born model with a simple smoothing function*. J Comput Chem, 2003. **24**(14): p. 1691-702.
121. Donninger, H., et al., *Ras signaling through RASSF proteins*. Semin Cell Dev Biol, 2016.
122. Zanconato, F., M. Cordenonsi, and S. Piccolo, *YAP/TAZ at the Roots of Cancer*. Cancer Cell, 2016. **29**(6): p. 783-803.
123. Sun, S. and K.D. Irvine, *Cellular Organization and Cytoskeletal Regulation of the Hippo Signaling Network*. Trends Cell Biol, 2016.
124. Fallahi, E., N.A. O'Driscoll, and D. Matallanas, *The MST/Hippo Pathway and Cell Death: A Non-Canonical Affair*. Genes (Basel), 2016. **7**(6).
125. Kapoor, A., et al., *Yap1 activation enables bypass of oncogenic Kras addiction in pancreatic cancer*. Cell, 2014. **158**(1): p. 185-97.
126. Shao, D.D., et al., *KRAS and YAP1 converge to regulate EMT and tumor survival*. Cell, 2014. **158**(1): p. 171-84.
127. Lin, L. and T.G. Bivona, *The Hippo effector YAP regulates the response of cancer cells to MAPK pathway inhibitors*. Mol Cell Oncol, 2016. **3**(1): p. e1021441.
128. Nussinov, R., et al., *Oncogenic KRAS signaling and YAP1/beta-catenin: Similar cell cycle control in tumor initiation*. Semin Cell Dev Biol, 2016.
129. Nussinov, R., C.J. Tsai, and H. Jang, *Independent and core pathways in oncogenic KRAS signaling*. Expert Rev Proteomics, 2016: p. 1-6.
130. Lu, S., et al., *Drugging Ras GTPase: a comprehensive mechanistic and signaling structural view*. Chem Soc Rev, 2016.
131. Sharif, G.M. and A. Wellstein, *Cell density regulates cancer metastasis via the Hippo pathway*. Future Oncol, 2015. **11**(24): p. 3253-60.

132. Nussinov, R., et al., *Principles of K-Ras effector organization and the role of oncogenic K-Ras in cancer initiation through G1 cell cycle deregulation*. Expert Rev Proteomics, 2015. **12**(6): p. 669-82.
133. Muratcioglu, S., et al., *GTP-Dependent K-Ras Dimerization*. Structure, 2015. **23**(7): p. 1325-35.
134. Smith, M.J. and M. Ikura, *Integrated RAS signaling defined by parallel NMR detection of effectors and regulators*. Nat Chem Biol, 2014. **10**(3): p. 223-30.
135. Chen, M., et al., *Ras Dimer Formation as a New Signaling Mechanism and Potential Cancer Therapeutic Target*. Mini Rev Med Chem, 2016. **16**(5): p. 391-403.
136. Jambrina, P.G., et al., *Phosphorylation of RAF Kinase Dimers Drives Conformational Changes that Facilitate Transactivation*. Angew Chem Int Ed Engl, 2016. **55**(3): p. 983-6.
137. Peng, S.B., et al., *Inhibition of RAF Isoforms and Active Dimers by LY3009120 Leads to Anti-tumor Activities in RAS or BRAF Mutant Cancers*. Cancer Cell, 2015. **28**(3): p. 384-98.
138. Sayyed-Ahmad, A., et al., *Computational Equilibrium Thermodynamic and Kinetic Analysis of K-Ras Dimerization through an Effector Binding Surface Suggests Limited Functional Role*. J Phys Chem B, 2016.
139. Banerjee, A., et al., *The disordered hypervariable region and the folded catalytic domain of oncogenic K-Ras4B partner in phospholipid binding*. Curr Opin Struct Biol, 2016. **36**: p. 10-7.
140. Hwang, E., et al., *Structural basis of the heterodimerization of the MST and RASSF SARAH domains in the Hippo signalling pathway*. Acta Crystallogr D Biol Crystallogr, 2014. **70**(Pt 7): p. 1944-53.
141. Makbul, C., et al., *Structural and thermodynamic characterization of Nore1-SARAH: a small, helical module important in signal transduction networks*. Biochemistry, 2013. **52**(6): p. 1045-54.
142. Stieglitz, B., et al., *Novel type of Ras effector interaction established between tumour suppressor NORE1A and Ras switch II*. EMBO J, 2008. **27**(14): p. 1995-2005.
143. Nan, X., et al., *Ras-GTP dimers activate the Mitogen-Activated Protein Kinase (MAPK) pathway*. Proc Natl Acad Sci U S A, 2015. **112**(26): p. 7996-8001.
144. Kovrigina, E.A., A.R. Galiakhmetov, and E.L. Kovrigin, *The Ras G Domain Lacks the Intrinsic Propensity to Form Dimers*. Biophys J, 2015. **109**(5): p. 1000-8.
145. Zhou, Y. and J.F. Hancock, *Ras nanoclusters: Versatile lipid-based signaling platforms*. Biochim Biophys Acta, 2015. **1853**(4): p. 841-9.
146. Ni, L., et al., *Structural basis for autoactivation of human Mst2 kinase and its regulation by RASSF5*. Structure, 2013. **21**(10): p. 1757-68.
147. Avruch, J., et al., *Protein kinases of the Hippo pathway: regulation and substrates*. Semin Cell Dev Biol, 2012. **23**(7): p. 770-84.
148. Creasy, C.L., D.M. Ambrose, and J. Chernoff, *The Ste20-like protein kinase, Mst1, dimerizes and contains an inhibitory domain*. J Biol Chem, 1996. **271**(35): p. 21049-53.

149. Hwang, E., et al., *Structural insight into dimeric interaction of the SARAH domains from Mst1 and RASSF family proteins in the apoptosis pathway*. Proc Natl Acad Sci U S A, 2007. **104**(22): p. 9236-41.
150. Richter, A.M., G.P. Pfeifer, and R.H. Dammann, *The RASSF proteins in cancer; from epigenetic silencing to functional characterization*. Biochim Biophys Acta, 2009. **1796**(2): p. 114-28.
151. Avruch, J., et al., *Rassf family of tumor suppressor polypeptides*. J Biol Chem, 2009. **284**(17): p. 11001-5.
152. Aoyama, Y., J. Avruch, and X.F. Zhang, *Nore1 inhibits tumor cell growth independent of Ras or the MST1/2 kinases*. Oncogene, 2004. **23**(19): p. 3426-33.
153. Guo, C., et al., *RASSF1A is part of a complex similar to the Drosophila Hippo/Salvador/Lats tumor-suppressor network*. Curr Biol, 2007. **17**(8): p. 700-5.
154. Oh, H.J., et al., *Role of the tumor suppressor RASSF1A in Mst1-mediated apoptosis*. Cancer Res, 2006. **66**(5): p. 2562-9.
155. Ikeda, M., et al., *Hippo pathway-dependent and -independent roles of RASSF6*. Sci Signal, 2009. **2**(90): p. ra59.
156. Khokhlatchev, A., et al., *Identification of a novel Ras-regulated proapoptotic pathway*. Curr Biol, 2002. **12**(4): p. 253-65.
157. Praskova, M., et al., *Regulation of the MST1 kinase by autophosphorylation, by the growth inhibitory proteins, RASSF1 and NORE1, and by Ras*. Biochem J, 2004. **381**(Pt 2): p. 453-62.
158. Cooper, W.N., et al., *RASSF2 associates with and stabilizes the proapoptotic kinase MST2*. Oncogene, 2009. **28**(33): p. 2988-98.
159. Song, H., et al., *Role of the tumor suppressor RASSF2 in regulation of MST1 kinase activity*. Biochem Biophys Res Commun, 2010. **391**(1): p. 969-73.
160. Vos, M.D., et al., *The pro-apoptotic Ras effector Nore1 may serve as a Ras-regulated tumor suppressor in the lung*. J Biol Chem, 2003. **278**(24): p. 21938-43.
161. Calvisi, D.F., et al., *NORE1A tumor suppressor candidate modulates p21CIP1 via p53*. Cancer Res, 2009. **69**(11): p. 4629-37.
162. Park, J., et al., *Tumor suppressor ras association domain family 5 (RASSF5/NORE1) mediates death receptor ligand-induced apoptosis*. J Biol Chem, 2010. **285**(45): p. 35029-38.
163. Calvisi, D.F., et al., *Ubiquitous activation of Ras and Jak/Stat pathways in human HCC*. Gastroenterology, 2006. **130**(4): p. 1117-28.
164. Calvisi, D.F., et al., *Ras-driven proliferation and apoptosis signaling during rat liver carcinogenesis is under genetic control*. Int J Cancer, 2008. **123**(9): p. 2057-64.
165. Sanchez-Sanz, G., et al., *MST2-RASSF protein-protein interactions through SARAH domains*. Brief Bioinform, 2016. **17**(4): p. 593-602.
166. Chavan, T.S., et al., *Plasma membrane regulates Ras signaling networks*. Cell Logist, 2015. **5**(4): p. e1136374.
167. Cho, K.J. and J.F. Hancock, *Ras nanoclusters: a new drug target? Small GTPases*, 2013. **4**(1): p. 57-60.

168. Kerkhoff, E. and U.R. Rapp, *Cell cycle targets of Ras/Raf signalling*. *Oncogene*, 1998. **17**(11 Reviews): p. 1457-62.
169. Ewen, M.E., *Relationship between Ras pathways and cell cycle control*. *Prog Cell Cycle Res*, 2000. **4**: p. 1-17.
170. Zhang, W. and H.T. Liu, *MAPK signal pathways in the regulation of cell proliferation in mammalian cells*. *Cell Res*, 2002. **12**(1): p. 9-18.
171. Donninger, H., et al., *Cell cycle restriction is more important than apoptosis induction for RASSF1A protein tumor suppression*. *J Biol Chem*, 2014. **289**(45): p. 31287-95.
172. Zhang, S., et al., *Wnt/beta-catenin signaling pathway upregulates c-Myc expression to promote cell proliferation of P19 teratocarcinoma cells*. *Anat Rec (Hoboken)*, 2012. **295**(12): p. 2104-13.
173. Kaveri, D., et al., *beta-Catenin activation synergizes with Pten loss and Myc overexpression in Notch-independent T-ALL*. *Blood*, 2013. **122**(5): p. 694-704.
174. Juan, J., et al., *Diminished WNT -> beta-catenin -> c-MYC signaling is a barrier for malignant progression of BRAFV600E-induced lung tumors*. *Genes Dev*, 2014. **28**(6): p. 561-75.
175. Palomero, T., et al., *NOTCH1 directly regulates c-MYC and activates a feed-forward-loop transcriptional network promoting leukemic cell growth*. *Proc Natl Acad Sci U S A*, 2006. **103**(48): p. 18261-6.
176. Dongre, A., et al., *Non-Canonical Notch Signaling Drives Activation and Differentiation of Peripheral CD4(+) T Cells*. *Front Immunol*, 2014. **5**: p. 54.
177. Sharma, N., et al., *PI3K/AKT/mTOR and sonic hedgehog pathways cooperate together to inhibit human pancreatic cancer stem cell characteristics and tumor growth*. *Oncotarget*, 2015. **6**(31): p. 32039-60.
178. Ramaswamy, B., et al., *Hedgehog signaling is a novel therapeutic target in tamoxifen-resistant breast cancer aberrantly activated by PI3K/AKT pathway*. *Cancer Res*, 2012. **72**(19): p. 5048-59.
179. Kern, D., et al., *Hedgehog/GLI and PI3K signaling in the initiation and maintenance of chronic lymphocytic leukemia*. *Oncogene*, 2015. **34**(42): p. 5341-51.
180. Ilic, N., et al., *PI3K-targeted therapy can be evaded by gene amplification along the MYC-eukaryotic translation initiation factor 4E (eIF4E) axis*. *Proc Natl Acad Sci U S A*, 2011. **108**(37): p. E699-708.
181. Cope, C.L., et al., *Adaptation to mTOR kinase inhibitors by amplification of eIF4E to maintain cap-dependent translation*. *J Cell Sci*, 2014. **127**(Pt 4): p. 788-800.
182. Boussemart, L., et al., *eIF4F is a nexus of resistance to anti-BRAF and anti-MEK cancer therapies*. *Nature*, 2014. **513**(7516): p. 105-9.
183. Muellner, M.K., et al., *A chemical-genetic screen reveals a mechanism of resistance to PI3K inhibitors in cancer*. *Nat Chem Biol*, 2011. **7**(11): p. 787-93.
184. Nussinov, R., B. Ma, and C.J. Tsai, *A broad view of scaffolding suggests that scaffolding proteins can actively control regulation and signaling of multienzyme complexes through allostery*. *Biochim Biophys Acta*, 2013. **1834**(5): p. 820-9.

185. Nussinov, R., C.J. Tsai, and C. Mattos, *'Pathway drug cocktail': targeting Ras signaling based on structural pathways*. Trends Mol Med, 2013. **19**(11): p. 695-704.
186. Volodko, N., et al., *RASSF tumor suppressor gene family: biological functions and regulation*. FEBS Lett, 2014. **588**(16): p. 2671-84.
187. Donniger, H., et al., *Ras signaling through RASSF proteins*. Semin Cell Dev Biol, 2016. **58**: p. 86-95.
188. Sanchez-Sanz, G., et al., *SARAH Domain-Mediated MST2-RASSF Dimeric Interactions*. PLoS Comput Biol, 2016. **12**(10): p. e1005051.
189. Sun, S. and K.D. Irvine, *Cellular Organization and Cytoskeletal Regulation of the Hippo Signaling Network*. Trends Cell Biol, 2016. **26**(9): p. 694-704.
190. Plouffe, S.W., A.W. Hong, and K.L. Guan, *Disease implications of the Hippo/YAP pathway*. Trends Mol Med, 2015. **21**(4): p. 212-22.
191. Vassilev, A., et al., *TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm*. Genes Dev, 2001. **15**(10): p. 1229-41.
192. Nussinov, R., et al., *Oncogenic KRAS signaling and YAP1/ $\beta$ -catenin: Similar cell cycle control in tumor initiation*. Semin Cell Dev Biol, 2016. **58**: p. 79-85.
193. Nussinov, R., C.J. Tsai, and H. Jang, *Independent and core pathways in oncogenic KRAS signaling*. Expert Rev Proteomics, 2016. **13**(8): p. 711-6.
194. Emekli, U., et al., *HingeProt: automated prediction of hinges in protein structures*. Proteins, 2008. **70**(4): p. 1219-27.
195. Frishman, D. and P. Argos, *Knowledge-based protein secondary structure assignment*. Proteins, 1995. **23**(4): p. 566-79.
196. Fiser, A., R.K. Do, and A. Sali, *Modeling of loops in protein structures*. Protein Sci, 2000. **9**(9): p. 1753-73.
197. Ogmen, U., et al., *PRISM: protein interactions by structural matching*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W331-6.
198. Tuncbag, N., et al., *Predicting protein-protein interactions on a proteome scale by matching evolutionary and structural similarities at interfaces using PRISM*. Nat Protoc, 2011. **6**(9): p. 1341-54.
199. Duhovny, D., R. Nussinov, and H.J. Wolfson, *Efficient unbound docking of rigid molecules*. Algorithms in Bioinformatics, Proceedings, 2002. **2452**: p. 185-200.
200. Schneidman-Duhovny, D., et al., *PatchDock and SymmDock: servers for rigid and symmetric docking*. Nucleic Acids Res, 2005. **33**(Web Server issue): p. W363-7.
201. Neal, S., et al., *Rapid and accurate calculation of protein  $^1H$ ,  $^{13}C$  and  $^{15}N$  chemical shifts*. J Biomol NMR, 2003. **26**(3): p. 215-40.
202. Liu, G., et al., *Structure of MST2 SARAH domain provides insights into its interaction with RAPL*. J Struct Biol, 2014. **185**(3): p. 366-74.
203. Das, J. and G.M. Rahman, *CI domains: structure and ligand-binding properties*. Chem Rev, 2014. **114**(24): p. 12108-31.